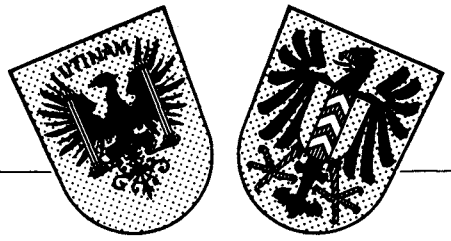


EFTF 93

Seventh European Frequency and Time Forum

PROCEEDINGS

16, 17, 18 March 1993
NEUCHÂTEL/Switzerland



7th European Frequency and Time Forum

Neuchâtel, Switzerland

March 16-18, 1993

Conference organiser and proceeding publisher:

FSRM
Swiss Foundation for Research
in Microtechnology
Rue de l'Orangerie 8
CH-2000 **Neuchâtel**, Switzerland

Phone: int. ++41 38 24 52 00
Fax: int. ++41 38 24 71 45

1993 Conference Location: Neuchâtel University, Espace Louis-Agassiz,
Neuchâtel/Switzerland

Permanent EFTF Secretariat:

FSRM
Swiss Foundation for Research in Microtechnology
Rue de l'Orangerie 8
CH-2000 Neuchâtel

Phone: int. ++41 38 24 52 00
Fax: int. ++41 38 24 71 45

7 TH EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 16 - 17 - 18 MARCH 1993

Table of contents

Introduction

Local Organisation Committee, Chairman XIX

Executive Committee, Acknowledgements XX

Scientific Committee XXI

Exhibitors XXII

Opening Session

Chairman: K. Hilty

- Opening 3
K. Hilty, 1993 Chairman of the Scientific Committee
Swiss Telecom (PTT) - Bern - Switzerland

- Welcome and official address of the Swiss Federal Office of Metrology 7
O. Piller
Swiss Federal Office of Metrology - Wabern - Switzerland

- Driving forces affecting precise time and frequency (*keynote*) 9
H. Hellwig
Air Force Office of Scientific Research - Washington DC - USA

- Timing of digital signals and telecommunications network synchronisation (*invited*) 23
A. Reid
British Telecom - London - UK

- Time-Frequency synchronisation in radioastron orbital VLBI projects (*invited*) 33
V.V. Andreyanov
Astro Space Center - Moscow - Russia

Quartz Crystals 1

Chairman: R. Besson

1. The influence of dimensions on the resonant frequency temperature dependence of AT-cut quartz strips 41
J. Zelenka, J. Pavlovec, J. Sedlacek, J. Suchanek
Technical University of Liberec - Liberec - Czechoslovakia

2. Finite element analysis of AT strip quartz resonators 45
K. Trümpy, W. Zingg
ETA S.A. - Grenchen - Switzerland

3. An adjustment method of resonant frequency and frequency temperature coefficients for NS-GT cut quartz crystal resonators 51
H. Kawashima
Seiko Electronic Components Ltd. - Tochigi-shi - Japan

4. Characterization of BVA in industrialization process 57
J.J. Boy, A. Berthaut, G. Renard, R. Roy, D. Thiébaud, G. Robichon, C. Longet
ENSMM - LCEP - Besançon - France

5. Complex mass-loading effects in AT-cut quartz crystal resonators 63
I. Mateescu, I. Mateescu
Inst. of Physics and Technology of Materials - Bucarest-Magurele - Rumania

Telecom Applications

Chairman: A. Reid

1. Stability requirements of oscillators used for synchronization loops in communications systems (*invited*) 69
A.L. Welti
ETH Zentrum-IKT - Zürich - Switzerland

2. Clock generation and synchronization for high-speed serial data systems 77
D.T. Comer
Pennsylvania State University - Middletown, PA - USA

3. Accurate network synchronization for telecommunications networks using paired paths 83
M. Kihara, A. Imaoka
NTT Transmission Systems Laboratories - Yokosuka - Japan

4. Concatenated clock performance within a synchronous optical transport hierarchy (*abstract*) 89
R. Avery, G. Zampetti
Telecom Solutions - San Jose, CA - USA

Oscillators

Chairman: D. Hauden

1. Designing for low acceleration sensitivity in crystal resonators and oscillators 93
J.A. Kosinski, A. Ballato
US Army Research Laboratory - Fort Monmouth, NJ - USA

2. High quality quartz crystal oscillators compatible with surface mounting technology 99
C. Gelle, L. Levine
Thomson / CEPE - Argenteuil - France

3. Improving operational and technical parameters of quartz oscillators
A.G. Gevorkyan, S.P. Kovita, A. Myasnikov
Russian Institute of Radionavigation and Time - St. Petersburg - Russia 105

Time Transfer

Chairman: C. Thomas

1. Characterization of the signal delays in a ground station designed for satellite two way time transfer 113
J.A. Davis, P.R. Pearce
National Physical Laboratory - Teddington - UK

2. Frequency and time transfer with geodetic GPS receivers 119
T. Schildknecht, P. Baeriswyl, G. Beutler, W. Gurtner
Astronomical Institute University Bern - Bern - Switzerland

3. Two-way time comparisons via DFS-Kopernikus 2 between FTZ and PTB 125
P. Hetzel, A. Söring
Physikalisch-Technische Bundesanstalt - Braunschweig - Germany

4. Improved time transfer using a geostationary direct TV satellite..... 129
F. Meyer, M. Granveaud, B. Laporte, F. Vernotte, M. Vincent
Observatoire de Besançon - Besançon - France

5. Two-way satellite time transfer between USNO and TUG using the Intelsat satellite at 307 degree East 135
D. Kirchner, H. Ressler, R. Robnik, A. Davis, W. Klepczynski, P. Wheeler, W. Powell, L. Veenstra
Technische Universität Graz - Graz - Austria

6. Real-time restitution of GPS time 141
C. Thomas
Bureau International des Poids et Mesures - Sèvres - France

7. Glonass common-view time transfer 147
W. Lewandowski, P. Moussay, G.T. Cherenkov, N.B. Koshelyaevsky, S.B. Pushkin
Bureau International des Poids et Mesures - Sèvres - France

Quartz Crystals 2

Chairman: B. Neubig

1. Physical limitation on the quality factor of quartz resonators; quartz oscillator at 4.2 K 155
F. Bastien, A. El Habti, A. Zarka, F. Lardet-Vieudrin
CNRS-LPMO - Besançon - France

2. Correlation between resonators performances and raw material quality - A review 161
J.P. Aubry, J.P. Michel, M. Zafrany, O. Cambon
Thomson CEPE - Argenteuil - France

3. Quartz resonators for Mars planetary missions 171
J.L. Josset, M.E. Roulet
CSEM S.A. - Neuchâtel - Switzerland

4. Measurement of crystal parameters using a network analyzer in direct impedance mode 177
H. Matthey
Vibro-Meter S.A. - Bienne - Switzerland

5. Small frequency/temperature perturbations in resonators for crystal oscillators: Their effects on system applications and how to measure them 181
J. Dowsett, P.E. Morley, F. Ross
C-MAC Quartz Crystals Ltd. - Harlow Essex - UK

Time Scales and Instrumentation

Chairman: A. Bauch

1. Practical standards for PM and AM noise at 5, 10, and 100 MHz (*invited*) 189
F.L. Walls
National Inst. of Standards & Technology - Boulder, CO - USA

2. Sifting through nine years of NIST clock data with TA2 199
M.A. Weiss, T.P. Weissert
National Inst. of Standards & Technology - Boulder, CO - USA

3. Russian national time scale long term stability estimation using GLONASS time link 211
N.B. Koshelyaevsky, A. Al'shina, P. Daly, S.B. Pushkin
NPO "VNIIFTRI" - Moscow Region - Russia

4. Trends in international timing 221
D.W. Allan, A. Lepek
National Inst. of Standards & Technology - Boulder, CO - USA

Poster 1 - Frequency Control Components and Materials

Chairman: W. Zingg

1. Piezoelectric material constants for surface wave research 231
J.A. Kosinski, J.G. Gualtieri, A. Ballato
US Army Research Laboratory, Fort Monmouth, NJ - USA

2. Characterization and temperature sensitivity of shear horizontal acoustic waves in quartz plates 235
M. Hoummady, E. Bigler, G. Martin
CNRS-LPMO - Besançon - France

3. Thermal transient frequency behavior of SAW devices with a laser beam as a thermal probe 239
N. Oura, N. Kuramochi, M. Takahashi, T. Nakase, I. Kawasaki
Tokyo Institute of Technology - Yokohama - Japan

4. Development of quartz-resonant-sensors using FEM 243
Ch. Reichinger
Technische Universität München - München - Germany

5. Ion exchange and radiation response of hydrogen-related point defects in crystalline natural quartz 247
H. Bahadur
National Physical Laboratory - New Delhi - India

6. Controlled crystal dissolution applied to quartz 255
E. Philippot, M. Deleuze, A. Goiffon, A. Ibanez
CNRS URA 407 - Montpellier - France

7. A high overtone bulk acoustic wave resonator as a tool for acoustic characterisation of thin films and layers 261
G.D. Mansfeld, A.D. Freik, B.N. Krutov
Russian Academy of Sciences - Moscow - Russia

8. The temperature variation on the surface and along the centerline of autoclave during quartz crystals growth (*abstract*) 267
W. Hofman
Institute of Electronic Materials Technology - Warszawa - Poland

9. A semi-quantitative model of the dynamic behaviour of space USOs under fragmented and continuous irradiations 269
J. Bourrieau, M. Brunet, J.F. Dutrey, B. Laporte
CERT-ONERA/DERTS CNES - Toulouse - France

10. Behaviour and measurement of drive level dependency on quartz crystals 279
J.S. Yerna
YeTech Applications b.v. - Doetinchem - The Netherlands

11. A new approach to aging investigation of quartz resonators 283
M. Wojcicki, A. Smolarski, A. Lisowiec
Tele- and Radio Research Institute - Warszawa - Poland

12. Characteristics of TC-cut crystal units excited by lateral field 287
I.V. Abramson
Omsk Research Instrumentation Institute - Omsk - Russia

13. Finite elements method modelling of accelerometric and barometric sensitivity of quartz resonators 291
F. Deyzac
ONERA - Châtillon-sous-Bagneux - France

14. Low ageing HC-36/U encapsulated third-overtone AT-cut 16.384 MHz crystal resonators and oscillators 297
K.K. Tuladhar, G. Jenni
Oscilloquartz S.A. - Neuchâtel - Switzerland

15. Stability of Cr/Au layer on quartz substrates 305
M. Zafrany, F. Meyer
Thomson/CEPE - Argenteuil - France

16. VCTCXO for synchronous digital hierarchy (SDH) applications.....311
W. Balzer
KVG Kristall-Verarbeitung - Neckarbischofsheim - Germany

17. High-stability quartz oscillators for SARSAT/COSPAS project and other satellite-based global positioning systems.....315
Y.L. Vorokhovsky, I.G. Petrosian, B.G. Drakhlis, A.A. Volkov
Morion Mfg Co. - St. Petersburg - Russia

18. Design of high frequency oscillators used in telecommunications.....319
V. Candelier, Ph. Cathelin, B. Nasreddine, G. Guitel
Thomson/CEPE - Argenteuil - France

19. Miniature, low power consumption OCXO oscillator.....323
A. Masiukiewicz, B. Kalinowska
Tele- and Radio Research Institute - Warszawa - Poland

20. Quartz frequency control products for modern telecommunications systems.....327
M.P. Cracknell
C-MAC Quartz Crystals Ltd. - Harlow, Essex - UK

Poster 2 - Frequency Control and Timing Systems

Chairman: L. Prost

1. Spectrum characteristics of GPS time comparison affected by SA effect.....335
H. Jinlun, J. Shuguang
Shanghai Observatory, Academia Sinica - Shanghai - China

2. Instrumental delay calibration with zero baseline interferometer for VLBI time transfer.....341
H. Kiuchi, S. Hama, J. Amagai, T. Yoshino, M. Imae
Communications Research Laboratory - Kashima-Ibaraki - Japan

3. High accuracy two-way time transfer and ranging experiments using SATRE modems.....345
C. Baumer, M. Dienert
Timetech GmbH - Stuttgart - Germany

5. Measures of frequency locking in oscillators with application to ion cooling in a radio-frequency trap.....349
M. Planat
CNRS-LPMO - Besançon - France

6. Passive hydrogen maser frequency stability and accuracy investigations.....353
N.A. Demidov, A.A. Belyaev, B.A. Sakharov, A.A. Uljanov, F.L. Walls
KVARZ Measuring Instruments and Systems - Nizhni Novgorod - Russia

7. High performance hydrogen flow regulator for hydrogen masers 359
L.C. Giurgiu, M.P. Dinca, O.C. Gheorghiu
University of Bucharest - Bucharest Magurele - Romania

9. Possibilities and limits of performance improvements for the diode-laser-pumped Rb maser 363
Z. Xiwen, D. Jinquan, L. Jinting, A. Shaofeng, T. Yongfang
Wuhan Institute of Physics - Wuhan - China

10. A high uniformity magnet for a high C-field cesium beam resonator 367
A. De Marchi, G. Costanzo, M. Repetto, M. Chiampi, C. Lombardi, F. Maddaleno, E. Rubiola, A. Del Casale
Politecnico di Torino - Torino - Italy

11. Optimization of the escape rate for best frequency stability in the atomic hydrogen maser 373
L.G. Bernier
Observatoire Cantonal - Neuchâtel - Switzerland

12. Characteristics of fiber optic frequency reference distribution systems 379
J.J. Suter, R.A. Dragonette
Johns Hopkins University - Laurel, MD - USA

13. TV-frequency standard as a traceability source in accredited calibration laboratories 383
K. Kalliomäki, T. Mansten
Technical Research Centre of Finland (VTT) - Espoo - Finland

14. A precise GPS-based time and frequency system 387
J. Mc Nabb, E. Fossler
TRAK Microwave Corporation - Tampa, Florida - USA

15. Design and performance of the NIST satellite Two-way time transfer modem (*abstract*) 391
D. Davis, D. Howe, M. Lombardi, M. Weiss
National Inst. of Standards & Technology - Boulder, CO - USA

16. Temporal and spatial temperature variations in the surroundings of a clock 393
J. Mannermaa, K. Kalliomäki, T. Mansten
University of Oulu - Oulu - Finland

17. Precise frequency distribution using fiber optics 399
R.L. Sydnor, M. Calhoun
California Institute of Technology - Pasadena, Ca - USA

18. Calibration experiments of GPS time transfer receivers and ionospheric delay measuring receivers between Japan and France 409
M. Imae, M. Aida, T. Goto, C. Miki, C. Thomas, G. Petit, M. Granveaud, W. Lewandowski
Communication Research Laboratory - Kashima-machi, Ibaraki - Japan

19. Application of meteor burst equipment for high precision comparisons of time and frequency standards 413
L.A. Epictetov, R.R. Merzakreev, V.V. Sidorov
Kazan State University - Kazan, Tatarstan - Russia

21. Precise GPS ephemerides for time transfer 417
G. Petit, W. Lewandowski, C. Thomas
Bureau International des Poids et Mesures - Sèvres - France

22. Use of Loran-C signals for time/frequency applications 423
D. Sennedot, Y. Minguy, B. Thomas
Direction des Constructions Navales - Brest Naval - France

23. The first months operation of a Maser CH1-75 and two HP 5071A Cesium 429
F. Collin, P. Paquet
Observatoire Royal de Belgique - Bruxelles - Belgium

24. Transfer of reference signals for space VLBI: short and medium terms 433
G. Daigne
Observatoire de Bordeaux - Floirac - France

SAW 1

Chairman: J. Heighway

1. State-of-the-art SAW oscillators: Design and performance (*invited*) 439
G.K. Montress, T.E. Parker
Raytheon Research Division - Lexington, MA - USA

2. Modeling of a general structure of SAW acoustically and electrically coupled resonator filters 455
D.P. Chen, J. Heighway, C.S. Hartmann
Advanced SAW Products S.A. - Bevaix - Switzerland

3. Design and performance of GHz range, low cost oscillators using shear horizontal surface acoustic waves 459
I.D. Avramov
Institute of Solid State Physics - Sofia - Bulgaria

4. A new technique for SAW resonator frequency trimming 465
L. Roselli, G. Avitabile, G. Manes, E. Verona, C. Caliendo
University of Florence - Florence - Italy

Cesium 1

Chairman: L. Cutler

1. Optical pumping in PTB's frequency standard CSX 473
S. Weyers, R. Schröder, T. Heindorff, A. Bauch
Physikalisch-Technische Bundesanstalt - Braunschweig - Germany

2. Microwave spectrum asymmetry in an optically pumped cesium beam clock 479
V. Giordano, N. Dimarcq, G. Théobald, P. Cérez, C. Audoin
CNRS-LHA - Orsay - France

3. The effect of majorana transitions on the frequency of a cesium atomic clock 483
R. Schröder, A. Bauch
Physikalisch-Technische Bundesanstalt - Braunschweig - Germany

4. Accuracy analysis of the atom-microwave interaction in a TE_{017} cylindrical Rabi Cavity for a high C-field cesium beam resonator 489
A. De Marchi, E. Bava, P. Tavella
Politecnico di Torino - Torino - Italy

SAW 2

Chairman: Ph. Defranould

1. Comparison of low loss SAW technologies for mobile telecommunication applications 497
C.S. Hartmann, J. Heighway
Hartmann Research Inc. - Dallas, TX - USA

2. A new compact SAW low loss filter for mobile radio 499
M. Solal, J.M. Hodé
Thomson Sintra ASM DTAS - Valbonne - France

3. The modeling of heavy electrode gratings for narrow band SAW devices 505
V.P. Plessky, T. Thorvaldsson
Advanced SAW Products S.A. - Bevaix - Switzerland

4. The design and development of a 950 MHz low loss SAW coupled resonator filter for a mobile communication receiver application 511
M.A. Schwab, P.A. Durrant
Advanced SAW Products S.A. - Bevaix - Switzerland

5. Effects of surface electrical boundary condition on excitation and propagation of highly piezoelectric leaky surface acoustic waves 517
K. Hashimoto, M. Yamaguchi
Chiba University - Chiba-shi - Japan

Cesium 2

Chairman: P. Thomann

1. Realisation of a frequency standard using an atomic fountain (*invited*) 525
A. Michaud, M. Chowdhury, C.J. Cooper, K.P. Zetie, G. Hillenbrand, V. Lorent, A. Steane, C.J. Foot
University of Oxford - Oxford - UK

2. The fourth primary cesium atomic clock of the PTB 531
A. Bauch, B. Fischer, T. Heindorff, R. Schröder
Physikalisch-Technische Bundesanstalt - Braunschweig - Germany

3. Properties of an optically pumped cesium beam frequency standard with $\phi = \pi$ between the two oscillatory fields 537
C. Audoin, V. Giordano, N. Dimarcq, P. Cérez, P. Petit, G. Théobald
CNRS-LHA - Orsay - France

4. The high C-field concept for an accurate cesium beam resonator 541
A. De Marchi
Politecnico di Torino - Torino - Italy

5. NIST-7, The new US primary frequency standard 549
R.E. Drullinger, J.P. Lowe, D.J. Glaze, J. Shirley
National Inst. of Standards & Technology - Boulder, CO - USA

Filters

Chairman: J. Dowsett

1. Extension of monolithic filter application field and their miniaturization 555
S.A. Sakharov, A.V. Medvedev, I.M. Larionov, Y.A. Samsonov
Scientific Research Institute "FOMOS" - Moscow - Russia

2. Design of energy trapping integrated filters 561
J. Détaint, J. Schwartzel, A. Zarka, B. Capelle, C. Joly
CNET - Bagneux - France

3. New techniques used for 200 MHz wide band filters manufacturing 567
J.P. Michel, M. Zafrany, T. Soubirous, L. Schwartz, Y. Billon, D. Mouton, M. Berte, S. Lechopier, J.L. Le Corre
Thomson/CEPE - Argenteuil - France

New Atomic Frequency Standards

Chairman: C. Audoin

1. Laser-cooled metastable xenon: Progress toward an optical frequency standard (*invited*) 577
W.D. Phillips, S. Rolston, M. Walhout, A. Witte, H. Megens
National Inst. of Standards & Technology - Gaithersburg, MD - USA

2. Frequency measurement of the hyperfine splitting of $^{171}Yb^{+}$ ions stored in a Paul trap 583
A. Bauch, D. Schnier, Chr. Tamm
Physikalisch-Technische Bundesanstalt - Braunschweig - Germany

3. Repeatability of the magnesium frequency standard 589
A. Godone, C. Novero, P. Tavella, K. Rahimullah
IENGF - Torino - Italy

4. Diode-laser pumped, rubidium cell frequency standards 593
C. Szekely, R.E. Drullinger, F.L. Walls, J.P. Lowe, A. Novick
National Inst. of Standards & Technology - Boulder, CO - USA

Sensors

Chairman: P. Krempl

1. Lamb waves pressure sensor 601
N. Tirole, A. Choujaa, D. Hauden, P. Blind, J.C. Pommier, A. Cachard, G. Martin, M. Froelicher
LPMO-CNRS - Besançon - France

2. Force-frequency transducer using a multimode doubly rotated quartz resonator 607
B. Dulmet, R. Bourquin, N. Shibanova
ENSMM-LCEP - Besançon - France

3. Quartz crystal rotation rate sensors (*abstract*) 613
S. Dalla Piazza, M. Christen
Asulab S.A. - Neuchâtel - Switzerland

4. A flow sensor based on shear horizontal acoustic wave oscillator 615
C. Déjous, D. Rebière, J. Pistré, M. Hoummady
Université de Bordeaux I - Talence - France

Space 1

Chairman: S. Starker

1. European complement to GPS: Presentation of the concept and experimental system 623
T. Trémas, J. Barbier
Centre National d'Etudes Spatiales - Toulouse - France

2. RIRT's contribution to the solution of metrological support problem with regard to native radio systems and aids including GLONASS 633
A.G. Gevorkyan, Y.G. Gouzhva, P.P. Bogdanov, I.G. Pushkina, S.N. Klyushnikov
Russian Institute of Radionavigation and Time - St-Petersburg - Russia

4. Time transfer between USA and France through LASSO 637
C. Veillet, P. Fridelance
Observatoire de la Côte-d'Azur - Grasse - France

5. Preliminary report on the comparison of LASSO and GPS time transfers 641
F. Baumont, P. Grudler, C. Veillet, P. Fridelance, J. Wiant, W. Lewandowski, G. Petit
Observatoire de la Côte d'Azur - Grasse - France

Materials

Chairman: J.P. Aubry

1. Observations of defects and surface acoustic waves in lithium niobate 647
A. Zarka, B. Capelle, J. Détaint, J. Schwartzel, J.M. Hodé
Université P. et M. Curie - Paris - France

2. Temperature dependence of elastic constants of $GaPO_4$ and its influence on BAW and SAW devices 653
W. Wallnöfer, J. Stadler, P. Krempl
AVL List GmbH - Graz - Austria

3. X-ray topography study of gallium phosphate crystals and resonators 659
B. Capelle, A. Zarka, J. Schwartzel, J. Détaint, Y. Zheng, A. Ibanez, E. Philippot
Université P. et M. Curie - Paris - France

4. Electric relaxation by movement of impurities in quartz 667
C. Poignon, G. Jeandel, G. Morlot
Laboratoire Infrarouge - Vandoeuvre les Nancy - France

5. Optical study (I.R. and visible) of the X-rays irradiation effects on swept quartz (AT and SC cuts) 671
J.P. Bachheimer
Université J. Fourier - St-Martin d'Hères - France

Space 2

Chairman: S. Feltham

1. Ion standards for space (*invited*),(*abstract*) 679
L. Maleki
Jet Propulsion Laboratory - Pasadena, CA - USA

2. The Argos data collection and location system. System description, performances and applications 681
M. Brunet, G. Portas, A. Goasguen
CNES - Toulouse - France

3. Microwave links for precise time and frequency transfer between ground- and space-based clocks 693
S. Starker, E. Sappl, W. Schäfer
DLR - Wessling - Germany

4. GPS precise time frequency and ephemeris determination for manned space flight 699
A. Anderman
Rockwell International Corporation - Downey, CA - USA

5. Design of a hydrogen maser for space and progress on cryogenic hydrogen maser research 707
R.F. Vessot, E.M. Mattison, G.U. Nystrom, L.M. Coyle, R.L. Walsworth, R. Decher, R. Nicoll, D. Graveline
Smithsonian Astrophysical Observatory - Cambridge, MA - USA

Postdeadline Poster

- Temporal characterization of 100 MHz picosecond laser pulses. Application to phase noise measurement of a crystal controlled oscillator 715
B. Cunin, C. Eckert, A. Haessler, F. Heisel, J.A. Miehé, M. Planat
Groupe d'Optique appliquée, Centre de Recherches Nucléaires - Strasbourg - France

Banquet events

Summary of the banquet speech 721
C. Cardinal
Institut "L'Homme et le Temps" - La Chaux-de-Fonds - Switzerland

Award of the **european time and frequency prize** 723

Author index 725

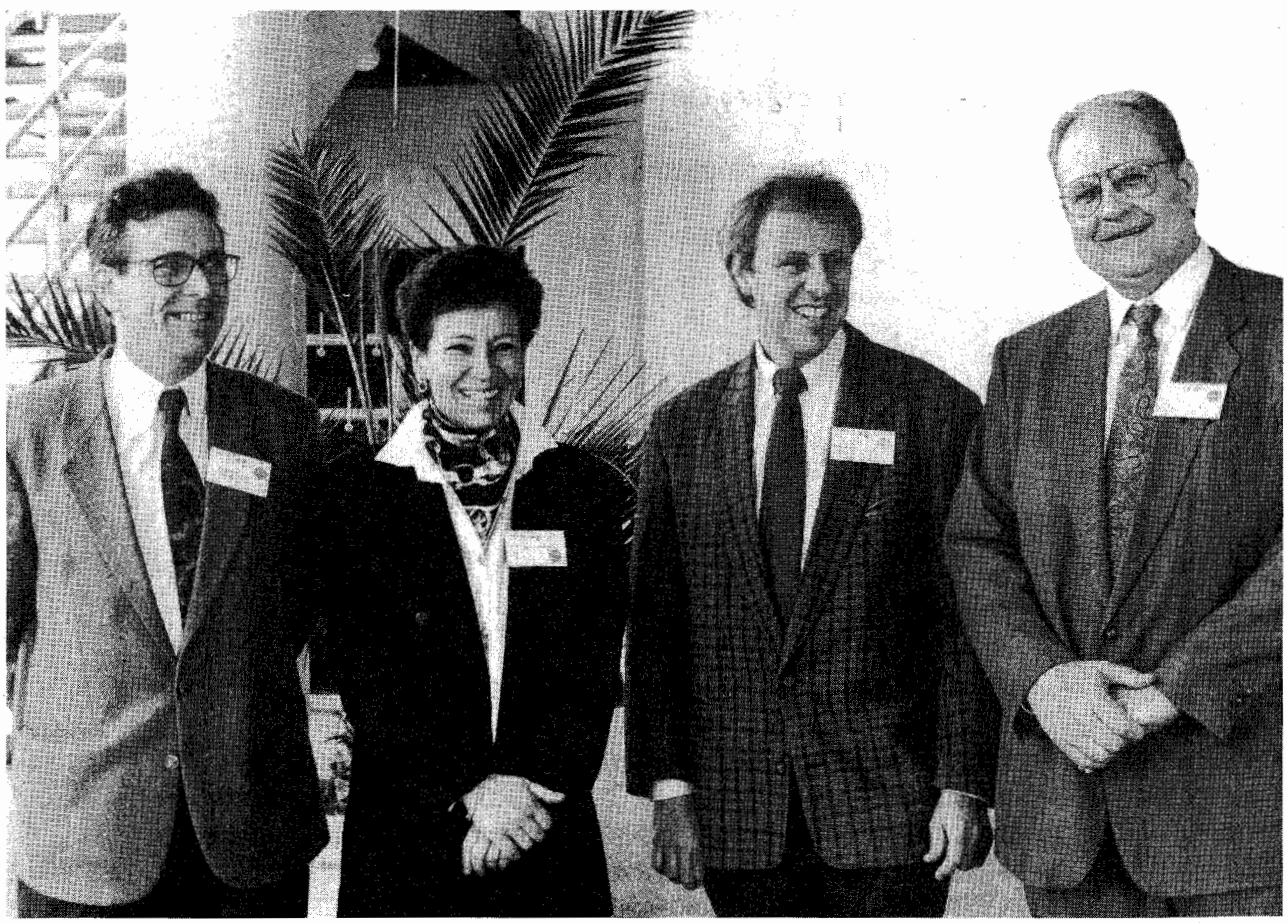
Meeting Report 729

1993 Local Organisation Committee

Marcel Ecabert, FSRM, Neuchâtel
Verena Rutschmann, FSRM, Neuchâtel
Bernard Schlueter, Keres Engineering Ltd., Neuchâtel

1993 Chairman of the Scientific Committee

Kurt Hilty, Head of the section metrology and calibration, Swiss Telecom (PTT), Bern



From the left: M. Ecabert, V. Rutschmann, K. Hilty, B. Schlueter

Executive Committee

C. Audoin	Laboratoire de l'Horloge Atomique, CNRS, Orsay, France
R. Besson	Laboratoire de Chronométrie, Electronique et Piézoélectricité, and Ecole Nationale Supérieure de Mécanique et des Microtechniques, Besançon, France
M. Ecabert	Fondation suisse pour la recherche en microtechnique - FSRM, Neuchâtel, Switzerland
J.-J. Gagnepain	Département Sciences pour l'ingénieur, CNRS, Paris, France
P. Kartaschoff	Section Recherche et Développement de la Direction Générale des PTT, Bern, Switzerland
B. Schlueter	Keres Engineering Ltd., Neuchâtel, Switzerland

Honorary member of the Executive Committee:

A. Rémond	Service Chronométrique de l'Observatoire de Besançon, France

Acknowledgements

We wish to thank the following institutions and companies for their sponsorship:

Republic and Canton of Neuchâtel / City of Neuchâtel / Swiss Academy of Engineering Sciences (SATW) / ESA-ESTEC / United Tobacco Factories / Federal office of education and science / Swiss federal office of Metrology / University of Neuchâtel / Swiss Bank Corporation, Neuchâtel / Imprimerie Krattiger, Corcelles / Jean-Claude Facchinetti, Café au Moka, Neuchâtel.

SCIENTIFIC COMMITTEE

Hilty K.	*Chairman 1993, Swiss PTT, Bern*	*Switzerland*
Aubry J.-P.	Thomson CEPE	France
Audoin C.	C.N.R.S. - L.H.A., Orsay	France
Ballato A.	US Army Research Laboratory, Fort Monmouth	USA
Bauch A.	Physikalisch-Technische Bundesanstalt Braunschweig	Germany
Besson R.	E.N.S.M.M. - L.C.E.P, Besançon	France
Boloix R.	Real Observatorio Armada, San Fernando	Spain
Brunet M.	C.N.E.S, Toulouse	France
Busca G.	Observatoire Cantonal, Neuchâtel	Switzerland
Cutler L.	Hewlett-Packard Co., Palo Alto	USA
De Jong G.	Van Swinden Laboratory, Delft	The Netherlands
De Marchi A.	Politecnico di Torino, Torino	Italy
Defranould Ph.	Thomson Sintra ASM, Valbonne	France
Deyzac F.	ONERA, Châtillon-sous-Bagneux	France
Dowsett J.	C-MAC Quartz Crystals Ltd., Harlow Essex	UK
Duchaussoy H.	D.R.E.T., Paris Armées	France
Feltham S.	ESA/ESTEC, Noordwijk	The Netherlands
Graf E.	Oscilloquartz S.A., Neuchâtel	Switzerland
Granveaud M.	LPTF-Observatoire de Paris, Paris	France
Handrek H.	Quarzkeramik GmbH, Stockdorf/München	Germany
Hartl Ph.	Universität Stuttgart, Stuttgart	Germany
Hauden D.	L.P.M.O. - C.N.R.S., Besançon	France
Hruska C. K.	York University, North York, Toronto	Canada
Kalliomäki K.	Technical Research Centre of Finland, Espoo	Finland
Kartaschoff P.	Swiss PTT, Bern	Switzerland
Kirchner D.	Technische Universität Graz, Graz	Austria
Knight D.	National Physical Laboratory, Teddington	UK
Krempl P.	AVL List GmbH, Graz	Austria
Leschiutta S.	Polytecnico, Torino	Italy
Nakagiri K.	Radio Research Laboratory, Tokyo	Japan
Neubig B.	Tele Quarz GmbH, Neckarbischofsheim	Germany
Nieuwejaar P.	Norwegian Naval Materiel Command, Haakonsvern	Norway
Nottarp K.	Institut für Angewandte Geodaesie, Frankfurt-am-Main	Germany
Prost L.	Office Fédéral de Métrologie, Wabern	Switzerland
Quinn T. J.	Bureau Intern. des Poids et Mesures, Sèvres	France
Rutman J.	FIEE Féd. Ind. Electr. et Electroniques, Paris	France
Starker S.	DLR, Wessling	Germany
Steele J.	National Physical Laboratory, Teddington	UK
Thomann P.	Observatoire cantonal, Neuchâtel	Switzerland
Uljanov A.	Kvarz Institute of Electronic Measurements, Nizhni Novgorod,	Russia
Unverdross R.	Unverdross Technik, Starnberg	Germany
Vanier J.	National Research Council Canada, Ottawa	Canada
Walls F.	National Inst. of Standards & Technology, Boulder	USA
Winkler G.	US Naval Observatory, Washington DC	USA
Yannoni N.	Hanscom AFB, Bedford	USA
Zingg W.	Microcrystal div. of ETA, Grenchen	Switzerland

EFTF 93 - Exhibitors

Individual booths:

ASP Advanced SAW Products SA	Case postale 415	CH-2022 Bevaix
CEPE	44, av. de la Glacière	F- 95100 Argenteuil
C-MAC Quartz Crystals Ltd.	Edinburgh Way	GB-Harlow, CM20 2DE
CQE Composants Quartz et Electronique	BP16, rue Keller	F- 10150 Pont Ste Marie
E.S.T. Equipments Services Techniques	85bis, route de Grigny	F- 91130 Ris Orangis
FTS/Austron Inc.	P.O. Box 14766	USA-Austin, TX 78761
Oscilloquartz S.A.	Brévards 16	CH-2002 Neuchâtel
Roditi International Corp. Ltd.	130 Regent St.	UK-London W1R 6BR
Scientific Research Institute FONON	Krasnobogatyrskaya 44, PO Box 63	107076 Moscow , Russia
Telecom Solutions	85 West Tasman Drive	USA-San Jose 95134
TimeTech GmbH	Nobelstrasse 15	D-7000 Stuttgart 80
Unverdross Technik	Am Pfeifenberg 5	D-8031 Wörthsee

Common booths:

Cambio International	31 Dixon Street	USA-Laconia, NH 03246
Colorado Crystal Corporation	2303 West 8th Street	USA-Loveland, CO 80537
Sokol Crystal Products	3001 Latham Drive	USA-Madison, WI 53713
Morion Mfg Co.	13a KIMa Ave.	199155 St.Petersburg Russia
Staudte Engineering Inc.	1579 North Main Street	USA-Cedar City, Utah 84720
P.R. Hoffman Materials	321 Cherry Street	USA-Carlisle, 17013
Quartzlock Instruments	Gothic, Plymouth Road	UK-Totnes, Devon TQ9 5LH
Time & Frequency Ltd.	14, Habanai St., POB 1792	IL-Holon 58117
TRAK Europe	29 Dunsinane Avenue	UK-Dundee, DD2 3PN
TRAK Microwave Corporation	4726 Eisenhower Blvd.	USA-Tampa, FL 33634-6391

OPENING SESSION

Chairman: K. Hilty

Opening

by Kurt Hilty, Chairman of the scientific committee 1993

Ladies and gentleman I have the pleasure to welcome you all to the Seventh European Frequency and Time Forum in Neuchâtel.

J'ai le plaisir de saluer **Mme Dusong**, qui représente la ville de Neuchâtel
et **M. Dubois** du gouvernement de la République et Canton de Neuchâtel.

I would like to express a special welcome to **Mrs Dusong** as a representative of the city of Neuchâtel,
to **Mr. Dubois** who represents the Républic and Canton of Neuchâtel
and to **Dr. Piller**, as president of the Council of States and director of the Swiss Federal Office of Metrology in Berne.

I welcome our keynote speaker **Dr. Hellwig** of the Airforce Office of Scientific Research in Washington DC,
Mr. Reid of British Telecom
and **Prof. Andreyanov** of Astro Space Center in Moscow.

The Seventh European Frequency and Time Forum is held for the 3rd time in Neuchâtel.

For the first time we combined the symposium with a **Tutorial** held on Monday 15th, whose programme chairman has been **Prof. Hauden** of the Laboratoire de Physique et Métrologie des Oscillateurs, of Besançon. Three working sessions were dedicated to the problems of quartz oscillators, their applications in control loops and appropriate measurement techniques. More than 50 scientists and engineers took the chance to dive into these specific fields.

The present symposium is covering a wide range of topics ranging from **quartz crystal oscillators**, to **atomic frequency standards, time transfer, filters, sensors** and **space applications.**

You are given the possibility to choose your favourite topics out of 120 different presentations.

Due to the rapidly growing demand for **telecommunications**, there is a tremendous need for excellent frequency stability,
a demand of extremely different kinds of oscillators, from high stability atomic clocks to very fast controllable oscillators
and filters for servo loops.
In order to point to these specific requirements we set up a **special session for telecommunications applications** and two **sessions for surface acoustic wave technology.**

I strongly recommend the two **poster sessions on Wednessday morning** from 10h30 to 12h to which all the authors are asked to be present in order to discuss their posters.

I would also like to draw your attention to our **exhibition** which will be open during the whole symposium. 16 companies and institutes are presenting their latest products.

A lot of thanks go to the **Swiss Academy of Technical Science,**
to the **city of Neuchâtel and to the Canton of Neuchâtel** for their generous financial support of the **EFTF 93,**
to the **University of Neuchâtel** who is our host for the next days
and to **Mrs Rutschmann** and **Mr. Ecabert** of **Swiss Foundation for Research in Microtechnology** who did all the administrative and organizational work related to this conference.

While we were organizing this conference I asked myself several times,

what Time means!

I do not know if you can give me an answer or if this conference will lead me to a solution.

Is Time as Albert Einstein said once:

" time is what we see on our watch "

or is it more as Helmar Nahr defined:

" The essence of time resides in the change of the things ".

Finally I consulted the encyclopaedia and found apart from the technical definition the following:

" Time is in the human consciousness the differently experienced elapse from the present to the past or from the expected future to the present ".

All of us remember occasions when we had to wait for someone for 5 minutes, how long these 5 minutes got to be and on the other hand how quickly the same 5 minutes passed when we had an appointment.

If I compare this definition with what I read in all the summaries submitted for this meeting I see some essential discrepancy. In the definition the **differently experienced elapse** is stressed and we look for an exact timing. A long time ago we used for this purpose **hour glasses** and now most of us use sophisticated **electronic watches.** We scientists always look for a **better, more precise definition of time** and how to measure it **most accurately** anywhere in the world. But deep inside us we all have our own time, we all realize a certain time period in very different ways, according to our momentary feeling.

I am glad to be able to announce already the next

European Frequency and Time Forum 1994,

which will be held from the 8th to 11th March, 1994 in Munich.

And now let me push the handle of **Wells Time Machine** and travel into time!

I have the pleasure to introduce as the first speaker **Mrs. Dusong** as a representative of the City of Neuchâtel. In 1992 Mrs. Dusong was elected as the **first woman** to the City Council of Neuchâtel and she is responsible for the financial department and economic promotion.

Madame je vous remercie infiniment d'avoir bien voulu nous honorer de votre présence et je vous prie également de transmettre nos remerciements sincère aux autorités de la ville de Neuchâtel pour leurs hospitalité.

Our next speaker **Mr. Dubois** is bringing us the greetings of the **Republic and Canton of Neuchâtel.** Since 1980 Mr. Dubois is a member of the **State Council** of Neuchâtel and head of the **administration of justice** and **economical affairs**.

Monsieur je vous suis très reconnaissant d' avoire trouvé le temps de participer à
l' ouverture de notre symposium et de nous honorer de votre présence. Nous apprécions grandement votre générosité à notre égard et nous vous prions de transmettre nos remerciements aux autorités de la république et canton de Neuchâtel.

Now let us travel slowly from the more political level to our well known scientific world.

I would like to introduce **Dr. Piller**, who is since 1992 **president of the Council of the States,** which is the upper chamber of the Swiss Parliament.
As a **scientist** he is **director** of the **Swiss Federal Office of Metrology** in Berne. His institute is responsible for the **basic metrology** and the **national standards** of Switzerland.

I have the great honour to introduce **Dr. Hellwig, Director of the US Air Force Office of Scientific Research.**
As Director he is since 1990 responsible for the entire **basic research programme of the US Air Force.**
He is well known for his research work on atomic clocks, on masers, on lasers and molecular beams while he was with the **Heinrich Hertz Institute in Berlin,** the **US Army Electronics Commands,** the **National Bureau of Standards in Boulder Colorado** and as **President of Frequency and Time Systems Inc.,** Beverly, Massachusetts.

At NIST he managed the **Nation's Primary References** for **Time and Frequency.**

Dr. Hellwig will talk to us about: " **Driving Forces Affecting Precise Time and Frequency** ".

I am glad to introduce a colleague of **British Telecom, Mr. Reid,** who is working on **Network Strategy.**
He is an expert on timing and synchronisation in digital telecommunications systems. Mr. Reid will talk about the requirements in new telecommunications technologies as Synchronous Digital Hierarchy and Asynchronous Transmission Mode.

It is a special honour to introduce as the last speaker of the opening Session **Prof. Andreyanov** of Moscow.
Prof Andreyanov is a radiophysicist and head of **Space Electronics Department of the ASTRO SPACE CENTER,** which is part of the **LEBEDEV PHYSICAL INSTITUTE** in Moscow.
He is an expert on **Very Long Baseline Interferometers** for **Radioastronomy**
and is working on problems on
Time and Frequency Synchronisation of Space and Ground-Based Telescopes.

Swiss PTT-Telecom
Research and Development
Metrology and Calibration

Kurt Hilty

Berne, March 16, 1993

Welcome and official address of the Swiss Federal Office of Metrology

O. Piller, Président du Conseil des Etats et Directeur de
l'Office fédéral de métrologie

More than 100 years ago, Werner von Siemens (1816 - 1892) had already reached the conclusion that high precision measurement is essential to the continued advancement of science and technology. Today, it seems perfectly clear how important this vision was. Improvements in measurement science have led to new insights into fundamental physics and these, in turn, have contributed to advances in measurement techniques. This positive feedback is certainly one of the driving forces behind the huge developments in measurement science and technology that have occurred in this century. This is especially so in the domain of time and frequency measurement.

"To find a system of fundamental units", James Clerk Maxwell wrote 125 years ago in his famous work on electricity and magnetism, "we must look not to the dimensions, movement, or mass of our planet, but to the wavelength, frequency and mass of the eternal, unchangeable and identical atoms". It is because Maxwell's recommendation was put into practice so effectively in the measurement of time and frequency that the highest accuracies have been obtained for these quantities: the beat of time is found in the humming of atoms.

Consider, for a moment, the falling of autumn leaves. Their movement, towards the Earth's centre, causes the Earth to turn slightly faster, just as an ice skater accelerates, while in a pirouette, by drawing her arms in. From the point of view of physics, what is happening is clear, and one could even measure this incredibly small effect using a modern atomic clock. However, such accuracies are not playthings; they are essential in today's world for navigation, telecommunications, geodesic measurements from geostationary satellites, and for space research.

Precise measurement of time and frequency was, and is, the basis of fascinating new technical developments. Nevertheless, we should not loose sight of the fact that time also has a human scale. It is especially worthwhile, in today's technical world, pausing a moment to reflect on this. Were we to take time for ourselves, and our fellows, without always marching to the beat of the atomic clock, then, in addition to the blessings given us by technical progress, we might enjoy, fully, life itself.

Driving Forces Affecting Precise Time and Frequency

Helmut Hellwig
Air Force Office of Scientific Research
Washington DC, USA

Abstract

In today's rapidly changing world two major developments have begun to profoundly affect time and frequency device development and application: (1) shifts in military requirements, and (2) the deployment of truly global, time-ordered navigation and communication networks.

This paper analyzes the effects of both developments and predicts a further increase in the importance of a sound time and frequency infrastructure. The underlying three driving forces are discussed: Performance, reliability and affordability. Of these, reliability and affordability, not performance, will strongly influence clock designs and systems concepts. Clock and systems performance is increasingly meaningful only in the context of these two primary drivers. This paper details the derived driving forces such as manufacturability, redundancy, and environmental insensitivity as they impact research, design, and applications.

Historical Perspective

Since the early days of world exploration, beginning with Columbus and the voyages of Magellan and Cook, navigation and time have been inseparable. Advances in timekeeping have near-instantaneously been used for better position determination and mapping. The currently deploying systems of globally networked time [1], satellite-based global navigation [2], and high data-rate global communication [3] via satellites, cable, and optical fiber are summarized in Table 1. Their demands are changing the balance of the three major functional driving forces that govern the development of hardware and systems: performance, reliability, and affordability. Affordability and reliability have become more important than performance, and, in most cases, either one of them is actually the dominant driving force. This should come as no surprise because time and frequency devices have graduated from laboratory applications and systems feasibility demonstrations to critical components in global networks that are essential to the functioning of modern, industrialized societies.

Reliability, Affordability, and Performance

In order to better understand the relationship between reliability, affordability, and performance, it is instructive to study the comparison of various types of clocks and oscillators. Table 2 depicts such a comparison ranging from the wristwatch to the hydrogen maser. As indicator for performance, time/frequency precision was chosen as reflected by retrace [4]; i.e., an oscillator's frequency offset following a powered-down period.

Affordability is reflected by order of magnitude purchase price in US dollars. Reliability is given as an estimate of the mean time between failures (MTBF - in days). In addition, Table 2 provides an overall, qualitative rating of reliability that includes not only the MTBF but some valuation of reliability factors such as environmental insensitivity and start-up complexities. Affordability is complemented with a statement about perceived advantages (+) and disadvantages (-) as regards producibility; i.e., ease of manufacturing. It may surprise some readers that the gas cell fill (isotopic optical filtering), in the case of rubidium, and the state selecting magnets and the hot-wire detector, in the case of cesium, are listed as producibility assets; such readers must realize that these particular design elements are the very reason that these two device types made it commercially in comparison to other high- (or higher-) performance concepts such as the thallium beam, ammonia maser, hydrogen cyanide laser, barium oxide beam, cesium maser, rubidium beam, and numerous other ideas [5], [6]. Table 2 clearly indicates that performance and affordability are nearly proportional to each other except at the high end. Reliability, on the other hand, degrades only by about one order of magnitude across the whole spectrum.

Figures 1 through 3 are an attempt to depict, in a very generalized manner, the relationships between the three driving forces as regards devices (clocks, oscillators, receivers). In each of the figures, pairs of driving forces are plotted against each other while the third is assumed to be constant. Figure 1 shows the linear relationship between performance and affordability until the "state-of-the-art barrier" (dotted vertical line) is approached where affordability rapidly approaches zero; only technology breakthroughs can change this terminal character. In Figure 2 it is shown that reliability and performance have no substantive dependence on each other at the lower performance spectrum; but reliability rapidly drops to reach a new, much lower level near the "state-of-the-art-barrier." The reasons for this behavior may be found in increased design complexity and parts count at the high performance limit.

A very important aspect of the interplay between reliability and affordability is shown in Figure 3 which depicts two linear relationships: In general, reliability includes not only production cost but also turn-on and environmental characteristics; thus, affordability tends to decrease with increasing reliability since improved turn-on and environmental desensitization generally require increased device complexity. If, however, only MTBF is considered, the modifications in management and engineering approaches, that today are called "total quality," have demonstrated that an inverse relationship can be achieved: increased reliability (MTBF, quality) go hand in hand with improved producibility (decreased cost) [7]. Thus, powerful incentives exist that can be marshaled to increase simultaneously reliability and affordability.

Systems Considerations

Networks are established to provide a common service to a substantial number of users with geographic distributions that range from local areas to global. Once operational, the prime objective of any network is to remain operational and to never disappoint a (dependent) user.

Thus, the primary driver in networks is reliability as depicted in Table 3. Reliability concerns may be motivated by national security, safety, business, or other factors. Networks of local regional, national, and global coverage exist, or are being readied, in all three application areas: time/frequency, navigation, and communication.

As shown in Table 3, the first three subordinate elements of reliability are device related; i.e., they are attributes of the clock, oscillator, or receiver: MTBF, turn-on and environmental sensitivity. They can be modified by the technology used, by the design concept, and by detailed design, as well as through parts and materials, manufacturing processes, and storage and handling. The remaining three elements, listed in Table 3, are systems related. Redundancy can be achieved by device back-up with one or more spares and often sophisticated switchover that may have to preserve phase and/or frequency. Graceful degradation refers to a careful balance between network and device functions such that network signals can be used to interpolate or even mimic device functions should particular devices fail. Last, but not least, autonomy is a critical concept in networks allowing stand-alone operation of network subnodes, nodes, or subnetworks in case of breakdown of either devices or network links. Local, regional or even global autonomy is the driving force for the deployment of precision clocks and oscillators within networks; a good example for this is the triple-redundant atomic clocks on board the GPS satellites.

For users, the primary driver is affordability. The user will seek alternatives to time/frequency based solutions if the required performance cannot be afforded. One must keep in mind that affordability is a relative term; e.g., the user will not acquire a clock, oscillator, receiver, or related subsystems if their acquisition or operational costs approach the cost of the host system or platform. Many years ago a proposal for an aircraft collision avoidance system floundered because the cost of a cesium clock exceeded that of some small aircraft! Table 4 lists the subelements of affordability. The first three relate to hardware acquisition; manufacturability and design complexity are different in that the former refers to the manufacturing process, the latter to the predictability (and testability) of performance and functionality. Production rate is important in that there appears to be no substitute for production line volume to assure low parts cost, adequate manufacturing and test equipment, and a highly skilled work force. The last three entries in Table 4 relate to the device operating as part of a larger system. To be affordable, the need for skilled operator (or maintenance) personnel must be eliminated; the device must conform to commonly accepted mechanical, electrical, and control interfaces. Finally, the totality of the reliability issues discussed above bears on affordability after acquisition.

Military needs in time/frequency, position determination, and communication/command/control/intelligence (C^3I) have had a dominant place in device and network research and development. The maturing of deployed systems in all three application areas together with their coordinated use very recently has revolutionized warfare as demonstrated in the Gulf War of 1991. This is depicted in Figure 4 as the focused application of overwhelming force,

concentrated in space and time on an identified target. Sometimes called "hyperwar," its ingredients are knowledge of the target (what it is; its signature; providing data to the warfighters), its location (pinpoint positioning in relation to the platforms employed), and battle management (simultaneity and sequence of operations and actions). Since the military need for massive destruction by conventional or nuclear weapons was rooted in lack of precision as well as lack of knowledge about the battlefield, the new capabilities offered by time/frequency-based navigation and communication will allow warfare to concentrate on military targets with no collateral damage.

Conclusions

With the advent of global networks, reliability and affordability have, in most cases, become more important than performance. Thus, without due concern for these driving forces, businesses may lose their competitive edge, time/frequency ordered systems may be displaced by other technologies, researchers may not reap the rewards of their intellectual investment, and national security may be compromised. With proper attention, however, participation in positioning and data communication may become as much individual common place as today's quartz watch. Furthermore, the military capabilities based on time and frequency will continue to obsolete nuclear war and other forms of mass destruction, an achievement which the time and frequency community can justly be proud of. Finally, a key ingredient in achieving these benefits will be a continuing belief by all concerned that reliability and affordability can be improved simultaneously as complementary aspects of a focus on total quality.

REFERENCES

[1] Kirchner, D., "Two-way Satellite Time Transfer Via Communications Satellites," Proc. IEEE 79, (1991), pp. 983-990.

[2] Barndt, L., "US Coast Guard GPS Information Center and its Function within the Civil GPS Service," Proc. 23rd Annual PTTI Applications and Planning Meeting Dec 1991, NASA Conference Publication 3159.

[3] Sårch, D., "Data Communications," IEEE Spectrum, Jan 1993, pp. 42-45.

[4] H. Hellwig, "Environmental Sensitivities of Precision Frequency Sources," IEEE Trans. I&M 39, (1990), pp. 301-306.

[5] Gerber, E.A., and Ballato, A., eds., "Precision Frequency Control," Academic Press, New York, (1985).

[6] Vanier, J., and Audoin, C. "The Quantum Physics of Atomic Frequency Standards," Adam Hilger, Bristol, England, and Philadelphia, Pennsylvania (1989).

[7] Walton, M., "The Deming Management Method," Perigee Books, Putnam Publishing Group, New York.

TABLE 1 MAJOR HAPPENINGS: 1990's

- **TIME/FREQUENCY:**

  GLOBALLY NETWORKED TIME (10^{-14})

- **NAVIGATION:**

  GPS, GLONASS (METER ACCURACY GLOBALLY)

- **COMMUNICATION:**

  ISDN, SONET (GBIT GLOBAL NETWORKS)

TABLE 2 COMPARISON OF DEVICES

DEVICE	PRECISION (RETRACE)	COST ($)	PRODUCIBILITY (+)	PRODUCIBILITY (-)	MTBF (DAYS)	RELIABILITY (RATED)
WATCH	10^{-5}	10^{0}	IC	-	10^{4}	+++
XO	10^{-6}	10^{1}	IC	-	10^{4}	+++
TCXO	10^{-7}	10^{2}	IC	-	10^{3}-10^{4}	+++
OCXO	10^{-9}	10^{3}	IC	OVEN	10^{3}-10^{4}	++
RUBIDIUM GAS CELL	10^{-11}	10^{3}-10^{4}	GAS FILL	OVEN, CAVITY	10^{3}	++
CESIUM BEAM	10^{-12}	10^{4}-10^{5}	MAGNETS, DETECTOR	OVEN, EM, CAVITY	10^{3}	++
HYDROGEN MASER	10^{-12}	10^{5}-10^{6}	-	CAVITY, BULB, SOURCE	10^{2}-10^{3}	+

TABLE 3 NETWORKS

TIME/FREQUENCY - NAVIGATION - COMMUNICATION

PRIMARY DRIVER: RELIABILITY

- MTBF (DEVICE RELATED)
- START UP (DEVICE RELATED)
- ENVIRONMENTAL SENSITIVITY (DEVICE RELATED)
- REDUNDANCY (DEVICE SWITCH OVER)
- GRACEFUL DEGRADATION (DEVICE HIERARCHY)
- AUTONOMY (DEVICE DISTRIBUTION: GLOBAL, REGIONAL, LOCAL)

TABLE 4 USER

OSCILLATORS - CLOCKS - RECEIVERS

PRIMARY DRIVER: AFFORDABILITY

- MANUFACTURABILITY
- PRODUCTION RATE
- DESIGN COMPLEXITY
- OPERATOR INTERFACE
- SYSTEMS INTERFACE
- RELIABILITY

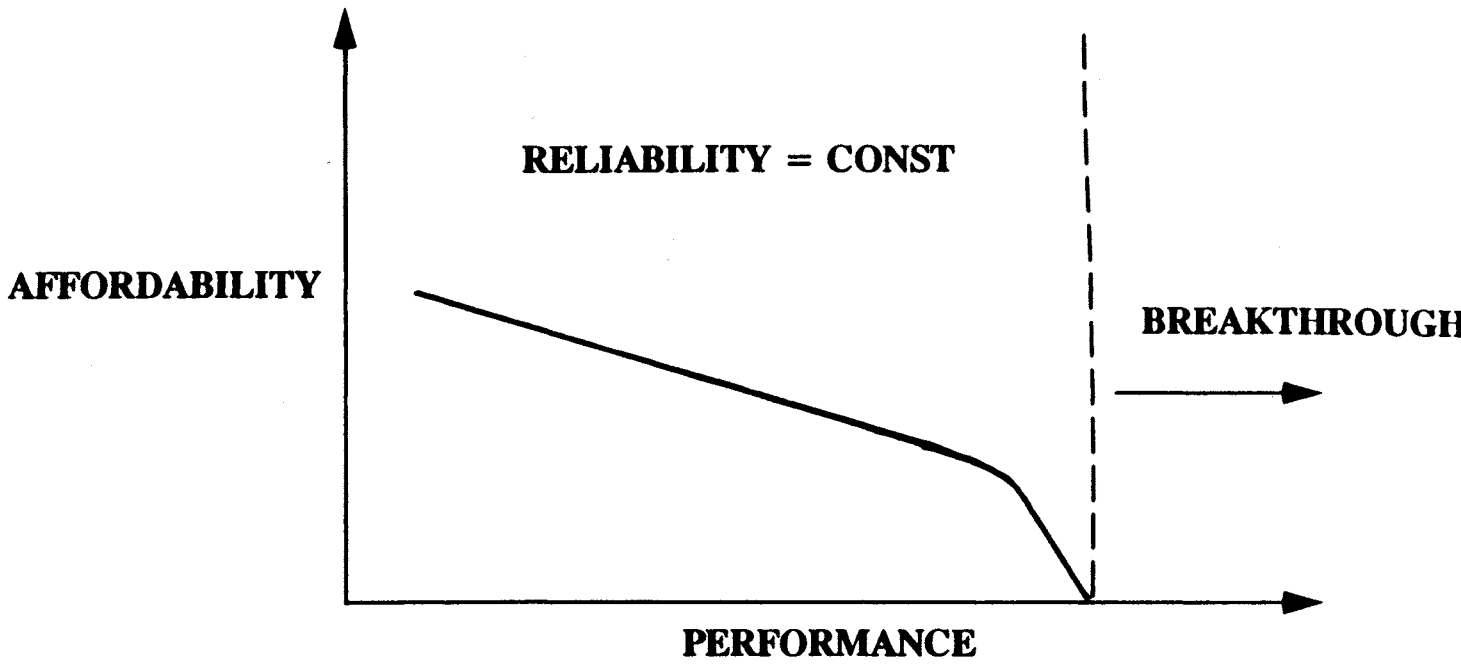


Fig 1 Affordability as a function of performance for constant reliability The vertical, dashed line represents the state-of-the-art.

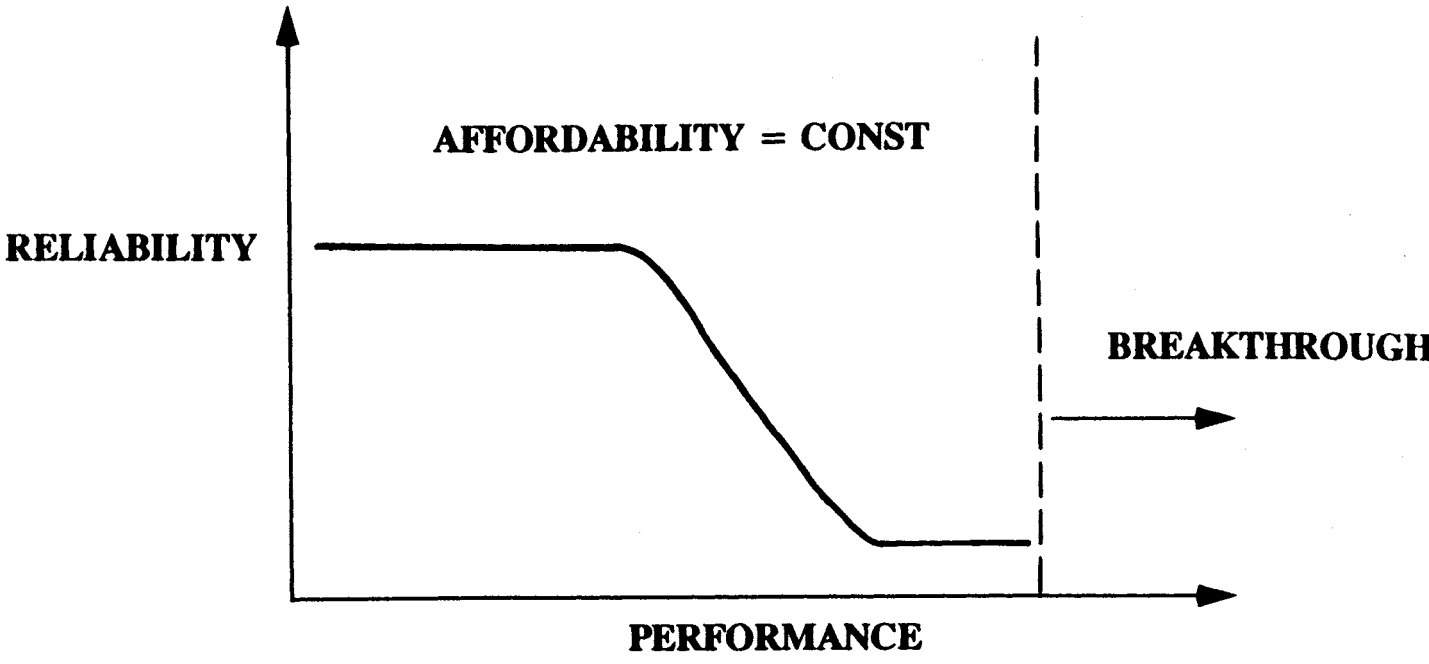


Fig 2 **Reliability as a function of performance for constant affordability (cost). The vertical, dashed line represents the state-of-the-art.**

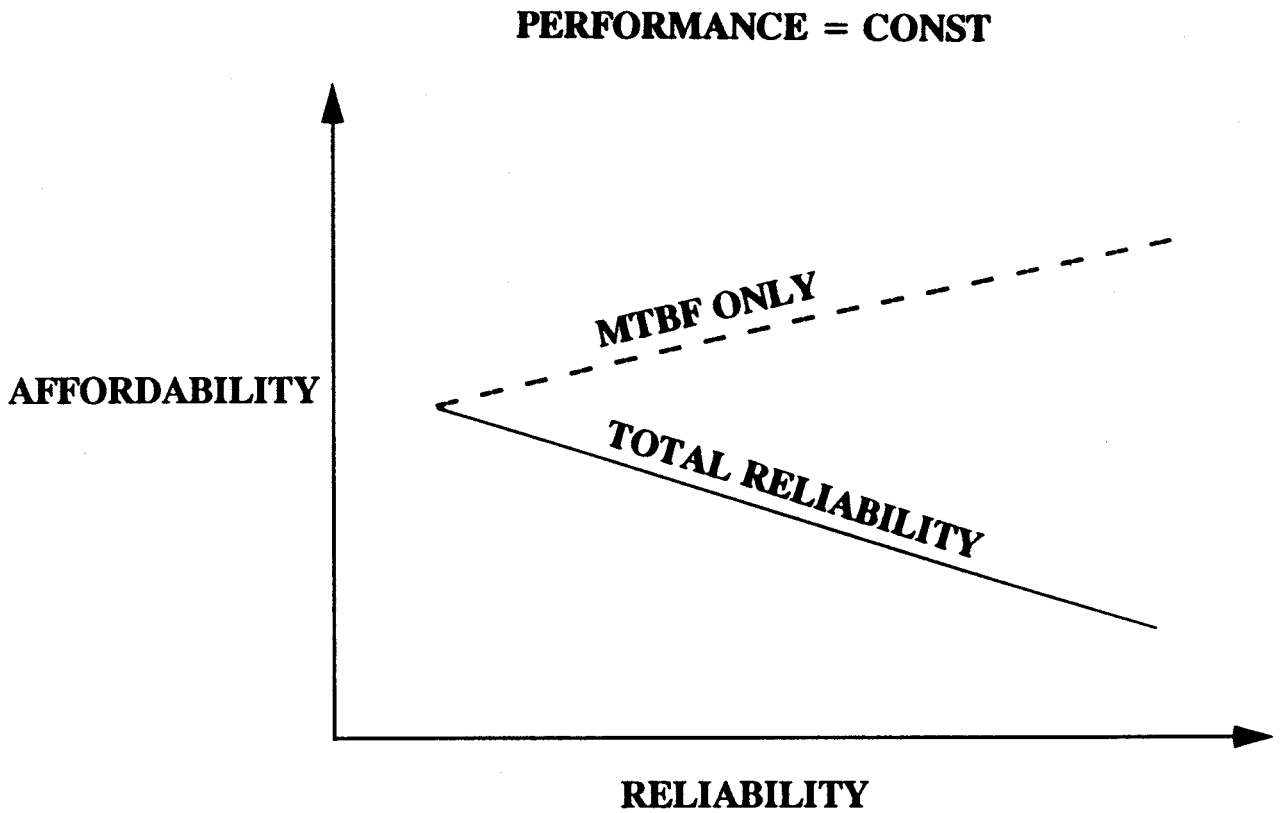
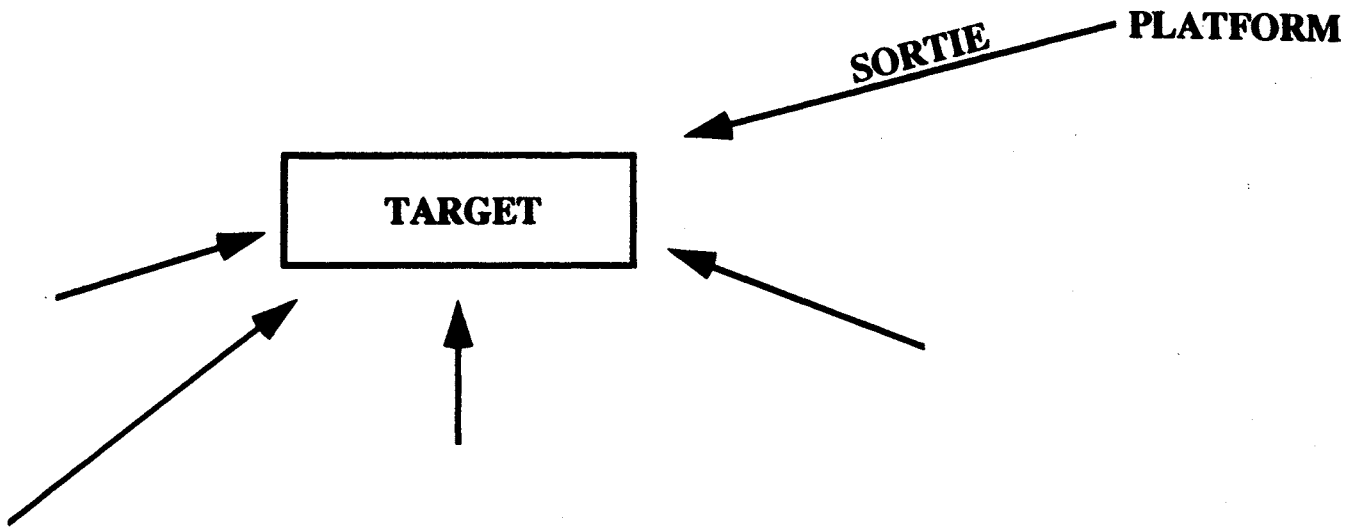


Fig 3 **Affordability as a function of reliability for constant performance. If the total reliability (including turn-on and environmental factors) is considered, affordability drops if reliability is increased; however, if only mean-time between failures (MTBF) is considered, modern approaches to quality allow increased MTBF at reduced cost.**



OVERWHELMING FORCE CONCENTRATION (HYPERWAR).

- **WHERE (POSITION OF TARGET AND SELF)**
- **WHAT (C^3I)**
- **WHEN (SIMULTANEITY AND SEQUENCE)**

Fig 4 Today's Military Paradigm.

TIMING OF DIGITAL SIGNALS AND TELECOMMUNICATIONS NETWORK SYNCHRONISATION

Andy Reid

BT

ABSTRACT

This paper describes the effects of recent developements in technology and international standards on the synchonisation of telecommunications networks. The international standards for both SDH and ATM take full advantage on the great advances in optical and microelectronics technologies to build a telecommunications network with very much greater flexibility and service bandwidths than available on the current public telephony network. The introduction of these technologies alters the way synchronisation can be carried around the network, and while the introduction of SDH need not cause any redesign of network synchronisation, it makes a redesign desirable. The types of synchronisation impairment to which SDH is sensitive are different to those that the current telephony network is sensitive. In addition to the work of defining SDH and ATM, the international standards bodies have been defining new parameters and test methods to control these impairments.

1 INTRODUCTION

During 1988, the International Telephone and Telegraph Consultative Committee (CCITT), agreed a new set of standards for transmission systems. The Synchronous Digital Hierarchy (SDH) [1][2][3] represents a major step forward from the existing set of standards (Plesiochronous Digital Hierarchy - PDH). and is designed to maximise the potential from gigabit optical technology and high speed, high density microelectronics. The bit rate achievable over a fibre has roughly increased by a factor of four every five years, and today systems operating at 2.5Gbit/s are standard. Advances in microelectronics, especially in the speed and density at which CMOS technology will operate, have allowed very high levels of functional integration of the multiplexers that build the 2.5Gbit/s signal from basic telephony channels at 64kbit/s. In addition to this, the desire to manage and control the transmission network remotely using software based network management systems has meant that all telecommunications equipment must be capable of control via an electronic

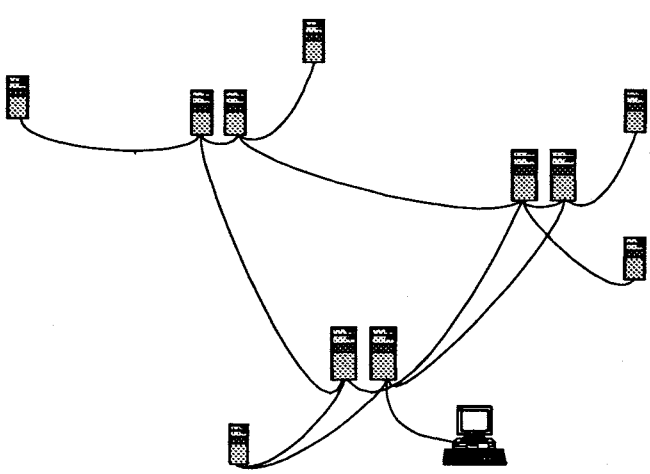


Figure 1 The vision - highly functional equipment directly interconnected with fibre

management and control interface. The vision of the SDH network is illustrated in figure 1. This has highly integrated equipment directly connected by fibre and all controlled by a software management and control system.

The introduction of the SDH, however, into telecommunications networks has a subtle and yet profound influence on the way these networks are synchronised. In this paper, we explore the reason this has come about, look at some of the means of quantifying the effects, and discuss some of the ways telecommunications networks operators can successfully introduce SDH.

More recently, CCITT have agreed recommendations which lay the foundation for a Broadband Integrated Services Digital Network (BISDN) based on Asynchronous Transfer Mode (ATM) [4]. ATM has the ability to multiplex services which have an inherently varying bit rate, for example many interactive data applications, and takes advantage of periods when one service requires less capacity to give it to another which requires more at that moment. This is usually referred to a statistical multiplexing as it depends on the statistical probabilities of when each service might have a peak requirement for capacity. Existing packet networks based on X.25 use a form of statistical multiplexing, however, ATM is capable of operating at very much higher bit rates - hundreds of megabit/s rather than hundreds of kilobit/s. ATM, in many ways, has an even more profound effect on network synchronisation. While this paper concentrates on SDH where work is more mature, the implications for ATM networks are also addressed.

2 SYNCHRONISATION IN TODAY'S DIGITAL PSTN

Prior to the digitalisation of the telephone network, the only possible way of switching signals from one port of a switch matrix to another was by physically making a contact between the ports, known as space switching. Digitalisation made possible a second means of switching - time switching. Time switching makes use of a time division multiplex where many channels are each allocated a short time, or time slot, on a physical port, and switching can take place by switching a channel from one time slot on an input port to a different time slot on an output port. In Europe, all time switches are based on ports with 30 time slots (operating at 2048kbit/s). In addition, digitalisation makes it possible to build a form of space switch where the physical connection is made electronically and only for a very short period of time, the duration of a time slot, known as a time-shared space switch. All digital switches in today's PSTN are a mixture of time switches and time-shared space switches.

In order to work properly, all time switches and time-shared space switches must know when a time-slot is present on a port. The whole digital switch must be in synchronism with itself in order to correctly pass time slots from one stage to another. Moreover, a digital switch must be in synchronism with its neighbours if it correctly track the time slots it receives on its input ports. This is the basis for the synchronisation requirements in today's networks.

The way the network is synchronised was designed to be consistent with the architectural design of the network. This describes the network in terms of nodes and links where the nodes are the switches and the links are the transmission systems as illustrated in figure 2. The great majority of network synchronisation designs use the transmission links in order to take the synchronisation from one node to the next. The nodes are generally laid out in a hierarchy and the links are then used to transport the timing to the nodes according to the hierarchy. Clearly not all links are needed to distribute the synchronisation throughout the network, however, other links can be used to maintain the synchronisation when a link fails.

The design of synchronisation in the existing network is based on the requirements of the electronic switches and the architecture of the PSTN. These two points have underpinned all telecommunications network synchronisation to date.

It can be seen that in this view of the network, one of the prime responsibilities of the transmission links is to carry the timing and indeed, this has been one of the main design criteria of existing PDH digital transmission systems.

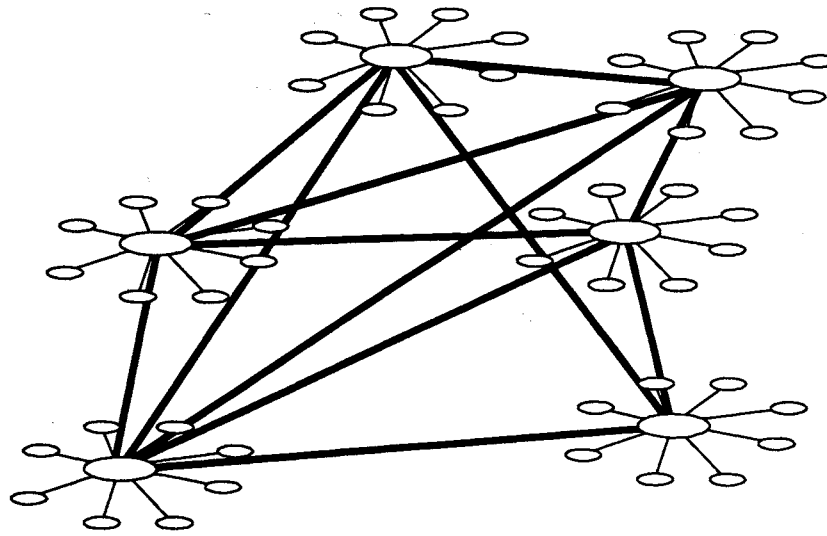


Figure 2 Architecture of the the public switched telephone network (PSTN)

2.1 Current Parameters Used to Describe the Quality of Synchronisation

These fall into two categories: those describing the quality of the transport of the timing across the links and those describing the quality of the clock utilities at the nodes which keep track of the network timing.

Two important parameters were developed to describe the impairments caused by the transmission systems on transport of the synchronisation signal [5]. All cable based transmission systems will be affected to a greater or lesser extent by the daily temperature cycling which can cause cables to expand during the day and contract again at night. As the cable expands, it "holds" more bits and it appears that the delay through the cable increases and similarly when the cable contracts, it appears to "hold" fewer bits and so the delay through the cable appears to decrease. This affects the ability of the clock on the end of the transmission system to stay precisely locked to the clock at the other end. The parameter used to describe this is wander.

The transmission system must also interleave different bits from different digital signals and when this process is reversed at the far end, there is a "waiting time" between the bits of any one tributary signal. In most transmission systems, the characteristics of this waiting time are very complex and the parameter, waiting time jitter, is used to describe and quantify it.

Two parameters are used to describe impairments to the quality of the clock at the node which is used to track the network synchronisation at that node. The first reflects the noise in the clock and is based on statistical measurement parameters. The second is to measure the way the clock responses to errors and breaks in the transmission system which cause the clock to misalign to the incoming signal from the transmission link [6][7].

These four parameters have been the main method of specifying and controlling the quality of network synchronisation at each node. They are based directly on the impairments expected in the system and as will become clear, SDH and ATM introduce different forms of impairments. This has been one of the major work items in the international standardisation process as a new understanding of the parameters has been required. Not all of this work is complete and this paper summarises the progress to date.

3 NETWORK ARCHITECTURE WITH SDH AND ATM

In recent years, many network operators have found that the introduction of new services and especially the great expansion in major companies private networks has been greatly impeded by the architectural view of the network which describes the network in terms of switches and transmission systems. In particular, the main transmission units of 2Mbit/s are all designed to simply connect between main nodes. A 2Mbit/s provided for a customer's private network, which must transit many nodes, is manually patched across the nodes leading to expensive error prone processes.

SDH allows a full 2Mbit/s network based on 2Mbit/s switches at nodes and allows a 2Mbit/s network topology to be constructed which is optimised to all the differing requirements for 2Mbit/s. The same is true of a network at 150Mbit/s. This had lead to the development of a new network architecture which is based on the concept of many layers of networks stacked onto of each other. For example, in SDH, the 64kbit/s PSTN is stacked on top of the 2Mbit/s network which is stacked on top of the 150Mbit/s network as illustrated in figure 3. In ATM, a virtual channel network, is stacked on top of a virtual path network which is stacked on top of the SDH 150Mbit/s network.

This layered network structure was one of the founding principles of SONET and it has now been developed into a rigorous network architecture which has been used to define the functionality of equipment, define and specify the management and control interface of equipment, as well as to design specific SDH networks [8].

3.1 The Impact of Layered Network Architecture on Synchronisation

The introduction of layered network architecture, per se, has no effect on the way networks are synchronised if the PSTN were to remain the only electronically switched layer network, as is the case at present. However, the primary reason for its introduction is to allow the construction of electronically switched layer networks at higher bit rates. The impact of electronic switching in these higher order layer networks on the synchronisation must now be taken into consideration in the design of the synchronisation network

Most fundamentally, any electronic digital switch, in order to operate efficiently, must split up the signal into cell and switch the cell as a whole. Generally the smallest viable cell is 8 bits of data, normally called a byte or octet. SDH uses the byte as its main switched cell. This means that the basic method of multiplexing the signal is interleaving bytes

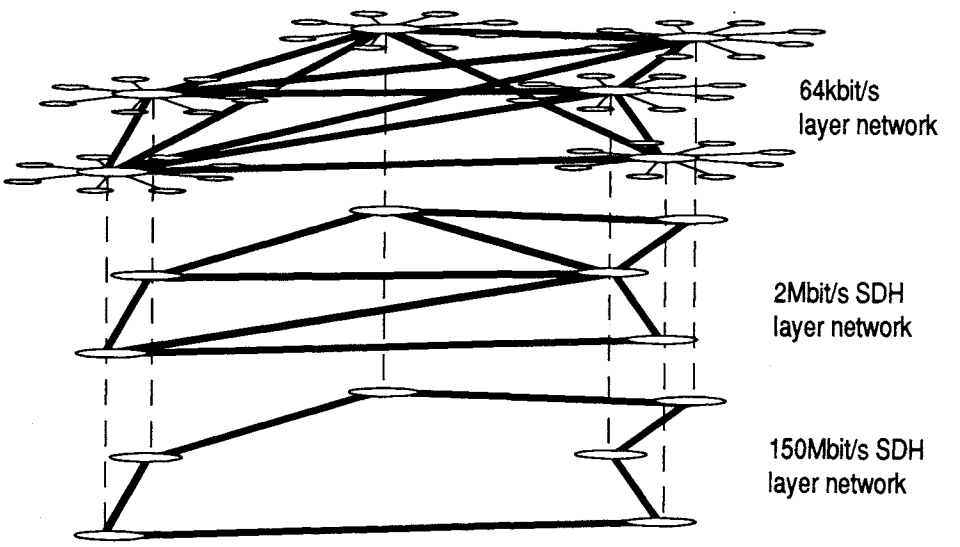


Figure 3 SDH layered network architecture

from many users and the switch can then move the bytes from one user from one port to another according to the desired routing. ATM uses a basic cell size of 48 bytes together an additional 5 bytes which are required to dynamically control the routing of the cells across the switch. The basic multiplex in ATM is therefore the interleaving of 53 byte cells.

If the digital signals are considered to be a stream of continuously advancing phase, these cells have the effect of quantising the phase information carried in the digital signals to levels of the cell size. The phase advances in discrete step of the cell size.

While the introduction of SDH and ATM affects the way synchronisation can be carried in the network as a result of this phase quantisation, the SDH and ATM network elements also require to be synchronised. The rest of this paper is devoted to a more detailed discussion of these two aspects.

3.2 Effects of Quantising Phase

SDH works by assigning a basic number of bytes in a 125μs period for the transport of a tributary signal. For example a 2048kbit/s channel is allocated 32 bytes. This number should be nominally accurate, however, there are several impairments to the synchronisation which means that the incoming phase of the 2048kbit/s is marginally more or marginally less than that of the SDH cross-connect. SDH provides a mechanism that enables an occasional byte to be added or subtracted within the 125μs period in order to accommodate these slight variation in phase. In SDH, this is called the pointer mechanism and the occasional adjustments in the number of bytes in the 125μs period is called a pointer adjustment. A pointer processor is illustrated in figure 4. The far end will therefore see phase normally arriving as a synchronous stream and very occasionally there may be an adjustment to this synchronous stream.

In ATM, the rate of cells in the multiplex is not fixed in the same way as in SDH. The number of cells required by a tributary in a given period of time, eg 125μs, is determined simply by the 48 bytes available in the cell and will normally not be an integer. In all, this means that the time gap between cells will vary in a pseudo-random pattern. The far end will therefore see discrete lumps of phase arrive at a regular rate when averaged over a long time, however, over a short period of time, the lumps of phase arrive irregularly.

These two mechanism for quantising the phase are inherent in the technologies and so their implication on synchronisation must be understood. However, at this point, it is important to realise that the quantisation is a direct consequence of the new architecture which is bringing essential benefits to the network and not simply peculiarities of SDH and ATM.

In a world that has evolved to SDH and ATM, where there is no PDH transmission, and all interfacing to equipment is by optical fibre, then the quantisation has no impact since the signals are always embedded in the SDH or ATM multiplex structures and there is never any need to separate the signal from its method of quantisation. However, in the long intervening period, PDH interfaces will persist

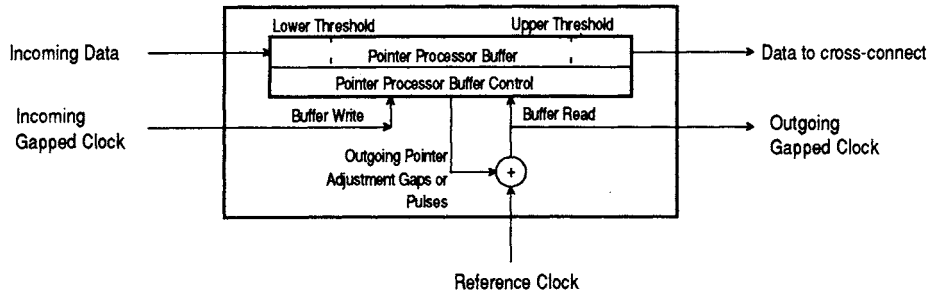


Figure 4 An SDH pointer processor

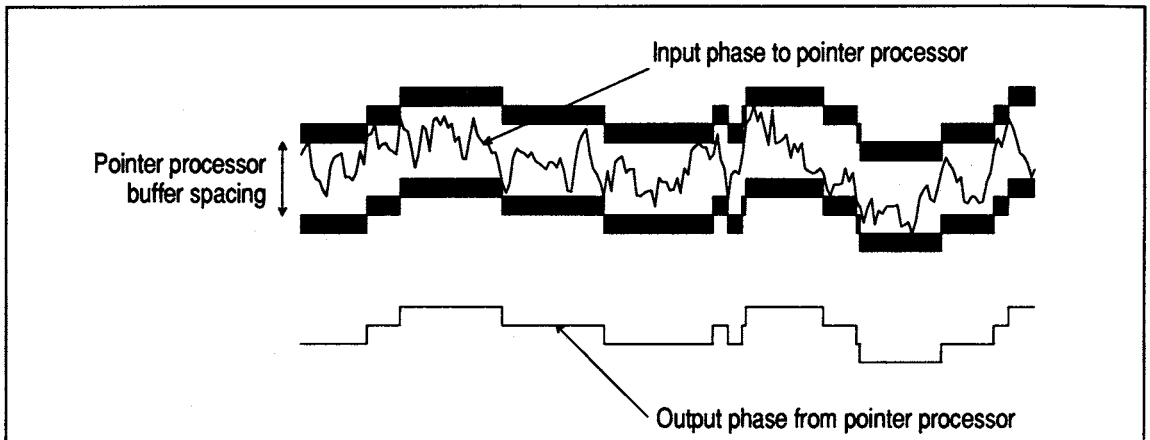


Figure 5 Pointer adjustments as a result of clock noise

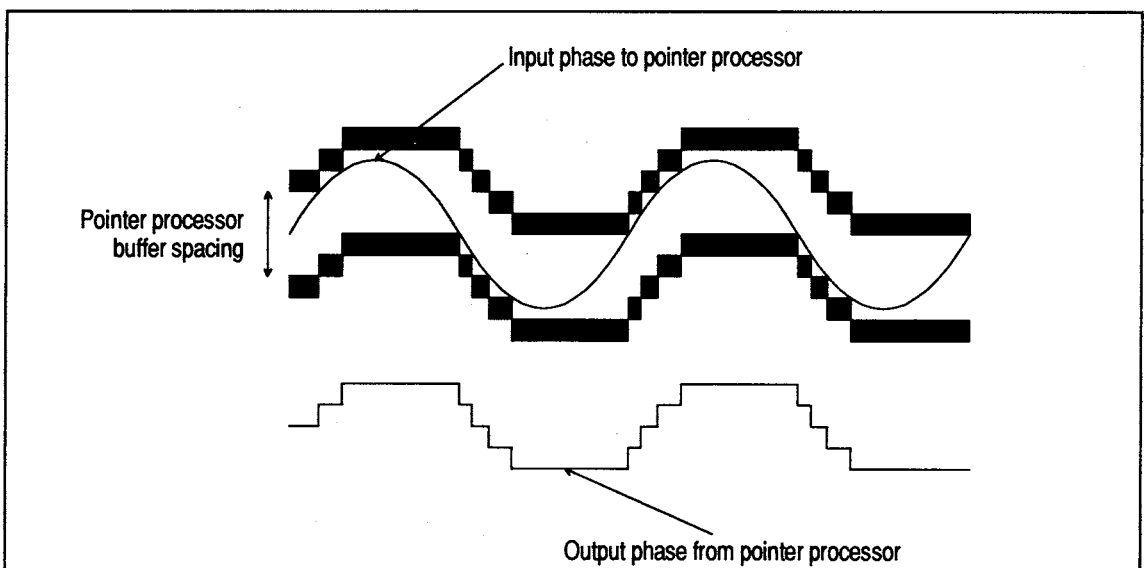


Figure 6 Pointer adjustments from daily wander

as the means of interworking between SDH/ATM and PDH transmission. For example, a 2048kbit/s PDH interface will be used extensively to interwork SDH transport systems with existing 64kbit/s switching equipment. These PDH interfaces are all based on regular digital signals where the phase advances linearly within very close tolerances. Successful interworking therefore requires that SDH and ATM reconstruct these interfaces from the quantised phase signal.

Only some work has been done on the way in which the PDH interface can be constructed from ATM, however, a great deal has been done on how to construct the PDH interface from SDH. The next section briefly details how this is achieved.

3.3 Reconstructing a PDH Interface From an SDH Tributary With Pointer Adjustments

The first step in reconstructing the PDH signal is a good understanding of the statistics of the pointer adjustments. These arise from a number of impairments to network synchronisation particularly clock noise, daily wander, and any drift in the frequency of a clock as a result of it losing its reference it standard timing source. Each of these will result in quantised phase as illustrated in figure 5, figure 6, and figure 7. The probability of the pointer adjustments depends largely on

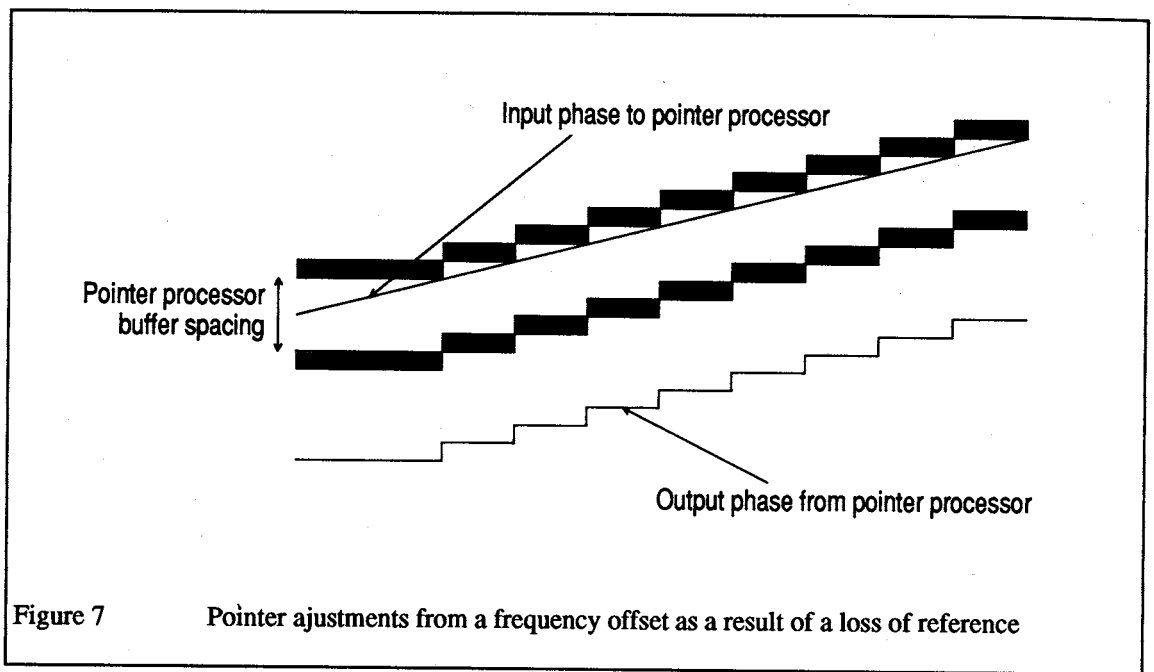


Figure 7 Pointer ajustments from a frequency offset as a result of a loss of reference

the bit rate of the tributary. For example, the byte adjustment at 2Mbit/s represents 4μs of phase, while at 50Mbit/s represents only 160ns of phase. In practise, this means that while pointer adjustments at 50Mbit/s and above will be relatively common, they will be extremely rare at 2Mbit/s. Indeed, extensive simulation of SDH networks has been carried out and it has been almost impossible to generate and 2Mbit/s pointer adjustments, even in the most adverse circumstances.

The PDH interface circuitry must now take these pointer adjustments and feed out the additional - or extract the reduced - phase over a sufficient period of time that it meets the specification for the PDH interface. Much time was spent in the standards bodies on the trade of between the complexity of this circuitry and the number of adverse circumstances and associated pointer statistics that it should successfully handle. Final agreement has been reached on this matter [9], however, in the interim, it has become clear that the specification of the PDH interfaces is currently inadequate and these are currently being

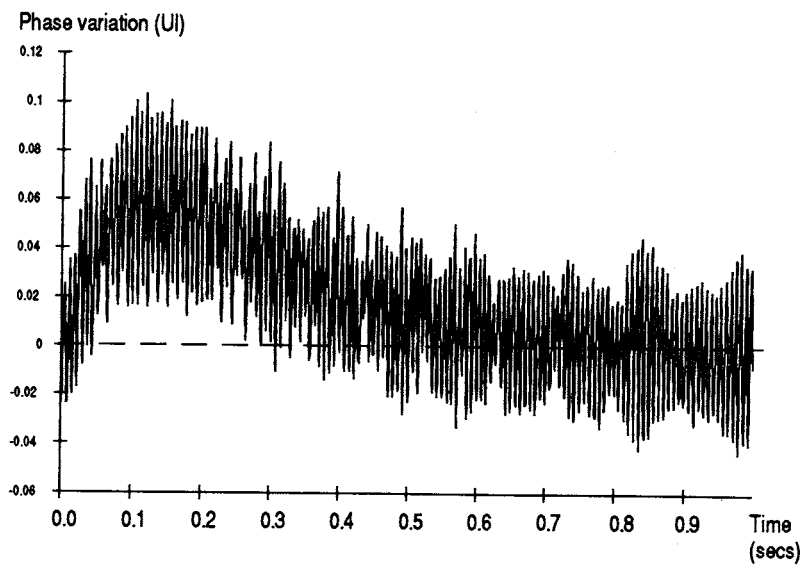


Figure 8 Pointer adjustment smoothed out on a PDH interface also with waiting time jitter

enhanced. Figure 8 illustrates a pointer adjustment being leaked out over a period of a second. It also shows a level of waiting time jitter also present and generated by the interface circuitry.

4 SYNCHRONISING SDH AND ATM EQUIPMENT

The concept of synchronising Synchronous Digital Hierarchy equipment would seem obvious from the name, however, the concept of synchronising Asynchronous Transfer Mode equipment may not! The reasons are the same as outlined above both rely on synchronous structures of bytes or cells. It is only the tributary of ATM that can appear asynchronous when it is carried in the cells which can be dynamically allocated. SDH allocates the bytes on a predetermined basis and so the tributaries are synchronous within the limits of the pointer mechanism.

Before considering the architectural requirements of synchronising SDH and ATM networks, it is first important to understand what levels and characteristics of synchronisation impairments SDH and ATM networks are sensitive to. With SDH impairments to the synchronisation will result in pointer adjustments and so the objective in reducing the levels of the impairments to the synchronisation is to reduce the number of pointer adjustments. In ATM the effects are less clear. They potentially affect the control of the statistical multiplex. More importantly, current proposals suggest using the network synchronisation as a reference for generating the regular clock required for PDH interfaces.

4.1 Reducing Clock Noise

The main cause of pointer adjustments is clock noise. While daily wander will cause pointer adjustments, these will be relatively rare - a few a day and only at the higher bit rates. Clock noise can potentially produce pointer adjustments at 150Mbit/s every few seconds, however, the number is very sensitive to the level of the noise. The 64kbit/s switches of the PSTN were essentially unaffected by clock noise and so the specification and measurement of clocks has not tended to concentrate on this parameter. On the other hand, the PSTN switches are very sensitive to a drift in the clock frequency caused by a loss of the reference input. In SDH this, like wander, will only cause occasional adjustments. In short, the key parameter for synchronising PSTN switches, frequency drift, is not critical for the synchronisation of SDH equipment while the key parameter for synchronising SDH equipment, clock noise, is not critical for the synchronisation of the PSTN switches.

The standards bodies which are considering the specification of clocks suitable for the synchronisation of SDH equipment have given much consideration to the specification of clock noise [10]. This is a random phenomenon and no test can be repeated with certainty and so a degree of confidence must be established. Historically the specification has been based on the parameter MRTIE

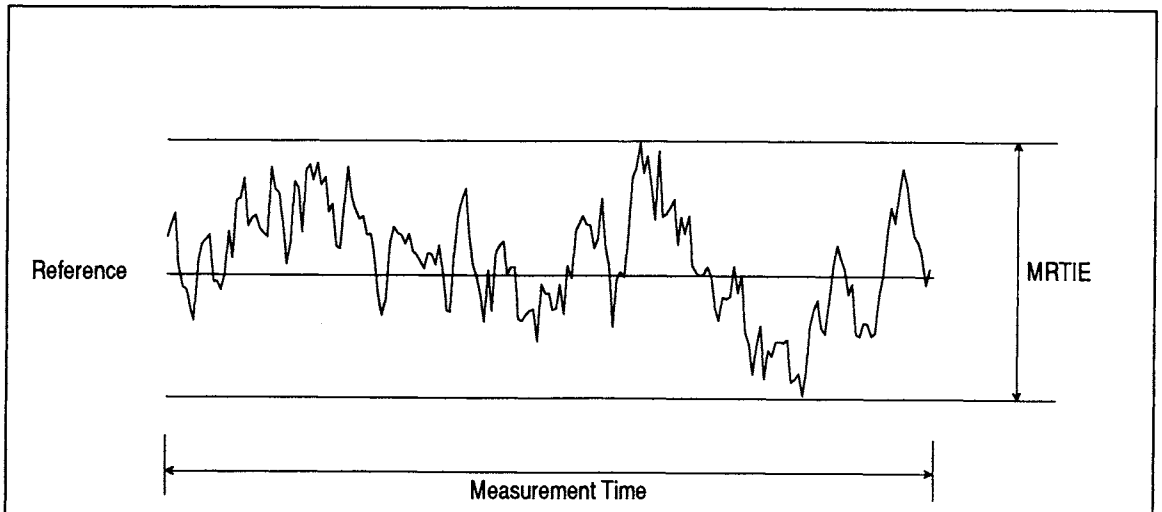


Figure 9 Measurement of Maximum Relative Time Interval Error (MRTIE)

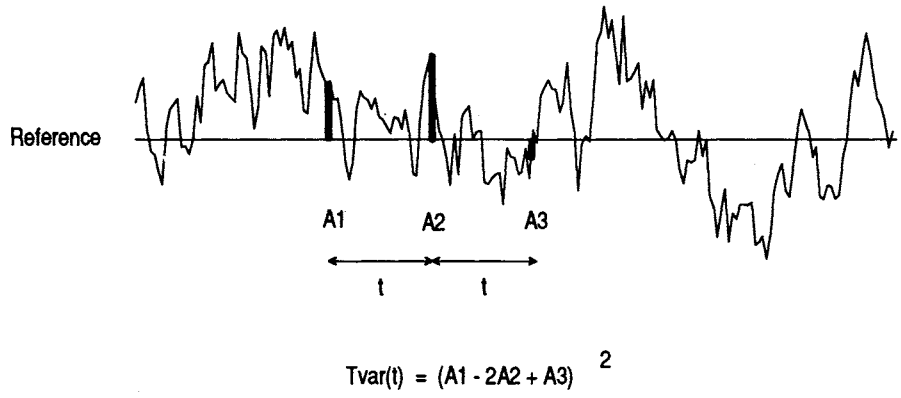


Figure 10 Measurement of Tvar or Allan variance

(maximum relative time interval error) which is the maximum peak-to-peak phase excursion measured in a given measurement time interval relative to the synchronising input reference as shown in figure 9. While this parameter is very easy to measure, it has very little statistical basis. The proposed parameter is Tvar (time variance) also know as Allan Variance. Allan variance is based on the square of the second differences between samples of the noise signal and improves the stability of the test when the clock has flicker or 1/f noise and is illustrated in figure 10. There are still discussions on what reference synchronisation signal should be used for the test and the precise method of averaging used.

In all, the main design objective of the synchronisation for SDH is low clock noise.

4.2 Possible Solutions to the Distribution of Synchronisation to SDH Equipment

There are several options, however, the simple use of the an existing synchronisation network based on 2Mbit/s links between nodes is likely to lead to problems as a timing loop can be established,, at least in the short term. This arises when an SDH equipment cross-connects the 2Mbit/s to which it is synchronised before is reaches the clock utility. This is not an unlikely event in the development of SDH networks and so an alternative strategy is needed.

The most readily implementable option is to use the optical line rate clock, which is synchronous, to carry the timing between nodes. This can give simple and effective distribution of timing. The number of clocks through which the timing passes must be carefully considered as the clock noise can build up. As each clock will filter out noise as well as add noise, the use of different types of clock can control the build of the clock noise in different parts of the spectrum. Again the standards bodies have agreed guidelines on the number and type of clocks than can be linked.

New and interesting options now exist using off air timing. Both GPS (global positioning satellite system) and LORANS-C offer a very high stability timing reference which can be recovered from off air with a relatively cheap timing utility. This offers several advantages. The timing is very stable with very low clock noise and no wander. The timing is not affected by errors and breaks in the transmission links. There is no design required for a complex synchronisation network. And particularly relevant for private networks operators, private network equipment can be synchronised with relying on timing from the public network operator.

5 CONCLUSIONS

Both SDH and ATM have caused a re-examination of synchronisation in telecommunications networks which was first developed for the digitalisation of the PSTN network. The effects have been both ways. SDH and ATM have affected the ways in which synchronisation is provided to elements in telecommunications network and has started

many network operators to consider a more radical solution based on off air timing as increasing viable. In addition, the development of SDH and ATM has been constrained by the limits on synchronisation and historical synchronised interfaces at 2Mbit/s and 1.5Mbit/s. Several new test parameters have been required to measure phenomena which either previously did not exist or were not important to the PSTN network. Standards bodies are still considering a few outstanding issues affecting SDH, however, sufficient work has now been completed to allow network operators to reliably deploy SDH equipment. With ATM there are still several outstanding issues on which there has been little discussion to date.

REFERENCES

[1] CCITT Recommendation G.707 Synchronous Digital Hierarchy Bit Rates

[2] CCITT Recommendation G.708, Network Node Interfaces for the Synchronous Digital Hierarchy

[3] CCITT Recommendation G.709, Synchronous Multiplexing Structure

[4] CCITT Recommendation I.431, Broadband User Network Interface

[5] CCITT Recommendation G.823 The Control of Jitter and Wander Within Digital Networks Based on the 2048kbit/s Hierarchy

[6] CCITT Recommendation G.811 Timing Requirements at the Outputs of Primary Reference Clocks Suitable for Plesiochronous Operation of International Digital Links

[7] CCITT Recommendation G.812 Timing Requirements at the Outputs of Slave Clocks Suitable for Plesiochronous Operation of International Digital Links

[8] CCITT Recommendation G.803 Architecture of Transport Network Based on the SDH.

[9] CCITT Recommendation G.783 Characteristics of Synchronous Digital Hierarchy Multiplexing Equipment Functional Blocks

[10] CCITT Draft Recommendation G.81s, Timing Characteristics of Slave Clocks Suitable for Operation of SDH Equipment

BIBLIOGRAPHY

Jitter; Patrick Trachitta and Eve Varma; Artech House

Transmission Networking: SONET and the Synchronous Digital Hierarchy; Mike Sexton and Andy Reid; Artech House

TIME-FREQUENCY SYNCHRONISATION IN RADIOASTRON OVLBI PROJECTS

V.V.Andreyanov

Astro Space Center of P.N.Lebedev Physical Institute, Moscow, Russia

ABSTRACT

Time and Frequencies synchronisation special features in RADIOASTRON Orbital Long Base Interferometer (OVLBI) are considered. To achieve successful cross-correlation two signals received from the same sky source it is necessary to keep their phase difference (some degrees) during integration time (100..1000 s) at the signal freguency up to 22 GHz.Spacecraft movement and the other reasons cause many phase uncertainties including high dynamical ones. This paper shows how to estimate and compensate them.

Keywords:Space Radiotelescope (SRT), Ground Radiotelescope (GRT), Phase uncertainty, Correlation, Coherence, Tracking station, Spacecraft (SC).

1. RADIOASTRON FUNCTIONAL CONFIGURATION AND CAPABILITIES

RADIOASTRON-1 and 2 are russian current and future OVLBI projects implemented by wide international cooperation. High SRT orbit (apogee 80 thousands km) and 4-5 long distance separated ground Tracking Stations (TS) provide long duration radiocontact between SRT and TS's. Fig.1 illustrates OVLBI configuration. Here, as distinct from ground-based VLBI, the coherent two-side radiocommunication is used additionaly for frequency and time sinchronization of SRT and GRT's data streams via TS's ("phase transfer link") and for transmission of SRT astronomical data to the Earth directly (no on-board recorder). Except of RADIOASTRON at present time Japaneese OVLBI (VSOP) is built also.

The main OVLBI goal is to reconstruct radio image and to investigate nature of the observed galactic or extragalactic source with the highest angular resolution that is determined $\lambda_s/\mathbf{B}'$, here λ_s - signal wave length, $\mathbf{B}'$ - base projection on the "picture" plane. For RADIOASTRON-1 the resolution will achieve 30 arcmicrosecond that is 10 times better that for VLBI (at the same λ_s).

When spacecraft together with SRT and the Earth together with GRT's are moving the base length and its position are changing relatively to the sky source forming so called synthesized aperture. It is equivalent on the angular resolution to the gigantic antenna (drop line, Fig.1).

The information for following radio map (image) reconstruction of the observed source can been derived from cross-correlation results of the signals (in digital form usually) from the all radiotelescopes.

Data for correlation processing are recorded normaly on high density magnetic tapes.Clocks and auxiliary data are recorded together with astronomical data streams. Therefore OVLBI raw data is series of correlation coefficients versus time for different base $\mathbf{B}$ position:

$$\mathbf{R(T,\ F)}_{\max,\max} \equiv \ell\left[a\cdot\frac{2\pi}{\lambda_s}B(t)\cos\varphi(f,t)\right],$$

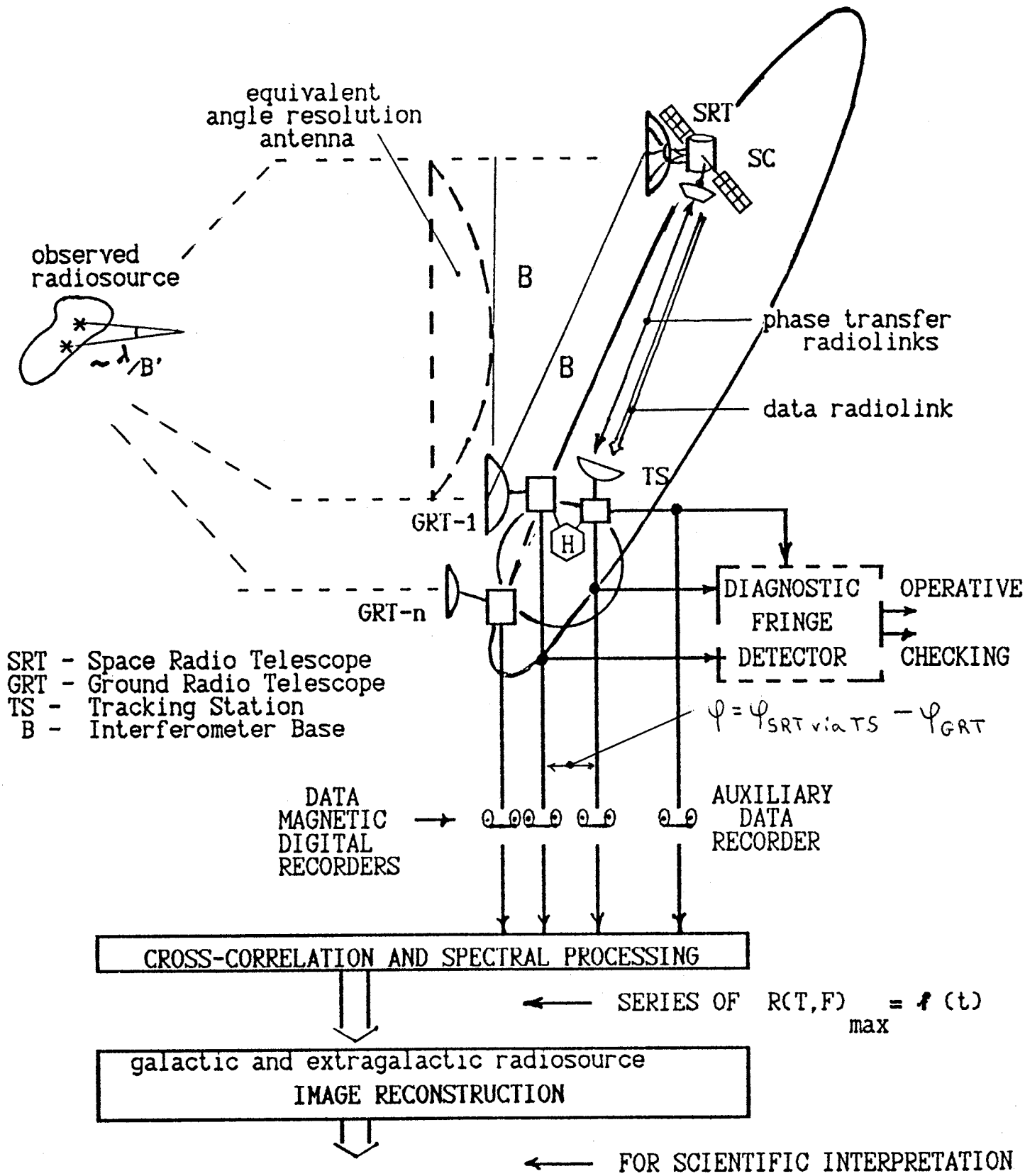


Fig.1 Generic Radioastron OVLBI configuration.

where: sign "max, max" means two-dimention maximum: in time and in interference, fringe, frequency;
a - interferometer signal to noise ratio.

2. SYNCHRONIZATION PURPOSES AND REQUIREMENTS

For successful cross-correlation it is necessary to coincide each pair of the signals (radiotelescopes's data streams from the same sky source) in time and to combine their frequency spectra as closely as possible taking into account that they are changing continuously.
There are many components of signals phase uncertainty $\Delta\varphi_{\oplus}$ when these signals (data streams) are deliveried to the correlation processor:

$$\Delta\varphi_{\oplus} = 2\pi (c/\lambda_s) \cdot \Delta T_{\oplus} + 2\pi \Delta F_{\oplus} \cdot \mathcal{T}$$

$$\Delta T_{\oplus} = \Delta T_s + (\Delta T_{r-l} + \Delta T_{clk} + \Delta T_{med} + \Delta T_{app})$$

$$\Delta F_{\oplus} = \Delta F_s + (\Delta F_{r-l} + \Delta F_{LO} + \Delta F_{med}),$$

Above it is marked:
$\oplus$ - total phase, delay or frequency uncertainty (geometric sum),
r-l - radio link between spacecraft and tracking station,
s - signal from radiosource,
LO - local oscillator, getherodyne,
clk - ground radiotelescope's and tracking station's clocks derived by H-maser oscillator,
med - propogation medium,
app - apparatus, electronics,
c - light velocity.

Some components are known badly or very variable: atmosphere changing, unpredicted spacecraft attitude and velocity changing, etc. In distinct of ground-based VLBI OVLBI has special features:
1) Radiotelescopes move not only relatively to the observed radio source (it is necessary to synthesize aperture) but relatively to each other with high instable velocity (up to 8 km/s);
2) Real SC movement parameters are determined more worse (many orders) than the Earth rotation ;
3) Some phase components depend in addition on radio links parameters, which are absent in VLBI. Radio link phase noise, doppler, doppler rate and delay are varying continuously and do not connected directly with observed signals.

Summarized phase requirement for cross-correlation result can be expressed in form of coherence C (this is degree of phase distiquish between two signals):

$$C \equiv \cos(\Delta\varphi_{\oplus}).$$

Desirable coherence for RADIOASTRON C = 0,99 that is 1% coherence loss ($\sigma \leq 8$ degree rms). The acceptable (the worst) meaning is C = 0,9 ($\sigma \leq$ 26 deg. rms). Of course, correlation coefficients themself will be less. To specify admitted meaning of each phase component t nees to analys so called radioetric improvemen that determines OVLBI sensitivi on flux density from a radiosource; it is proportional to

$$(\Delta f_s \cdot \mathcal{T})^{1/2},$$

Δf_s - max. used signal bandwidth, for RADIOASTRON - 32 MHz;
$\mathcal{T}$ - max. achieved coherent integration time.

Unfortunately, both Δf_s and $\mathcal{T}$ are constrained by technical and physical capabilities of OVLBI.
For example, maximum $\mathcal{T}$ is limited usually (in VLBI) by LO's frequency stability $\Delta f/f$:

$$\mathcal{T} = \Delta\varphi_{\oplus} [2\pi f_s \cdot \Delta f/f]^{-1},$$

if $\Delta f/f = 10^{-14}$ (H-maser), then

$$\mathcal{T} \leq 100...1000 \text{ s},$$

depending on f_s (22...0,33 GHz for RADIOASTRON-1).
It means that after phase uncertainty compensation, so as possible, residual delay must be less $1/2\,\Delta f_s \leq$ 10 ns and residual frequency shift

must be less 0,001 Hz during integration time.
In OVLBI case more strong limitation for $\mathcal{T}$ can arise from unknown part of spacecraft acceleration ΔW:

$$\mathcal{T} = \Delta\varphi_{\oplus} \, (c \cdot f_o / \Delta W)^{1/2} \, f_s^{-1} ,$$

f_o - phase transfer link frequency.

3. PHASE VARIATION DYNAMICS

As it was mentioned during each $\mathcal{T}$ seconds multichannel cross-correlator forms one indication of R(T,F), that corresponds to the current OVLBI base vector B'(t), and OVLBI aquisition (phase transfer loop and phase errors compensation) must keep delay and frequency changing within above noted constrains. It means that except of phase noise following phase terms ought to be taken into account:

$$\Delta\varphi_{\oplus} = \Delta\varphi_{o\oplus} + \Delta\dot{\varphi}_{\oplus} t + \Delta\ddot{\varphi}_{\oplus} t^2$$

where $\Delta\varphi_{o\oplus}$ depends on delay uncertainty itself and determines required correlator delay "window"; $\Delta\dot{\varphi}_{\oplus}$ is the first order term and determines required frequency "window" (essentially, that one doppler part depends on received signal frequency but the other - from communication radio link frequency); $\Delta\ddot{\varphi}_{\oplus}$ is the second order term that determines achieved integration time $\mathcal{T}$.
For example, component includes phase uncertainties due to unpredicted parts of the spacecraft attitude, velocity and acceleration.
Phase dynamics is illustrated by Table:

	uncertainties T, mk	F, Hz	$\dot{T}$ mks/	$\dot{F}$ Hz/s
OVLBI	10	5	20	500
VLBI	0,1	10^{-4}	0,05	0,5

Table shows that data processing must provide searching for T-F windows and tracking at rates $\dot{T}$ and $\dot{F}$ what are 2-3 orders more that in ground-based VLBI case.
It needs to note that phase changing due to phase transfer radiolink is estimated for two-side propogation time: for RADIOASTRON-1 - up to 0,5 s. Phase noise is defined by signal to noise ratio in the phase transfer loop.

4. PHASE TRANSFER AND COMPENSATION SCHEME

RADIOASTRON uses (see Fig.2) two-way coherent radio link 7,2 GHz -up /8,47 GHz-down for transmission reference signal to SRT (carrier only) and return to tracking station and wide-band 15 GHz down link for radi o astronomical data transmis sion (QPSK, two streams simultanuously, 72 Mbit/s each) to russian and NASA TSs. All tracking stations and ground radiotelescopes are equipted by H-maser USOs. All space getherodyne , sampling and timing signals are formed coherently from received reference signal.
It is clear that the reference signal phase at the spacecraft φ'_o differs from transmitted phase φ_{oSC} due to unaccurate prediction of SC movement parameters, ionosphere and troposphere influence and other causes. Down link adds the similar phase changings, so the signals phases received by TS will be:

φ''_o **kn/p** - for phase loop carrier,

φ''_o**km/p** - for astronomical data,

here and on Fig.2 **k, n, m, p, l** - known ratios of frequencies used in RADIOASTRON-1 system.
Ground radiotelescope derives reference frequency directly from

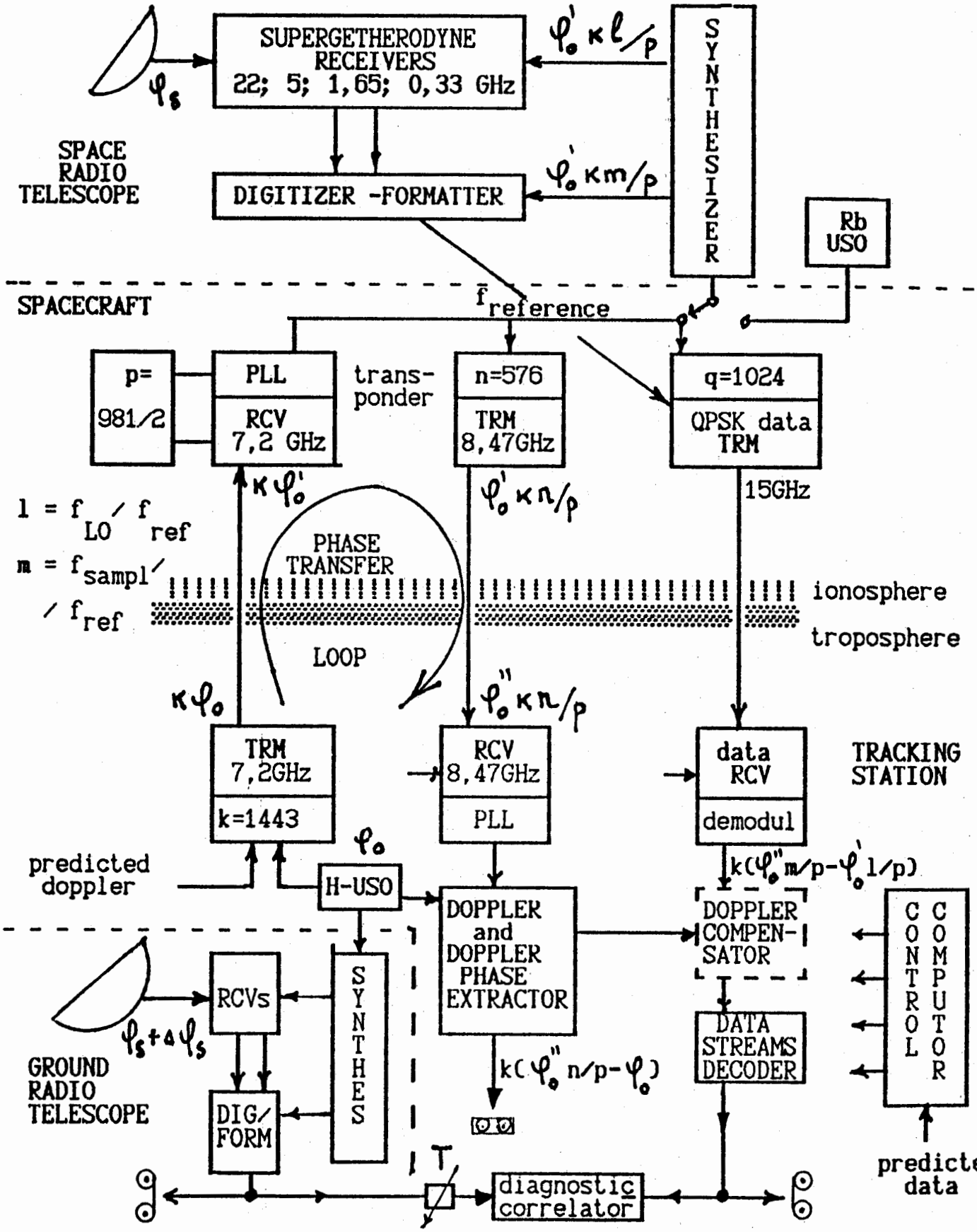


Fig.2 Radioastron synchronisation scheme.

the same H-maser as TS (if it is replaced near TS) with phase φ_o or from the similar independent H-maser.

Our task is to obtain compensation of the next phase differences:

$k(\varphi''_o n/p - \varphi_o)$ - to minimize 2-way phase influence in phase transfer loop, and

$k(\varphi''_o m/p - \varphi'_o l/p)$ - to minimize astronomical data phase shift (on sampling frequency) due to up-link and 2-way links.

This is achieved by the extracting doppler and doppler phase (versus time) of the received signals, comparison with transmitted phase and removal these phase differences.

Ideally, it would seem that data phase errors through space synchronisation can be excluded.

But really, this is not so: we do not know exactly troposphere and ionosphere behaviour in time, possible unmodeled SC acceleration, etc.

That is why, for example,

$$\varphi''_o \neq 2\,\varphi'_o ,$$

so residual, uncompensated phase remains. It is important as exactly as possible to provide tracking for delay and frequency changings.

Both procedures can be implemented during main correlation processing if phase differences noted above will be measured, recorded and used later for control of delay and phase rotator in processor together with analysis of useful (informatic) signal delay ΔT_s and fringe frequency ΔF_s.

But it is preferable to separate preliminary phase differences through 2-way radio links trip from one through radiosource position change relative to the interferometer base (see drop-line box, Fig.2).

Therefore, total phase uncertainty can be decreased before cross-correlation processing. In this case the radioastronomical data from SRT (together with corrected clock marks) would be recorded on the magnetic tape in the same form and manner as from any ground telescope.

CONCLUSION

Attentive calculation and estimate of all phase components behaviour have shown that:

-phase noise can be restricted 3-5 degrees rms for the longest distance (near SC apogee);

-residual frequency uncertainty due to, mainly, SC angular errors will not exceed 5 Hz for 22 GHz signal;

-residual freguency rate can be limited as so to keep frequency within 0,01...0,001 Hz bandwidth "filter" during 100...1000 s;

-residual delay uncertainty due to SRT orbit and the troposphere influence can achieve some mks but can be measured with 10 ns accuracy.

Some from OVLBI principles were practically confirmed in Space at 1986--88 by US-Australian-Japaneese specialists; they had used TDRSS satellite as small SRT and the communication radio links as "phase transfer" loop. But in that case TDRSS is geostationary satellite so its dynamics was poor and synchronisation conditions were more comfortable.

Therefore, difficult time --frequency synchronisation procedurs for OVLBI can be implemented by the integrated using of modern accurate timing equipment, coherent two-way space radio links and special two--dimention (time-frequency) processing of the fast changing data streams taking into account predicted spacecraft orbit and the atmosphere parameters.

QUARTZ CRYSTALS 1

Chairman: R. Besson

THE INFLUENCE OF DIMENSIONS ON THE RESONANT FREQUENCY TEMPERATURE DEPENDENCE OF AT-CUT QUARTZ STRIPS

Josef Pavlovec, Jan Sedláček, Josef Suchánek (1),
Jiří Zelenka (2)

(1) TESLA Hradec Králové AG, Okružní 1144, CS-50080 Hradec Králové,
(2) Technical University of Liberec, Electrical Engineering Department, Hálkova 6,
CS-46117 Liberec, Czech Republic.

ABSTRACT

The one-dimensional linear equations given by Lee and Wang (1992) are used for the description of small vibrations superposed on thermal deformations induced by steady uniform temperature change in quartz strips with monoclinic symetry. The obtained relations are used for the calculation of the resonant frequency temperature behaviour of the thickness-shear mode of vibration coupled with the thickness-length flexure mode of vibration for the different length to thickness ratio of the strip dimensions. The calculated frequency-temperature dependence is compared with the measured one.

INTRODUCTION

Results of the theoretical and experimental studies of the vibrations of AT-cut quartz strips has been published in many papers during the last decade and two starting points has been (Ref. 1) published a three dimensional mode-maching theory where the coupled thickness-shear, flexure and extensional vibrations in the direction of a width and thickness-shear, face-shear and flexure vibrations in the direction of a length of a strip were considered. The frequency-temperature performance and the electrical equivalent circuit parameters were derived for the partly electroded strips with the width parallel to x axis. By the derivation the guided modes of both plated and unplated infinite bars were used to find the trapped modes of waveguides containing alternate plated and unplated section of the strip. It follows from the analysis given in the mentioned paper that the thickness-shear and flexure vibration in the x axis direction are the dominant modes which have to be considered by the expression of the frequency teperature characteristics of the AT-cut strips.

Another method of the theoretical analysis of vibrations of AT-cut quartz strips has been applied by Lee and his cooperators. Lee, Syngellakis and Hou (Ref. 2) deduced first of all the two-dimensional equations of motion of succesively higher-order approximations for piezoelectric crystal plates with triclinic symmetry from the three-dimensional equations of linear piezoelectricity by expansion in series of trigonometric functions of the thickness coordinates of the plate. Then Lee and Wang (Ref. 3) derived a system of one-dimensional equations of motions for AT-cut quartz strips resonators with narrow width and for frequencies upto and including the fundamental thickness shear from the two-dimensional one. By neglecting the piezoelectric coupling and week mechanical coupling through c_{56}, four groups of coupled equations of motion they obtained for the strips with length parallel to x axis of the quartz. For the equations of each group, closed form solutions has been obtained and the traction-free conditions at four edges has been accomodated.

In the present paper the Lee's and Wang's solutions of the vibrations of AT-cut quartz strips of narrow width and finite length (Ref. 3) are used for the calculation of the resonant frequency temperature characteristics of the AT-cut x-length strips. The influence of the length to thickness dimensions ratio is calculated from the group of coupled thickness-shear and thickness length flexure vibrations. The frequency spectrum of the further three groups of coupled vibration considering coupled first twist-overtone of thickness shear and thickness-length flexure vibrations, coupled length extension, width-stretch and symmetric width-shear vibrations and coupled width-shear, width-length flexural and anti-symmetric width stretch vibrations has been used by finding the suitable width to length dimmensions ratio of the strip. The obtained results are compared with the measured values.

ONE-DIMENSIONAL EQUATIONS OF MOTION FOR SMALL AMPLITUDE WAVES SUPERPOSED ON HOMOGENIUS THERMAL STRAIN

The AT-cut quartz strip resonators shown with its coordinates and dimensions in Fig. 1 is considered in subsequent discussion. Similarly as in references (Refs. 4, 5, 6) let us express the influence of the thermally biased homogenius strain by means of terms β_{ki} and D_{ijkl}.

The term β_{ki} is given by the relation (Refs. 4, 5)

$$\beta_{ki} = \delta_{ki} + \alpha_{ki}, \tag{1}$$

where δ_{ki} is a *Kronecker delta*,

$$\alpha^{\Theta}_{ki} = \alpha^{(1)}_{ki} + \alpha^{(2)}_{ki} + \alpha^{(3)}_{ki}, \tag{2}$$

$\alpha^{(n)}_{ki}$ are n-th order linear thermal expansion coefficients (measured by Bechmann, Ballato and Lukaszek (Ref. 7) and corrected by Kosinski, Gualtieri and Ballato (Ref. 8)) and Θ is the temperature change, $\Theta = T - T_0$.

The tensors α_{ki} and β_{ki} are predominantly a diagonal tensors with the off-digonal terms smaller by a magnitude of 10^{-6}. Therefore, the off-digonal terms are neglected ($\beta_{ki} = 0$ for $k = i$).

$$\beta_k = \beta_{kk} = 1 + \alpha^{\Theta}_{kk}. \tag{3}$$

The term D_{ijkl} is given by the relation

$$D_{ijkl} = C_{ijkl} + D^{(1)}_{ijkl}\Theta + D^{(2)}_{ijkl}\Theta^2 + D^{(3)}_{ijkl}\Theta^3, \tag{4}$$

where

$$D^{(q)}_{ijkl} = \frac{1}{q!}C^{(q)}_{ijkl} + C_{ijklmn}, \tag{5}$$

C_{ijkl} and C_{ijklmn} are the second and third order elastic stiffness of quartz, while $C^{(1)}_{ijkl}$, $C^{(2)}_{ijkl}$ and $C^{(3)}_{ijkl}$ are respectively the first temperature derivatives, second temperature effective derivatives and third effective temperature derivatives. Values of the temperature derivatives were calculated and re-

ported in reference (Refs. 5, 6). The magnitudes of D_{ijkl} where reported by Yong and Lee (Ref. 5).

If the temperature dependent material properties D_{ijkl} and β_i are includ to the coupled one-dimensionsnal equations of motion, dispersion and stress-displacement relations, frequency equations and traction free ends of strip conditions derived by Lee and Wang (Ref. 3) we obtain the following constitutive equations for the coupled thickness-shear and thickness-length flexure vibrations:

From the displacement equations of motion

$$\tilde{c}_{11}\beta_1^2 u_{1,11}^{(10)} - \frac{2\alpha_1}{b}c_{66}(\beta_1\beta_2 u_{2,1}^{(00)} + \frac{\alpha_1}{b}\beta_1^2 u_1^{(10)}) = \varrho(1+2R)\ddot{u}_1^{(10)}$$

$$c_{66}(\beta_2^2 u_{2,11}^{(00)} + \frac{\alpha_1}{b}\beta_1\beta_2 u_{1,1}^{(10)}) = \varrho(1+R)\ddot{u}_2^{(00)} \qquad (6)$$

the dispersion relation

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = 0 \qquad (7)$$

where

$$\begin{aligned} a_{11} &= \beta_2^2 c_{66}^* \bar{\xi}_r^2 - (1+R)(\frac{\Omega}{\alpha_2})^2 \\ a_{12} &= \frac{1}{\sqrt{2}}\beta_1\beta_2 c_{66}^* \bar{\xi}_r \\ a_{21} &= \sqrt{2}c_{66}^*\beta_1\beta_2\bar{\xi}_r \\ a_{22} &= \beta_1^2(c_{11}^*\bar{\xi}_r^2 + c_{66}^*) - (1+2R)\Omega^2 \end{aligned} \qquad (8)$$

can be obtained if the modes of the vibrations to have with form

$$\begin{aligned} u_1^{(10)} &= \sum_{r=1}^{2} A_{1r}\cos\xi_r x_1 e^{i\omega t} \\ u_2^{(00)} &= \sum_{r=1}^{2} A_{2r}\sin\xi_r x_1 e^{i\omega t}. \end{aligned} \qquad (9)$$

Substitution of the stress-displacement relations

$$\begin{aligned} T_1^{(10)} &= 2\tilde{c}_{11}\beta_1 u_{1,1}^{(10)} \\ T_6^{(00)} &= 4c_{66}(\beta_2 u_{2,1}^{(00)} + \frac{\alpha_1}{b}\beta_1 u_1^{(10)}) \end{aligned} \qquad (10)$$

to the boundary conditions

$$\beta_1 T_1^{(10)} = 0 \text{ and } \beta_2 T_6^{(00)} = 0 \text{ at } x_1 = \pm a \qquad (11)$$

results in

$$\sum_{r=1}^{2}\gamma_r A_{2r}\bar{\xi}_1\sin(\xi_r a) = 0$$

$$\sum_{r=1}^{2}(\frac{1}{\sqrt{2}}\beta_1\beta_2\gamma_r + \beta_2^2\bar{\xi}_r)A_{2r}\cos(\xi_r a) = 0, \qquad (12)$$

where

$$\gamma_r = \frac{A_{1r}}{A_{2r}}, \quad r = 1,2. \qquad (13)$$

are the amplitude ratios which can be computed from (7).

The vanishing of the determinant of the coefficients matrix of (12) gives the frequency equation which must be solved in conjunction with the dispersion relation (7).

The elastic stiffnesses used in equations given above depend on the temperature and are defined by the relations

$$\begin{aligned} \bar{c}_{pq} &= D_{ppqq} - \frac{D_{pp22}D_{22qq}}{D_{2222}}, \quad p,q = 1,3 \\ \bar{c}_{p4} &= D_{pp23} - \frac{D_{pp22}D_{2223}}{D_{2222}}, \quad p = 1,3 \\ \bar{c}_{4q} &= D_{23qq} - \frac{D_{2322}D_{22qq}}{D_{2222}}, \quad q = 1,3 \\ \bar{c}_{44} &= D_{2323} - \frac{D_{2322}D_{2223}}{D_{2222}} \\ \hat{c}_{pq} &= \bar{c}_{pq} - \frac{\bar{c}_{p4}\bar{c}_{4q}}{\bar{c}_{44}}, \quad p,q = 1,3 \\ \tilde{c}_{11} &= \hat{c}_{11} - \frac{\hat{c}_{13}\hat{c}_{31}}{\hat{c}_{33}} \\ c_{11}^* &= \frac{\tilde{c}_{11}}{c_{66}}, \quad c_{66}^* = \frac{D_{1212}}{c_{66}} = 1, \end{aligned} \qquad (14)$$

where

$$c_{66} = [D_{1212}]_{\Theta=0}.$$

FREQUENCY-TEMPERATURE BEHAVIOUR OF THICKNESS-SHEAR VIBRATIONS OF AT-CUT QUARTZ STRIPS

Lee and Wang (Ref. 3) derived one-dimensional frequency equations for the four groups of coupled modes in AT-cut quartz strips. The coupled thickness-shear and thickness-length flexure vibrations, the coupled first twist-overtone of thickness-shear and thickness-length flexure vibrations, the coupled length-extensional, width-stretch and symmetric width-shear vibrations and the coupled width-shear, width-length flexural and antisymmetric width-stretch vibrations are include to this four groups.

The frequency spectrum of the mentioned mode of vibrations calculated from Lee's and Wang's frequency equations for the AT-cut strip resonator with width to thickness dimension ratio $c/b = 4.195$ is given in Fig. 2 as a function of the length to thickness dimension ratio l/b of the strip. The resonant frequency of the thickness-shear vibration is the main resonance frequency of the strips utilizing in the electronic devices. It is clear from Fig 2 that convenient dimension ratios l/b from the point of view ot thickness-shear vibration are approximately 10.7, 12.4, 14.05, 15.7, 17.25 and 19.0

The calculated resonant frequency temperature dependence of the thickness-shear mode of vibration is given in Fig. 3 and 4 for the dimension ratios of the strip mentioned above and the orientation angle of AT-cut is $\theta = 35°15'$ and $\theta = 35°12'$ respectively. It is clear from Fig. 3 and 4 that for the shorter strip the influence of the coupling between the thickness-length flexure and thickness-shear vibrations on the temperature behaviour can be compensated by a change of the AT-cut orientation angle θ. The measured resonant frequency temperature dependence of the AT-cut quartz strip with the dimensions $l = 8.00$ mm, $b = 0.41$ mm and $c = 1.73$ mm and the orientation angle $\theta = 35°10'$ is given in Fig. 5.

REFERENCES

[1] R.F. Milsom, D.F. Elliott, S. Terry-Wood and M.Redwood. "Analysis and design of coupled-mode miniature bar resonators and monolithic filters". *IEEE Trans. Son. Ultrason.*, 30, pp. 140-155, 1983.

[2] P.C.Y. Lee, S. Syngellakis and J.P. Hou. "A two-dimensional theory for high frequency vibrations of piezoelectric crystal plates with or without electrodes". *J. Appl. Phys.*, 61(4), pp. 1249-1262, 1987.

[3] P.C.Y. Lee and Ji. Wang. "Vibrations of AT-cut quartz strips of narrow width and finite length". *Proc. 46th Ann. Freq. Control Symposium*, 1992, pp.

[4] P.C.Y. Lee and Y.K. Yong. "Temperature derivatives of elastic stiffness derived from the frequency-temperature behavior of quartz plates". *J. Appl. Phys.*, 56, pp. 1514-1521, 1984.

[5] Y.K. Yong and P.C.Y. Lee. "Frequency temperature behavior of flexural and thickness-shear vibrations of rectangular rotated Y-cut quartz plates". *Proc. 39th Ann. Freq. Control Symposium*, 1985, pp. 415-426.

[6] P.C.Y. Lee and Y.K. Yong. "Frequency-temperature behavior of thickness vibration of doubly rotated quartz plates affected by plate dimensions and orientations". *J. Appl. Phys.*, 60, pp. 2327-2342.

[7] R. Bechmann, A.D. Ballato and T.J. Lukaszek. "Higher-order temperature coefficients of the elastic stiffnesses and compliances of alpha-quartz". *Proc.IRE*, 50, pp. 1812-1816, 1962.

[8] J.A. Kosinski, J.G. Gualtieri and A. Ballato. "Thermoelastic coefficients of alpha quartz". *IEEE Trans. Son. Ultrason. Freq. Control*, 39, pp. 502-507.

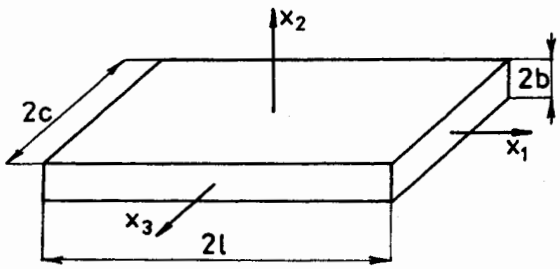


Fig. 1. An X-length AT-cut quartz strip resonator.

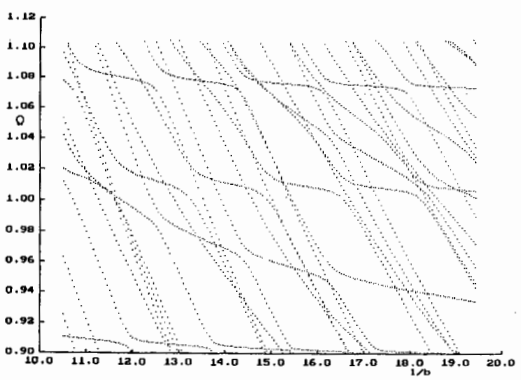


Fig. 2. Predict Ω vs. l/b for AT-cut quartz strip with $c/b = 4.195$.

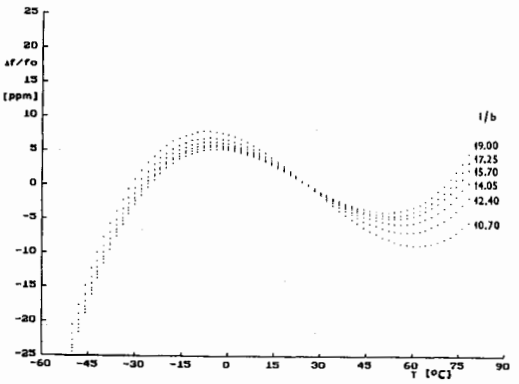


Fig. 3. Predict thickness-shear frequency-temperature curves for AT-cut quartz strips for l/b ratio variable, $c/b = 4.195$ and orientation angle $\theta = 35°15'$.

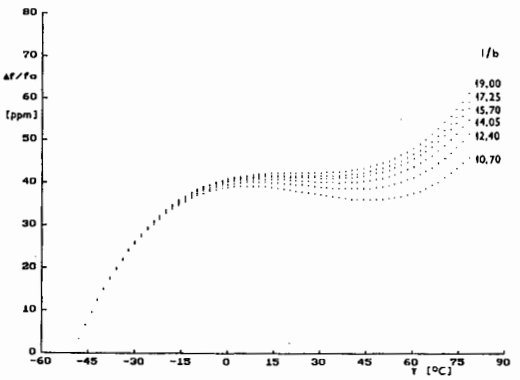


Fig. 4. Predict thickness-shear frequency-temperature curves for AT-cut quartz strips for l/b ratio variable, $c/b = 4.195$ and orientation angle $\theta = 35°12'$.

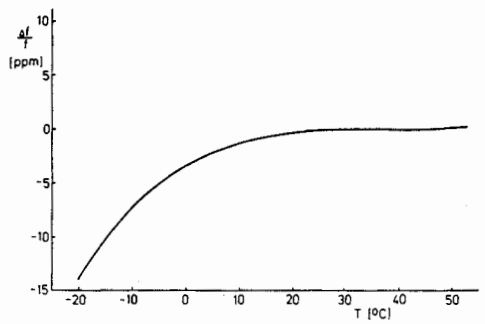


Fig. 5. Measured thickness-shear frequency-temperature curve for AT-cut quartz strip with the dimensions $l = 8.00$ mm, $b = 0.41$ mm and $c = 1.73$ mm and the orientation angle $\theta = 35°10'$.

Finite Element Analysis of AT Strip Quartz Resonators

K. Trümpy, W. Zingg

ETA SA , Micro Crystal ; CH-2540 Grenchen Switzerland

SUMMARY

In narrow rectangular AT strip quartz resonators we observe a mechanical coupling between the fundamental thickness-shear mode and flexural and other modes which mainly determines the performance of the device. This coupling can be influenced by the geometric properties of the resonator. The values of motional capacitance, the Q-factor and the absence of activity dips depend on a suitable choice of the resonator shape.

In order to better understand these perturbing phenomenas and to predict the device's behaviour, we analysed the AT strip resonator with a three-dimensional finite element (FE) model. The required high degree of discretization is obtained by the use of prismatic elements with cubic interpolation. The FE method allows to predict the predominant coupling to flexural modes as well as the weaker coupling to in-plane modes. We present measured and calculated values of frequencies and motional capacitances at variable temperature as a function of width/thickness ratio.

KEYWORDS: AT strip quartz resonator, Three-dimensional FE-analysis, chemically milled resonators, coupling to width flexure mode, activity dips.

INTRODUCTION

The modelled miniature resonators are of rectangular shape with length along z'-axis and width parallet to x-axis. In order to validate the FE approach, a systematical experimental study was necessary. AT strip resonators were manufactured on wafers with multiple crystal designs [1]. Each wafer contains 64 crystals with eight different values of width. At the end of the photolithographic and chemical milling process, electrical parameters of all crystals are measured on wafer. In addition to electrical parameters, frequency and resistance response with temperature are measured on selected resonators mounted into small ceramic packages.

The FE-results obtained by purely mechanical or piezoelectric calculations yield resonant frequencies and motional capacitance values at different temperatures which are plotted as a function of width/thickness ratio. Cases of different thicknesses of the rectangular electrodes (Cr/Au plating) have beeen considered. Mass loading, elastic and electric effects of the electrodes are taken into account. The FE-analysis also allows to predict the frequencies and motional capacitances of several anharmonic modes and mode splitting due to coupling of the thickness-shear mode with width-flexural as well as length-flexural, face-extension, face-flexural, face-shear and thickness-extension modes. A close agreement is obtained between the experimental values, determined within a certain manufacturing tolerance, and the FE-computations, provided that a sufficiently fine element partitioning is accomplished. We believe that the FE-program will prove helpful for the design optimization of miniature AT-strip quartz resonators.

In our FE-model we used subparametric prismatic elements with conforming cubic Hermite interpolation scheme. Compared to Lagrange interpolation, Hermite interpolation leads, at the same discretization level, to a considerably smaller number of degrees of freedom for the complete crystal structure. The eigenmodes are calculated with a subspace-iteration method with overrelaxation accelerated convergence. Although Lanczos algorithm is now widely used, we think that our subspace iteration method is efficient and very reliable in calculating solutions of large eigenvalue-equations with up to 48000 degrees of freedom.

THREE-DIMENSIONAL FE-ANALYSIS

Finite element method is a powerfull tool in analysing the physical properties of piezoelectric resonators. Two-dimensional thick plate theory (Reissner-Mindlin) allows to predict the predominant coupling of thickness shear mode to flexural modes occuring in strip resonators, whereas with a fully three-dimensional model interaction of the main mode with any perturbing vibration ca be analysed. Thickness shear deformation of plates may be described precisely enough by a polynomial expansion of third degree. We subdivide the quartz structure into a one-layer FE mesh with subparametric prismatic elements of conforming cubic Hermite interpolation. This interpolation scheme takes into account displacement gradients as nodal values in addition to displacement values. All values are concentrated at the element corners, however in Lagrange interpolation, we also have nodal displacement values located at the element edges. Hermite interpolation leads to a much lower number of degrees of freedom as compared to Lagrange interpolation because corner nodes are common to as many as four elements (one layer mesh), whereas edge middle nodes are common to at most two elements.

We also used elements with quartic Hermite interpolation, in order to take into account components of order four of thichness shear deformation. Quartic Hermite interpolation requires also edge middle, and even worse, side middle nodal values of displacement, leading to the unfavourable situation of a very large FE global stiffness matrix. Components of order four and above are essentially not excited in the vincinity of thichness shear vibration. We only give values obtained with cubic interpolation.

Piezoelectric models consider, besides the three displacement components, also the value of electric potential. This fourth component, required to describe static electric effects in quartz resonators, leads to computation times up to twice as high as thouse for a purely mechanical model. Since electromechanical coupling is very weak in quartz, the results obtained with the complete model differ, at least what concerns the strong coupling to width flexural modes, only little from values obtained by the purely mechanical model.

Except for the small electrode connections, the crystal resonator is

completely length symmetric and also for AT-cut the elastic, piezoelectric and dielectric constants are invariant against reflexion on the z'-y' plane. Thus the spectrum devides into antisymmetric and symmetric modes. Modelling only half of the crystal structure, with boundary conditions of antisymmetry for nodes located on the symmetric axis, gives the modes of interest.
The conforming interpolation yields descending frequency values with increasing element partitioning.

A. General FE-formulation

The geometrically isotropic cubic interpolation function contains, besides the 20 terms of the complete cubic polynom in three dimensions, 12 additional terms of higher degree which guarantee continuity of function values between adjacent element sides and edges [2] (conforming elements). Element coordinate interpolation is trilinear i.e. the shape of the undeformed element is an arbitrarily twisted eight node prism with straight edges (subparametric element). The three components of displacement u, v, w along the FE-coordinates x,y,z and the electrical potential Φ are obtained from the following interpolation relation

$$\begin{pmatrix} u \\ v \\ w \end{pmatrix} = \begin{bmatrix} \vec{N}^T & 0 & 0 \\ 0 & \vec{N}^T & 0 \\ 0 & 0 & \vec{N}^T \end{bmatrix} \begin{pmatrix} \vec{u} \\ \vec{v} \\ \vec{w} \end{pmatrix} = \mathbf{N}\,\vec{s}\ ; \quad \Phi = \vec{N}^T \vec{p} \tag{1}$$

where $\vec{N}(x,y,z)$ = vector of cubic interpolation functions,
$\vec{s} = (\vec{u}^T, \vec{v}^T, \vec{w}^T)^T$ = vector of element nodal displacements,
$\vec{p}$ = vector of element nodal electrical potential.

The mechanical strain $\vec{\varepsilon}$ and the electric field $\vec{E}$ are related to the nodal values through the derivatives of the interpolation functions

$$\vec{\varepsilon} = \mathbf{B}_S\,\vec{s}\ , \qquad \vec{E} = -\mathbf{B}_\Phi\,\vec{p} \tag{2}$$

$$\mathbf{B}_S = \begin{bmatrix} \vec{N}_x & 0 & 0 & 0 & \vec{N}_z & \vec{N}_y \\ 0 & \vec{N}_y & 0 & \vec{N}_z & 0 & \vec{N}_x \\ 0 & 0 & \vec{N}_z & \vec{N}_y & \vec{N}_x & 0 \end{bmatrix}^T ; \quad \mathbf{B}_\Phi = \begin{bmatrix} \vec{N}_x^T \\ \vec{N}_y^T \\ \vec{N}_z^T \end{bmatrix}$$

with $\vec{N}_x = \partial\vec{N}/\partial x$, $\vec{N}_y = \partial\vec{N}/\partial y$, $\vec{N}_z = \partial\vec{N}/\partial z$.
We have the following linear constitutive expression for piezoelectrical materials

$$\begin{pmatrix} \vec{\sigma} \\ \vec{D} \end{pmatrix} = \begin{pmatrix} \mathbf{C} & \mathbf{P} \\ \mathbf{P}^T & -\mathbf{D} \end{pmatrix} \begin{pmatrix} \vec{\varepsilon} \\ -\vec{E} \end{pmatrix} \tag{3}$$

$\vec{\varepsilon}$ = 6x1 mechanical strain vector, $\vec{\sigma}$ = 6x1 mechanical stress vector,
$\vec{E}$ = 3x1 electric field vector, $\vec{D}$ = 3x1 electric displacement vector
$\mathbf{C}$ = 6x6 elastic stiffness matrix, $\mathbf{P}$ = 6x3 piezoelectric matrix,
$\mathbf{D}$ = 3x3 dielectric matrix.
The matrices $\mathbf{C}, \mathbf{P}, \mathbf{D}$ are rotated by a fourth, third or second rank tensor transformation corresponding to the AT-cut angles.
The equation of motion for a piezoelectric body can be derived from a variational principle [3]. In the case of undamped free vibrations we obtain an expression for the natural frequencies ω

$$\left[\begin{pmatrix} \mathbf{K}_{SS} & \mathbf{K}_{S\Phi} \\ \mathbf{K}_{S\Phi}^T & \mathbf{K}_{\Phi\Phi} \end{pmatrix} - \omega^2 \begin{pmatrix} \mathbf{M}_{SS} & 0 \\ 0 & 0 \end{pmatrix} \right] \begin{pmatrix} \vec{S} \\ \vec{\Phi} \end{pmatrix} = \begin{pmatrix} \vec{F} \\ \vec{Q} \end{pmatrix} \ (= \vec{0}\,) \tag{4}$$

with the following global quantities :
$\vec{S}$ = vector of displacements, $\vec{\Phi}$ = vector of electric potentials,
$\vec{F}$ = vector of forces, $\vec{Q}$ = vector of electric charges,
$\mathbf{K}_{SS}$ = elastic stiffness matrix, $\mathbf{K}_{S\Phi}$ = piezoel. coupling matrix,
$\mathbf{K}_{\Phi\Phi}$ = dielectric stiffness matrix, $\mathbf{M}_{SS}$ = mass matrix.

The global matrices $\mathbf{K}$ and $\mathbf{M}$ are derived from the assembly of all element matrices $\mathbf{k}$ and $\mathbf{m}$ (ϱ = mass density)

$$\mathbf{k}_{SS} = \int \mathbf{B}_S^T\,\mathbf{C}\,\mathbf{B}_S\,dV \ ; \qquad \mathbf{k}_{S\Phi} = \int \mathbf{B}_S^T\,\mathbf{P}\,\mathbf{B}_\Phi\,dV$$
$$\mathbf{k}_{\Phi\Phi} = -\int \mathbf{B}_\Phi^T\,\mathbf{D}\,\mathbf{B}_\Phi\,dV \ ; \qquad \mathbf{m}_{SS} = \varrho \int \mathbf{N}^T\,\mathbf{N}\,dV \tag{5}$$

All components of vectors $\vec{F}$ and $\vec{Q}$ are zero for free node variables for which eq.(4) represents a general eigenvalue problem with eigenvalue ω^2 and eigenmodes ($\vec{S}^T, \vec{\Phi}^T$). It is possible to statically condense out all unspecified electric potential values. Unfortunately this procedure completly destroys the sparse structure of the stiffness matrix, thus rendering it totally unpracticable for large eigenvalue problems. Either one has to take into account piezoelectric stiffening with a perturbation method [4], or we have to directely solve eq.(4) with a subspace iteration method. We obtained best results with a Ritz orthogonalisation of the iterated vectors where only the projection of mass matrix is taken into account [2, (1st, 2nd ed.)] For the acceleration of convergence we used an overrelaxation method [5].In a purely mechanical model, we only consider the displacement vector, the elastic stiffness and mass matrices.

B. Electrode effects

The element mass matrix $\mathbf{m}_{SS} = \mathbf{m}_Q + \mathbf{m}_E$ contains besides the quartz structure term $\mathbf{m}_Q$ the electrode mass loading terms $\mathbf{m}_E$.

$$\mathbf{m}_E = \varrho_E\, d \int \mathbf{N}^T\,\mathbf{N}\ (x, y,\ z = \pm\,(t+d)/2\ )\ dxdy \tag{6}$$

ϱ_E = electrode mass density, d = thickness of the electrodes in the x-y plane, t = thickness of the quartz structure.
Analogous for stiffness we have the sum $\mathbf{k}_{SS} = \mathbf{k}_Q + \mathbf{k}_E$ of the quartz structure term $\mathbf{k}_Q$ and the electrode elastic terms $\mathbf{k}_E$.

$$\mathbf{k}_E = d \int \mathbf{B}_S^T\,\mathbf{C}_E\,\mathbf{B}_S\ (x, y,\ z = \pm\,(t+d)/2\ )\ dxdy \tag{7}$$

with the membrane plane stress condition isotropic stress-strain relation

$$\mathbf{C}_E = \frac{E}{1-\nu^2} \begin{bmatrix} 1 & \nu & 0 & 0 & 0 & 0 \\ \nu & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \gamma \end{bmatrix} ; \quad \gamma = (1-\nu)/2$$

E = Young's modulus, ν = Poisson's ratio.
In our program we utilize a formulation able to treat also the case of arbitrarely oriented electrodes [6].
The electrode electrical conductivity maintains constant electric potential of all nodes located inside the electrode area. The resonance frequencies are obtained by setting on the same level the potential of both electrodes (short-circuiting). Antiresonance frequencies are obtained by an unspecified potential difference between the electrodes (open electrodes).
The main electrode effect for thickness shear modes results from the mass loading term. The discontinuity of electric potential at the electrode borders can not be treated with the FE-model.

C. Determination of motional capacitance

The total kinetic energy T of the vibrating structure is $T = \frac{1}{2}\,\omega^2\,\vec{S}^T\mathbf{M}_{SS}\,\vec{S}$. We equate this energy to the electromagnetic energy stored in an equivalent inductance L_m created by an electric current $I = \omega\,Q$. The electric charge Q is obtained by summing the motional electric charges of all nodal points located on the grounded electrodes. In the resonance case with $\omega^2 = 1/L_mC_m$ an expression for the motional capacitance C_m results

$$T = \tfrac{1}{2}\,L_m\,I^2 = \tfrac{1}{2}\,L_m\,\omega^2\,Q^2 = \tfrac{1}{2}\,Q^2/C_m\ , \qquad C_m = \tfrac{1}{2}\,Q^2/T \tag{8}$$

In the purely mechanical model we calculate the electric displacement $\vec{D}$ by neglecting the small term $\mathbf{D}\vec{E}$ (eq.3), thus $\vec{D} = \mathbf{P}^T\,\vec{\varepsilon}$. The component normal to the electrodes $D_n = \mathbf{P}_n^T\,\vec{\varepsilon}$ induces a charge densitiy. Neglecting electrode boundary effects we have

$$Q = \int_{A_E} D_n\,dA = \frac{1}{h} \int_{V_E} \mathbf{P}_n^T\,\vec{\varepsilon}\,dV$$

where A_E = electrode surface, V_E = volume enclosed by the electrodes and h = electrode distance.

D. Effects of temperature

The properties of quartz resonators are influenced by the temperature T. The elastic coefficients c_{ij} of matrix **C** in eq.(3) vary with $\Delta T = T - 25°C$ through the temperature coefficients $Tc_{ij}^{(k)}$

$$c_{ij}(T) = c_{ij}(25°C)\,(1 + Tc_{ij}^{(1)}\Delta T + Tc_{ij}^{(2)}\Delta T^2 + Tc_{ij}^{(3)}\Delta T^3)$$

Similar relations are valid for the coefficients of dielectric and piezoelectric matrices.

By expansion the mass densitiy ϱ and the resonator shape change with temperature

$$\varrho(T) = \varrho(25°C)/((1 + d_{11}(T))(1 + d_{22}(T))(1 + d_{33}(T)))$$

with the principal expansion coefficients

$$d_{ii}(T) = a_{ii}^{(1)}\Delta T + a_{ii}^{(2)}\Delta T^2 + a_{ii}^{(3)}\Delta T^3$$

forming the diagonal expansion matrix **B**. By a second-rank tensor transformation **B** has to be rotated according to the cutting angles of the quartz resonator to $\mathbf{B}' = [b'_{ij}]$. We have for the coordinates $\vec{K}$ of each node

$$\vec{K}(T) = \vec{K}(25°C) + \mathbf{B}'(T)\,\vec{K}(25°C)$$

The mass density of thin electrodes parallel to the x-y plane is

$$\varrho_E(T) = \varrho_E(25°C)(1 - b'_{11} - b'_{22}).$$

For every temperature value we have to carry out a complete FE-analysis but normaly with just one eigenmode, assuming that we know the frequency value at 25°C.

EXPERIMENTAL RESULTS

The simplified or ideal geometry on an AT-strip resonator is shown in figure1. For 16.8 MHz crystals, the plate thickness t is typically 95 micrometers. In this study, the length in z'-direction was kept constant (5.8 mm) whereas the width w varied between 1.4 and 1.85 mm. In terms of width/thickness ratio, this corresponds to the range from w/t = 14.7 to 19.4. While the electrode size was constant (3mm x 1.35 mm), its thickness was either 200 nm or 400 nm of gold over 15 nm of chromium. In order to obtain frequencies in the same range for both plating values, the wafer nominal thicknesses w were 96 and 93.5 microns respectively. The AT-cutting angle θ was 35.2°.

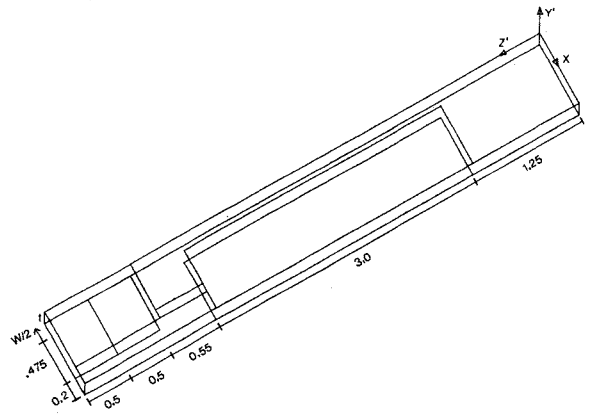


Figure 1. Geometric layout of one half of the resonator structure.

AT-strip resonators were milled chemically using multidesign masks [7],[8] which yield very accurate differences of width between crystals with different designs on the same wafer. Three sets of masks were used each with eight width parameters. The typical width step amounts to 20 or 25 microns (0.21 to 0.25 in terms of w/t ratio).

Width measurements of resonators on wafer were done with an accuracy better than 5 microns and wafer thickness was controlled to +/- 0.3 micrometer by electro-mechanical gauge measurements at five points (corners and center). Since wafers differ in thickness (variation typically half a micron), all measured frequencies were corrected by a multiplicative factor equal to the ratio between the mean thickness of each wafer and the nominal thichness value (96 and 93.5 microns respectively).

A schematic representation of the x-y' resonator cross section, as it is obtained by the chemical milling process, is given in figure 2. The etching facets are not considered in the measured width values.

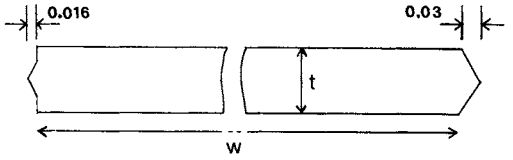


Figure 2. Cross section (x-y') resulting from chemical milling process.

In figures 3 and 4, the corrected frequencies are plotted against width/thickness ratio w/t together with motional capacitance Cm with standard mean deviation σCm, resistance R (f, Cm, σCm and R at 25°C) and the linear frequency-temperature coefficient α (eq.9). Each flexure branch Fm is identified by the number m of half wavelengths of flexure across the resonator width [9].

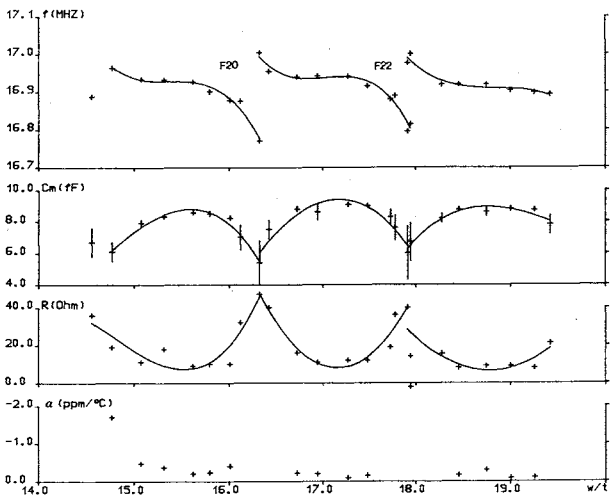


Figure 3. Measured values f, Cm, σCm, R and α versus w/t. t nominal = 96.0 microns, 200 nm plating.

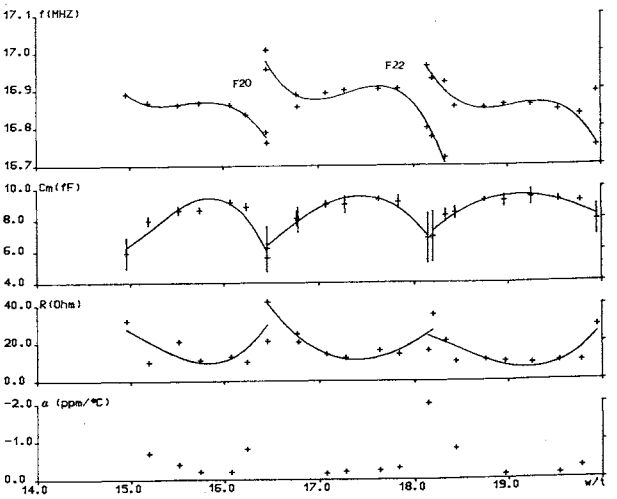


Figure 4. Measured values f, Cm, σCm, R and α versus w/t. t nominal = 93.5 microns, 400 nm plating.

The striking feature of these figures is the pseudo-periodicity of f,

Cm, R and a . with respect to w/t. A comparison between both values of electrode thickness (200 and 400 nm) shows a shift of the critical ratios w/t where transition between adjacent flexure branches occurs. This shift amounts to 0.3 as confirmed in figure 5, where the same wafers (t = 93.5 microns) have been metallized consecutively at 200 nm and 400 nm of gold. The normalized frequency values of figure 5 are ratios of measured frequencies f to the frequency $\overline{fo}$ of an infinitely extended metallized plate

$$\overline{fo} = \frac{1655(MHz\ \mu m)}{93.5(1+R)} \quad ; \quad R = \frac{2\ el.thickn.}{t}\ \frac{\varrho_{PLATING}}{\varrho_{QUARTZ}}$$

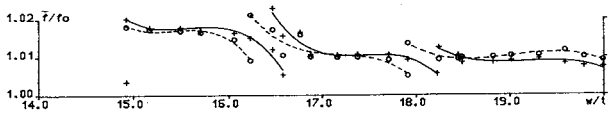


Figure 5. Normalized frequencies $f/\overline{fo}$, - - -200 nm, —400 nm. plating, t = 93.5 microns.

All drawn curves represent best fit approximations of a polynom of third degree.

NUMERICAL RESULTS

The FE-model should be able to predict the w/t ratio locations where maximum coupling between thickness shear mode and width flexural modes occur. Compared to a subdivision of resonator half width into 15 elements, the location near w/t = 15 moved upwards 0.06 and 0.2 for a element subdivision of 13 and 11 respectively. Our computations had to be carried out on a VAX 4000 90 workstation, constraining us to the following half width element discretization : w < 1.43mm 15 elem., w < 1.53 mm 16 elem., w < 1.63 17 elem., w < 1.73 mm 18 elem. and 19 elements above. The resonator length dimension is subdivided into 104 elements.

Nodes placed on the lower resonator face, with distances x from the symmetric axis greater or equal 0.2 mm and z' values of less or equal 0.5 mm from the left small resonator side (fig.1), are mechanically and electrically clamped. This surface corresponds to one of the mounting areas.

Material parameters and temperature coefficients of quartz are taken from [10]. Isotropic elastic constants and mass density of gold plating are from [11].

We usually calculated for each resonator 20 eigenmodes near a specified frequency shift value. (typically 17 hours computation time for 19x104 mech. elements with 47800 d.o.f.) The main thickness shear mode can easily be identified by it's large motional capacitance value, whereas anharmonic thickness shear overtones have to be identified by their characteristic variation over thickness of the integral values of inplane displacements (or simply by plotting the modes). The results computed with the purely mechanical FE-model, represented in figures 6 and 7 were obtained by taking into account only the main part of the electrodes (fig.1). Considering the small electrode connections produces a frequency change of only some 100 Hz. The experimental values from figures 3 and 4 are slightly shifted to higher w/t values, corresponding to a correction of width of 0.015 mm, as suggested by figure 2.

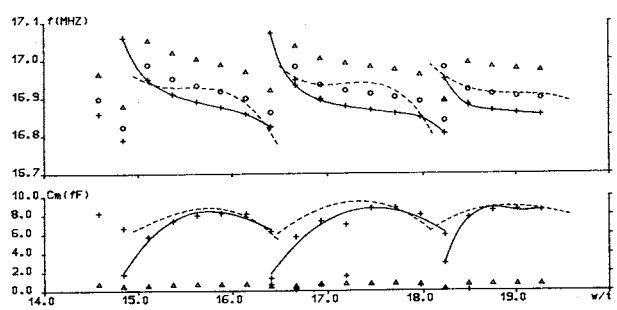


Figure 6. Computed values f and Cm versus w/t. (mechanical model)
t = 96 microns, 200 nm plating.
+ : fundamental thickness shear mode
o : first anharmonic thickness shear overtone
Δ : second anharmonic thickness shear overtone
- - -experimental values versus w'/t.

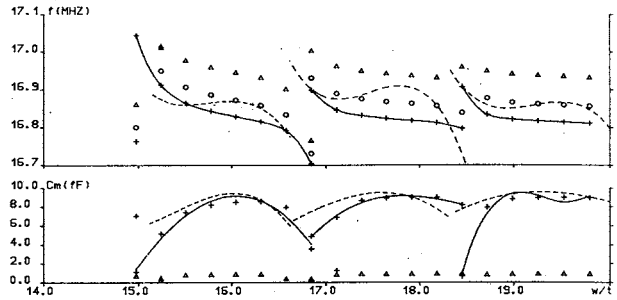


Figure 7. Computed values f and Cm versus w/t. (mechanical model)
t = 93.5 microns, 400 nm plating.
+ : fundamental thickness shear mode
o : first anharmonic thickness shear overtone
Δ : second anharmonic thickness shear overtone
- - -experimental values versus w'/t.

Results obtained with a complete piezoelectric FE-model are represented in figures 8 and 9, where the complete electrode structure has been modelled. Because of the limited size of virtual memory available on our workstation (320 MB), we could only compute values for resonators of widths up to 1.525 mm. Frequencies of the piezoelectric model are about 10 kHz higher than those of the mechanical model.

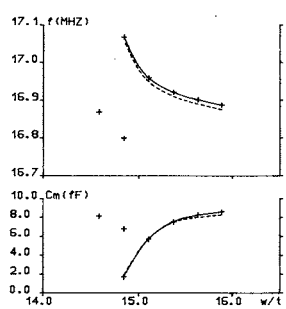


Fig.8. t = 96 μm, 200 nm plating

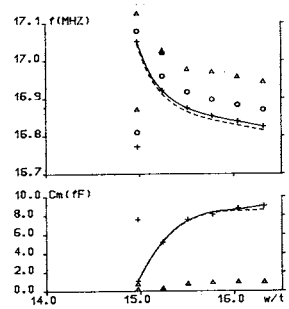


Fig. 9. t = 93.5 μm, 400 nm plating

Computed values f and Cm versus w/t. (piezoelectric model),
+ : fundamental thickness shear mode
o : first anharmonic thickness shear overtone
Δ : second anharmonic thickness shear overtone
- - -values of figures 6 and 7 respectively.

The fully three-dimensional FE-model allows to analyse coupling of the main mode with any perturbing (spurious) mode. For this purpose we determined resonator characteristics in function of varying temperature. Besides the coefficients a, γ, T_{INF} of the frequency-temperature expression

$$f(T) = f(25°C)\ (\ 1 + a\ (T - T_{INF}) + \gamma\ (T - T_{INF})^3\ ) \qquad (9)$$

we also obtain the interesting information, whether the motional capacitance values are fairly constant with temperature or large variations occur. This latter case means that a spurious mode interacts with the main mode (activity dip).

In figures 10 and 11 are represented the variations δCm (calculated form the standard mean deviation formula) and the linear frequency-temperature coefficient a versus w/t ratios (modelled inclusive

electrode connections). Resonator properties were calculated at the temperatures 0°, 25°, 50°, 75° and 100°C.

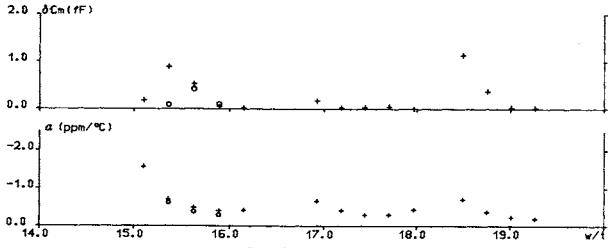


Figure 10. Computed values δCm and α versus w/t.
t = 96 microns, 200 nm plating.
+ : mechanical FE-model
o : piezoelectric FE-model.

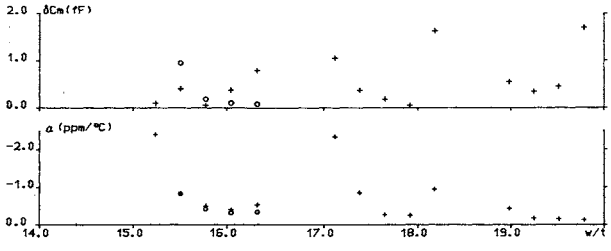


Figure 11. Computed values δCm and α versus w/t.
t = 93.5 microns, 400 nm plating.
+ : mechanical FE-model
o : piezoelectric FE-model.

The calculated inflexion temperature values T_{INF} typically lie between 25°C and 33°C. The values γ typically range from 0.088 to 0.1 ppb/°C^3. This is in good agreement with the experimental values.

DISCUSSION

We first present FE-mode pictures of a resonator with a well trapped thickness shear mode and weak coupling to the width flexure mode. The magnified amplitudes refer to a total resonator energy of 1 mWs.

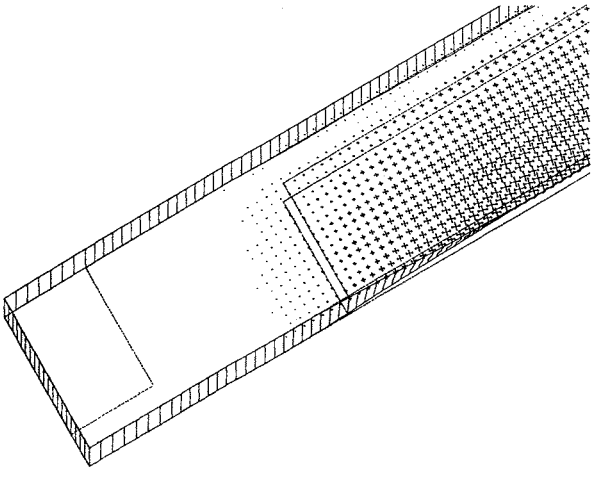


Figure 12. Thickness shear fundamental mode with distributions of principal stresses in the resonator mid-plane.
t = 96 microns, 200 nm plating, w/t = 18.75.

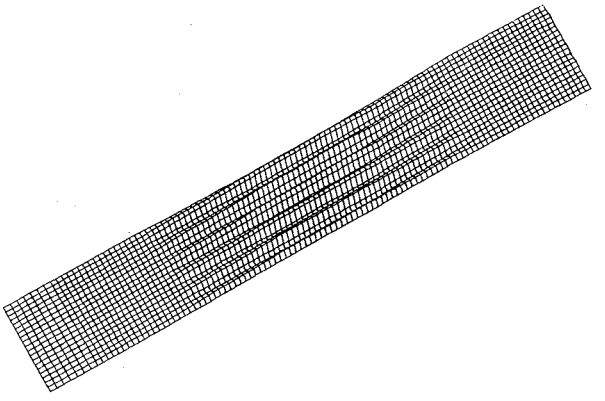


Figure 13. Out of plane displacement of resonator mid-plane surface. Magnification 250. t = 96 microns, 200 nm plating, w/t = 18.75.

Spectral response of two selected measured resonators are shown in figures 14 and 15. Impedance values abs(Z), measured with a HP 4194 A network analyzer, are plotted as a function of frequency. In figure 14 we represent the impedance curve of an excellent resonator, with resonance, antiresonance as well as first, second and fourth anharmonic overtone (thickness twist) modes. From our FE-model we obtain an electrically inactive first anharmonic overtone (Cm = 0). The small resonance in figure 14 results from imperfect length symmetry of the real quartz resonators. In figure 15 we represent the impedance curve of a resonator with an activity dip at 90°C.

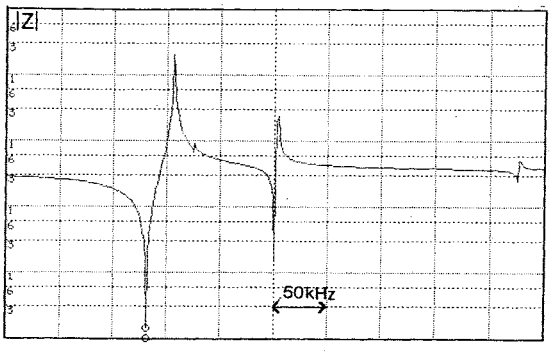


Figure 14. Impedance curve of an excellent resonator.
t = 96.2 microns, 200 nm plating, w/t = 15.61.

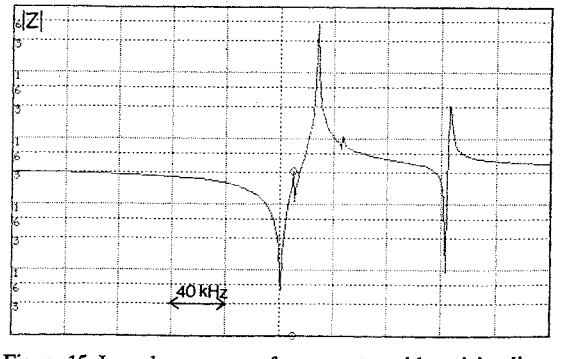


Figure 15. Impedance curve of a resonator with activity dip.
t = 95.8 microns, 200 nm plating, w/t = 16.94.

The FE-model unveils mode splitting for t = 96 microns at w/t = 17.19 (fig.6). Pictures of this mode are represented in figures 16 and 17 (mechanical model).

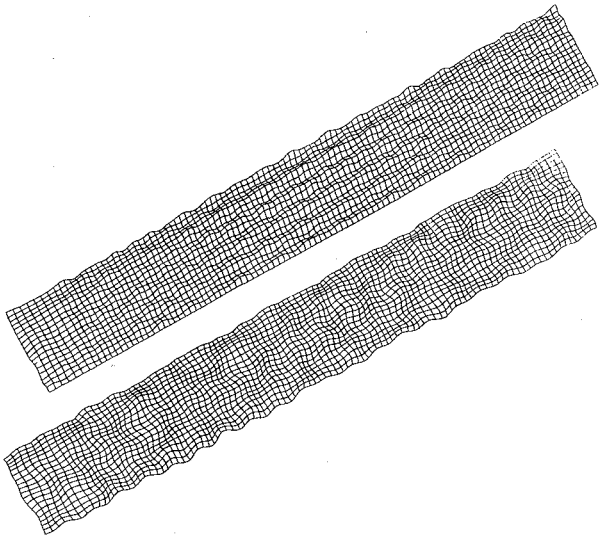


Figure 16. Resonator mid-plane surface displacement.
t = 96 microns, 200 nm plating, w/t = 17.19,
Mode 908: f = 16.87745 MHz, Cm = 6.98 fF.
above : deform. perpendicular to mid-plane, magn. 100
below : deformation parallel to mid-plane, magn. 100.

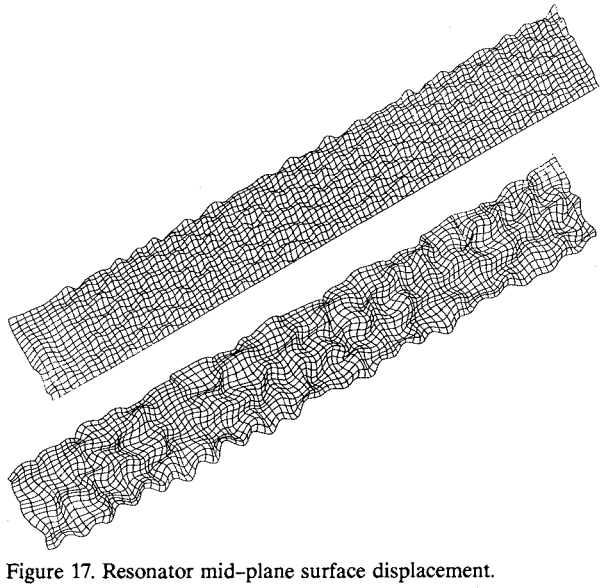


Figure 17. Resonator mid-plane surface displacement.
t = 96 microns, 200 nm plating, w/t = 17.19,
Mode 909: f = 16.87824 MHz, Cm = 1.54 fF.
above : deform. perpendicular to mid-plane, magn. 100
below : deformation parallel to mid-plane, magn. 100.

The fairly good coincidence of the appearance of an activity dip between the experiment and the theory is rather incidental, because on the other hand, from figure 10 we would expect just the opposite (small δCm value at w/t = 17.19). First we mention that, in contradiction to values of figure 6 and 7, the results in figures 10 and 11 have been obtained in taking into account also the small electrode connections. This indicates that spurious modes respond sensitively to small changes in resonator shape. Secondly, what concernes coincidence between experiment and theory, the real resonator thickness t (95.8 microns) is not exactly equal to the value (96 microns) of the FE-model. As a third point, it is wothwhile to mention the facts that the values of figure 6 result from the purely mechanical model and that probably we would need a more refined FE-discretization in order to obtain such an accurate resolution of the modal spectrum. Certainly it is also necessary to choose temperature increments smaller than 25°C, to obtain realistic values δCm of the figures 10 and 11.

Comparing the values σCm and δCm respectively of figures 3 and 10 with those of figures 4 and 11, it may be observed however that motional capacitance dispersion respectively variation is smaller, both experimental and theoretical, for resonators with 200 nm electrodes than for 400 nm platings. Apparently heavy electrodes promote coupling of the thickness shear mode with perturbing modes as shown in figure 17. This kind of rather chaotic vibration depend, besides width and thichness, also on resonator length.

So far we have not found a reliable theoretical criterion concerning the Q-factor of the resonator. We did calculate the deformation energy within the mounting area or the rigid body translational and rotational kinetic energy (for the complete structure), resulting from the reaction forces of clamped nodes within the mounting area. In both cases we obtained no clear correspondance to the measured R values of figures 3 and 4.

CONCLUSION

The analysis of narrow rectangular AT strip quartz resonators is important, because for this type of resonator there are various mechanisms provoking coupling to spurious modes. We developped a three-dimensional FE-method allowing us to analyse theese complicated phenomena. For the coupling to width flexure we obtained a clear correspondance between experiment and theory. The more delicate coupling to spurious modes, responsible for activity dips, can be analyzed with satisfactory qualitative agreement. The program is also usefull for the identification of resonances of thickness twist modes as well as the analysis of the frequency-temeperature behaviour.

We think that with computers of the new generation it will be possible to carry out FE-analysis with even greater precision (discretization) and shorter computation times. Also we envisage to analyze third harmonic thickness shear overtones.

REFERENCES

[1] B. Studer, W. Zingg, 4th EFTF 1990, p. 653-658, Technology and characteristics of chemically milled miniature quartz crystals.

[2] H.R. Schwarz, Methode der finiten Elemente, B.G.Teubner Stuttgart, 1st ed. 1980, 2nd ed. 1984, 3rd ed. 1991.

[3] M. Naillon, R.H. Coursant, F. Besnier, Acta electonica, 25, 4, 1983, p. 341-362. Analysé de structures piézoelectriques par une méthode d'éléments finis,

[4] Y.K. Yong, Z. Zhang, 46th ASFC 1992, p. 567-581, On straight crested wafes in a third overtone SC-cut quartz resonator.

[5] K.J. Bathe, S. Ramaswamy, Computer meth. in appl. mech. eng., 23 1980, p. 313-331, An accelerated subspace iteration method.

[6] K. Trümpy, 4th EFTF 1990, p. 286-294, 3-d nonlinear finite element analysis; determination of the static force induced frequency perturbation of quartz-crystal resonators.

[7] L.N. Dworsky, Motorola, 41st AFCS 1987, p. 295-302, AT-quartz strip resonators.

[8] W.T. Woodland, R.S. Witte, Motorola, 13th Piezoelectric devices conference 1991, Vol.1 p.70-74, AT-strip resonator applications and capabilities.

[9] R.D. Mindlin, W.J. Spencer, J. Accoustic soc. amer., Vol. 42 1967, p. 1268-1277, Anharmonic thickness-twist overtones of thichness-shear and flexural vibrations of rectangular AT-cut plates.

[10] B.J. James, 42nd AFCS 1988, p. 146-154, A new measurement of the basic elastic and dielectric constants of quartz.

[11] J. Zelenka, Piezoelectric resonators and their applications, Elsevier, Amsterdam 1986.

AN ADJUSTMENT METHOD OF RESONANT FREQUENCY AND FREQUENCY TEMPERATURE COEFFICIENTS FOR NS-GT CUT QUARTZ CRYSTAL RESONATORS

Hirofumi KAWASHIMA

Seiko Electronic Components Ltd.
Tochigi-shi, Japan 328

ABSTRACT

This paper describes a method for adjusting resonant frequency and frequency temperature coefficients for NS-GT cut quartz crystal resonators. An object of this paper is to provide NS-GT cut quartz crystal resonators with excellent frequency temperature behavior.

It is well known that NS-GT cut resonators (Refs. 1-6) show good frequency temperature behavior because the first-, second- and third-order temperature coefficients α_+ , β_+ , γ_+ for principal vibration reach zero approximately.

However, a distribution of frequency temperature behavior takes place in mass production. In order to obtain resonators with better and stable frequency temperature behavior within +/-2.5ppm over a wide temperature range of -30 to +85℃ in mass production, an absolute value of α_+ must be less than 5×10^{-8}/℃.

The present adjustment method makes it possible to control α_+ easily, so that the NS-GT cut resonators are found to be mostly within +/-1.0ppm in the above-mentioned temperature range.

Keywords: NS-GT cut resonator, Frequency temperature behavior, Frequency and temperature coefficient adjustment, Evaporation process.

§1 INTRODUCTION

A quartz crystal oscillator is widely used in communication equipment, consumer products and measurement apparatus as a source of a frequency standard and clock. Recently, a further miniaturized oscillator with low consumption power is being earnestly required with high precision.

As a resonator to meet these requirements, a NS-GT cut quartz crystal resonator which consists of a vibrational portion and two supporting portions, is available because its resonator miniaturized is of comparatively low frequency of 1 to 4 MHz and shows the most excellent frequency temperature behavior of all quartz crystal resonators. It is, therefore, characterized in that an oscillator extremely excellent in frequency temperature behavior is obtained without any external temperature compensation by thermister and so on for employment of a NS-GT cut quartz crystal resonator.

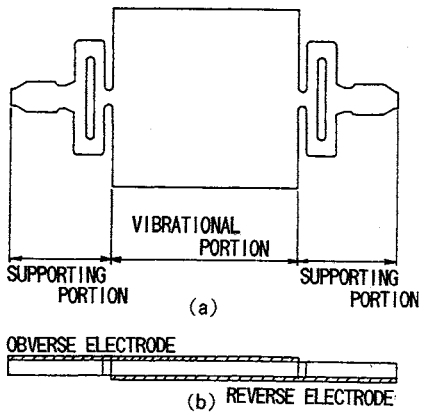


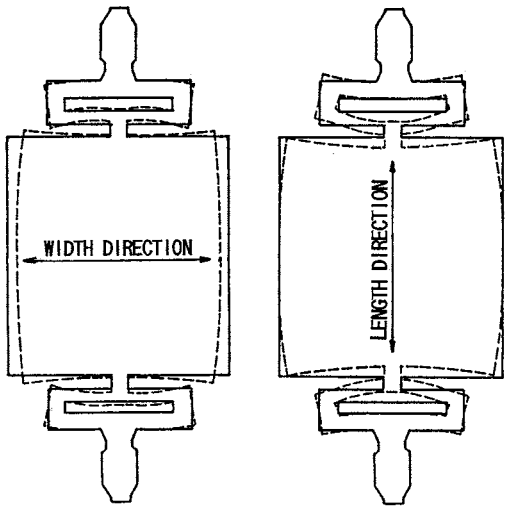
Fig. 1 NS-GT cut quartz crystal resonator(a) and its electrodes construction of section(b).

A distribution of frequency temperature behavior for principal vibration, however, takes place in production and the distribution is principally caused by the first order temperature coefficient α_+. A method for decreasing the distribution of α_+ is, therefore, desired. We have developed a method for absorbing a distribution of resonant frequency and frequency temperature coefficients caused in production.

§2 ADJUSTMENT PRINCIPLE

2.1 NS-GT cut resonator

Figure 1 shows the concrete shape of a NS-GT cut quartz crystal resonator consisting of a vibrational portion and two supporting portions, formed by an etching process. The present



(a) Principal vibration (b) Subvibration

Fig. 2 Mode shapes (a), (b) of principal- and sub-vibrations.

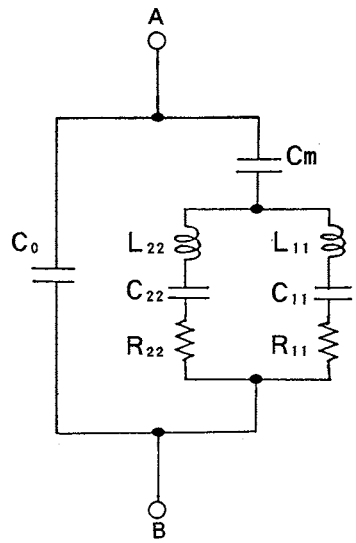


Fig. 3 Electrical equivalent circuit of coupling quartz crystal resonator.

resonator's shape is so designed that the vibration of the principal- and sub-vibrations is facilitated without suppression and the vibration energy of the resonator is trapped into the vibrational portion. The electrodes are deposited on obverse and reverse of the vibrational portion.

Figure 2 shows two modes (a), (b) of a width-extensional vibration for a principal vibration whose frequency is approximately proportional to width of the vibrational portion and a length-extensional vibration that is coupled with the width-extensional vibration. The frequency of the subvibration is also in approximate proportion to length of the vibrational portion. The modes were analyzed by the finite element method. A solid line indicates a static condition, and a wave line also indicates a vibrational condition.

2. 2 Electrical equivalent circuit

Figure 3 shows the electrical equivalent circuit of a coupling quartz crystal resonator in the case of capacitance coupling. Each sign shows motional arms of principal vibration (L_{11}, C_{11}, R_{11}), subvibration (L_{22}, C_{22}, R_{22}), coupling capacitance Cm and shunt capacitance C_0. Accordingly, the impedance **Z** between the electric terminals A, B of both electrodes is given as a summation of real part Rei and imaginary part jXe.

$$\mathbf{Z}=\mathrm{Rei}+j\mathrm{Xe} \qquad (i=1,\ 2) \qquad (1)$$

In addition, since vibration losses of quartz crystal resonators are very small, and by taking into account that series resistances R_{11} and R_{22} vanish, the imaginary part Xe is given as

$$\mathrm{Xe}=\frac{-1}{2\pi f C_0}\cdot\frac{\left(\frac{f^2}{f_1^2}-1\right)\left(\frac{f^2}{f_2^2}-1\right)-\frac{C_1C_2}{Cm^2}}{\left(\frac{f^2}{f_1^2}-1\right)\left(\frac{f^2}{f_2^2}-1\right)-\frac{C_1C_2}{Cm^2}-\left\{\frac{C_2}{C_0}\left(\frac{f^2}{f_1^2}-1+\frac{C_1}{Cm}\right)+\frac{C_1}{C_0}\left(\frac{f^2}{f_1^2}-1+\frac{C_2}{Cm}\right)\right\}} \qquad (2)$$

where f_1, f_2: uncoupled resonant frequencies for principal- and sub-vibrations

$$L_{11}=L_1,\quad L_{22}=L_2$$

$$\frac{1}{C_1}=\frac{1}{C_{11}}+\frac{1}{Cm},\qquad \frac{1}{C_2}=\frac{1}{C_{22}}+\frac{1}{Cm}$$

and the real part Rei is also given as

$$\mathrm{Rei}=\mathrm{Ri} \qquad (i=1,\ 2) \qquad (3)$$

where $R_{11}=R_1$, $R_{22}=R_2$

From eq. (2) each resonant frequencies f_+, f_- coupled between two modes of a principal vibration and a subvibration are derived under the condition that Xe = 0.

2. 3 Adjustment principle

From eq. (2) resonant frequencies f_+, f_- coupled are related to the equations:

$$f_1^2+f_2^2=f_+^2+f_-^2$$

$$(1-K_m^2)\,f_1^2 f_2^2=f_+^2 f_-^2 \qquad (4)$$

where C_1C_2/Cm^2 is replaced with K_m^2 and Km is called "coupling factor".

Furthermore, differentiating eq. (4) once with

temperature t ($\cdot \equiv \partial/\partial t$) and if we put

$$\alpha_1 = \dot{f}_1/f_1 \ , \quad \alpha_2 = \dot{f}_2/f_2$$

$$\alpha_+ = \dot{f}_+/f_+ \ , \quad \alpha_- = \dot{f}_-/f_- \qquad (5)$$

$$\alpha_{Km} = \dot{K}m/Km \ , \quad K = Km^2/(1-Km^2)$$

An important relations is obtained from eqs. (4) and (5) as

$$\alpha_+ + \alpha_- = \alpha_1 + \alpha_2 - K\alpha_{Km} \qquad (6)$$

namely, a summation of α_+ and α_- becomes constant theoretically. In addition, if we take $f_{MP} = f_-/f_+$ and $R = f_2/f_-$, α_+ for principal vibration is transformed into

$$\alpha_+ = \frac{\alpha_1 - \alpha_2 f_{MP}^2}{1 - f_{MP}^2} + \frac{f_{MP}^2}{1 - f_{MP}^2}\{K\alpha_{Km} - R^2(\alpha_1 - \alpha_2)\} \qquad (7)$$

with respect to α_+ of eq. (7) α_+ for NS–GT cut resonators is probably proportional to a normalized frequency of $\delta = f_{MP}^2/(1-f_{MP}^2)$ because the values of

$$\frac{\alpha_1 - \alpha_2 f_{MP}^2}{1 - f_{MP}^2} = n$$

and

$$K\alpha_{Km} - R^2(\alpha_1 - \alpha_2) = m \qquad (8)$$

may be approximately constant to a small variation of f_{MP}.

In particular, the sign m which represents the slope almost independent on resonant frequency coupled, can be easily obtained from the experiments.

§3 RESULTS AND DISCUSSION

Figure 4 shows a NS–GT cut quartz crystal resonator and the positions of weights to be deposited by an evaporation process so as to adjust α_+. The weights are deposited on the vibrational portion in the vicinity of both ends in the length direction. As will be described in Fig. 5, these positions are very convenient for adjustment of α_+ because α_+ can be adjusted without changing resonant frequency for a principal vibration.

Figure 5 shows a relationship of resonant frequencies coupled f_+, f_- versus weights of Ag deposited by an evaporation technology. As is easily predicted from Fig. 4, resonant frequency f_- for a subvibration is found to vary markedly as compared with that for a principal vibration. From the result, a coupling condition between the principal- and sub-vibrations can be controlled by

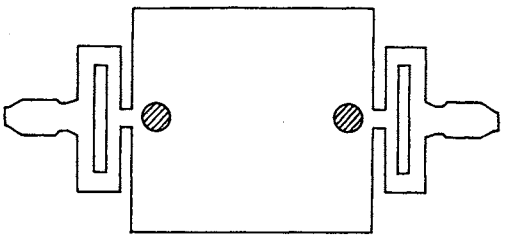


Fig. 4 Positions of weights to be deposited by an evaporation process to adjust α_+.

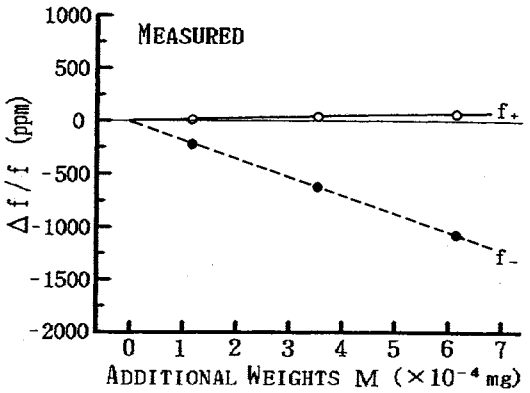


Fig. 5 Relationship of resonant frequencies f_+, f_- versus weights of Ag.

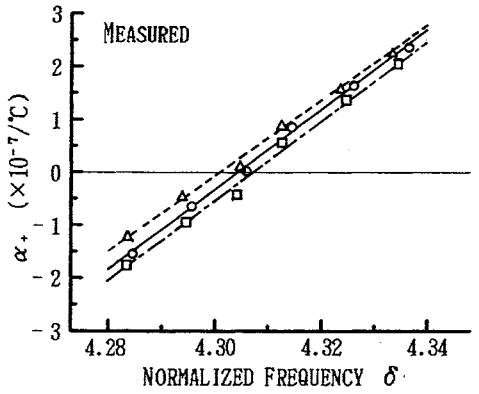


Fig. 6 Relationship between the first order temperature coefficient α_+ and a normalized frequency δ.

adding the weights. In general, the quantity of frequency adjustment that gives $\alpha_+ = 0$ is less than 1000ppm.

Figure 6 shows a relationship between the first order temperature coefficient α_+ and a normalized frequency δ. As δ decreases, α_+ also decreases. Since α_+ is so designed that it has a value of about 2×10^{-7}/℃ after formation of resonators, a variation of α_+ versus δ can be regarded as the straight line in a narrow range of δ which gives $\alpha_+ = 0$.

Figure 7 shows five samples of frequency temperature behavior for NS–GT cut resonators formed by an etching process and, needless to say,

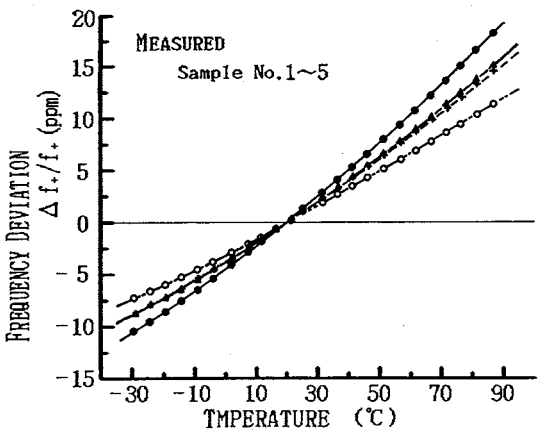


Fig. 7 Frequency temperature behavior for NS-GT cut resonators (Before the adjustment).

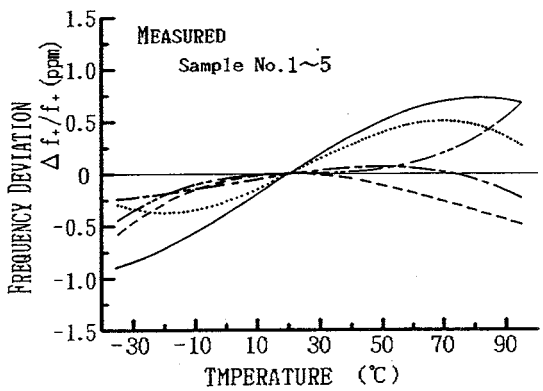


Fig. 8 Frequency temperature behavior for NS-GT cut resonators (After the adjustment).

the characteristics before the adjustment of frequency temperature behavior. In general, the resonators formed by an etching process are so designed that α_+ has a value of 1 to 3×10^{-7}/℃, as a consequence, frequency deviation for the resonators in frequency temperature behavior is within +/-20ppm in a temperature range of -30 to +85℃, because an absolute value of the second- and third-order temperature coefficients β_+, γ_+ that are very dependent on a cut angle(ϕ, θ) and masses of the supporting portions, is less than 5 $\times10^{-10}$/℃^{2}, and 6×10^{-12}/℃^{3}, respectively.

In a relation between α_+ and δ, the slope m has a mean value of 75×10^{-7}/℃ for NS-GT cut quartz crystal resonators with a frequency of 2.1 MHz. Since α_+ for resonators formed by an etching process are measured in detail, the frequency ratio f_{MP} that gives $\alpha_+=0$ is easily calculated from eq. (7).

Figure 8 shows frequency temperature behavior for the above-mentioned five resonators adjusted by the present method. As is apparent from Fig. 8, the frequency deviation after the adjustment is mostly within +/-1.0ppm in a wide temperature

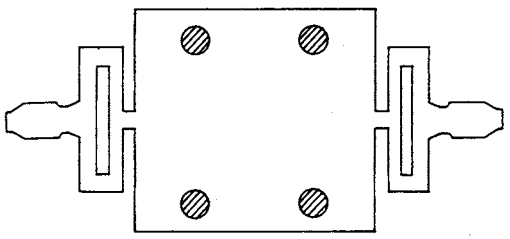


Fig. 9 Four positions of weights to be deposited by an evaporation process to adjust f_+.

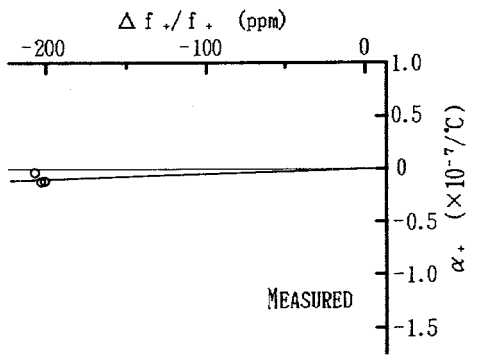


Fig. 10 Relationship of α_+ versus a variation of resonant frequency f_+.

range of -30 to +85℃.

It was shown that the first order temperature coefficient α_+ can be adjusted. As was reported at the previous Frequency Control Symposium by the author and coworkers, there exist some positions wherein α_+ is almost independent on variation of resonant frequency f_+ for a principal vibration.

Figure 9 shows four positions of weights to be deposited by an evaporation process so as to adjust resonant frequency f_+ without changing α_+. The four weights are deposited on the vibrational portion in the vicinity of both ends in the width direction.

Figure 10 shows a relationship of α_+ versus a variation $\Delta f_+/f_+$ of resonant frequency f_+ when weights are deposited on four positions of the vibrational portion. It is easily understood that α_+ is almost independent on a variation of f_+. Thus, resonators with very small frequency temperature coefficients α_+, β_+, γ_+, adjusted to the nominal frequency can be easily obtained, because α_+ and f_+ can be independently adjusted.

Table 1 shows the values of electrical equivalent circuit parameters for the five NS-GT cut quartz crystal resonators after the adjustment of α_+ and f_+. The series resistance R_1 and a quality factor have a value of 65.5 to 79.7Ω and 192,000 to 238,000, respectively. The resonators are found to be sufficiently satisfactory as a resonator.

Table 1 The values of electrical equivalent circuit parameters for the NS-GT cut resonators after the adjustment of α_+ and f_+.

No.	Frequency (MHz)		$R_{1,2}$ (Ω)	$L_{1,2}$ (H)	$C_{1,2}$ (fF)	C_0 (pF)	r	Q ($\times 10^3$)
1	Prin.	2.10	65.5	1.18	4.88	2.70	553	238
	Sub.	1.89	349	3.17	2.23		1211	108
2	Prin.	2.10	67.6	1.17	4.91	2.94	599	228
	Sub.	1.89	339	3.10	2.28		1289	109
3	Prin.	2.10	67.0	1.17	4.90	2.74	559	230
	Sub.	1.89	299	3.11	2.27		1207	124
4	Prin.	2.10	79.7	1.16	4.94	2.73	553	192
	Sub.	1.89	350	3.13	2.26		1208	106
5	Prin.	2.10	67.0	1.17	4.89	2.76	564	230
	Sub.	1.89	401	3.06	2.31		1195	91

§4 CONCLUSIONS

With a view of obtaining an adjustment method of resonant frequency and frequency temperature coefficients, particularly, the first order temperature coefficient for a NS-GT cut quartz crystal resonator, a study was performed theoretically and examined experimentally.

First, an electrical equivalent circuit of a coupling quartz crystal resonator which can be expressed as a capacitance coupling was shown. From the circuit the impedance of the resonator that consists of real part and imaginary part was calculated.

Second, resonant frequencies coupled are derived from the imaginary part of the impedance that must be zero, and also the frequency ratio f_{MP} is newly defined by employment of the resonant frequencies coupled, so that α_+ was found to be given as a function of f_{MP} and almost proportional to a normalized frequency δ.

Third, according to the theoretically obtained result, experimental investigation was performed, as a consequence, α_+ was found to be accurately adjusted to zero by using the related equation of α_+ and δ, and the frequency deviation was mostly within +/-1.0ppm over a wide temperature range of -30 to +85℃.

Finally, electrical equivalent circuit parameters was shown for five NS-GT cut resonators for which α_+ and resonant frequency of a principal vibration, was adjusted to zero approximately and nominal frequency. These resonators were found to be sufficiently satisfactory as a resonator.

ACKNOWLEDGEMENTS

The author would like to thank Mr S. Takigawa and Mr M. Matsuyama for their help in the experiments.

REFERENCES

(1) H. Kawashima, H. Sato and O. Ochiai : " New Frequency Temperature Characteristics of Miniaturized GT Cut Quartz Resonators", Proc. the 34th Annual Symp. on Frequency Control, pp. 131-139 (1980).

(2) H. Kawashima: " Vibrational Analysis of GT Cut Quartz Crystal Resonators with the Supporting Portions at the Ends" Proc. the 40th Annual Symp. on Frequency Control, pp. 193-200(1986).

(3) H. Kawashima : " An Analysis of Frequency Temperature Characteristics and Electrical Equivalent Circuit Parameters of a New Shape GT Cut Quartz Crystal Resonator Formed by an Etching Method", IEEE Ultrasonics Symp. Proc. pp. 465-470(1989).

(4) H. Kawashima: " An Analysis of Oscillation Frequency Characteristics in a CMOS Oscillating Circuit Using Coupling Quartz Crystal Resonator ", Proc. the 44th Annual Symp. on Frequency Control, pp. 585-592 (1990).

(5) H. Kawashima and K. Sunaga : " Temperature Compensated Crystal Oscillator Employing New Shape GT Cut Quartz Crystal Resonator", Proc. the 45th Annual Symp. on Frequency Control, pp. 410-417 (1991).

(6) H. Kawashima and K. Sunaga : " Frequency Temperature Behavior and Its Hysteresis of S-TCXOs Using New Shape GT Cut Quartz Crystal Resonators", Proc. the 1992 IEEE Frequency Control Symposium, pp. 434-441(1992).

CHARACTERIZATION OF BVA IN INDUSTRIALIZATION PROCESS

J.J. Boy*, A. Berthaut*, G. Renard*, R. Roy*, D. Thiebaud*
G. Robichon**, C. Longet**

* LCEP / ENSMM - Route de Gray - La Bouloie - 25030 Besançon Cedex - FRANCE
** bva INDUSTRIE - 1, boulevard Fleming - BP 1937 - 25020 Besançon - FRANCE

ABSTRACT

Miniaturized electrodeless 10 MHz SC-cut BVA are "daily" manufactured in the L.C.E.P. For each "serial" production, we collect results about quartz material quality on one hand, and about electrical parameters behaviour of the resonator, in other one. We attach us to vary the supplying of the raw material to tentativelly understand, first the performances under environmental parameters (acceleration, pressure, vibrations and radiations) and second the variations of the short and long term stability. Furthermore, we have taken into account other parameters as : reproductibility of the performances in one serie and from one serie to other one, reliability of the product, ...

Some key points are discussed and systematically tested in order to gather statistics and to define how the performances can be reproduced according to dispersion. The "retrace" of resonators submitted to ionizing radiations is also considered.

Finally, we present some additive detailed results about short term stability. Numerous $S_y(\phi)$ and Allan variance data of 10 MHz BVA are achieved on various kinds of measurements.

Keywords : Quartz crystal, BVA resonators, long term drift, short term stability, Q-factor, turn-over point.

1. INTRODUCTION

During the two last years, some improvements are implemented in the industrialization process of "tactical" mini-BVA (SC-cut 10 MHz third Overtone) in the way to transfert this production to the industry. Some keypoints, presented below, are systematically discussed and tested in order to define how the performances can be reproduced according to different specifications (for telecomunications or space applications, for example).

The technical characteristics of this "tactical" BVA developed for use in severe environments are the following:

*Nominal frequency : 9,999,983 Hz
*Tolerance at the lower T.O. point : [-10, +6 Hz]
*SC-cut, 3rd overtone
*Turn Over point : [72, 85 °C]
*Typical motional resistance : 95 Ω
*Q-factor : more than 1.2 10^6
*Can : HC 40/U (without pinch off)

We summurize (after discussion) the typical values of the other performances (aging, short term stability, accelerometric or barometric sensitivities) in the conclusion of this present paper.

2. BACKGROUND AND KEYPOINTS OF THE DESIGN

As LCEP realized usually some tens of SC-cut BVA resonators a year, a few of them are used for our researchs concerning as well the electronic parts than the resonator itself. Though the design of the BVA_2 is well known (1) (2) (3), we have decided to point out all the main parameters of manufacturing to tentativelly improve the realiability of the product as well as the performances.

Furthermore, as we said in (4) and as far back as possible, we have stored all the possible data about the raw -natural or synthetic- quartz material used.

The first diagram shows the principal keypoints of the design which work towards perfecting the symmetries around Z-axis and the median plane. Indeed, the resonator is composed of two "condensers" electrode-holders on both sides of the resonator, these three parts being holdered by four rigid clips associated to a symmetrical suspension consisting in 8 springs (i.e. 4 pairs). Furthermore, the vibrating part of the resonator is linked to its external support by four rigid and short quartz bridges.

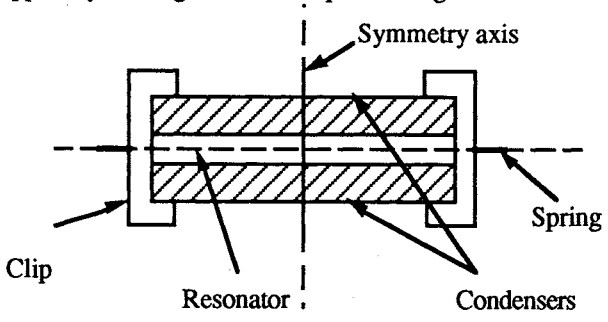


Fig. 1 : mini-BVA (diagram)

So, to maintain as far as possible the symmetries of this design, we had to check some keypoints as :

2.1- Angular orientation of the SC-cut blanks : After sawing and grinding, we performe some adjustements and check controls to obtain blanks for which the angular dispersion is better than ± 5". The direction of Z-axis in the plane of the blanks is also precisely determined to adjust the three parts together. Indeed, a finite element method permits us to calculate that a 5 deg.-error in the orientation of one condensor in comparison with the resonator creates a frequency shift at turn-over point of about 15 Hz (and 20 Hz for the AT-cut BVA working at 10 MHz to).

2.2 - During the final mounting : To avoid differential thermal stresses between condensors and resonator, we adjust (better than 0.5°)

along Z-axis the three parts of the "sandwich". And, a mechanical mount permits us to adjust and check correctly the position and the tightening of the clips.

Furthermore, we have pointed out other manufacturing parameters which are considered also as keypoints :

2.3 - Grinding, polishing and chemical etching: The alternated uses of grinding, polishing and chemical etching operations avoid, as far as possible, the "dead-lattice" and microcracks on the surface of the resonator.

2.4 - Long time burn-in : This process is assumed to cancel the internal stresses introduced during sawing or polishing and to refine the blank surfaces. In the past, these operations have shown large frequency shifts.

2.5 - Frequency adjustement : The frequency adjustement is realized by chemical etching to get off **just** the microlayer of dead lattice created by the last polishing step. So, we avoid to create the unavoidable etch pits which can appear when the raw material is not swept.

2.6 - Temperature of the final bake-out : During some tens hours, BVA are put at about 280°C and under high vaccum (between 3.10^{-8} and 10^{-7} millibar) just before coldweld enclosure (in a MC 40/U can without pinch off). These annealing and aging steps tend to cancel the frequency variations dues to the removal of gaz from the surface of the resonator and the internal stresses relaxation.

Finally, we obtain SC-cut BVA resonators working at 10 MHz, which are, furthermore, stored in a oven controlled at a temperature close to the turn-over point. Furthermore, we create fast thermal cycles 5 to 7 times a week. During this process, the resonators are, of course, included in a oscillator loop.

3. EXPERIMENTAL PROCEDURE

Then, the resonators are tested under temperature up to 125 °C : with a desk computer, we check the phase difference between the input and the output levels on a "PI" bridge within which the quartz resonator is inserted. We store the resonance frequency and the motional resistance. first **just after bake-out** (MPE1) and second a few **weeks later after thermal cycles** (MPE2).

For each characteristic, the three coefficients of the best fitted third degree polynomia as :

$$\frac{\Delta F}{F} = a.\Delta T + b.\Delta T^2 + c.\Delta T^3$$

are calculated and stored.

The first BVA of this "batch" (series 17 to 23) have been charaterized during a eight hours temperature ramp from 20 to 125 °C (0.2°C/mn) (process 1), the other ones (series 24 to 27) during a "step-by-step" characteristic (process 2). In this case, the frequency is only stored when a given temperature of the oven is stabilized within 0.1°C (and checked by a quartz thermometer put close to the resonator under tests). Below, the figure 2 shows the difference existing between the two kinds of frequency-temperature curves (MPE1, MPE2). Due to the relatively high time constant (a few tens of seconds), the turn over point is shifted to higher temperatures of about 1.5°C.

During measurements with the process 2, we can estimate the stability of the oven is better than ± 0.05°C and consequently we evaluate the error on the T.O. point at ± 0.2°C. The other performances of our measurement system are, in the worst case :

Frequency : $\pm 5.10^{-9}$
Motional resistance : 1%
Q-factor : 2%

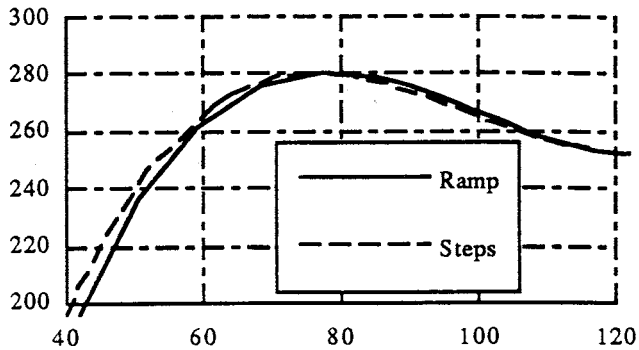


Fig.2 : MPE1 and MPE2 (in Hz/°C)

4. HISTOGRAMS : STATISTICAL DATA ON TWO-YEARS BVA PRODUCTION

All parameters are stored for statistical purposes. We present some histograms which are very interesting (particularly these on turnover temperature) to show slow changes (intentional or not...) in manufacturing parameters.

The first one (Fig.3) is to illustrate the variations of the resonance frequencies at turnover temperature. The standard deviation is slighly high because, up to now, the frequency adjustement operation depends only on the "patience" of the operator, the temperature control in the operating room being not really well accurate. We must remember that the frequency slope around ambiant temperature is about 8 Hz/°C !

Finally, we have not took in account the 18 resonators realized in natural quartz (series 17 and 18) whichfor the frequency adjustement was very difficult because the chemical etching time is very different for natural and synthetic raw material (about three times higher for natural quartz).

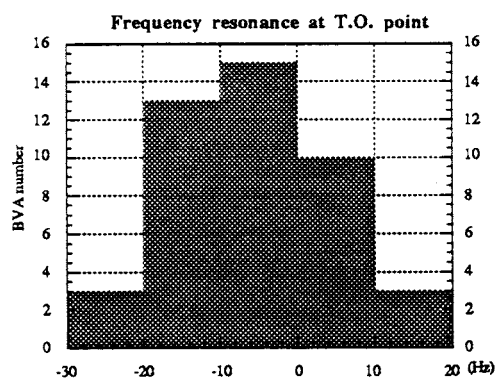


Fig. 3 : Resonance frequency for 45 BVA around 9,999,983 Hz
(Mean value : -4.78 Hz and Std dev.: 11.48 Hz)

4.2 - Series 20 to 27 : Turn-over temperatures

The two next diagrams show how certain modifications implemented in the process have improved the standard deviation of the turn-over temperatures. For the first one (SN 20 to 23), though the angular orientation

of the SC-cut blanks are precisely adjusted, the tightening of the clips was not well defined. So the residual stresses due to the dissymmetry of the mounting can introduce a small shift in the turn-over temperature. For the second one (Fig. 5), we have sorted the clips in function of their tightening.

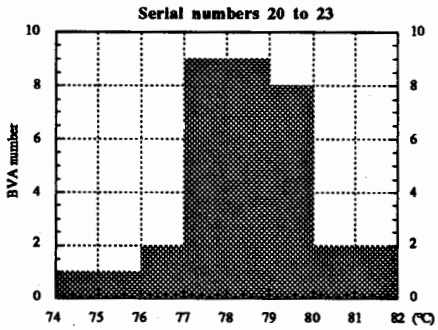


Fig.4: T.O. determined by temp. ramp (37 res.) (Mean value : 78.03°C and Std dev.: 2.03°C)

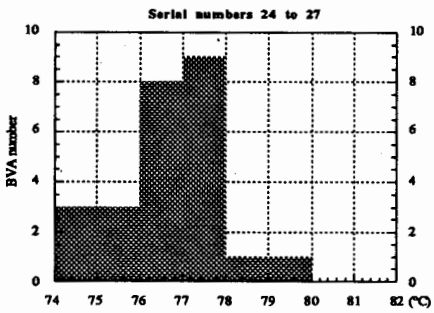


Fig.5 : T.O. determined step by step (29 res.) (Mean value : 76.17°C and Std dev.: 1.58°C)

The next table (Table 2) presents some statistical results measured on 9 resonators belonging to the series 26. The frequency-temperature curves are performed step by step (with 30 measured points) between 15 and 130°C.

	Mean v.	Std dev.
F(T.O)-9,999,983Hz	1.23	7.13
Turn.Over (°C)	76.04	0.69
F(T.O.) - F(25°C) (Hz)	188.11	4.48
a ($*10^{-6}$)	0.899	0.015
b ($*10^{-9}$)	-13.60	0.173
c ($*10^{-12}$)	62.67	0.87
Rm (@ T.O.)	95.3	2.29
Q ($*10^{6}$) at 25°C	1.28	0.03

Table 2 : Series 26

4.3 - Frequency adjustements :

The problems encountered during the frequency adjustement are illustrated by the next table. For the 11 resonators belonging to series 27, the differences observed between the two measurements of the frequency differences (i.e. frequencies at turn-over point and at 25°C) are presented. So, these few Hz observed have to be added to the uncertainties of frequency measurements. Consequently, the standard deviation at turn-over temp. is worse.

	Mean v.	Std dev.
F(T.O.) - F(25°C) in test condensers	**179.7**	8
Turn Over (°C) in test condensers	76.3	1.34
F(T.O.) - F(25°C) after final bake-out	**184.4**	9.2
Turn Over (°C) after final bake-out	76.7	1.94

Table 3 : Parameters of series 27

4.4 - Resistance vs temparature and Q-factor :

The different behaviours existing between kinds of raw material can be also exhibited on the motional-resistance temperature curves. The following figure (n°. 6) presents typical behaviours for 5 quartz material. For example, to a resonator in natural quartz, the resistance increases slowly with the temperature. In contrary, for a swept synthetic one, the resistance is slighly constant. For all the resonators made in quartz noticed "US Stand", the resistance increases up to 50°C and decreases after.

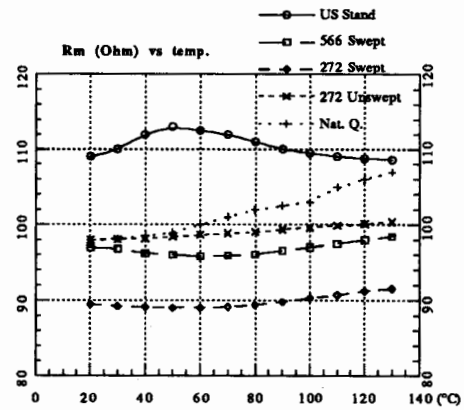


Fig. 6 : different resistance-temperature curves

Table 4 includes mean Q values calculated at 25°C for the previous five raw materials.

Series	17-18	20	21-23	23-25	26	27
Raw mat.	Nat.	272	272S	566S	?Uns.	US
Q ($*10^{6}$)	1.2	1.36	1.37	1.32	1.28	1.18
Std dev.	0.08	0.02	0.02	0.03	0.03	0.04

Table 4 : different Q-factors

5. FREQUENCY RETRACES

We have also characterized each series in performing some retraces in "standard" oscillators. Typically, after stopping the power supply of the oscillator and the oven during 16 hours, we record again the frequency up to the stabilization. For non irradiated synthetic quartz, the retrace is systematically better than 1.10^{-9} (see figures 7a and 7b).

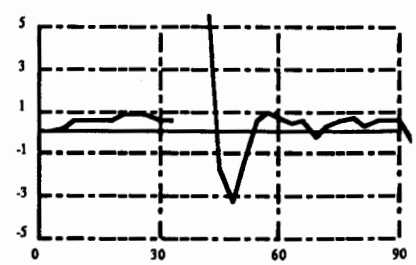


Fig. 7a : Retrace on serial number 26.10

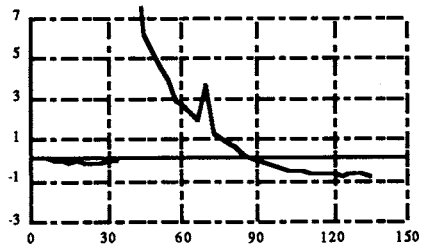


Fig. 7b:Freq. shift (10^{-9}) vs time (mn) (24.10)

For the resonators irradiated with γ-rays, the retrace is generally bad even for swept synthetic quartz as we can see in figures 8a and 8b.

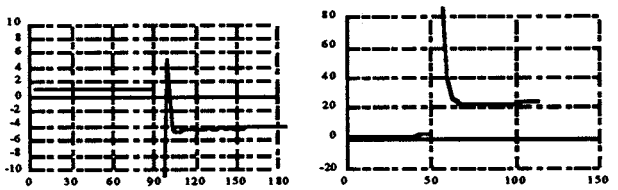


Fig.8a : 19.06 (swept synth. quartz) and Fig. 8b : 17.08 (natural quartz)

6. LONG TERM DRIFTS

Aging is defined as a change in frequency relative to the initial frequency of the unit. The mechanism that can cause aging are (5) (6) :

-mass transfer due to contamination inside the resonator inclosure

-stress relief in the resonator due to quartz material himself but also to mounts, bonds or electrodes (except, of course, in case of electrodeless BVA)

-other effects (as diffusion effects due to DC voltage driven for example, or leak and outgassing, pressure, thermal or electric field changes) can also modify the long term drift but can be, as far as possible, controlled and not modified.

-and the agings of the components of the oscillator circuit which can be considered as negligeable.

So, we think that the change in frequency is primarily influenced by two factors : the **change in mass distribution** and **stress relaxation** (7).

The change in mass distribution arises during the processes of absorption and desorption of physisorbed monolayers is expressed by the relation :

$$\frac{df(t)}{dt} = -K_1.f(t)$$

and then the dependance of frequency on time takes an exponential form.

The other processes which involve mass changes are **chemisorption** and **oxidation** described by :

$$f(t) = f(t_0) + K_r.Log(1+B.t)$$

where K_r **is the recovery rate constant** for a particular type of crystal unit and **B represents a constant related to the initial conditions.**

This last relationship can be used to describe also the relaxation of internal stresses in the quartz, but here the Kr constant is proportional to the absolute temperature (5).

The long term drifts of our BVA are only recorded at about 80°C which is heir turn-over temperatures. The next figures show, first, the raw data (the mean value of 20 records a day) and then the best fits calculated from the first day of records) which is considered as the mathematical aging law.

As we can see (Figures 9), all the long term drifts can be fitted by a **Log function.** So, mass distribution changes can be the most probable event which infuences the frequency shift. Other tests at different temperatures are not yet finished to evaluate the role of the relaxation of internal stresses in quartz itself.

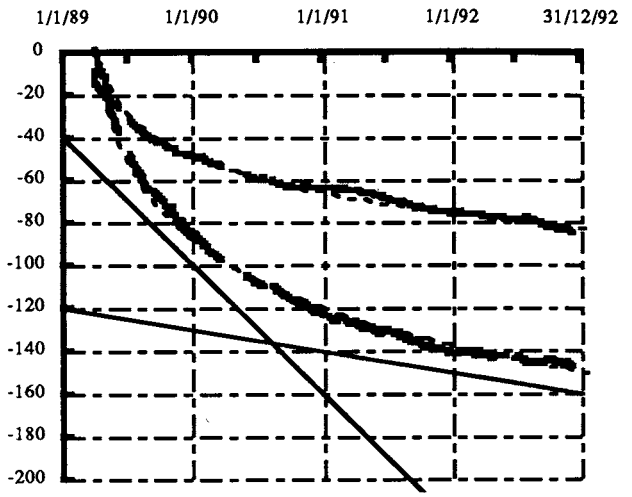


Fig.9a: L.T. drift on 07.04/11 (4 years)(10^{-10}) <$2*10^{-11}$/day after 3 months and <$3*10^{-12}$/day after 1 year
Best fits : -22*Ln(0.03t + 1) and -40*Ln(0.03t + 1)

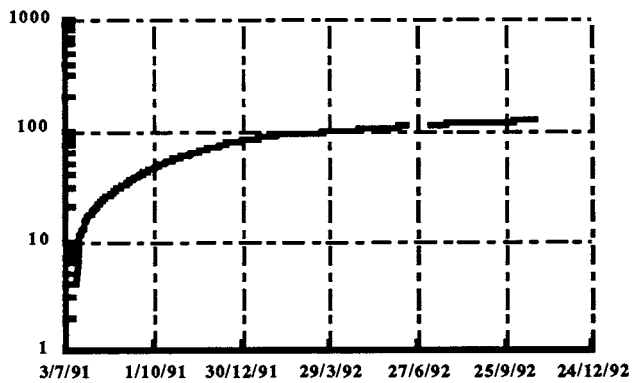


Fig.9b: L.T. drift on 16.00 (1.5 year) in 10^{-10}

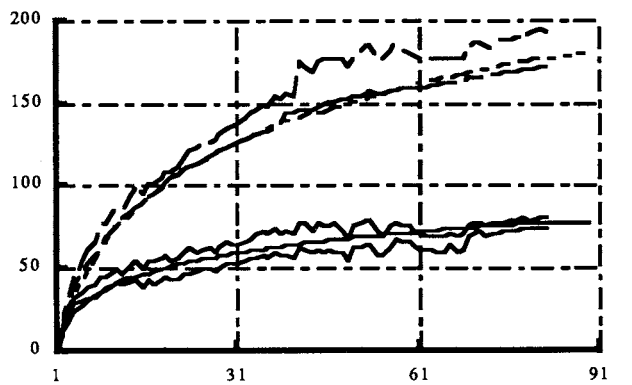


Fig.9c: L.T. drift (in 10^{-10} vs day number) for synthetic quartz (SN24.xx and SN26.xx)

The above figures present only recorded data, the resonators being in a thermostated room not controlled in atmospheric pressure and hygrometry.

The typical values for two resonators of each series are summarized in the table 5 (expressed in 10^{-10}).

SN	30 days	90 d	1 year	K_r	B	Com.
07	-20	-40	-70	-30	0.03	Uns.
14	45	100	220	117	0.015	Swept
16	15	45	100			Swept
17	850	1200				Nat Q.
21	50	75	110	23	0.3	Swept
24	125	180	250	55	0.3	Swept
26	65	85	115	20	0.8	Uns.

Table 5 : Long term drifts

The mathematical aging law can be also used to represent the daily rate of various BVA as in figure 10.

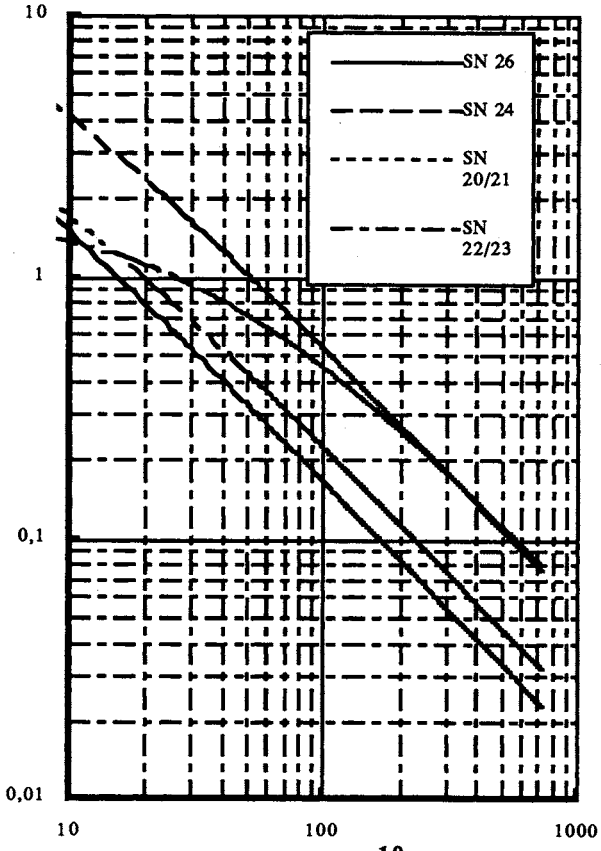


Fig.10 : Daily rate in 10^{-10} vs day number

In conclusion, we can said that higher is the B constant, better is the pre-aging performed. As K_r constant, its value depends on raw material quality, of course, but also on the defaults of symmetry of the structure (as discussed above). The long term drifts of the quartz belonging to the series 26 are the best because the tightenings of the clips have been particularly well defined. So, the internal stresses introduced by eventual defaults of symmetry are minimized.

7. SHORT TERM STABILITY

The measurement system used is based on which is presented by F. Walls in (8). We have to use a pair of similar resonators excited by the same low noise source (see in Figure 14 the diagram of the measurement system), the differential phase fluctuations being extracting through a mixer.

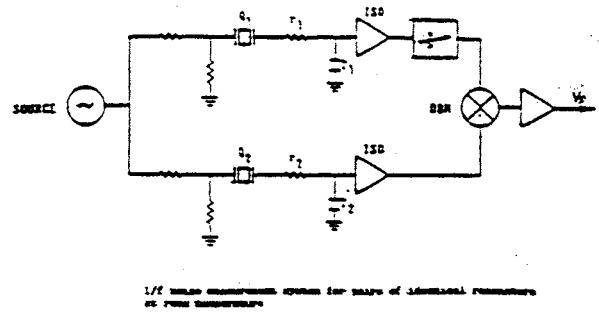


Fig. 14 : Measurement system

The signal is then amplified and read by a Dynamic Signal Analyzer driven by the software "Phi-noise", developped by both LCEP and bva Industrie. This software allows to plot the entire representation of the real data recorded. This method applies on several resonators gives results presented in figure 15 and the corresponding $\sigma_y(\tau)$ for 1 second in table 6.

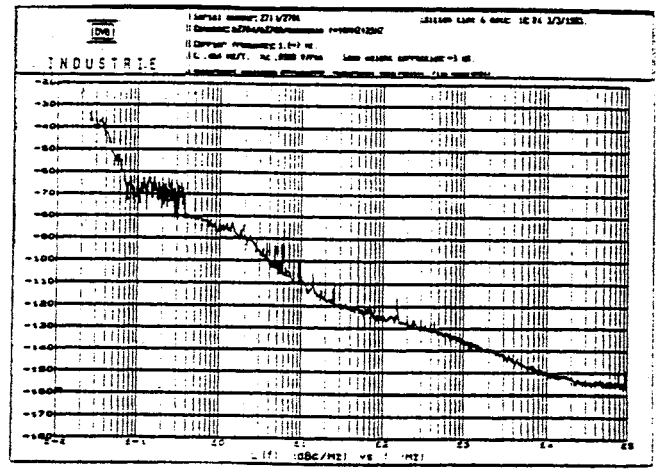


Fig. 15 : L(f) for 27.08/27.11

	Q-factor ($*10^6$)	σ_y(1s)
07.08/19.05	0.77	$1.2*10^{-11}$
26.01/22.10	1.28	$2.2*10^{-13}$
27.08/27.11	1.15	$6*10^{-13}$
2 µAT-cut	0.7	$3.5*10^{-12}$
23.xx (9)	1.32	$< 2*10^{-13}$ 100s Allan Var.

Table 6 : $\sigma_y(\tau)$ for various kinds of resonators

It appears that there are not direct relationships between Q-factor and σ_y(1s) for several reasons :

- The obligation in using two as much as possible identical resonators (in frequency and Q-factor) and the compatibility with the frequency of our low noise source imply that the optima conditions are rarely fulfiled. So, we can just affirm that the real short term stabilities can not be higher than those presented in the table.
- The Q-factor of the first BVA tested is bad because the quartz have been irradiated (4).

- The two "μAT-cut" resonators (working at 10 MHz), realized by "bva INDUSTRIE" are different than "tactical" BVA. Indeed, their entire volumes are about 7 times smaller. So, it is not possible to compare their short term stabilities with these of other ones having the same Q-factor (10).
- At least, the last result presented is extracted from (9) and is measured in oscillator by J.R. Norton. So, it is not directly comparable with the others.

CONCLUSION

A systematic study of the design and the manufacturing of the "tactical" mini-BVA has permitted us to implement yet some improvements in its realization.

The determination of the keypoints of the design is completed. They concern, for almost, the problems of symmetry of the structure, around Z-axis and the median plane of the resonator. On the other hand, their influences on certain performances are not achieved. We have, particularly, to performe other measurements about the influence of the raw material on the long term drift. The initial conditions, i.e. the pre-aging (included in the manufacturing process), appear as essential to accelerate this aging, the structure being able, as for its, to exhibit low drifts.

If the frequency adjustement is not, up to now, well defined it is only du to the measurement of the temperature room. U. Peir presents, in (11), some encourageing results which let hope improvements in production.

The last results are just typical values :

the accelerometric sensitivity is less than $10^{-10}/g$

the barometric sensitivity is about 5.10^{-11}/bar.

As for the sensitivity to a magnetic field, some measurements are now in progress.

REFERENCES

1. R.J. BESSON : "A new electrodeless resonator design" - 31st Annual Freq. Control Symp. (1977), pp. 147, 152

2. R.J. BESSON and al., French patent 8509097 and US patent 4.705 982 (11-10-1987)

3. R. DELAITE : "Special miniaturized SC-cut quartz crystal resonator for severe environnements applications", Proc. 3rd European Frequency and Time Forum. (1989), pp. 296, 300

4. J.J. BOY, A. BERTHAUT, D. THIEBAUD, K. YAMNI : "Long term drift and γ-irradiation effect", Proc. 6th European Frequency and Time Forum. (1992), pp. 93, 98

5. A.W. WARNER, D.B. FRASER, C.D. STCKBRIDGE : "Fundamental studies of aging in quartz resonators", IEEE Trans. on Sonics and Ultrasonics (1965)

6. R.L. FILLER, J.A. KOSINSKI, V.J. ROSATI, J.R. VIG : "Aging studies on quartz crystal resonators and oscillators", 38th Annual Freq. Control Symp. (1984), pp. 225, 231

7. M.R. MIJKOVIC et al. : "Aging prediction of quartz crystal units", 42nd Annual Freq. Control Symp. (1988), pp 404, 411

8. F.L. WALLS, A.E. WAINWRIGHT : "Measurement of the short term stability of quartz crystal resonators and the implication for crystal oscillator design and apllications", IEEE trans. on instrumentation and measurements - Vol. IM-24 n°1, March 1975

9. J.R. NORTON, R.J. BESSON : to be published in proc. AFCS (1993)

10. F.L. WALLS, R.J. BESSON, J.J. GAGNEPAIN : "A new model of 1/F noise in BAW quartz resonators", IEEE frequency control symposium (1992), pp. 327, 333.

11. U.R. PEIER : "Automatic temperature test system for BVA resonators based on IEC 444", Proc. 1st European Frequency and Time Forum. (1987), pp. 136, 142

COMPLEX MASS-LOADING EFFECTS IN AT-CUT QUARTZ CRYSTAL RESONATORS

Irina Mateescu, Ioan Mateescu

Institute of Physics and Technology of Materials
PO BOX MG-7, Bucharest - Magurele, ROMANIA

1. ABSTRACT

The harmonic dependence of the equivalent electrical inductance of AT-cut quartz resonators for different mass-loading was studied.
CrNi, CrNiAu and CrAu electrodes have been used with thickness between 750 Å - 3000 Å, which were deposited on optically polished and unpolished quartz plates.
Electrode deposition was made by 2 methods : thermal evaporation in vacuum and radiofrequency plasma sputtering.
The assigning of inductance dependence to the non-uniform distribution of motion of AT-cut quartz plates under electrodes, is supported by the correlation of the results for these electrode depositions, with mass-loading, adhesive strength and residual stress.
The harmonic dependence of inductance for CrAu electrodes has also been analyzed for various temperature treatments (150°C, 250°C, 350°C).

2. INTRODUCTION

Single crystal quartz has been successfully used as material for crystal resonators. Among the different cuts of the quartz crystal, which offer different vibration modes patterns over a frequency range, AT-cut remains one of the most important due its specific characteristics:
To be able to fulfill specific application demands, one has to know very well the influence of different technological factors on parameters of resonators.
Of main importance are the effects associated to deposition of thin metallic electrodes on the piezoelectric plates.
The interesting studies were made about effects of mass-loading on crystal plates and mass-loading measurements of the resistivity, adhesive strength and residual stress of the film metallizations on crystal quartz [Ref. 3], about the effects of intrinsic stresses in the electrode films, the changing of the temperature coefficient of frequency as a function of increasing mass-loading [Refs. 4 - 7].
The electrode resistivity was used as a tool in the investigation of the physical properties of the thin metallic films [Ref. 8].
In our previous papers we have analyzed effects of mass-loading on electrical parameters of AT-cut quartz crystal resonators [Ref. 9] and the non-uniform distribution of motion influence on the effective mass-loading in AT-cut quartz resonators [Ref. 10].
In the present paper we try to discuss the harmonic dependence of the equivalent electrical inductance L of the AT-cut quartz resonators for different mass-loading.
We used different conditions for deposition, some electrode materials, polished and unpolished plates and, for CrAu electrodes, various temperature treatments.
We suggest a connection between harmonic inductance dependence on the adhesive strength and residual stress in thin metallic electrodes deposited on the surface of a piezoelectric plate.
This dependence is made through the harmonic variation of active area $\mathbf{A_a}$, the effective thickness of the plate $\mathbf{h_{eff}}$ and the piezoelectric stress constant **e** variation.

3. EXPERIMENTAL

We used AT-cut cultured Sawyer quartz, plan-parallel plates with $\theta = 35°17'$
The plates thickness has been 0.34 mm, the diameter 14 mm, the electrode diameter 7 mm.
Some plates were optically polished and the other only grinned with 12 μ abrasive.
CrNi, CrNiAu electrodes have been deposited on these plates through thermal evaporation in vacuum with thickness between 1000 Å and 3000 Å.
The deposition was monitored in each run by a MIKI - MSV 1841 vibrational quartz thickness / rate monitor.
The same electrodes with the same thickness were deposited by radio-frequency sputtering in Argon atmosphere ($3 \cdot 10^{-3}$ Torr and 50 W power) with the sputtering targets Cr 20 %, Ni 80% and Au.
CrAu electrodes with thickness between 750 Å - 2000 Å have also been deposited on quartz plates by evaporation in vacuum.
After the first set these CrAu electrodes plates were subsequently been annealed at 150°C, 250°C, 350°C.
The resonance and antiresonance frequencies series (1, 3, 5, 7) of plates mounted in supports, were measured for every mass-loading and after each step of annealing, for the plates with CrAu electrodes.
L, C parameters have been computed for all overtones.
Measurements of series of frequencies and equivalent resistance were made using a HP 4194 A analyzer and for $\mathbf{C_0}$ was used a Wayne-Kerr bridge.

4. RESULTS AND DISCUSSION

Analyzing the harmonic and mass-loading dependency of equivalent dynamic inductance **L** of AT-cut crystal resonators we observed the following :
- the equivalent inductance **L** increases with overtone order;

the increase is greater when the mass-loading is higher, but the main frequency inductance decreases slowly with the mass-loading;

- a higher inductance increase was observed when the unpolished quartz plates were replaced by polished plates (Fig. 1).

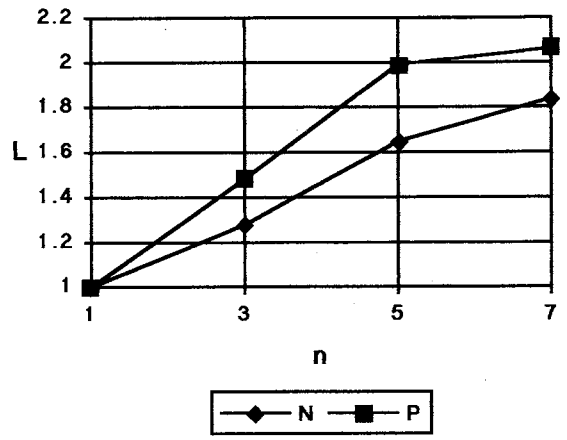


Fig. 1a - Normed inductance for polished (P) and unpolished plates; CrNi electrodes

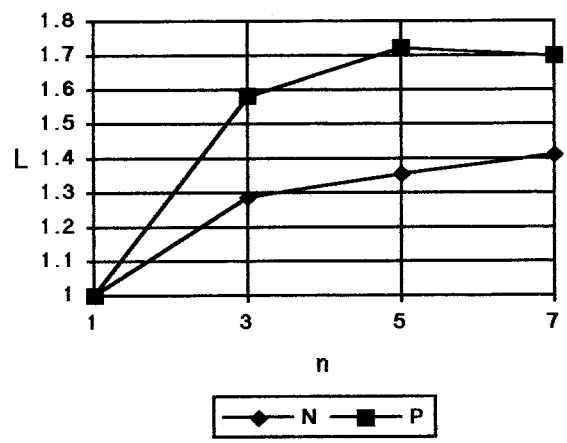


Fig. 1b - Normed inductance for polished (P) and unpolished plates; CrNiAu electrodes

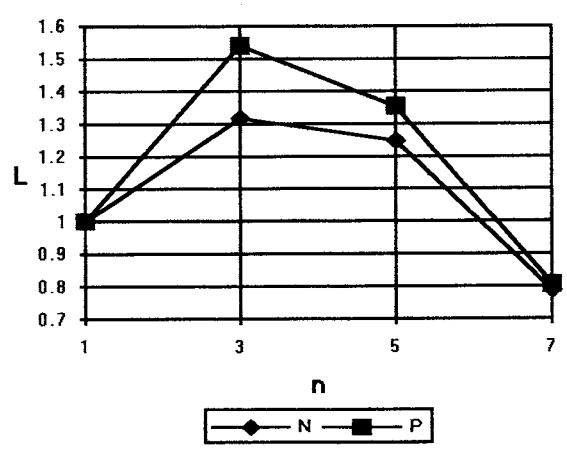


Fig. 1c - Normed inductance for polished (P) and unpolished plates; CrAu electrodes

The same effect was obtained when the electrode deposition by evaporation in vacuum was used instead of deposition by radiofrequency sputtering (Fig. 2).

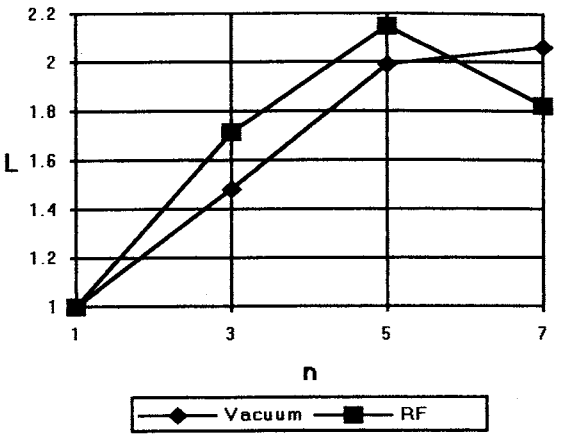


Fig. 2a - Normed inductance for vacuum (V) and sputtering (RF) deposition; CrNi electrodes

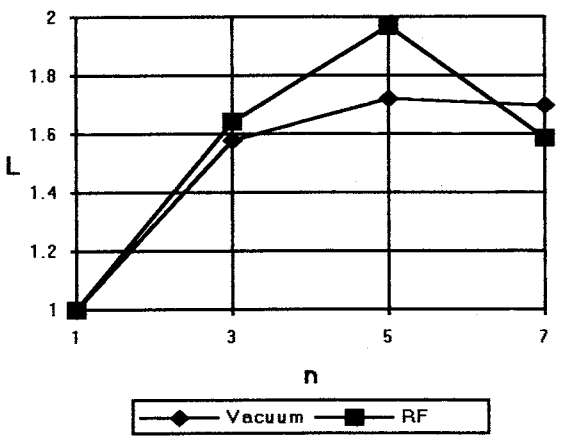


Fig. 2b - Normed inductance for vacuum (V) and sputtering (RF) deposition; CrNiAu electrodes

Relating these observations with previous results [Ref. 10], we assign this tendency of dynamic inductance **L** to increase with the order of overtone, to the decrease of the active area of AT-cut quartz resonators due to the non-uniformity of motion of quartz plates under electrodes. However, in certain conditions the inductance increase is followed by a decrease at high harmonics. In many cases a stronger initial increase is associated as well with stronger decrease.

To establish the electrode effects on this decrease-tendency we made measurements on resonators with electrodes made of different materials but the same thickness (2000 Å).

Figure 3 presents harmonic inductance dependency for three electrode materials, CrNi, CrNiAu, CrAu.

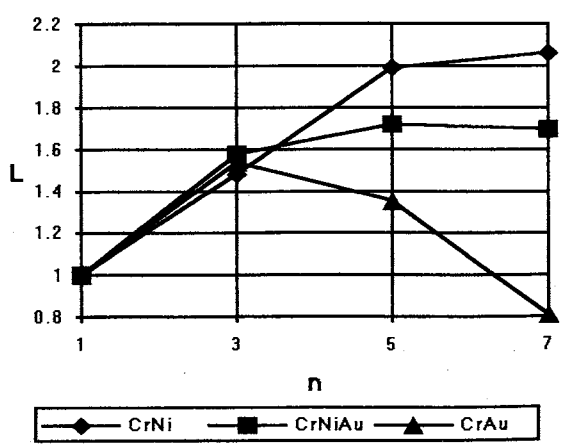
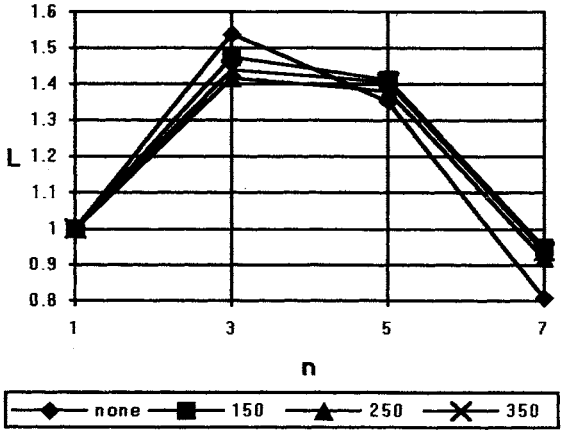


Fig. 3 - Normed inductance for CrNi, CrNiAu, CrAu electrodes

In Figure 4 and Figure 5 we observe inductance dependency on the thermal treatment of resonators with CrAu (2000 Å thickness) deposited by evaporation in vacuum on polished and unpolished plates.



Fi

g. 4- Normed inductance for different treatment temperature; CrAu electrodes; polished plates

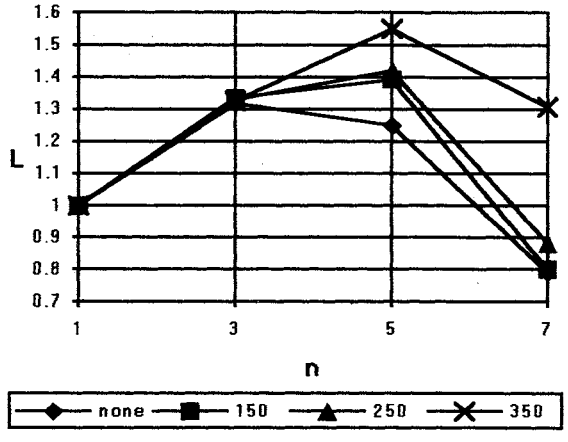


Fig. 5 - Normed inductance for different treatment temperature; CrAu electrodes; unpolished plates

It is known [Ref. 3] that the as-deposited CrNiAu electrode is more adherent than CrAu electrode and that the heat treatment improves the adherence of CrAu electrodes. With the treatment temperature increase the residual stress of the as-deposited CrAu film, which was tensile, becomes comprehensive.

Our results suggest that the delay of appearance of decreasing tendency is associated with improvement of the adherence.

We tried to explain the obtained results by using the relation between the equivalent electrical inductance **L** and the physical parameters [Ref. 11].

$$L = \frac{h_{eff}^3 \cdot \rho}{A_a \cdot e^2}$$

where

h_{eff} - a half of effective plate thickness;

ρ - quartz density;

A_a - active area of the electrode;

e - stress piezoelectric constant of AT-cut crystal.

According to the previously mentioned assumptions, we associate the increasing tendency of the inductance with the decreasing of the active area $\mathbf{A_a}$ with overtone order. This decrease is associated to the non-uniform distribution of motion of the quartz plates under the electrodes.

The mass-loading, the polishing of the plates and the radiofrequency sputtering of the electrodes have similar effects an active area.

However, the decreasing tendency cannot be attributed to this variation of the active area, which is continuously decreasing with harmonics [Ref. 10].

To explain this decreasing inductance we have to take into account the following effects :

- in many cases the decreasing tendency of the inductance is associated with fast previous increase;
- the increase tendency is enhanced by mass-loading increase;
- in contrast with the active area, harmonic mass-loading dependency presents a limitation or a decay at high overtones.

Our results show that inductance decreases at high overtones seems to be correlated with the decrease of the electrode support stresses from compression to tension.

According to inductance equation this decrease tendency should be connected with the increase of piezoelectric stress constant **e** or with the decrease of the effective plate thickness $\mathbf{h_{eff}}$ for high overtones.

In conclusion, we suggest that the inductance decrease effect is related with mass-loading dependency an adhesion and stress conditions which can influence the effective thickness of the plates and piezoelectric stress constant for harmonic orders.

The presented results should be taken into account in any model for the effects of the electrode in AT-cut resonators at fundamental and overtones frequencies.

ACKNOWLEDGMENTS

The authors wish to thank Dr. C. Popescu for useful discussions concerning this work and Dr. D. Serbanescu and G. Korony for their help in sputtering and vacuum depositions.

We thank also M. Barbos and C. Bora for computer processing of data and the piezoelectric devices team for their specific contribution.

REFERENCES

1. Kosinski, S. Mallikarjun, A. Ballato, Mass-loading measurements of quartz crystal plates, Proc. 43rd ASFC, 1989
2. Vig, T. R. Meeker, The aging of bulk acoustic wave resonators, filters and oscillators, Proc. 45th ASFC, 1991
3. T. Vianco, W. R. Conley, J. K. G. Panitz, Resistivity, adhesive strength and residual stress measurement of thin film metallizations on single crystal quartz, Proc. 44th ASFC, 1990
4. P. EerNisse, Quartz resonator frequency shifts arising from electrode stress, Proc. 29th ASFC, 1975
5. H. Sherman, Temperature coefficient of the frequency shift arising from electrode film stress, Trans. on Sonics and Ultrasonics, vol. 30, no. 2 pp. 104 - 110, March 1988.
6. L. Filler, J. R .Vig, Fundamental mode SC-cut resonators, Proc. 34th ASFC, 1980
7. Ballato, T. Lukaszek, Higher order temperature coefficients of frequency of mass loaded piezoelectric crystal plates, Proc. 29th ASFC 1978
8. I. Maissel, R. Glang, Handbook of Thin Film Technology, McGraw Hill, 1970
9. Mateescu, Effects of mass-loading on electrical parameters of AT-cut quartz resonators, 6th EFTF, Noordwijk, 1992

Noordwijk, 1992
10. Mateescu, E. Candet, Non-uniform distribution of motion influence on the effective mass-loading in AT-cut quartz resonators, Proc. 46th ASFC, 1992
11. Hafner, Crystal Resonators, Trans. on Sonics and Ultrasonics vol. SU-21, no. 4, 1974

TELECOM APPLICATIONS

Chairman: A. Reid

Stability Requirements of Oscillators Used for Synchronization Loops in Communications Systems

A. L. Welti
Swiss Federal Institute of Technology
Communication Technology Laboratory, ETH Zentrum
8092 Zurich, Switzerland

ABSTRACT—The imperfect reproductions of reference clocks in receivers of modern communications systems degrade the performance of synchronization loops and hence, the overall transmission quality. In this paper, the phase-locked loop (PLL) for the carrier recovery and the delay-locked loop (DLL) for the timing recovery of the pseudonoise (PN) code in spread spectrum systems are analyzed regarding the impacts of imperfect oscillators (having a phase noise and a frequency offset between transmitter and receiver) and channel noise on the mean exit time. Optimal design criterions for the second-order loops depending on the accuracy and stability of the oscillators are given to achieve the maximum mean exit time.

I. Introduction

Almost every modern communication or transmission system employs some kind of synchronization device in the receiver to guarantee a "synchronous" operation with respect to the incoming signal or to generate a frequency or timing reference. A very common device is the PLL which estimates and tracks, depending on the loop structure, the unknown time varying phase of the received carrier signal. Another important tracking loop represents the DLL which is widely used in positioning and navigation systems (e.g., global positioning system, GPS) or in spread spectrum communication systems [2], [4], [9], [18] to guarantee that the replica of the spreading code in the receiver is time synchronous to the received code subject to a time varying propagation delay time, e.g., in a mobile communication channel.

The PLL and DLL are both highly nonlinear devices. Because of this fact it is possible that these tracking loops can lose their synchronism if they are disturbed by some interference signal, e.g., disturbances in the communication channel. The mean time to lose synchronism (in the following called mean exit time) serves as a performance criterion which takes the nonlinear behavior of such loops fully into account.

These loops employ voltage controlled oscillators (VCO). In practice, such a VCO has a certain inaccuracy in its nominal frequency or some frequency drift (long-term frequency instability). Depending on the loop structure such an uncertainty with respect to the reference clock in the transmitter can result in a nonzero steady state synchronization error. This error reduces dramatically the mean exit time of the tracking loop and will also increase the probability of bit errors in the data demodulator, see [14] for the case of a first-order DLL. Another impact on the performance of the loop is the VCO phase noise (short-term instability). Besides the noise in the transmission channel this adds another noise component to the loop error signal and reduces the performance even more. Therefore, the synchronization loops should be designed optimally under the constraints of channel disturbances *and* imperfect oscillators.

Another problem arises from the facts that on one side the VCO must be tunable within some (may be large) range depending on the application, and that on the other side it should have a low frequency instability to get a satisfactory tracking performance. These are contradictory conditions because a VCO with large tuning range is usually not very accurate in its frequency and vice versa. For a practical system, a good compromise between tuning range and stability must be found.

In this paper, it will be shown for the second-order PLL and DLL how drastically the mean exit time decreases due to channel noise, phase noise of the VCO, and frequency offset between transmitter and receiver. Simple rules for the requirement of the VCO stability are given such that the mean exit time of the loop is maximum.

II. Synchronization Loops

In this Section, the mathematical models of the second-order PLL and DLL will be derived resulting in a set of nonlinear stochastic differential equations.

A. Phase-Locked Loop (PLL)

The PLL is widely used in communications systems to generate a phase coherent local reference carrier in the receiver. Figure 1 shows the block diagram of this loop with the received signal $r(t)$ given by

$$r(t) = \sqrt{2P}\,\sin\left(\omega_0 t + \varphi(t)\right) + \sqrt{\frac{N_0}{2}}\,n(t) \qquad (1)$$

where P denotes the average signal power, ω_0 stands for the carrier radian frequency, φ represents an unknown time varying phase subject to carrier frequency instabilities (long-term and short-term) and Doppler effects between the transmitter and receiver. This paper focuses only on the oscillator instabilities and discards Doppler effects. The disturbance to the transmitted signal in the communication channel is modeled in (1) by additive white Gaussian noise (AWGN)

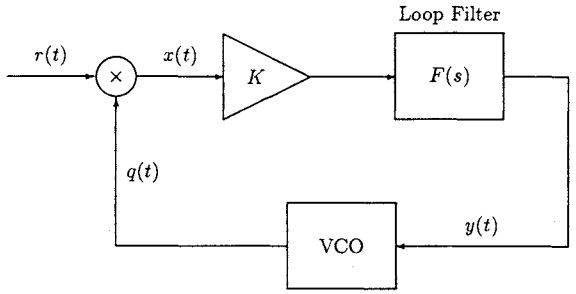


Figure 1: Block diagram of the PLL.

with a two-sided power spectral density of $N_0/2$ W/Hz and $n(t)$ describes a standard white Gaussian noise, i.e., having unit intensity. The output signal of the VCO is of the form

$$q(t) = \sqrt{2}\cos(\omega_0 t + \hat{\varphi}(t)) \tag{2}$$

with $\hat{\varphi}(t)$ denoting the phase estimate of $\varphi(t)$. Multiplying this by $r(t)$ results in [11]

$$x(t) = \sqrt{P}\, S(\varepsilon) + \sqrt{\frac{N_0}{2}}\,\tilde{n}(t) \tag{3}$$

with the phase error (synchronization error) $\varepsilon(t)$ and the discriminator characteristic (S-curve) $S(\varepsilon)$ given by

$$\varepsilon(t) = \varphi(t) - \hat{\varphi}(t)\,, \qquad S(\varepsilon) = \sin(\varepsilon)\,, \tag{4}$$

and $\tilde{n}(t)$ denotes again a standard white Gaussian noise process. By using a so-called passive loop filter with time constant τ of the form

$$F(s) = \frac{1}{s\tau + 1} \tag{5}$$

and the gain factors of the individual blocks lumped into K we arrive at the baseband model of the PLL shown in Fig. 2. Such a loop filter is commonly used in practice to suppress parasitic harmonics of the carrier frequency leaking through the mixer, for example. The VCO is modeled as a pure integrator and hence,

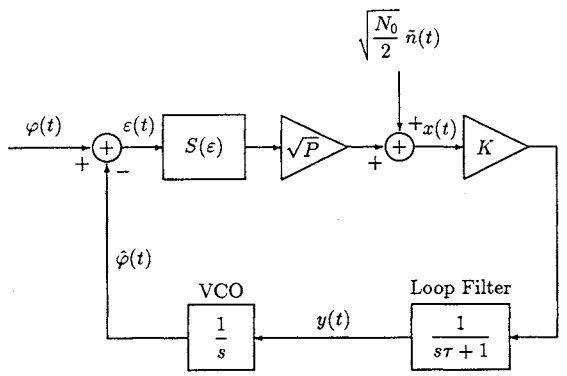


Figure 2: Baseband model of the PLL.

the PLL turns out to be a second-order type I (having one pure integrator) loop. From this Figure it can be easily recognized that the PLL is a nonlinear control device under the influence of a stochastic disturbance (in this case white Gaussian noise). From Fig. 2 we get

$$\left.\begin{aligned} \dot{y}(t)\tau + y(t) &= Kx(t) \\ y(t) &= \dot{\varphi}(t) - \dot{\varepsilon}(t)\,. \end{aligned}\right\} \tag{6}$$

The received phase $\varphi(t)$ can vary in time because of Doppler effects influencing the transmitted signal. Also, a linear time dependent phase shift between the transmitter and receiver will occur, if there is a constant offset $\Delta\omega_0$ between the transmitted carrier radian frequency and the locally generated VCO frequency. In practice, this is always the case since two oscillators can never be built such that the frequencies match precisely. Of course, there are some other means causing a time varying phase. A constant frequency offset results in

$$\varphi(t) = \varphi_0 + \Delta\omega_0 t \tag{7}$$

where φ_0 describes some constant phase. From (6) and (7) we get

$$\left.\begin{aligned} \dot{\varepsilon}(t) &= \Delta\omega_0 - y(t) \\ \dot{y}(t) &= -\frac{y(t)}{\tau} + \frac{K\sqrt{P}}{\tau}S(\varepsilon) + \frac{K}{\tau}\sqrt{\frac{N_0}{2}}\,\tilde{n}(t) \end{aligned}\right\} \tag{8}$$

which describe the baseband model of the PLL in Fig. 2 by two nonlinear stochastic differential equations of first-order. By applying the same time scaling $t^\star = \alpha t$ as in [13] and following similar steps as in [17] with the definitions

$$\left.\begin{aligned} &z(t^\star) = \frac{1}{\alpha}\left[\Delta\omega_0 - y(t^\star)\right] \\ &\nu(t^\star) = -\tilde{n}(t^\star)\,, \qquad \tilde{K} = K\sqrt{P}\,, \qquad \alpha = \sqrt{\frac{\tilde{K}}{\tau}} \\ &u = \frac{\Delta\omega_0}{\tilde{K}}\,, \qquad \beta = \frac{1}{\sqrt{\tau\tilde{K}}}\,, \qquad \rho = \frac{\tilde{K}N_0}{4P} \end{aligned}\right\} \tag{9}$$

we finally arrive at the time scaled nonlinear stochastic differential equations:

$$\left.\begin{aligned} \dot{\varepsilon}(t^\star) &= z(t^\star) \\ \dot{z}(t^\star) &= -\beta z(t^\star) - S(\varepsilon) + u + \sqrt{2\beta\rho}\,\nu(t^\star) \end{aligned}\right\} \tag{10}$$

or in the form of a Langevin equation:

$$\ddot{\varepsilon}(t^\star) + \beta\dot{\varepsilon}(t^\star) + S(\varepsilon) - u = \sqrt{2\beta\rho}\,\nu(t^\star)\,. \tag{11}$$

The PLL with $S(\varepsilon) = \sin(\varepsilon)$ has theoretically infinitely many stable equilibrium points (ε_a, z_a) which are called attractors and infinitely many unstable equilibrium points (ε_b, z_b) which turn out to be saddle points. They can be found from (10) for $\rho = 0$, $\dot{\varepsilon} = 0$, $\dot{z} = 0$, and they exist only if $|u| \leq 1$. For $k = 0, 1, 2, \ldots$

it turns out:

$$\left.\begin{array}{l} z_a = z_b = 0 \\ \varepsilon_a = \pm 2k\pi + \arcsin(u) \end{array}\right\}, \quad |u| \le 1$$

$$\varepsilon_b = \begin{cases} (2k+1)\pi - \arcsin(u), & u > 0 \\ -(2k+1)\pi - \arcsin(u), & u < 0 . \end{cases} \qquad (12)$$

B. Delay-Locked Loop (DLL)

In spread spectrum systems, the synchronization of the locally generated PN-code to the received code is of fundamental importance to perform the despreading and data demodulation in the receiver. After code acquisition (coarse synchronization) the PN-code tracking loop (fine synchronization) takes over to track small changes in the code delay due to Doppler effects, channel noise, and relative drifts in the code clock between the transmitter and receiver. A well known PN-code tracking loop is the DLL, [2], [9], [10], [18]. In this paper we will consider the coherent DLL as shown in Fig. 3 with the same loop filter $F(s)$ as in (5) in order to suppress some leakage harmonics of the carrier frequency. Such a loop generates an error signal $e(t)$ which on the average corresponds to the difference of two with respect to each other shifted autocorrelation functions (ACF) of the PN-code and thus, having a zero crossing midpoint. The error signal will be amplified by a factor K (gain factors of the individual blocks lumped into it) and after passing the loop filter it is fed to the VCO to generate the clock for the PN-code generator. The received signal is modeled as

$$r(t) = \sqrt{P}\, c(t - T(t)) + \sqrt{\frac{N_0}{2}}\, n(t) \qquad (13)$$

where P names the average signal power, $c(t)$ represents the transmitted spreading code with code rate $R = 1/T_c$, T_c stands for the duration of a code chip, and $T(t)$ denotes the instantaneous propagation delay time of the channel. The additive noise term has a two-sided power spectral density of $N_0/2$ W/Hz where $n(t)$ is a standard white Gaussian noise. A locally generated early code $c(t - \hat{T} + 0.5T_c)$ and the late code $c(t - \hat{T} - 0.5T_c)$ which are in this example shifted by T_c relative to each other (in fact, it can be any shift in the range $0 \ldots 2T_c$) are multiplied by $r(t)$ and then

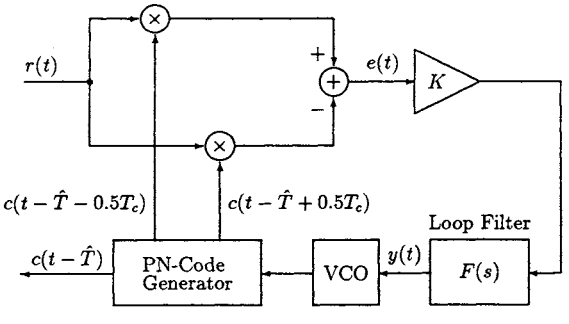


Figure 3: Block diagram of the coherent DLL.

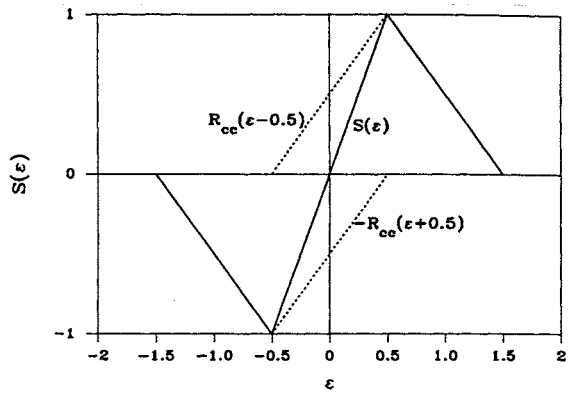


Figure 4: Discriminator characteristic of the DLL.

subtracted to give the loop error signal, which bears a baseband signal [18]

$$e(t) = \sqrt{P} S(\varepsilon) + \sqrt{N_0}\, \tilde{n}(t) \qquad (14)$$

and $\tilde{n}(t)$ denotes a standard white Gaussian noise. The timing error of the loop (synchronization error) normalized by the code chip time is given by

$$\varepsilon(t) = \frac{T(t) - \hat{T}(t)}{T_c} \qquad (15)$$

where $\hat{T}(t)$ defines the loop estimate of $T(t)$. It turns out that the nonlinearity $S(\varepsilon)$ describing the discriminator characteristic (S-curve) results in

$$S(\varepsilon) = R_{cc}(\varepsilon - 0.5) - R_{cc}(\varepsilon + 0.5) \qquad (16)$$

and $R_{cc}(\varepsilon \pm 0.5)$ denotes a shifted version (by $0.5T_c$) of the ACF of the spreading code $c(t)$, Fig. 4. The periodicity of the ACF of the PN-code plays not an important role since the code tracking loop works only in a limited range of ε. Outside this range the restoring force becomes zero and hence, a lock detector should indicate that a new acquisition process must be started.

In Fig. 5 the corresponding baseband model of the DLL is shown. Considering the relative propagation delay time $T(t)/T_c$ to change linearly in time due to

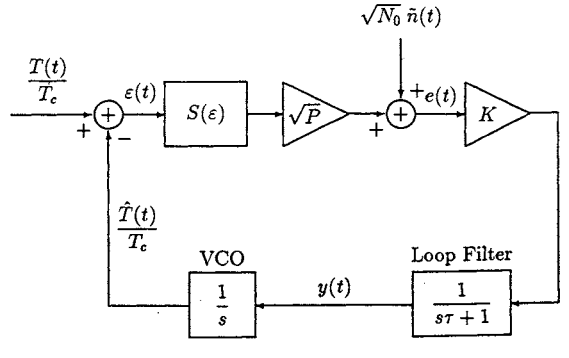


Figure 5: Baseband model of the DLL.

some constant offset ΔR (inaccuracy) between the transmitted and locally generated code clock rate we get

$$\frac{T(t)}{T_c} = a_0 + \Delta R\, t \tag{17}$$

where a_0 represents a constant delay time. Due to the similarity of the baseband models in Fig. 2 and Fig. 5 we can follow the same derivation steps as in the previous Section to arrive at the two stochastic nonlinear differential equations:

$$\left.\begin{aligned} \dot{\varepsilon}(t) &= \Delta R - y(t) \\ \dot{y}(t) &= -\frac{y(t)}{\tau} + \frac{K\sqrt{P}}{\tau} S(\varepsilon) + \frac{K}{\tau}\sqrt{N_0}\,\tilde{n}(t)\,. \end{aligned}\right\} \tag{18}$$

Using the same time scaling $t^\star = \alpha t$ and new definitions

$$z(t^\star) = \frac{1}{\alpha}\left[\Delta R - y(t^\star)\right]\,,\quad u = \frac{\Delta R}{\tilde{K}}\,,\quad \rho = \frac{\tilde{K} N_0}{2P} \tag{19}$$

we finally get equation (10) or the Langevin equation (11). For the DLL these equations have equilibrium points for $\rho = 0$, $\dot{\varepsilon} = 0$, $\dot{z} = 0$, and they exist only if $|u| \leq 1$. It turns out that there is one stable point attractor at (ε_a, z_a) given by

$$\varepsilon_a = \frac{u}{S'(\varepsilon_a)}\,,\qquad z_a = 0 \tag{20}$$

and for $u \neq 0$ there exists one saddle point at (ε_b, z_b) with

$$\varepsilon_b = \begin{cases} 1.5 + \dfrac{u}{S'(\varepsilon_b)}\,, & u > 0 \\ -1.5 + \dfrac{u}{S'(\varepsilon_b)}\,, & u < 0 \end{cases}\,,\qquad z_b = 0 \tag{21}$$

and from Fig. 4 we conclude that $S'(\varepsilon_a) = 2$, $S'(\varepsilon_b) = -1$. For $u = 0$ there are two saddle points with $\varepsilon_b = \pm 1.5$ and $z_b = 0$.

III. Mean Exit Time

In the previous Section, we derived the mathematical models for the PLL and DLL which can be described by the same nonlinear stochastic differential equations. The only difference lies in the definitions of the parameters and the shapes of the S-curves.

Considering first the DLL, the stable equilibrium point represents an attractor to all trajectories starting in some neighborhood D to it (domain of attraction) and the boundary of D is called separatrix ∂D [12], [13], [17]. A trajectory of the nonlinear dynamical system disturbed by noise starting at the attractor will fluctuate for some finite time (may be very long) around this equilibrium point and eventually will reach ∂D. If the trajectory crosses ∂D the system state exits from its locked region and the DLL loses synchronism (loss of lock). Without a new acquisition process there is no way of returning to the synchronous operation. For the DLL the average time to lose synchronism from its attractor is called mean time to lose lock.

In the case of the PLL, if the perturbation propels the loop operating point to cross ∂D, the loop state will settle in a stable region with a new attractor 2π (or a multiple thereof) separated from the initial domain because of the PLL's periodic discriminator characteristic, i.e., $S(\varepsilon) = \sin(\varepsilon)$. In this case the average time to leave one attractor is called the mean time to first cycle slip. Here, in the context of dealing with both loops let us just call this time *mean exit time*.

In 1940, H. A. Kramers [5] (see also [7], [12], [17]) showed that for systems described by the Langevin equation the mean exit time $\bar{t}_L$ is for $\beta \gg 1$ given by

$$\bar{t}_L = \frac{C_1}{\alpha}\frac{2\pi\beta}{\omega_A \omega_C}\exp\left(\frac{E_c}{\rho}\right) \tag{22}$$

with α, β, and ρ defined in (9), (19) and

$$\left.\begin{aligned} \omega_A = \sqrt{S'(\varepsilon_a)} &= \begin{cases} \sqrt{\cos(\varepsilon_a)}\,, & \text{PLL} \\ \sqrt{2}\,, & \text{DLL} \end{cases} \\ \omega_C = \sqrt{-S'(\varepsilon_b)} &= \begin{cases} \sqrt{\cos(\varepsilon_a)}\,, & \text{PLL} \\ 1\,, & \text{DLL}\,, \end{cases} \end{aligned}\right\} \tag{23}$$

whereas $C_1 = 0.5$ for $u = 0$ and $C_1 = 1$ for $u \neq 0$. For simplicity, we use (22) also for intermediate values of β because it appears in the pre-exponent only and hence, its influence to the exit time is not dominant. The expression E_c is obtained by

$$E_c = \int_{\varepsilon_a}^{\varepsilon_b} \left[S(\varepsilon) - u\right] d\varepsilon\,. \tag{24}$$

Therefore, E_c denotes the area between the S-curve and the normalized loop offset u as shown in Fig. 6a for the PLL and Fig. 6b for the DLL.

Rule of thumb: From (22) we can immediately deduce a rule of thumb on how accurate the oscillator should be. In order that the mean exit time takes on reasonable values, the exponent in (22) must be dominant, i.e., the area in Fig. 6 corresponding to E_c should be as large as possible, i.e., for $u = 0$, or equivalently $\Delta\omega_0 = 0$ for the PLL or $\Delta R = 0$ for the DLL meaning that the transmitter and receiver oscillators match precisely in the frequency which is never the case in practice. In numerous investigations [12], [13], [14], [17] it was shown that a reasonable tradeoff for an acceptable mean exit time is $|u| \approx 0.3$. With this condition the needed accuracy of the oscillators follow from (9) and (19):

$$\left.\begin{aligned} |\Delta\omega_0| &\leq 0.3K\sqrt{P}\,, \quad \text{PLL} \\ |\Delta R| &\leq 0.3K\sqrt{P}\,, \quad \text{DLL}\,. \end{aligned}\right\} \tag{25}$$

Hence, knowing the received signal power P, which can be kept at some constant level by a proper automatic gain control (AGC) [15], [16], and adjusting the necessary loop gain we can immediately "buy" the oscillator with the appropriate accuracy that fits our purpose.

Maximum Mean Exit Time: Next, it is shown that the loop parameters can be chosen such that the exponent of the mean exit time in (22) has a maximum leading to optimal parameters. Carrying out the integration in (24) and making use of (12), (20) and (21) it turns out that

$$E_c = \begin{cases} 2\cos(|\varepsilon_a|) - |u|\,(\pi - 2|\varepsilon_a|)\,, & \text{PLL} \\ \dfrac{3}{4}(1 - |u|)^2\,, & \text{DLL}\,. \end{cases} \tag{26}$$

Replacing $\tilde{K}$ in the definitions of ρ by $\tilde{K} = \Delta\omega/u$ for the PLL or $\tilde{K} = \Delta R/u$ for the DLL and maximizing E_c/ρ with respect to u results in

$$|u^{\text{opt}}| = \begin{cases} 0.394, & \text{PLL} \\ \frac{1}{3}, & \text{DLL} \end{cases} \tag{27}$$

and

$$E_c^{\text{opt}} = \begin{cases} 0.919, & \text{PLL} \\ \frac{1}{3}, & \text{DLL} \end{cases}. \tag{28}$$

Inserting these optimal values and using the definitions for α, β, and ρ from (9) and (19) and for ω_A and ω_C from (23) we get

$$\tilde{t}_L^{\text{opt}} = \begin{cases} \dfrac{2.69}{|\Delta\omega_0|} \exp\left(\dfrac{1.45}{|\Delta\omega_0|}\dfrac{P}{N_0}\right), & \text{PLL} \\ \dfrac{\sqrt{2}\pi}{3\,|\Delta R|} \exp\left(\dfrac{2}{9\,|\Delta R|}\dfrac{P}{N_0}\right), & \text{DLL}. \end{cases} \tag{29}$$

Results: Plotting the logarithm of (22) versus $\tilde{K}$ by inserting the definitions of α, β, ω_A, ω_C, and ρ gives the results in Fig. 7. It can be recognized that for a certain value of $\tilde{K}^{\text{opt}}$ according to

$$\tilde{K}^{\text{opt}} = \begin{cases} \dfrac{\Delta\omega_0}{u^{\text{opt}}} = 2.54\,|\Delta\omega_0|, & \text{PLL} \\ \dfrac{\Delta R}{u^{\text{opt}}} = 3\,|\Delta R|, & \text{DLL} \end{cases} \tag{30}$$

the loop has an optimal performance. Hence, very simple rules to make an optimal loop design were found. Also, it shows how sensitive the loop performs with respect to its optimum. Figure 8 represents the maximum achievable mean exit time versus the clock inaccuracy $|\Delta\omega_0|$ for the PLL or $|\Delta R|$ for the DLL for different values of P/N_0. On one side, for larger values of P/N_0 the mean exit time is increasing whereas on the other side an increasing oscillator frequency mismatch results in a large degradation in performance. Hence, to assure some minimum exit time, let us say at least 1000 seconds, it can be calculated from (29) or seen from Fig. 8 how much inaccuracy is tolerable for a given P/N_0. For example, if P/N_0 = 13dB Hz

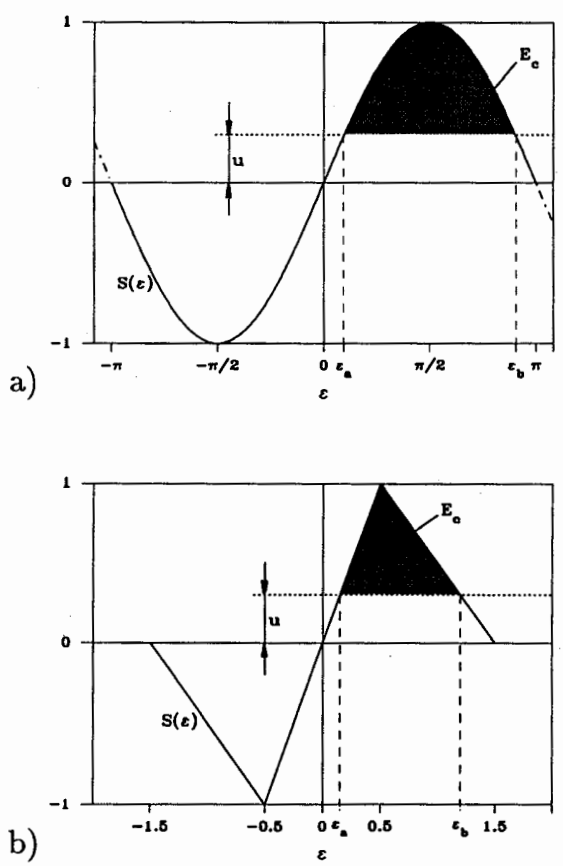


Figure 6: The exponential term E_c as shaded area between the S-curve and the normalized loop offset u in a) for the PLL and in b) for the DLL.

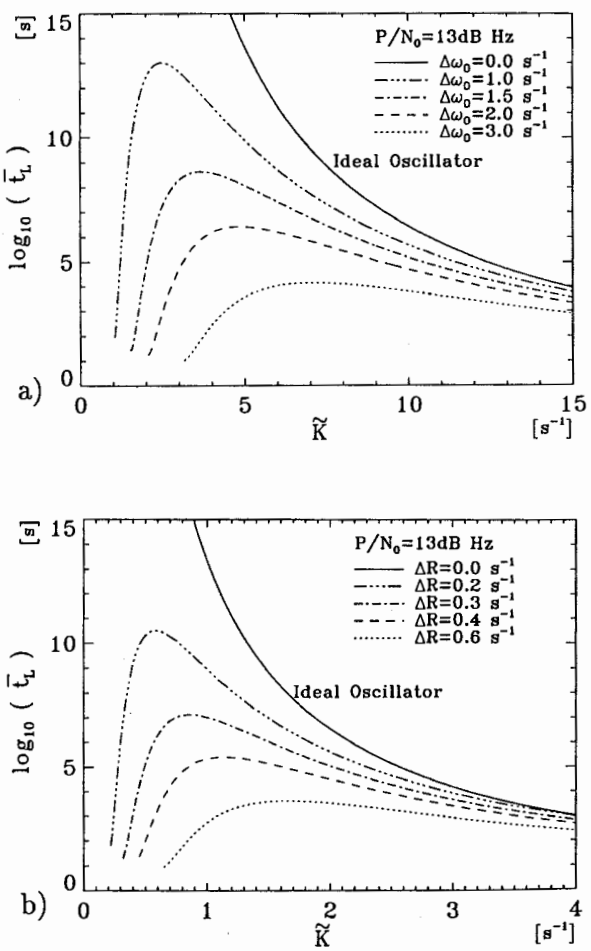


Figure 7: Logarithm of $\tilde{t}_L$ versus $\tilde{K}$ for different values of a) $\Delta\omega_0$ for the PLL and b) ΔR for the DLL.

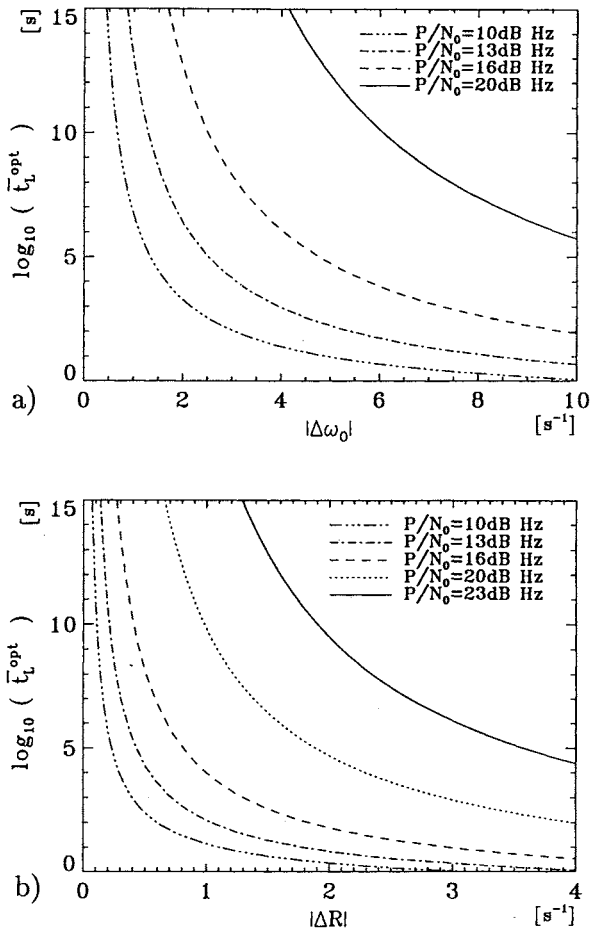


Figure 8: Logarithm of $\bar{t}_L^{\text{opt}}$ versus the clock inaccuracy a) $|\Delta\omega_0|$ for the PLL and b) $|\Delta R|$ for the DLL, for different values of P/N_0.

the inaccuracy must be better than $|\Delta\omega_0| < 4.0\ \text{s}^{-1}$ for the PLL or $|\Delta R| < 0.7\ \text{s}^{-1}$ for the DLL. Hence, this shows also that the DLL acts more sensitive to clock frequency mismatch (absolutely speaking) than the PLL. Of course, the nominal oscillator frequency ω_0 in a practical system is usually much higher than the nominal code rate R. Thus, the resulting relative inaccuracy $|\Delta\omega_0|/\omega_0$ is a more severe problem for the PLL than the relative code clock mismatch $|\Delta R|/R$ for the DLL.

IV. Oscillator Phase Noise and Tracking Performance

An oscillator reference signal can be modeled as [10], [18]

$$x_{\text{osc}}(t) = A\,[1+\mu(t)]\cos\left(\omega_0 t + \varphi(t) + \frac{\gamma t^2}{2}\right) \quad (31)$$

where $\varphi(t)$ is a random phase noise and the term $\gamma t^2/2$ represents the contribution to the phase due to long-term frequency drift (aging, for example). It is further

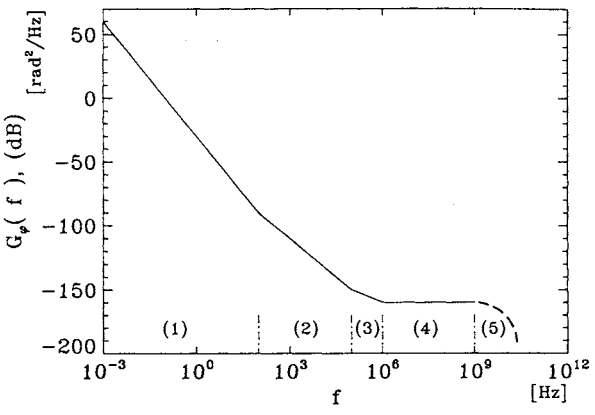


Figure 9: One-sided spectral density of the phase noise of oscillators in the baseband.

assumed that the unwanted amplitude noise $\mu(t)$ can be neglected, i.e., $\mu(t) = 0$. Also, let us focus on the the phase noise $\varphi(t)$ rather than on aging effects, i.e., $\gamma = 0$. In the previous Sections it was assumed that the phase accumulation was governed by a frequency mismatch between the transmitted and locally generated reference frequency. Now, let us consider $\varphi(t)$ to be some random zero mean phase noise produced by the oscillator having a one-sided power spectral density in the baseband denoted by $G_\varphi(f)$ in rad^2 / Hz. A spectral model [1] that has been found widely accepted represents $G_\varphi(f)$ as shown in Fig. 9. This model consists of straight-line asymptotes with several regions defined as:

region 1: frequency flicker, $G_\varphi(f) = k_1/f^3$
region 2: random phase walk, $G_\varphi(f) = k_2/f^2$
(white frequency noise)
region 3: phase flicker, $G_\varphi(f) = k_3/f$
region 4: white phase noise, $G_\varphi(f) = k_4$
region 5: cut off, $G_\varphi(f) \approx 0,\ f > f_m$

where for the example taken from [18] and shown in Fig. 9 the following values are valid:

$$\left.\begin{array}{ll} k_1 = 10^{-3}\,\text{rad}^2\,\text{Hz}^2\,, & k_2 = 10^{-5}\,\text{rad}^2\,\text{Hz} \\ k_3 = 10^{-10}\,\text{rad}^2\,, & k_4 = 10^{-16}\,\text{rad}^2\,/\,\text{Hz} \\ f_m = 10^9\,\text{Hz}\ . & \end{array}\right\} \quad (32)$$

The mechanism leading to flicker type of noise is still open to discussions if it represents a basic physical phenomenon or if it results because of the measuring experiments. Some discussions [6] point out that flicker noise (in fact, any noise with a spectral density of $1/f^B$, where B can be any fractional number) can be explained by the geometry of fractals. With such kind of spectra we arrive at mathematical difficulties since the autocorrelation function does not exist.

The loops considered in the previous Sections were of second-order type I having an imperfect integrator in the loop filter (passive loop filter). In such loops an oscillator frequency mismatch will cause a steady state synchronization error ε_a which degradates the performance compared to the case that this error is zero. Instead, using a so-called active loop filter (with

perfect integrator) described by

$$F(s) = 1 + \frac{K_1}{s} \tag{33}$$

results in a second-order type II (two pure integrators) tracking loop. Now, the frequency mismatch of the oscillator results in a zero steady state error in the loop which of course is a big advantage. But still the oscillator will have some phase noise which influences the tracking behavior. Let us study this phase noise influence in the case of a second-order type II DLL as shown by the baseband model in Fig. 10 (see also [13]). In this Figure, the phase noise of the VCO is added as a random process $\varphi(t)$ to the output signal of a perfect VCO [3], [8]. There are now two noise sources (channel noise and oscillator phase noise) that can cause loss of lock for the DLL. The derivation of the mean exit time for this DLL is given in [12], [13]. Since we do not consider here Doppler effects it turns out that the normalized loop offset a due to a Doppler rate in [13] vanishes, i.e., $a = 0$. Let us also assume that the linearized loop has a damping factor $\zeta = 0.707$. Hence, the mean exit time results approximately in, [13],

$$\bar{t}_L \approx \frac{0.75}{B_L} \exp\left(\frac{0.75\,\hat{\Psi}}{\sigma_\varepsilon^2}\right) \tag{34}$$

where it turns out that $\hat{\Psi} \approx 0.354$ and $\sigma_\varepsilon^2 = \sigma_c^2 + \sigma_\varphi^2$ with σ_c^2 denoting the jitter component penetrating the loop due to the channel noise, i.e.,

$$\sigma_c^2 = \frac{B_L N_0}{2P} \tag{35}$$

and σ_φ^2 due to the oscillator noise. The equivalent loop noise bandwidth B_L is given with $\zeta = 0.707$ by

$$B_L = \frac{\omega_n}{2}\left(\zeta + \frac{1}{4\zeta}\right) = 0.53\,\omega_n \tag{36}$$

and ω_n stands for the loop natural radian frequency. Using the power spectral density from Fig. 9 and the transfer function $H(f)$ of the linearized DLL, σ_φ^2 can be calculated as, [18],

$$\sigma_\varphi^2 = \int_0^\infty G_\varphi(f)\,|1 - H(f)|^2\,df \tag{37}$$

where for $\zeta = 0.707$ it turns out that

$$|1 - H(f)|^2 = \frac{(f/f_n)^4}{1 + (f/f_n)^4} \tag{38}$$

with

$$f_n = \omega_n/(2\pi)\,. \tag{39}$$

In the following, we consider the contribution of frequency flicker, random phase walk and white phase noise of the oscillator. The phase flicker contribution is neglected because we run into mathematical difficulties. Anyway, the major contribution will be governed by the frequency flicker (as will be seen later) which turns out to be

$$\sigma_{\varphi 1}^2 = \int_0^\infty \frac{k_1}{f^3}\frac{(f/f_n)^4}{1 + (f/f_n)^4}\,df = \frac{k_1\pi^3}{\omega_n^2}\,. \tag{40}$$

Similar, the contribution of the random phase walk results in

$$\sigma_{\varphi 2}^2 = \int_0^\infty \frac{k_2}{f^2}\frac{(f/f_n)^4}{1 + (f/f_n)^4}\,df = \frac{k_2\pi^2}{\sqrt{2}\,\omega_n} \tag{41}$$

and the white phase noise assumed that $f_m \gg f_n$ contributes in

$$\sigma_{\varphi 4}^2 \approx k_4 f_m\,. \tag{42}$$

Hence, the total jitter of the DLL is

$$\sigma_\varepsilon^2 = \frac{B_L N_0}{2P} + \frac{8.72 k_1}{B_L^2} + \frac{3.7 k_2}{B_L} + k_4 f_m\,. \tag{43}$$

Inserting this into (34) gives the mean exit time as a function of B_L with P/N_0 as a parameter for the

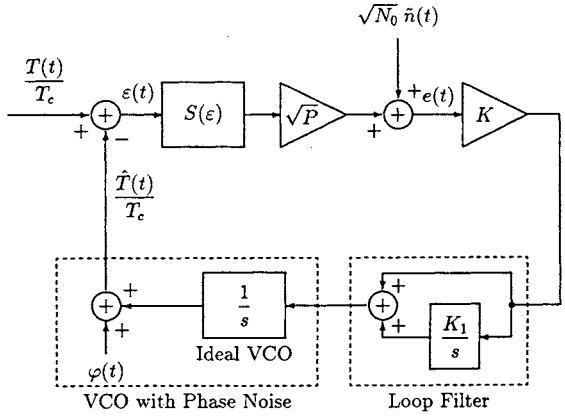


Figure 10: Baseband model of the second-order type II DLL having a VCO with phase noise.

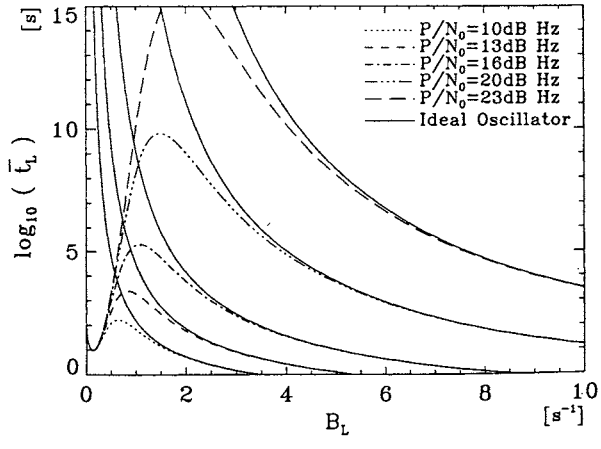


Figure 11: Logarithm of the mean exit time versus B_L for different values of P/N_0 for the second-order type II DLL.

considered example (32), Fig. 11. For a specific P/N_0, there exists (using an imperfect oscillator) an optimal B_L^{opt} which maximizes the exit time or equivalently minimizes σ_ϵ^2 with respect to B_L:

$$B_L^{\text{opt}} = 2.59 \sqrt[3]{k_1 \frac{P}{N_0}} \times \left(\sqrt[3]{1-\sqrt{1-Q}} + \sqrt[3]{1+\sqrt{1-Q}} \right) \approx 3.26 \sqrt[3]{k_1 \frac{P}{N_0}} \quad (44)$$

where Q is defined as:

$$Q = \frac{k_2^3}{20.3\, k_1^2} \frac{P}{N_0} . \quad (45)$$

The approximate choice of B_L^{opt} is valid if $Q \ll 1$, i.e., the frequency flicker is dominant. In this case, the jitter components due to channel noise and oscillator phase noise balance in the following way

$$\left(\sigma_c^{\text{opt}}\right)^2 \approx 2 \left(\sigma_\varphi^{\text{opt}}\right)^2 . \quad (46)$$

Hence, we conclude that the optimal loop bandwidth depends basically on the frequency flicker of the oscillator and on the received signal-to-noise ratio P/N_0.

V. Conclusions

This paper shows how drastically the frequency inaccuracy, the phase noise of the VCO, and the channel noise affect the mean exit time (performance criterion) of the phase-locked loop and the delay-locked loop. Optimal loop parameters depending on these constraints and maximizing the exit time are given in (30). The choice of accurate and stable oscillators is a severe and critical problem in practice. In this context, some simple rules designing the loops optimally were obtained.

A common believe that a loop with narrow bandwidth will perform always better than compared to a loop with larger bandwidth is mistaken, except if the bandwidth approaches zero. But from the system point of view such a loop does not have a practical meaning because the restoring force in the loop tends also to zero. Therefore, the dynamical behavior will be very bad. Hence, in the case of channel noise and oscillator phase noise the choice of $B_L = B_L^{\text{opt}}$ in (44) is essential, and as can be seen from Fig. 11, very critical.

Acknowledgment

I would like to thank U. P. Bernhard for numerous fruitful suggestions and Prof. P. E. Leuthold, Director of the Communication Technology Laboratory, ETH, Zurich, for his support of this research. Finally, the discussions with Dr. P. Kartaschoff, Swiss PTT General Directorate, on flicker noise are gratefully acknowledged.

References

[1] J. A. Barnes, A. R. Chi, L. S. Cutler, D. J. Healey, D. B. Leeson, T. E. McGunigal, J. A. Mullen, W. L. Smith, R. Snydor, R. F. C. Vessot, G. M. R. Winkler, "Characterization of Frequency Stability," NBS Technical Note 394, U.S. Govt. Printing Office, Washington, D.C., October 1970; see also IEEE Trans. Instrum. Meas., IM-20 (2), pp. 105-120, 1971.

[2] R. C. Dixon, *Spread Spectrum Systems*, Wiley, 1984.

[3] F. M. Gardner, *Phaselock Techniques*, Wiley, 1979.

[4] J. K. Holmes, *Coherent Spread Spectrum Systems*, Wiley, 1982.

[5] H. A. Kramers, "Brownian motion in field of force and the diffusion model of chemical reaction," Physica 7, 284-304, 1940.

[6] B. B. Mandelbrot, "The Fractal Geometry of $1/f$ Fluctuations," Proceedings of the Symposium on 1/f Fluctuations, pp. 206-212, Tokyo, July 1977.

[7] B. J. Matkowsky, Z. Schuss, C. Tier, "Uniform Expansion of the Transition Rate in Kramers' Problem," Journal of Statistical Physics, Vol. 35, Nos. 3/4, pp. 443-456, 1984.

[8] H. Meyr, G. Ascheid, *Synchronization in Digital Communications*, Vol. 1, Wiley, 1990.

[9] M. K. Simon, J. Omura, R. A. Scholtz, B. K. Levitt, *Spread Spectrum Communications*, Vols. I-III, Computer Science Press, Rockville, Maryland, 1985.

[10] J. J. Spilker, *Digital Communications by Satellite*, Prentice–Hall, Englewood Cliffs, 1977.

[11] A. J. Viterbi, *Principles of Coherent Communication*, McGraw-Hill, 1966.

[12] A. L. Welti, *Ausrastverhalten von Synchronisationsschaltungen zweiter Ordnung für pseudozufällige Sequenzen*, Ph.D. Dissertation, No. 9770, ETH Zürich, 1992.

[13] A. L. Welti, B. Z. Bobrovsky, "Mean Time to Lose Lock for a Coherent Second Order PN-Code Tracking Loop — The Singular Perturbation Approach," IEEE Journal on Selected Areas in Communications, JSAC, Vol. 8, No. 5, pp. 809-818, June 1990.

[14] A. L. Welti, B. Z. Bobrovsky, "Long-Term Frequency Stability Requirements and Optimal Design of a Code Tracking Delay-Lock Loop," IEEE Proceedings of GLOBECOM '88, pp. 551-555, Hollywood, November 1988.

[15] A. L. Welti, B. Z. Bobrovsky, "Doppler Acceleration Influence on Code Tracking in Direct Sequence Spread Spectrum Systems: Threshold Calculation and AGC Algorithms," IEEE Proceedings of GLOBECOM '89, pp. 1624-1628, Dallas, November 1989.

[16] A. L. Welti, B. Z. Bobrovsky, "On Optimal AGC Structure for Direct Sequence Spread Spectrum PN-Code Tracking," to appear in the IEEE Transactions on Communications.

[17] A. L. Welti, B. Z. Bobrovsky, "Mean Time to Lose Lock for the "Langevin"-Type Delay-Locked Loop," to appear in the IEEE Transactions on Communications.

[18] R. E. Ziemer, R. L. Peterson, *Digital Communications and Spread Spectrum Systems*, Macmillan, N.Y., 1985.

CLOCK GENERATION AND SYNCHRONIZATION FOR HIGH-SPEED SERIAL DATA SYSTEMS

Donald T. Comer

The Pennsylvania State University at Harrisburg
777 W. Harrisburg Pike
Middletown, PA 17057-4898 USA

ABSTRACT

This paper presents methods and circuits which are intended to improve the speed, jitter performance, and acquisition time of integrated circuit phase locked loops in serial data clock recovery applications. The approach presented is capable of achieving an acquisition time of 10 μsec at 100 MHz data rate with a maximum jitter of 100 psec, using a conventional 3-micron *P*-well CMOS process. Improved performance approaching that of SAW filter design is projected using a 1-micron *N*-well process.

clock recovery, phase locked loop, serial data, jitter

1. INTRODUCTION

Modern serial data transmission systems such as optical fiber links and disk memory systems send encoded data with the clock only implied. In order to recover the data and perform conventional logical operations, it is necessary to generate the clock from the incoming data.

There are two widely used methods of clock recovery from serial data. The first uses a nonlinear element to accomplish frequency doubling followed by a highly selective bandpass filter tuned to the nominal clock frequency. Fig. 1 shows a block diagram of this approach, which generally is the preferred method of clock recovery for high speed digital telecommunication systems (Refs. 1, 2).

A second method of clock recovery, which is more widely used in disk memory systems (Refs. 3,4) is shown in the block diagram of Fig. 2.

In this system, a local oscillator (VCO) is phase locked to incoming data. Although conventional phase locked loop theory can be applied to the design of this type of system, the clock recovery application requires special features. Because data is typically received asynchronously in fields, it is necessary for the loop to recapture and lock each time a new data field is received. Depending on the coding scheme, data transitions may be separated by several clock cycles. In the disk memory application a "header" is transmitted at the beginning of each new data field to allow the loop to capture and lock prior to the arrival of actual data. The loop must be designed for fast synchronization to avoid wasted memory capacity due to excessive header length and must maintain lock in the absence of data transitions.

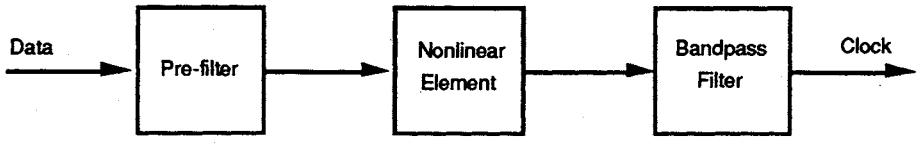


Fig. 1 Block diagram of high-speed digital clock recovery method

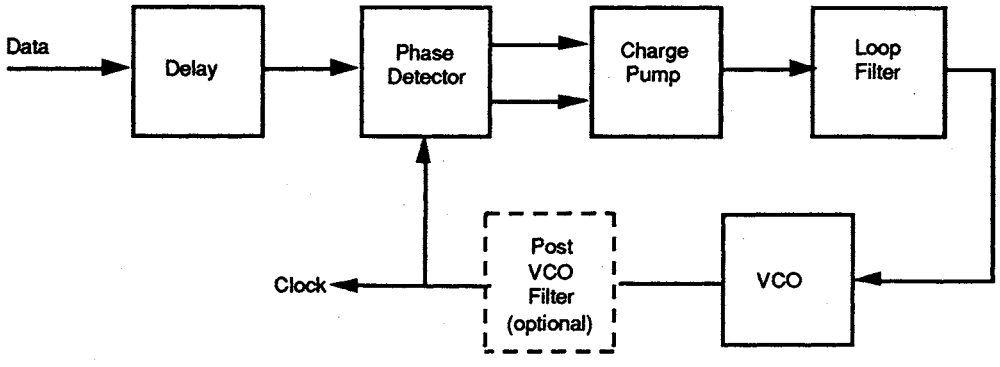


Fig. 2 Phase locked loop clock recovery method

Table I shows a comparison of performance of typical clock recovery circuits which are commercially available today for the two applications just discussed. Column A lists performance characteristics for a clock recovery module intended for fiber-optic data link applications. Column B lists similar data for a phase locked loop (PLL) based integrated circuit intended for disk memory clock recovery.

Specific ideas presented in this paper to improve the performance of PLL clock recovery circuits include:

1) improved phase detector
2) a fast sync VCO
3) post VCO filtering
4) implementation of logic functions using a new higher speed circuit approach

	A	B
TECHNICAL APPROACH	NONLINEAR DOUBLER, SAW FILTER	PHASE LOCKED LOOP
Data rate	200 Mbits/sec	25 Mbits/sec
Jitter	±30 psec	±250 psec
Acquisition Time	2 msec	10 μsec
Package technology	Hybrid module	Integrated circuit
Package	28 pin DIP	24 pin DIP

Table I Performance of commercial clock recovery circuits for (a) communication and (b) disk memory systems

Table I illustrates the fact that disk memory systems tend to run at slower data rates than telecommunications systems, have less stringent jitter requirements but require faster acquisition times. However, the trend for increasing bit rates in both applications creates the possibility that clock recovery requirements for the two applications may overlap in certain cases.

The primary purpose of this paper is to present methods of improving the PLL clock recovery circuit in the disk memory application. Because the proposed methods can provide significant improvements in performance, it is perhaps worthwhile to reconsider the PLL as a competitive method of clock recovery for telecommunications systems.

The general objectives of the methods proposed for the integrated PLL circuit are to:

1) allow for higher frequency of operation
2) maintain fast acquisition time
3) reduce jitter and alignment errors
4) maintain a fully integrated circuit/low cost product compatibility

2. CURRENT PLL CLOCK RECOVERY APPROACHES

Prior to discussing suggested improvements, it is worthwhile to review approaches to PLL clock recovery currently used in integrated circuit products. Fig. 3 shows a logic level schematic of a phase detector typical of that employed in a PLL system such as Fig. 2, while Fig. 4 shows a typical implementation of the charge-pump and loop filter circuit.

The phase detector of Fig. 3 is designed to work in conjunction with a system where the VCO runs at twice the expected clock/data rate. The delay circuit, typically implemented as a precision "one-shot" multivibrator is adjusted to produce a delay equal to one half the VCO period or one quarter of the period of the actual clock rate. The system is intended to produce a phase lock between the trailing edge of the delayed data and the leading edge of the VCO. Fig. 5 illustrates the operation of the phase detector, showing examples of error signal generated when the VCO and delayed data edges are not locked.

In this example, the first occurring data pulse arrives *late* relative to the VCO, a second data pulse is missing, and the next data pulse arrives *early* relative to the VCO. The desired function of the phase detector circuit is to enable on the leading edge of the data and then provide an error signal to force lock between the trailing edge of the data and the leading edge of the VCO.

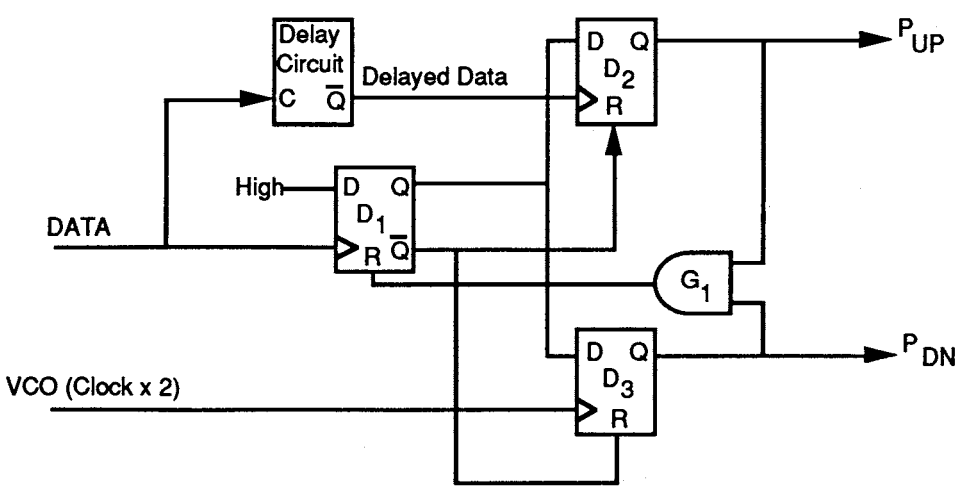


Fig. 3 Typical phase detector for PLL

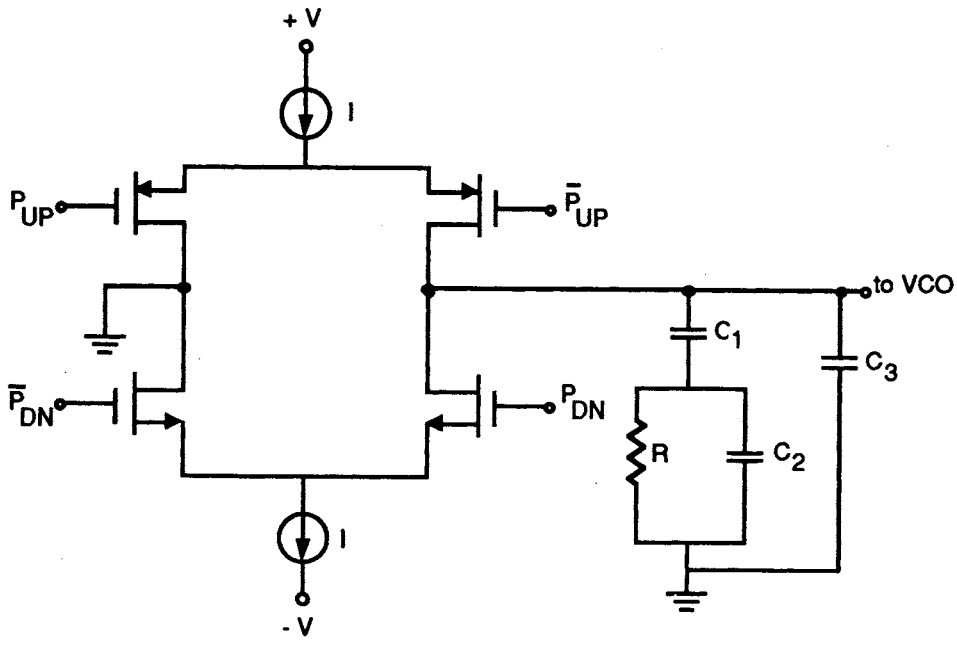


Fig. 4 Typical charge pump and loop filter circuit

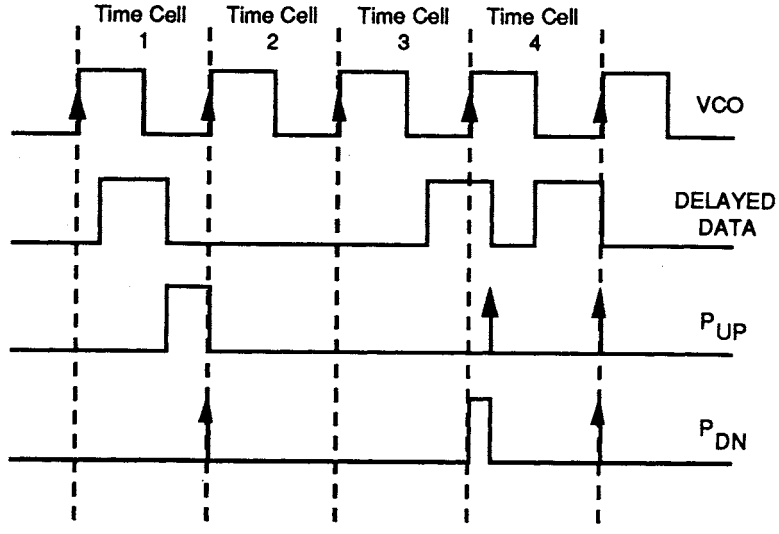


Fig. 5 Phase detector timing waveforms

This requires that an early arriving data pulse activate the charge pump with an increment of charge that increases the frequency of the VCO while a late arriving data pulse should activate the charge pump with an increment of charge that decreases the frequency of the VCO. The pulses P_{UP} and P_{DN} are assumed to be applied to the charge pump to respectively increase or decrease the VCO frequency. In time intervals where no data pulse is present the charge pump should receive no correction signal. If the data pulse arrives at exactly the correct time as illustrated in time cell 4, both P_{UP} and P_{DN} will generate a timing glitch with a duration determined by the set and reset delays of the latches D_2 and D_3 and will provide no net charge delivered by the charge pump.

The PLL system of Fig. 2 using the building blocks suggested in Fig. 3 and 4 can be analyzed as a linear control system where the input signal is the phase of the delayed data θ_i and the feedback signal is the phase of the VCO θ_O.

The closed loop system function may be written

$$\frac{\theta_O}{\theta_i} = \frac{K_P F(s) K_V}{s + K_P F(s) K_V} \tag{1}$$

where K_P is the gain of the phase detector/charge pump in amps per radian, $F(s)$ is the filter transfer function in volts/amp, and K_V is the sensitivity of the voltage controlled oscillator which is assumed to have a transfer function K_V/s in radians per volt.

The charge pump circuit of Fig. 4 will source or sink a current I into the filter for a time determined by the control pulses. The average output current over a half cycle interval (π radians) is $I/2$. Thus, the gain may be expressed as

$$K_P = \frac{I}{2\pi} \tag{2}$$

The filter function $F(s)$ may be computed from the circuit of Fig. 4 as

$$F(s) = \frac{sR(C_1 + C_2) + 1}{s^2R[C_1C_3 + C_2C_3 + C_1C_2] + s(C_1 + C_3)} \quad (3)$$

For the typical case, values are selected such that

$$C_2 << C_1$$
$$C_3 << C_1$$

The filter function then becomes

$$F(s) = \frac{\frac{1}{(C_2 + C_3)}\left(s + \frac{1}{RC_1}\right)}{s\left(s + \frac{1}{R(C_2 + C_3)}\right)} \quad (4)$$

An example of typical parameter values for 100 MHz system are as follows:

$K_P = 1.59$ mA/rad
$R = 1200\ \Omega$
$C_1 = 3000$ pF
$C_2 = 165$ pF
$C_3 = 112$ pF
$K_V = 2.5 \times 10^5$

A system with these values, with frequency normalized by a factor of 10^5, will have a closed loop system function

$$\frac{\theta_o}{\theta_i} = \frac{143.5\ (s + 2.78)}{s^3 + 30s^2 + 143.5s + 398} \quad (5)$$

This system will exhibit an acquisition time in the range of 10 μsec and will have adequate sensitivity for most applications.

3. CRITIQUE OF PLL APPROACH

PLL systems such as described in previous paragraphs have been used to create integrated circuit products intended for disk memory clock recovery applications and generally result in performance as shown in Column B of Table I. Most early products were fabricated on bipolar processes and more recently, a CMOS version compatible with conventional 3-micron *P*-well cases has been designed and found to exhibit similar performance characteristics (Ref. 5). The weakness of these products may be listed as follows:

3.1 Data Rates

It has been difficult to achieve data rates beyond 25-50 MHz using present approaches. This is due to problems of propagation delays of latches in the phase detector and switches in the charge pump. It has been also proven difficult to design a reliable VCO for this application

3.2 Jitter

Products to date have used either a multivibrator or ring delay oscillator to implement the VCO function. These circuits prove to have an advantage over tuned circuit and quartz crystal oscillators in the area of fast acquisition because of the fact that they can be gated or synchronized with incoming data and thus will begin oscillation very close to phase lock. However, these oscillator types generally exhibit poor jitter performance and have a very poor immunity to capacitive coupled impulse noise (Ref. 6).

3.3 Delayed Data

Because the VCO is locked to a delayed data edge, generated by a precision delay circuit, rather than to the actual data, any variation in the delay will effect clock alignment with data. Noise and short term variations of delay will produce a second source of jitter, whereas fixed errors will produce a reduced error budget and a tighter alignment window between data and clock.

4. IMPROVED PLL SYSTEM

The primary strengths of the PLL method of clock recovery are that by using a data synced VCO it is capable of very fast acquisition times and it is easily implemented in monolithic form using conventional VLSI technology. The object of the proposed improvements of this paper is to maintain these strengths while improving upon existing weaknesses. The improvements proposed are

1) the use of higher speed logic circuits
2) an improved phase detector
3) modification of precision delay requirement
4) employing post VCO filtering to decrease clock jitter

The suggested improvements are all compatible with present day CMOS wafer processing and can be implemented on a variety of processes.

4.1 Higher Speed Logic Circuits

The logic circuits required to implement a phase detector such as shown in Fig. 3 are typically constructed in CMOS from NAND/NOR gate building such as shown in Fig. 6. This type of gate, fabricated on a 3-micron gate *P*-well process will produce a propagation delay of approximately 500 psec when operated on a 5 volt supply. The same gate, fabricated on a 1-micron process will produce a delay of approximately 170 psec. Fig. 7 shows a schematic of a similar NOR gate constructed with NMOS devices but following an architecture similar to bipolar current mode logic. This type of circuit offers several advantages over conventional CMOS circuits in the PLL application. Because it produces a lower logic swing it produces less coupling noise to adjacent analog circuits. It is also easy to interface with conventional ECL levels, which are becoming more common for higher speed applications. However, its most prominent advantage is its speed improvement, which is approximately a factor of 2.5 over the conventional CMOS logic circuit of Fig. 6.

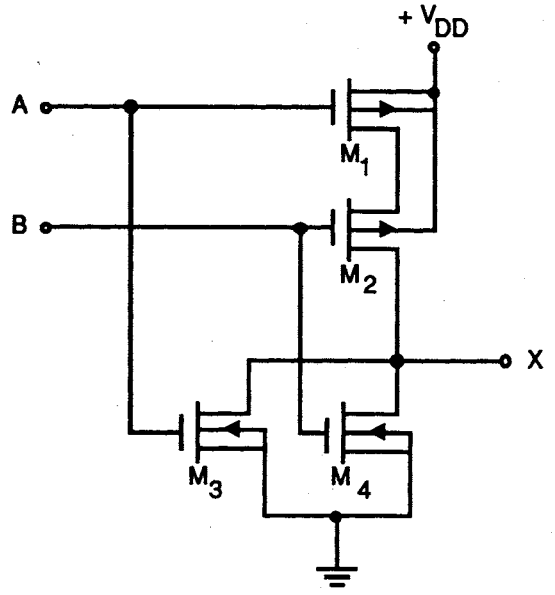


Fig. 6 Conventional CMOS NOR gate

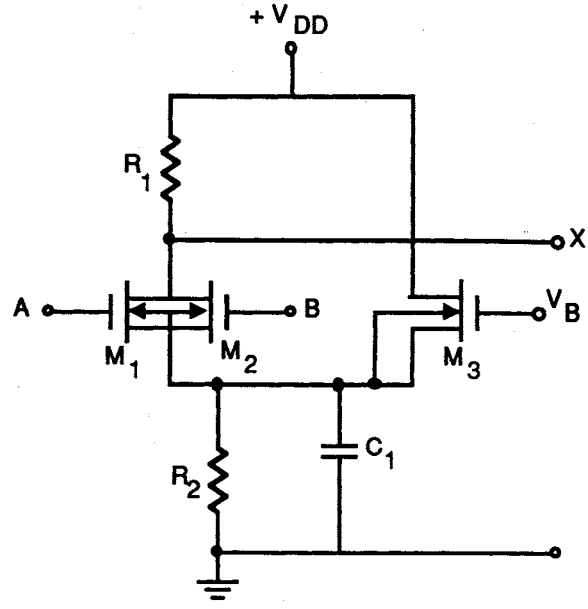


Fig. 7 Current mode NOR gate

4.2 Improved Phase Detector

Fig. 8 shows a schematic of a proposed phase detector which may be used to replace that of Fig. 3 in the PLL application. The phase detector of Fig. 8 employs a delay circuit similar to that in Fig. 3. However, the placement of the delay circuit is such that it is not in series with the data signal and its precision only effects the duration of the P_{DN} signal. The direct dependence of the accuracy of the phase lock on the delay period is thus eliminated. Fig. 9 shows expected timing waveforms for the circuit of Fig. 8 for a case where the first data transition arrives late relative to the VCO, no data transition occurs at the second clock interval and the data transition in the third time cell occurs early relative to the VCO.

It should be noted that the phase detector circuit of Fig. 8 requires a VCO that operates at twice the data clock rate. The phase detector is also designed to enable on either positive or negative transitions and thus is directly compatible with NRZ data. In the lock condition as illustrated in the 4th time cell, the leading edge of the VCO is phase locked to the data transition.

4.3 Post VCO Jitter Filtering

As mentioned earlier a multivibrator or ring delay type of oscillator has the advantage that it can be phase synchronized with incoming data so that the required capture time between the data and clock is minimized and false lock and ambiguous states can be eliminated. This advantage is provided at the expense of poor jitter performance compared to a crystal or tuned circuit oscillator. However, the addition of a bandpass filter of even modest selectivity in the loop following the VCO can provide a large improvement in jitter performance with very little penalty in acquisition time. See placement of post VCO filter in Fig. 2.

The bandpass filter function is assumed to be of the form

$$T(s) = \frac{Ks}{s^2 + \frac{\omega_o}{Q}s + \omega_o^2} \tag{6}$$

where ω_o is the center frequency, is chosen at the nominal VCO frequency and ω_o/Q is the circuit bandwidth. The jitter reduction of the bandpass filter may be expressed by (Ref. 6)

$$J_{out} \approx \frac{J_{in}}{Q} \tag{7}$$

where J_{in}/J_{out} are respectively input and output jitter. The closed loop gain of the system including the bandpass filter may be expressed as

$$\frac{\theta_o}{\theta_i} = \frac{2\frac{Q}{\omega_o}K_P F(s) K_V}{s + 2\frac{Q}{\omega_o}K_P F(s) K_V} \tag{8}$$

If the conditions $\omega_o/Q >> RK_PK_V$ and $\omega_o/Q >> RC_1$ hold, then the addition of the bandpass filter will have negligible effects on the system dynamics but nevertheless will result in jitter improvement. For example a VCO running at a frequency of 200 MHz with a jitter of ±500 psec in the system previously defined by eq. 5 can be filtered with a Q of 20 to produce a jitter of ± 25 psec with no observable change in acquisition time. For higher Q filters and even larger jitter reduction, the system dynamics can be optimized by a redesign of the loop filter using eq. 8.

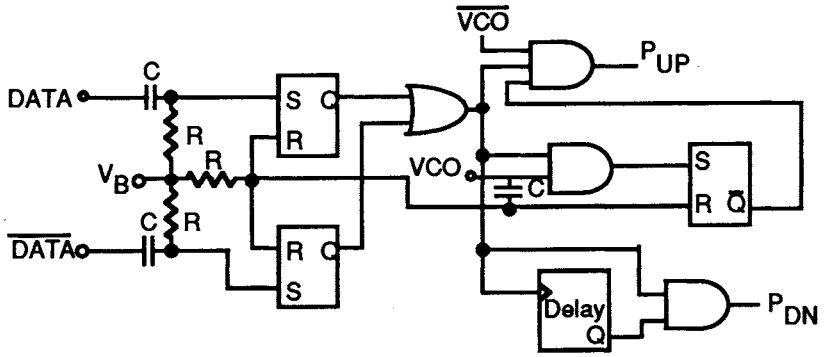


Fig. 8 Improved phase detector circuit

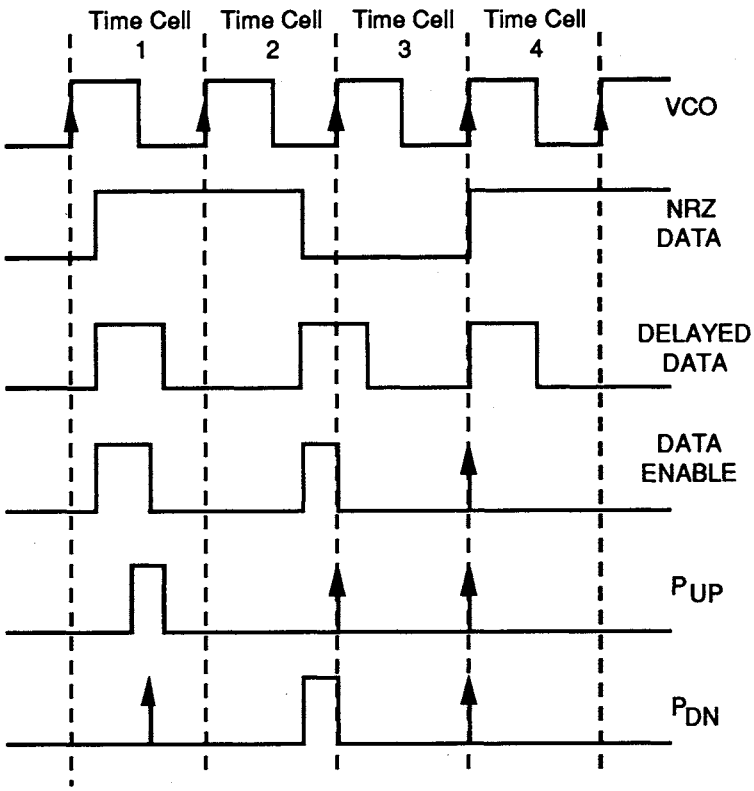


Fig. 9 New phase detector timing waveforms

5. CONCLUSION

A PLL clock recovery system constructed on a 3-micron *P*-well CMOS process using the techniques described in this paper has been designed to operate at data rates of 50 - 100 MHz and is currently being evaluated. SPICE simulations indicate that the same design fabricated on a 1-micron gate length process can result in operation at 150 - 300 MHz data rates, with performance that is competitive with SAW filter devices.

REFERENCES

1. Takasaki, Y., *Digital Transmission Design and Jitter Analysis*, Boston:Artech House, 1991.

2. Trischitta, P. and E. Varma, *Jitter in Digital Transmission Systems*, Boston:Artech House 1989.

3. National Semiconductor, *DP8460 Data Separator*, Sunnyvale, CA, USA:National Product Data Book, 1987.

4. Silicon Systems Inc., *HDD Data Recovery*, Tustin, CA USA:Microperipheral Products Data Book, 1991.

5. Comer, D. T., "PLL Clock Recovery Circuits for Serial Data Systems," Penn State Research Report SET-VLSI 104.0, Sept. 15, 1992.

6 Comer, D. T., "A Stable, Low Jitter, Triggered VCO for Phase-lock Application," Penn State Research Report SET-VLSI 106.0, Nov. 6, 1992.

ACCURATE NETWORK SYNCHRONIZATION FOR TELECOMMUNICATIONS NETWORKS USING PAIRED PATHS

Masami Kihara and Atushi Imaoka
NTT Transmission Systems Laboratories
1-2356 Take, Yokosuka-shi, Kanagawa, 238-03 Japan
Tel:+81-468-59-3164 Fax:+81-468-59-3197
e-mail:kihara@ntttsd.ntt.jp

ABSTRACT

Network expansion and Synchronous Digital Hierarchy (SDH) system introduction require high performance network synchronization. This paper introduces the idea of a network synchronization system that uses paired transmission paths, instead of the traditional single distribution paths.

The system can synchronize frequency with departures of 10^{-13}-10^{-14} and time within the order of sub-microseconds over more 2000 km. Furthermore, we describe a method to eliminate the effect of cable length differences on time distribution. The method creates bi-directional transmission paths on single cables. Overall time accuracy within 10ns is possible.

The system also permits autonomous operation and maintenance, in which it automatically determines synchronization topology and transfers system alarm information through round-trip transmission paths without requiring the intervention of the network operation system. Finally, the possibility of global network synchronization is addressed.

Keywords: network synchronization, frequency, time and SDH

1. INTRODUCTION

The second generation of network synchronization is coming to the telecommunications field. Digital network expansion and broad band ISDN services require a more mature network synchronization system that distributes accurate reference signals to a wide area and to keep high accuracy and stability of reference signals at all times. These trends are being fostered by the digitalization of local nodes and by the aggressive installation of SDH[1]. One recent effort for wide reference signal distribution is the application of the GPS system at telecommunication nodes[2]. The other is the improvement of slave clock performance using rubidium oscillators[3].

These circumstances encourage the adoption of a new accurate network synchronization system to synchronize and distribute frequency signals and phase controlled reference signals, so-called time reference signals. The system measures transmission delay and its variation, and compensates them. This system can also provide special communication capability between nodes. This paper describes the configuration and performance of the new system, an autonomous operation and management system for network synchronization, and an international synchronization method which is one application of the proposed system.

2. TELECOMMUNICATION TRENDS

Digital facilities such as cross-connect systems and exchanges, have already been installed in main nodes, and are now being introduced to local nodes. This digitalization is encouraged by ISDN services which are extending throughout the networks. Local node digitalization yields new slave nodes at a lower level at the end slave node of the current reference signal distribution topology. This means that the number of slave nodes is greatly increasing.

The aggressive introduction of SDH is also accelerating the increase in slave node number. In the SDH system, a frequency signal is terminated at devices having multiplexing and demultiplexing functions, at the same place at which the transmitted data are terminated. In particular, when such devices are used to replace the conventional pulse-justification multiplex/demultiplex systems to add/drop to/from multiplexed data, the number of frequency terminations, that is the number of slave nodes, is augmented.

In intra-cities such as metropolitan areas, the SDH system can constitute a ring network. In the SDH ring, frequency signal distribution topology is usually the basic ring configuration. Therefore, even if the reference signal is directly transferred from a primary reference clock at the entrance of the ring net-

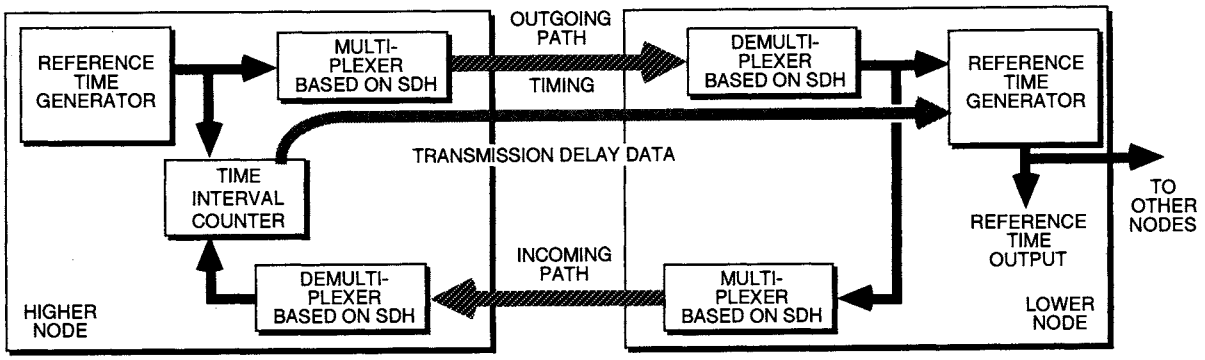


Fig.1 The basic configuration of the time distribution chain

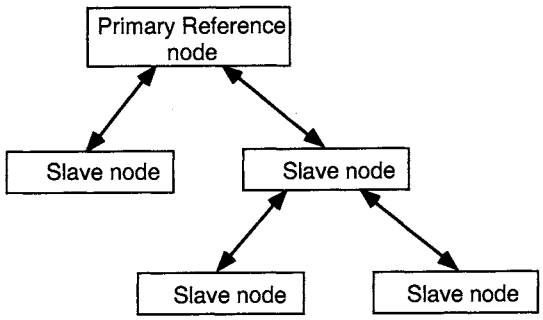


Fig. 2 System topology

work, the number of chaining slave nodes drastically increases according to the number of ring nodes.

The increase in the number of slave nodes mentioned above causes jitter and wander performance degradation and makes maintenance of the network synchronization system more difficult.

3. PERFORMANCE IMPROVEMENT

Of the performance degradation components, the wander generated by the transmission line system is dominant. The wander compensation achieved by the proposed synchronization system very effectively offsets the degradation problem. The method can extend the safety margin for worst case design, decrease the error probability in transmission systems, and also enables networks to synchronize both frequency and time.

3.1 System configuration

The conventional reference signal distribution is a one-way transfer from a higher node to a lower node in the reference signal distribution topology. Wander of the distributed signal at the lower node is identical to transmission delay variation accumulated in the transmission lines joining the primary reference and the slave node. Figure 1 explains the transmission delay compensation method. This method uses round-trip paths that consists of paired outgoing and incoming paths. The transmission delay of the round-trip path is measured by the time interval counter at the higher node. The delay data, together with frequency, time tag, and time data, is transferred to the lower node over an outgoing path. The lower node slave clock controls the transferred reference signal according to the delay data and then generates the compensated reference signal of frequency and time. The new synchronization system topology is a form of master-slave synchronization (Fig.2).

3.2 Performance

If the fibers used for outgoing and incoming paths are in the same cable, the wander compensation capability is of the order of sub-micro seconds for nodes separated by 2400 km over a two year period[4][5]. This performance was experimentally confirmed using paired paths with pulse-justification multiplex/demultiplex systems at both path ends.

Only SDH paired paths are considered due to the aggressive introduction of SDH systems. Our proposal is to utilize Virtual Container (VC) - 3 of the SDH information structure for frequency and time transfer. VC-3 can transport the information which is not synchronized to STM-N (N's of 1, 4 and 16 are defined in CCITT[1]). VC-3 phase (frame) does not have a fixed relationship with STM-N frames (Fig. 3 (a)), and its frequency is also allowed to be different from that of STM-N within a specified range.

These flexible characteristics permit VC-3 to transport the timing (phase) point of the time tag for time data and transmission delay measurement as well as pulse-justification multiplex/demultiplex systems. However the timing point may be quantized to 3 byte step of STM-1 (155Mbit/s) when VC-3 information is embedded in STM-1. This corresponds to byte *a* shown in Fig.3 (c). Byte *a* is interleaved to STM-1 in Fig.3 (b). When

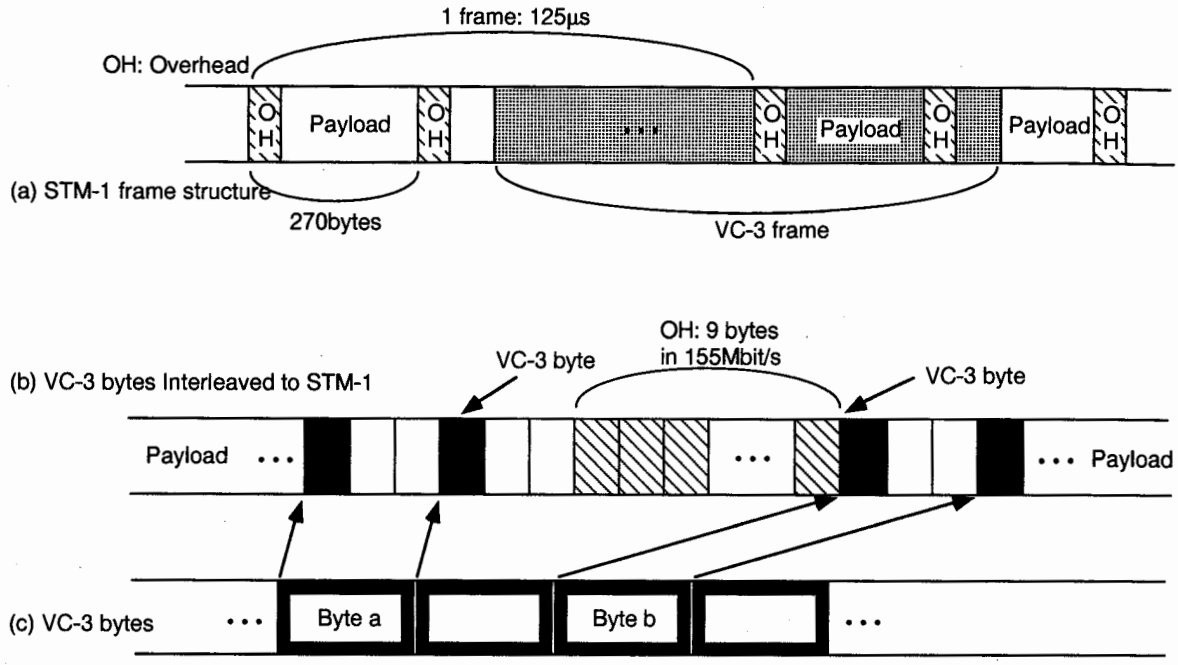


Fig.3 STM-1 frame structure and VC-3 interleave process

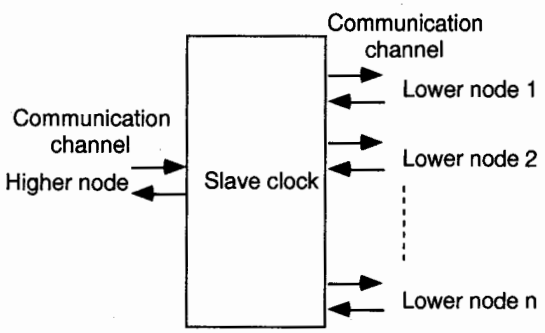


Fig. 4 Slave clock

the timing point of VC-3 is Section OverHead (SOH) of STM-1, as is true for VC-3 byte *b* in Fig.3 (c), the timing point is delayed by up to 3-4 VC-3 bytes. These quantum errors yield a time error that corresponds to the time difference between nodes. Our system controls the phase difference between the VC-3 signal and the STM-1 frame signal, and determines the optimum relative phase relationship for them to minimize the time error. The maximum time error of 1.2 μs before control was diminished to less than 15 ns with control in a preliminary experiment[6]. This implies that the system can synchronize frequency with departures of 10^{-13}-10^{-14}.

4. IMPACT ON OPERATION AND MANAGEMENT

The number of nodes in Japan has already exceeded 1000. In large synchronization networks, the distribution paths of reference signals are very complex because path exchanges are sometimes executed to improve network reliability, and because ring networks are being introduced. Operation and management activities for network synchronization is a new theme. CCITT recommends a method that uses the Section OverHead bytes of the SDH system to transfer synchronization status. We present more a positive method that establishes and maintains the synchronization system.

The proposed time and frequency synchronization system has a communication channel for time data transfer between primary reference and slave nodes (Fig. 4). The communication channel can be used at the same time when round-trip transmission paths are assigned to synchronize time and frequency. Primary reference and slave nodes can communicate with each other for synchronization network operation and management.

The new method enables the synchronization system to maintain by itself without requiring the intervention of other network operation systems. The synchronization system information transferred between higher and lower nodes is:

- node identification data (node ID)
- reference signal quality
- primary reference clock (PRC) and slave clock status

(1) Downstream from the primary reference to lower nodes

The reference signal distribution topology is one of the most important management items for master-slave synchronization. To confirm the integrity of the reference signal topology, the

new system transfers node IDs. Nodes in downstream links transfer the regenerated node information data including node ID to lower nodes after attaching their own node ID to the node information data transferred from the higher node. In the example shown in Fig. 5, node n receives node information data as shown in Fig. 6 (a) and then transmits the new node information data merging node n information as shown in Fig. 6 (c).

The node can check the reference signal distribution topology as far as the node itself. The node receives the node information data from higher node and first extracts node ID's from the data stream. Node ID's are stacked and the reference signal distribution path is verified (Fig. 6 (b)). The topology check rule is whether the node receives the reference signal that originates in the PRC, that is whether the first stack content is the node ID of the primary reference node. The number of chained nodes can be also checked in this process. Nodes can detect improper reference distribution paths assigned by unexpected path and transmission system reconfigurations.

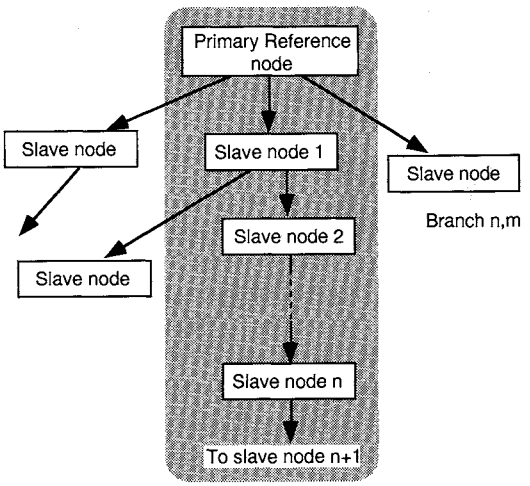


Fig. 5 Downstream node information data transfer

The other transferred item is synchronization system conditions. This includes quality and status such as a reference signal distribution path condition, and PRC and slave clock conditions. The slave clock can detect errors and transmission failures of the reference signal distribution path, and decide by itself whether the slave clock should be slaved to the other higher node for reference or whether clock operation should be switched to the holdover condition. Many rubidium oscillators have been recently introduced into the slave clock, and these devices can hold their frequency for a long time without a reference signal. Therefore, the holdover state is definitely preferred to slaving to an abnormal reference signal.

(2) Upstream from the lower to the primary reference node

In upstream links as shown in Fig. 7(b), each node transfers the

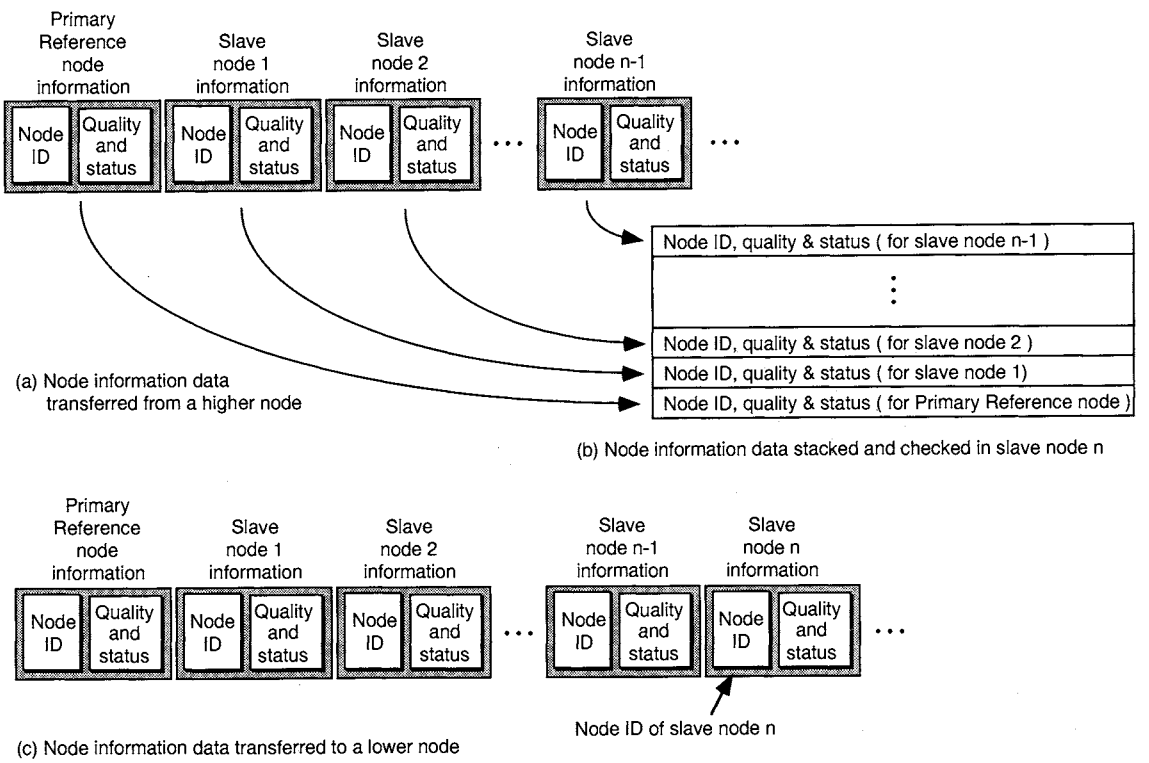


Fig. 6 Downstream node information data transfer

of the MITREX modem. Interface standardization is very effective to enhance time and frequency transfer. Furthermore, optical fiber cable transmission systems will provide standardized, long-distance, high-capacity digital paths that are also inexpensive. Their timing jitter is within several hundreds of picoseconds for the STM-16 bit stream (2.4Gbit/s).

- The SDH data format (STM) provides two areas: the OverHead (OH) includes data for maintenance and operation, while the Payload transmits user data. Therefore, an SDH system has the capability to transfer administration data such as time and date without occupying the Payload. This means that there is no need to lease a complete channel to transfer time and frequency. In time and frequency transfer via satellites, especially communication satellites, the experimental term is limited to just one week, even if the satellite company supports the experiment. Time and frequency transfer via SDH systems does not disturb in-service transmission paths. If spare OH bytes can be assigned to information for time and frequency transfer, long experiments in time and frequency transfer would be relatively inexpensive.

- The memory capacity of SDH system equipment is much lower than that of conventional synchronous multiplexers/demultiplexers. This very effectively reduces the transmission delay. The delay can be minimized by the new method we are proposing. SDH systems are promising for precise time and frequency transfer, even if they suffer from excessive jitter and wander.

Thus, consecutive, stable, and accurate international time transfer can be achieved. Time comparison, based on the common time information, also becomes possible.

6. CONCLUSION

The operation and management for the new network synchronization system using paired paths is presented. The improvement of the reference signal reliability for the network synchronization system is the most important subject. The operation and management of the system is the main future theme.

As to the merging of national and international network synchronization, it is indispensable to coordinate the time and frequency functions being discussed in CCIR with those emerging in telecommunication networks[7]. Our duty is to prevent the adoption of incompatible specifications for timing and frequency transfer.

ACKNOWLEDGEMENTS

We are grateful to Dr. Haruo Yamaguchi and Mr. Ikuo Tokizawa of NTT Transmission Systems Laboratories for their encouragement.

REFERENCES

[1]CCITT Recommendations of the G.700-series:G.707, G.708 and G.709

[2]E. W. Butterline and J. E. Abate and G. P. Zampetti:"Use of GPS to synchronize the AT&T national telecommunications network," Proc. of the 20th PTTI, pp.65-74, 1988

[3]M. Kihara and K. Hisadome:"frequency and time synchronization in digital communications networks," Proc. of the 20th PTTI, pp.183-191, 1988

[4]A. Imaoka and M. Kihara:"Long-term propagation delay characteristics of telecommunication lines," IEEE Trans. on Instru. and Meas., vol. 41, no. 5, pp.653-656, 1992

[5]M. Kihara and A. Imaoka:"Timing and time signal distribution in digital communications networks," Proc. of the 6th EFTF, pp.489-492, 1992

[6]A. Imaoka and M. Kihara:"Time signal distribution in communication networks based on synchronous digital hierarchy", Proc. of the 24th PTTI, 1992

[7]CCIR delayed contribution:"Draft new recommendation,"

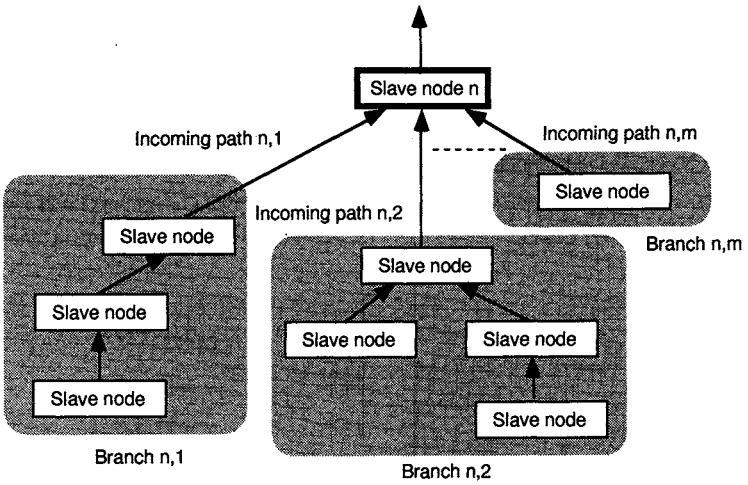


Fig. 7 Upstream node information data flow

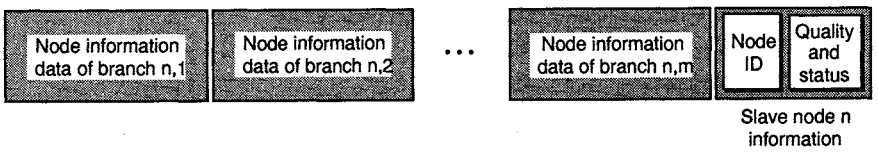


Fig. 8 Upstream node information data transfer

regenerated node information data, which includes its own data, to the higher node. The regenerated node information data sent to the higher node includes the lower node branch data categorized branch by branch as shown in Fig. 7(a). This upward transfer of node information permits the downstream topology and conditions to be accurately determined. A node can check whether the reference signal is properly distributed to all lower nodes within the pre-assigned maximum node chain length, and whether lower nodes are accurately synchronized. All node information data are consequently concentrated at the primary reference node. Therefore, the whole network synchronization system can be maintained as a unique topology, even if local areas design and update the reference signal distribution topology by themselves.

5. GLOBAL NETWORK SYNCHRONIZATION

Network synchronization performance as described in Section 3 can be upgraded so that it is equivalent to a standard for frequency and time dissemination, although the scope of telecommunication companies is merely to trace the UTC maintained by government facilities. Moreover, national/international network expansion is about to permit national/international time and frequency transfer.

With the SDH scheme, every network can offer the same function due to the adoption of the Network Node Interface (NNI) recommended in CCITT. The SDH system has the following characteristics:

- Both multiplexer/demultiplexer and transmission line system functions are contained, and even transmission line bit rates are standardized,
- Transmission line (hierarchical) bit rates are recommended to be 155Mbit/s for Synchronous Transport Module (STM) -1, 622Mbit/s for STM-4 and 2.4Gbit/s for STM-16.
- Global communication is possible because the NNI has been standardized and long distance communication is now possible with single mode fiber cables,
- Bit streams generated by different clocks having different frequency deviations can be multiplexed, transferred and demultiplexed by using data containers,
- Maintenance and operation capabilities are augmented, as overhead bytes for maintenance and operation are already embedded in the STM format,
- SDH system introduction is being encouraged by the CCITT Recommendations.

The above features of SDH provide the following advantages when an SDH system is used to transfer time and frequency:

- The SDH system looks to become the global multiplex/demultiplex and transmission system. This means that interfaces for time and frequency transfer can be standardized internationally if the NNI is adopted. In conventional experiments, the transmission path for standard time and frequency transfer is composed of various interfaces with the exception

Concatenated Clock Performance within a Synchronous Optical Transport Hierarchy

R. Avery & G. Zampetti
Telecom Solutions
85 West Tasman Drive
San Jose, CA 95134 USA
+1 408/428-7862, fax +1 408/428-7897

Today's digital communications network is evolving towards the implementation of synchronous optical transport (SDH and SONET). In doing so, many additional network element clocks will be added within the network topology. Linear network models of clock designs comprising of type I and type II loops, when concatenated, yield results that can be deceiving.

This paper is a follow up to "Synopsis of Timing Measurement Techniques used in Telecommunications" presented at the 6th EF&TF. Using the clock measurement methods such as TVAR, ZTIE, MVAR and MTIE outlined in a previous paper, and clock simulation techniques, a comparison will be provided between measured non-linear clock performance and simulated performance with a linear model.

These results will aid in the design decision when considering SDH or SONET based networks

OSCILLATORS

Chairman: D. Hauden

DESIGNING FOR LOW ACCELERATION SENSITIVITY IN CRYSTAL RESONATORS AND OSCILLATORS

John A. Kosinski and Arthur Ballato

US Army Research Laboratory
ATTN: AMSRL-EP-MD
Fort Monmouth, New Jersey 07703-5601

ABSTRACT

In this paper we discuss recent advances in the understanding of acceleration sensitivity with an emphasis on what these advances indicate in terms of designing for low acceleration sensitivity. We separate the design suggestions into two parts, namely the crystal resonator as a mechanical vibrator and the crystal oscillator as an electronic circuit. The resonator section includes such topics as symmetry considerations, metallization, mounting, etc., in both BAW and SAW devices, while the oscillator circuit section discusses the equivalent circuit modeling of crystal resonators and other loop components.

Key words: acceleration sensitivity, BAW resonator, SAW resonator, crystal oscillator.

1. INTRODUCTION

Acceleration sensitivity remains one of the most difficult problems faced by the piezoelectric devices industry. After a decade of intensive research, many questions still remain as to how to achieve reproducibly low acceleration sensitivity (10^{-10}/g) even in the laboratory. In this paper we discuss recent advances in the understanding of acceleration sensitivity with an emphasis on what these advances indicate in terms of designing for low acceleration sensitivity.

We separate the discussion into two parts, namely the crystal resonator as a mechanical vibrator and the crystal oscillator as an electronic circuit. The resonator section focuses on minimizing the acceleration-induced change in crystal resonator eigenfrequency, and includes such topics as symmetry considerations, metallization, mounting, etc. in both bulk acoustic wave (BAW) and surface acoustic wave (SAW) devices. The oscillator circuit section focuses on the analysis of the resonator transimpedance, and discusses the equivalent circuit modeling of BAW and SAW crystal resonators and other loop components.

2. THE CRYSTAL RESONATOR AS A MECHANICAL VIBRATOR

2.1 The Nature of the Acceleration-Induced Frequency Shift

The crystal resonator, whether BAW or SAW, quartz or some other piezoelectric material, operates as a high-Q mechanical vibrator whose behavior is determined jointly by the geometry and physical properties of the piezoelectric material. When the resonator is subjected to an external acceleration, the resonance frequency of the crystal is perturbed as described by the familiar vector equation

$$\delta f/f_o = \Gamma \cdot \mathbf{a} \ , \qquad (1)$$

where Γ is the acceleration sensitivity vector of the crystal resonator and $\mathbf{a}$ is the applied external acceleration. In order to design for low acceleration sensitivity, it is necessary to understand properly the cause of the frequency shift:

The frequency shifts because the external acceleration causes the crystal resonator to deform.

The frequency shift due to the deformation has two parts, namely 1) a 'linear' part arising solely from changes in the resonator dimensions and 2) a 'nonlinear' part arising from nonlinear elasticity in the deformed piezoelectric material. This understanding is critical, in that the solution to the problem of acceleration sensitivity lies not in eliminating material nonlinearities (the 'linear' shift would still remain), but in either eliminating the deformation or reducing the sensitivity of the resonator to it.

2.2 Resonator Design Considerations

The success of certain designs in reducing the resonator deformation and resulting frequency shift has been documented in the literature [1,2]. For BAW resonators, a good example is found in the superiority of 4-point, equally-spaced mounts over 3-point, cantilever mounts [1]. A good example in SAW is the performance improvement achieved in AQP-based SAW oscillators through the application of

ceramic stiffeners to the oscillator structure [2].

Many recent publications have focused on the role of symmetry in both mounting and mode shape with respect to eliminating acceleration sensitivity [3-20]. The general premise is that a perfectly symmetric mode shape (described mathematically by an even function) in conjunction with a perfectly symmetric mounting structure (which leads to acceleration-induced stresses described mathematically by an odd function) has no net acceleration sensitivity (the integral of an odd function over an even interval is always zero). There is a growing body of evidence supporting this contention. For example, the importance of the mode shape in determining the acceleration sensitivity has been documented in measurements of the differing acceleration sensitivities of various harmonics of SC-cut BAW resonators [1,21]. There is also evidence substantiating the importance of mass-loading in both BAW and SAW devices in this regard [5,7,22]. With respect to mounting symmetry, data exist which indicate that certain resonator-to-support attachment methods achieve the desired symmetry more accurately and more repeatably than other methods, with a corresponding improvement in the yields of low acceleration sensitivity units [23]. These are important points with many practical implications, as the resonator design and manufacturing processes must be driven to yield reproducible symmetry.

In certain cases such as SAW devices where an intrinsic lack of symmetry does not allow complete elimination of the acceleration-induced frequency shift, the contributions of specific types of stresses may be canceled by adjusting the lateral dimensions of the substrate [24-27]. This technique, known as aspect-ratio compensation, has been successfully applied in SAW devices [28] and has also been suggested for BAW devices [13-20].

The various analyses and experimental results clarify certain aspects of resonator design and fabrication for low acceleration sensitivity:

1) Since the root cause of resonator acceleration sensitivity is the deformation of the resonator, the first and foremost solution is the minimization of the resonator deformation.

2) Mounting and mode shape symmetry *can* be used to minimize acceleration sensitivity.

3) Aspect-ratio compensation has been successfully applied in SAW devices and provides a very practical approach to low acceleration sensitivity design.

The key to implementing any of the above design suggestions lies in achieving a high degree of reproducibility in the fabrication process. This requires a careful analysis of the reproducibility of each of the individual steps in the fabrication process, with particular emphasis on the contouring, mounting, and electroding steps. For example, energy trapping in BAW resonators can be achieved via either contouring or mass-loading. When designing for low acceleration sensitivity, whichever of these two processes is known in practice to yield more reproducible symmetry would be the preferred energy trapping mechanism. Similar considerations would apply in the case of mounting and bonding.

2.3 General Conditions for Intrinsically Acceleration Insensitive Crystal Resonators

The practical limitation of designs which rely upon symmetry in resonator mounting and mode shape in order to reduce the acceleration sensitivity is how well the required symmetry conditions can be reproduced in a production environment. It would be preferable to develop a resonator design wherein acceleration insensitivity is obtained independently of both mounting and mode shape. We are thus led to search for a set of general conditions, independent of the biasing state and mode shape, wherein the resulting frequency perturbation is identically zero.

The shift in eigenfrequency of a piezoelectric resonator subjected to a biasing condition, such as an external acceleration, may be calculated using the perturbation integral formulation of Tiersten [29], which we reproduce here for completeness as

$$\omega = \omega_\mu - \Delta_\mu \tag{2}$$

$$\Delta_\mu = H_\mu / 2\omega_\mu \tag{3}$$

$$H_\mu = -\int_V \hat{c}_{L\gamma M\alpha}\, g^\mu_{\alpha,M}\, g^\mu_{\gamma,L}\, dV \,. \tag{4}$$

In equations (2) through (4), ω and ω_μ are respectively the perturbed and unperturbed eigenfrequencies, the $\hat{c}_{L\gamma M\alpha}$ are coefficients dependent on the biasing field, g^μ_α denotes the normalized mode shape, and the integral is taken over the undeformed volume of the piezoelectric plate. If the biasing deformation and mode shape are known, the frequency shift can readily be calculated.

We note that for any mode of motion, there must exist some non-zero mode shape g^μ_α. Thus any zeros of the perturbation integral must be associated with the biasing coefficients $\hat{c}_{L\gamma M\alpha}$. The defining relation for the biasing coefficients is

$$\hat{c}_{L\gamma M\alpha} = T^1_{LM}\,\delta_{\gamma\alpha} + \zeta_{L\gamma M\alpha KN}\, E^1_{KN} + \zeta_{L\gamma KM}\, w_{\alpha,K} + \zeta_{LKM\alpha}\, w_{\gamma,K} \,. \tag{5}$$

In equation (5), the $c_{L\gamma KM}$ and $c_{LKM\alpha}$ are the "ordinary" second order elastic stiffnesses while the $c_{L\gamma M\alpha KN}$ are the "non-linear" third order stiffnesses. The terms $w_{\alpha,K}$ and $w_{\gamma,K}$ represent the biasing deformation gradients, and describe the physical deformation of the piezoelectric plate in response to the bias. T^1_{LM} and E^1_{KN} represent respectively the static biasing stress and strain and are defined as

$$T^1_{LM} = c_{LMKN}\, E^1_{KN} \qquad (6)$$

and

$$E^1_{KN} = \frac{1}{2}\,(w_{K,N} + w_{N,K}) \qquad (7)$$

After some manipulation, we may rewrite equation (5) as

$$\hat{c}_{L\gamma M\alpha} = w_{N,K}\, k_{L\gamma M\alpha KN} \qquad (8)$$

where we now define a set of stiffness cofactors dependent only upon the properties of the piezoelectric material as

$$k_{L\gamma M\alpha KN} = [c_{L\gamma M\alpha KN} + c_{LMKN}\,\delta_{\gamma\alpha} + c_{L\gamma KM}\,\delta_{\alpha N} + c_{LKM\alpha}\,\delta_{\gamma N}] \,, \qquad (9)$$

The general conditions for an intrinsically acceleration-insensitive crystal resonator are now clearly understood:

If for each coefficient $\hat{c}_{L\gamma M\alpha}$, the stiffness cofactors $k_{L\gamma M\alpha KN}$ of the deformation gradients $w_{N,K}$ are zero, then the eigenfrequency of the resonator will be independent of the biasing state and mode shape.

In the most general case, there are 81 $\hat{c}_{L\gamma M\alpha}$ each of which contains 9 stiffness cofactors for a total of 729 independent conditions which must be met for the resonator to be intrinsically acceleration-insensitive. The number of independent conditions which must be met for any particular resonator will be reduced by the symmetry class of the piezoelectric material, and may be further reduced by both mounting structure and mode shape.

As an example of the mounting structure reducing the number of required conditions, we consider the case of a SAW substrate rigidly and uniformly supported along its back surface. The substrate deformation in response to an acceleration is then limited to thickness extension or thickness shear only. In this case, the number of allowed deformation gradients is reduced from nine to three.

As an example of the mode shape reducing the number of required conditions, we consider the case of rotated Y-cut quartz BAW resonators operating in the slow thickness-shear mode. The mode shape varies slowly enough in the lateral dimensions that the predominant mode displacement gradient required for the perturbation integral is $g^{\mu}_{1,2}$ and thus the predominant biasing coefficient required is $\hat{c}_{1212}$ [8]. To a good approximation, there are then at most nine conditions required to achieve intrinsic acceleration insensitivity for an arbitrarily mounted resonator of this type. Crystal symmetry further reduces the number of required conditions (non-zero k_{1212KN}) to five.

There are a number of observations to be made regarding the intrinsic insensitivity conditions:

(1) The statement of the general conditions represents a set of criteria for which a material may be searched. It does not constitute proof that the conditions can be met.

(2) This analysis is valid only to the order of the approximations underlying the perturbation theory. The conditions as stated should be sufficient to determine the existence of zeros in a region of crystal orientations, however a higher order analysis may be required to precisely determine exact loci.

(3) The general conditions have been derived for an arbitrary static biasing deformation and an arbitrary mode shape. They are applicable not only to the problem of acceleration sensitivity, but to other stress- and deformation-induced phenomena such as the contributions of electrode and mounting stress relief to aging.

(4) The general conditions shed light on the misconception that acceleration sensitivity arises from the fact that the piezoelectric material is nonlinear. The cofactors $k_{L\gamma M\alpha KN}$ contain sums of both second- and third-order stiffnesses. In a perfectly linear material, the third-order stiffness terms vanish but not the second-order terms. Thus while material nonlinearity is a significant part of acceleration sensitivity, there is also a "linear" part arising solely from the physical deformation of the resonator shape.

3. THE CRYSTAL OSCILLATOR AS AN ELECTRONIC CIRCUIT

3.1 Crystal Resonator Transimpedance

The typical crystal controlled oscillator does not operate at the resonance frequency of the crystal resonator. Consequently, from the point of view of the crystal oscillator as an electronic circuit, the acceleration-induced shift in crystal resonator eigenfrequency is not of primary importance. Rather, the key to the behavior of the crystal oscillator lies in the perturbation in transimpedance of the crystal resonator at the actual operating point of the oscillator loop. This realization has led to the development of a modeling scheme wherein the effects of acceleration on the quartz resonator and oscillator may be examined separately [30]. It has been found that the acceleration-induced frequency perturbation in a quartz crystal oscillator is well represented by first properly modeling the effects of acceleration on the resonator transimpedance, and then examining the interaction of the perturbed resonator with the oscillator loop. The modeling procedure has been verified experimentally through a series of separate resonator and oscillator tests.

The Butterworth-Van Dyke (BVD) single-mode, lumped element, approximate equivalent electrical circuit for the BAW crystal resonator is well known. For the

case of the infinite flat plate, the equivalent circuit parameters may be determined from the crystal material properties using well-known relations obtained using the linear piezoelectric theory :

$$C_o = \frac{\epsilon A}{2 h} \tag{10}$$

$$R_1 = \frac{n^2 \pi^2 h}{4 A e^2} \eta_s = \frac{n^2 \pi^2 h \bar{c} \tau_1}{4 A e^2} \tag{11}$$

$$L_1 = \frac{h^3 \rho}{A e^2} \tag{12}$$

$$C_1 = \frac{4 A e^2}{n^2 \pi^2 h \bar{c}} \tag{13}$$

The point of note is that R_1 and C_1 depend on $\bar{c}$, while C_o and L_1 do not. Since the largest contribution to the acceleration-induced frequency shift is that due to the elastic nonlinearities, one is led to consider the effect of introducing an acceleration-sensitive elastic stiffness in these relations. Recognizing that the piezoelectric contribution to the stiffened elastic constant is small in the case of quartz resonators, we model the acceleration-sensitive elastic stiffness as

$$\bar{c}(\mathbf{a}) = \bar{c}(\mathbf{a}=0)*(1+\gamma \cdot \mathbf{a}) \tag{14}$$

wherein we use the lower case Greek gamma to denote the acceleration sensitivity of the elastic stiffness. The relationship between the acceleration sensitivity γ of the elastic stiffness and the acceleration sensitivity Γ of the resonance frequency is simply $\gamma \approx 2\ \Gamma$. The perturbations in the various BVD circuit elements follow directly from the substitution of equation (14) into equations (10) through (13):

$$C_0(\mathbf{a}) = C_0 \tag{15}$$

$$R_1(\mathbf{a}) = R_1(\mathbf{a}=0)*(1+\gamma \cdot \mathbf{a}) \tag{16}$$

$$L_1(\mathbf{a}) = L_1 \tag{17}$$

$$C_1(\mathbf{a}) = C_1(\mathbf{a}=0)*(1-\gamma \cdot \mathbf{a}) \tag{18}$$

These relations may then be used to calculate the BAW crystal resonator transimpedance under acceleration.

For SAW, SBAW, and similar resonators, the relations corresponding to equations (10) through (13) are more complicated, however the procedure for incorporating acceleration sensitivity in these equivalent circuit models follows directly from the derivation given here.

3.2 Oscillator Loop Analysis

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

$$|\Gamma| = - \frac{[\phi(\mathbf{a}) - \phi(\mathbf{a}=0)]}{2\ |\mathbf{a}|\ Q_L} \tag{19}$$

where $\phi(\mathbf{a})$ and $\phi(\mathbf{a}=0)$ are the perturbed and unperturbed phase of the resonator transimpedance, and Q_L is the loaded or phase slope Q of the resonator at the operating point. Implicit in equation (19) is that no other components in the oscillator loop are perturbed by the acceleration. For actual hardware implementations, this is often not true.

In the case where other loop components are also sensitive to the external acceleration, the effects on the other loop components are well represented by an ideal amplifier in series with an acceleration-sensitive load reactance. Based on experimental observations in discrete component/printed circuit board oscillators, we treat the load reactance as an ideal component with an acceleration-sensitive lead inductance as shown in Figure 1. The effect on the lead inductance may be modeled as a linear perturbation in similar fashion to equation (14), and the perturbation in transimpedance phase of the resonator/series load reactance combination may be employed in equation (19) to obtain the oscillator acceleration sensitivity under various operating conditions.

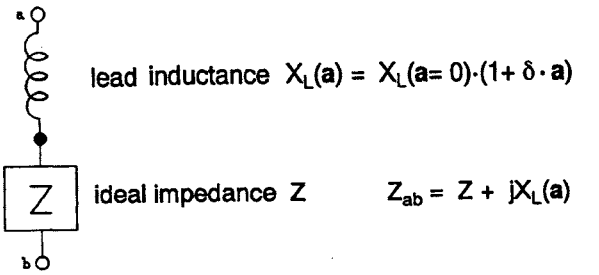


Figure 1. Empirically determined model of an acceleration sensitive loop component.

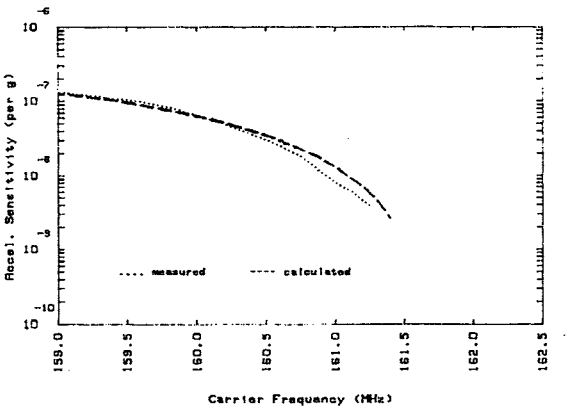


Figure 2. Measured acceleration sensitivity of an AT-cut BAW VCO compared to the acceleration sensitivity for the VCO as calculated using the resonator and loop component equivalent circuit models.

Figure 2 compares the application of this modeling technique to a crystal VCO with the measured data obtained for the unit. The results obtained using the equivalent circuit approach are in good agreement with the measured data.

3.3 Oscillator Design Considerations

The electronic circuit analysis clarifies certain aspects of oscillator design and construction for low acceleration sensitivity:

1) The acceleration sensitivity of the crystal oscillator depends on the location in frequency space of the oscillator operating point. Both the acceleration-induced phase perturbation and resonator Q_L may be functions of frequency, which leads to $|\Gamma|$ being a function of the loop operating conditions.

2) Crystal oscillator $|\Gamma|$ is minimized both by a) minimizing the acceleration-induced phase perturbations in the loop components (including the resonator) and b) maximizing Q_L of the resonator.

3) Acceleration-induced phase perturbations in loop components (other than the resonator) arise primarily from changes in lead inductances as the loop components shift and leads deform in response to the applied acceleration. Mechanically rigid circuit construction techniques (such as ceramic hybrids) should be employed and motion-sensitive components (components with leads, air-core inductors, trim pots, etc.) should be avoided.

4. CONCLUSIONS

The cause of acceleration sensitivity in crystal resonators is now well understood. At the present time, the most important practical points in designing crystal resonators for low acceleration sensitivity are the minimization of the resonator deformation and the maximization of mounting and mode shape symmetry. Aspect ratio compensation may also be applied in certain types of devices. In the future, application of the intrinsic acceleration insensitivity conditions may lead to additional design alternatives.

In the case of the acceleration sensitivity of the crystal oscillator, the interaction of the crystal resonator as an electronic component with the other loop components is only now coming to be appreciated. In designing oscillators for low acceleration sensitivity, the importance of maximizing resonator Q_L and eliminating motion-sensitive component types from both the design approach and manufacturing process cannot be overstated.

REFERENCES

[1] R. L. Filler, J. A. Kosinski, and J. R. Vig, "Further Studies on the Acceleration Sensitivity of Quartz Resonators," in Proceedings of the 37th Annual Frequency Control Symposium, Philadelphia, PA, June 1983, pp. 265-271.

[2] T. E. Parker, J. A. Greer, and G. K. Montress, "SAW Oscillators with Low Vibration Sensitivity," in Proceedings of the 45th Annual Frequency Control Symposium, Los Angeles, CA, May 1991, pp. 321-329.

[3] M. H. Watts, E. P. EerNisse, R. W. Ward, and R. B. Wiggins, "Technique for Measuring the Acceleration Sensitivity of SC-cut Quartz Resonators," in Proceedings of the 42nd Annual Frequency Control Symposium, Baltimore, MD, June 1988, pp.442-446.

[4] E. P. EerNisse, L. D. Clayton, and M. H. Watts, "Variational Method for Modeling Static and Dynamic Stresses in a Resonator Disc With Mounts," in Proceedings of the 43rd Annual Frequency Control Symposium, Denver, CO, May-June 1989, pp. 377-387.

[5] E. P. EerNisse, R. W. Ward, and O. L. Wood, "Acceleration-Induced Frequency Shifts in Quartz Resonators," in Proceedings of the 43rd Annual Frequency Control Symposium, Denver, CO, May-June 1989, pp. 388-395.

[6] E. P. EerNisse and R. W. Ward, "Crystal Resonator With Low Acceleration Sensitivity and Method of Manufacture Thereof," U.S. Patent 4,837,475, June 6, 1989.

[7] E. P. EerNisse, R. W. Ward, M. H. Watts, R. B. Wiggins, and O. L. Wood, "Experimental Evidence for Mode Shape Influence on Acceleration-Induced Frequency Shifts in Quartz Resonators," IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, Vol. UFFC-37, No. 6, pp. 566-570, November 1990.

[8] E. P. EerNisse, L. D. Clayton, and M. H. Watts, "Distortions of Thickness Shear Mode Shapes in Plano-Convex Quartz Resonators with Mass Perturbations," IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, Vol. UFFC-37, No. 6, pp. 571-576, November 1990.

[9] L. D. Clayton and E. P. EerNisse, "Frequency Shift Calculations For Quartz Resonators," in Proceedings of the 45th Annual Frequency Control Symposium, Los Angeles, CA, May 1991, pp. 309-320.

[10] P. C. Y. Lee and X. Guo, "Effect of Electrodes on the Acceleration Sensitivity of Crystal Resonators," in Proceedings of the 43rd Annual Frequency Control Symposium, Denver, CO, May-June 1989, pp. 416-418.

[11] P. C. Y. Lee and X. Guo, "Acceleration Sensitivity of Crystal Resonators Affected by the Mass and Location of Electrodes," in Proceedings of the 44th Annual Frequency Control Symposium, Baltimore, MD, May 1990, pp. 468-473.

[12] P. C. Y. Lee and X. Guo, "Acceleration Sensitivity of Crystal Resonators Affected by the Mass and Location of Electrodes," IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, Vol. UFFC-38, No. 4, pp. 358-365, July 1991.

[13] Y. S. Zhou, D. S. Shick, and H. F. Tiersten, "An Analysis of the Normal Acceleration Sensitivity of Contoured Quartz Resonators with Simple Rectangular Supports," in 1989 Ultrasonics Symposium Proceedings, Montreal, PQ, Canada, October 1989, pp. 393-400.

[14] Y. S. Zhou and H. F. Tiersten, "On the Influence of a Fabrication Imperfection on the Normal Acceleration Sensitivity of Contoured Quartz Resonators with Rectangular Supports," in Proceedings of the 44th Annual Frequency Control Symposium, Baltimore, MD, May 1990, pp. 452-460.

[15] H. F. Tiersten and Y. S. Zhou, "An Analysis of the In-Plane Acceleration Sensitivity of Contoured Quartz Resonators with Rectangular Supports," in Proceedings of the 44th Annual Frequency Control Symposium, Baltimore, MD, May 1990, pp. 461-467.

[16] H. F. Tiersten and D. V. Shick, "On the Normal Acceleration Sensitivity of Contoured Quartz Resonators Rigidly Supported Along Rectangular Edges," Journal of Applied Physics, Vol. 67, pp. 60-67, 1990.

[17] H. F. Tiersten and D. V. Shick, "On the In-Plane Acceleration Sensitivity of ST-Cut Quartz Surface Wave Resonators with Interior Rectangular Supports," Journal of Applied Physics, Vol. 67, pp. 2554-2566, 1990.

[18] Y. S. Zhou and H. F. Tiersten, "On the Normal Acceleration Sensitivity of Contoured Quartz Resonators with the Mode Shape Displaced with Respect to Rectangular Supports," Journal of Applied Physics, Vol. 69, pp. 2862-2870, 1991.

[19] H. F. Tiersten and Y. S. Zhou, "The Increase in the In-Plane Acceleration Sensitivity of the Plano-Convex Resonator Due to its Thickness Asymmetry," in Proceedings of the 45th Annual Frequency Control Symposium, Los Angeles, CA, May 1991, pp. 289-297.

[20] Y. S. Zhou and H. F. Tiersten, "An Analysis of the Normal Acceleration Sensitivity of Contoured Quartz Resonators Stiffened by Identical Top and Bottom Quartz Cover Plates Supported by Clips," in Proceedings of the 45th Annual Frequency Control Symposium, Los Angeles, CA, May 1991, pp. 298-308.

[21] A. Ballato, J. Kosinski, T. Lukaszek, M. Mizan, R. McGowan, and K. Klohn, "Acceleration Sensitivity Reduction in SAW and BAW Resonators by Electronic Means," in 1990 IEEE Ultrasonics Symposium Proceedings, Honolulu, HI, December 1990, pp. 573-576.

[22] J. Kosinski, A. Ballato, and T. Lukaszek, "Measurements of Acceleration-Induced Phase Noise in Surface Acoustic Wave Devices," in Proceedings of the 44th Annual Frequency Control Symposium, Baltimore, MD, May 1990, pp. 488-492.

[23] T. McClelland, B. Goldfrank, J. Korman, and D. Cocuzzi, "Manufacturing Methods and Technology for High Stability Vibration Resistant Quartz Crystal Units," U. S. Army LABCOM Research and Development Technical Report SLCET-TR-82-JO52-F, April 1991, 100 pp.

[24] D. V. Shick and H. F. Tiersten, "An Analysis of the Acceleration Sensitivity of ST-Cut Quartz Surface Wave Resonators Supported Along the Edges," in Proceedings of the 40th Annual Frequency Control Symposium, Philadelphia, PA, May 1986, pp. 262-268.

[25] H. F. Tiersten and D. V. Shick, "An Analysis of the Normal Acceleration Sensitivity of ST-Cut Quartz Surface Wave Resonators Rigidly Supported Along the Edges," in Proceedings of the 41st Annual Frequency Control Symposium, Philadelphia, PA, May 1987, pp. 282-288.

[26] D. V. Shick and H. F. Tiersten, "An Analysis of the In-Plane Acceleration Sensitivity of ST-Cut Quartz Surface Wave Resonators with the Substrate Extending Beyond the Surfaces," in Proceedings of the 42nd Annual Frequency Control Symposium, Baltimore, MD, June 1988, pp. 230-238.

[27] H. F. Tiersten and D. V. Shick, "An Analysis of the In-Plane Acceleration Sensitivity of ST-Cut Quartz Surface Wave Resonators with Interior Rectangular Supports," in Proceedings of the 43rd Annual Frequency Control Symposium, Denver, CO, May-June 1989, pp. 396-404.

[28] J. A. Greer and T. E. Parker, "Improved Vibration Sensitivity of the All Quartz Package Surface Acoustic Wave Resonator," in Proceedings of the 42nd Annual Frequency Control Symposium, Baltimore, MD, June 1988, pp. 239-251.

[29] H. F. Tiersten, "Perturbation Theory for Linear Electroelastic Equations for Small Fields Superposed on a Bias," Journal of the Acoustical Society of America, Vol. 64, No. 3, pp. 832-837, September 1978.

[30] J. Kosinski, A. Ballato, T. Lukaszek, and S. Mallikarjun, "A New Approach to Modeling Quartz Resonator and Oscillator Acceleration Sensitivity," in Proceedings of the 5th European Frequency and Time Forum, Besancon, France, March 1991, pp. 25-29.

HIGH QUALITY QUARTZ CRYSTAL OSCILLATORS COMPATIBLE WITH SURFACE MOUNTING TECHNOLOGY

CLAUDE GELLE, LAZARE LEVINE

THOMSON/CEPE
44, Av. de la Glacière - BP 165
95105 ARGENTEUIL - FRANCE

I/ INTRODUCTION

According to the electronic miniaturization trends CEPE has been studying for two years different hermetically sealed packages technologies to make high quality quartz crystal oscillators compatible with surface mounting technology (SMT) and requirements for military applications.

In this way, CEPE explored the capabilities of metal-ceramic package technology and tested several sealing method including soldering, laser and resistance welding.

Different test packages were chosen compatibly with sealing methods and possibility to include in them oscillator components such as integrated circuit.

At first we have to verify the bahaviour of the resonators considering there different sealing methods.

II/ ELECTRICAL PERFORMANCES

Tests have been performed on various configurations including different frequencies in order to evaluate the interest for a large range of applications such as TCXO, TMXO, smart munitions.

Table I gives the selected frequency values and the aims regarding electrical performances.

III/ DESCRIPTION OF TEST DEVICES

Tests devices will be presented according to sealing method.

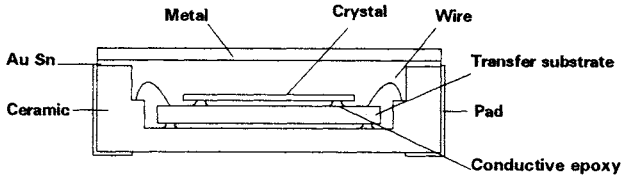
3.1./ SOLDERING METHOD

Package : Standard lead chip carrier type 40 LLCC

Crystal attachment : By conductive epoxy. At first the crystal is mounted on a transfer ceramic susbstrate. Then the assembly is attached on the bottom of the package. Gold wires assure the electrical interconnections with the chip carrier.

Atmosphere : Nitrogen

Solder : Tin gold preform (280°C)



3.2./ LASER METHOD

Package : Specific package developed by CEPE.

Crystal attachment : By conductive epoxy. The crystal is mounted directly on the bottom of the package. Electrical interconnexions are ensured by via holes.

Atmosphere : Primary vacuum.

MATERIAL	QUARTZ	QUARTZ	QUARTZ	QUARTZ
Frequency (MHz)	20	40	120	120
Cut	AT	AT	AT	SC
Crystal diametre m/m	6	6	6	6
Vibration mode	Fundamental	Fundamental	5^{th} overtone	3^{rd} overtone
Temperature range °C	- 40 + 85	- 40 + 85	- 40 + 85	- 60 + 110
Frequency stability ppm	± 50	± 50	± 50	± 5
Reference temperature °C	25	25	25	25
Frequency tolerance ppm	± 10	± 10	± 10	± 15
Atmosphere	N_2	N_2		
Atmosphere	Vacuum	Vacuum	Vacuum	Vacuum
Ageing/month	1.10^{-7}	1.10^{-7}		

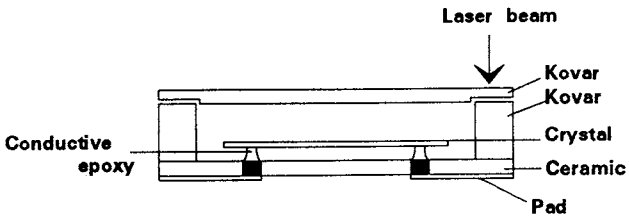


Figure 2

3.3./ RESISTANCE WELDING

Package : Metal ceramic package type (See fig. 3).

Crystal attachment : By conductive epoxy. The crystal is mounted on two of four metal supports brazed on the bottom of the package.

Atmosphere : Primary vacuum.

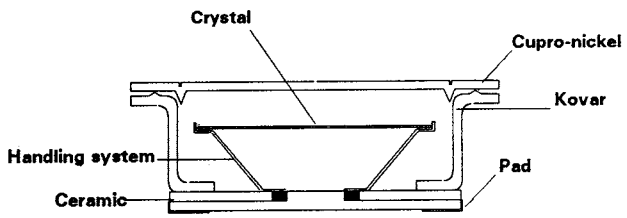
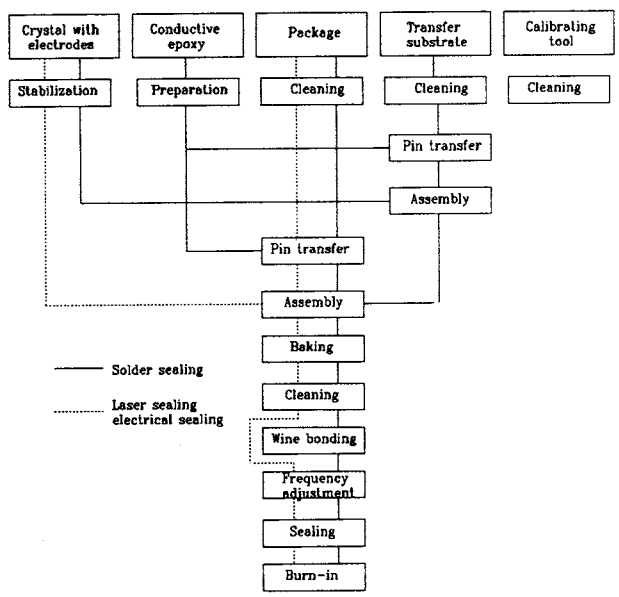


Figure 3

IV/ PROCESSING

The general outline of the processing flow chart is shown (Figure 4).



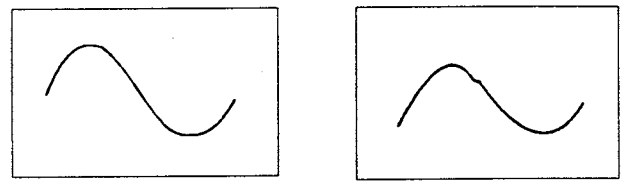
For these packaging technologies, three process steps need a particular attention :

- Crystal attachment
- Frequency adjustment
- Hermetical sealing

4.1./ CRYSTAL ATTACHMENT

When the crystal is attached by glue onto metal ribbons this handling system may cause a small environmental dependancy.

So, when the crystal is attached directly onto the bottom of the package, it's performances may be degraded by thermal and mechanical stresses leading to a distorting of the frequency temperature response of the resonator.



Normal curve Distorsion of curve

Figure 5

To minimize the effect of stress particular improvements had to be performed on the following subjects :

- Choice of the glue
- Calibration of the glue spot
- Orientation of the plate
- Curing profile

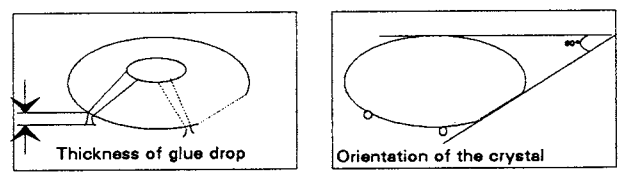


Figure 6

4.2./ TUNING

After the assembly of the crystal in the package only one electrode is accessible for the frequency adjustment.

For this step, two methods were successfully tested :

- Sputtering way : deposition of metal on the electrode under vacuum.

- Etching way : Etching of metal by ionic beam

Electrical monitoring of the crystal during processing is done in both cases.

At this step, we have to consider the drift of frequency caused by sealing and burn-in processes to determinate the aimed frequency value.

Table 1 gives the frequency drifts at 20 MHz for the differents sealing methods.

SEALING METHOD	AFTER SEALING	AFTER BURN-IN	TOTAL	
Soldering	- 4	0	- 4	ppm
Laser welding	- 5	± 2	- 5	ppm
Resistance welding	+ 3	± 0,3	+ 3	ppm

TABLE 1

4.3./ HERMETICAL SEALING

Characteristics of the various sealing equipments are presented successively.

4.3.1./ SOLDERING

Equipement : Infrared solder reflow furnace with four heating zones - PIEZO CERAM.

Temperature profil :

	T_{21}	T_{22}	T_{23}	T_{24}
°C	370	270	240	450

Conveyor speed : 20 cm min^{-1}
Temperature peak : 320°C

Atmosphere : Hydrogenated nitrogen

4.3.2./ LASER WELDING

Equipement : The experimental equipement developped for this study is presented fig. 7

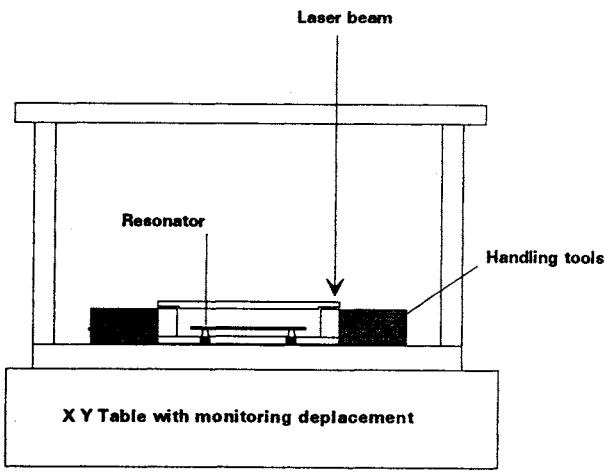


Figure 7

It was used a pulsed YAG laser from JK LASER with following adjustments :

Local lens	:	150 mm
Distance resonator/lens	:	132 mm
Pulse duration	:	3,8 ms
Frequency	:	10 Hertz
Energy	:	2,8 joules
Sealing speed	:	13,2 cm min^{-1}

Temperature profile : During the process the temperature of the bottom of the package remains under 120°C.

Atmosphere : Primary vacuum.

4.3.3./ RESISTANCE WELDING

Equipment : Standard resistance welding equipment type FP 500 K from PECO used by CEPE for T0 cases sealing.

Tools were adapted to the squarred geometry of test packages.

Atmosphere : Primary vacuum

V/ CHARACTERISATION

5.1./ SCREENING OF TESTS

The different technologies of resonators had been evaluated in thermal and mechanical environment.

The screening of tests is presend table 2.
Tests conditions were specified as following

- Second level mounting simulation

The devices are conveyed twice into an infra red reflow oven to simulate the mounting of resonator on PCB. The highest temperature is about 240°C.

- Mechanical shock : NFC 20727

. Level	:	1,500 g 1/2 sinus
. Duration of pulse	:	0,5 ns
. Procedure	:	6 directions

- Vibration, variable frequency : NFC 20706 Test FC

. Level	:	30 g
. Frequency range	:	10 to 2000 Hz
. Procedure	:	3 axis

- Temperature cycling : NFC 20714 - Test Na

. Level	:	- 55°C + 125°C
. Procedure	:	Two chambers
. Transfer time	:	< 1 minute
. Number of cycles	:	300

- ***Moisture resistance*** : NFC 20703 - Test Ca

  . Level : 40°C - 93% RH
  . Duration : 56 days

- ***Ageing*** : This parameter is evaluated for each type of resonator previously mounted in standard ovenized oscillators.

SCREENING OF TESTS

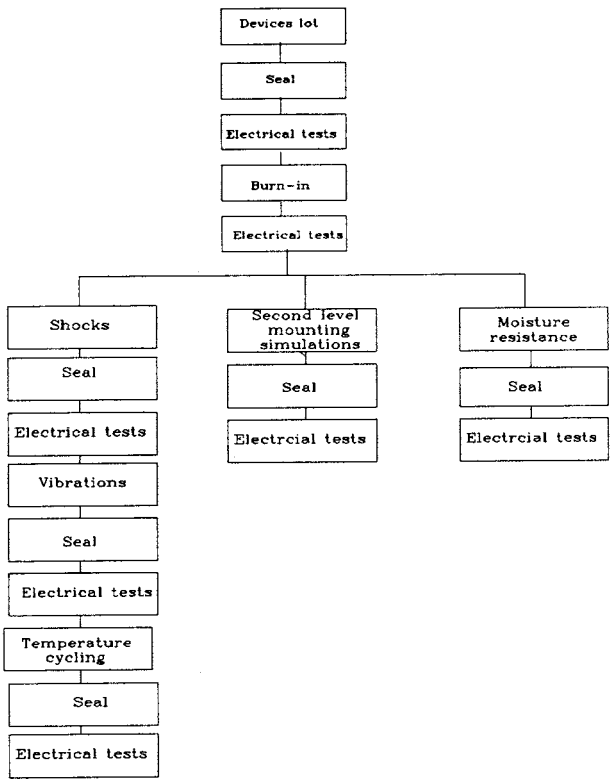


TABLE 2

5.2./ RESULTS

5.2.1./ SHIFT FREQUENCY AT 25°C AFTER EACH TEST

Table 3 presents the shift values in ppm

5.2.2./ AGEING

The results of ageing are presented in table 4 three months after the start.

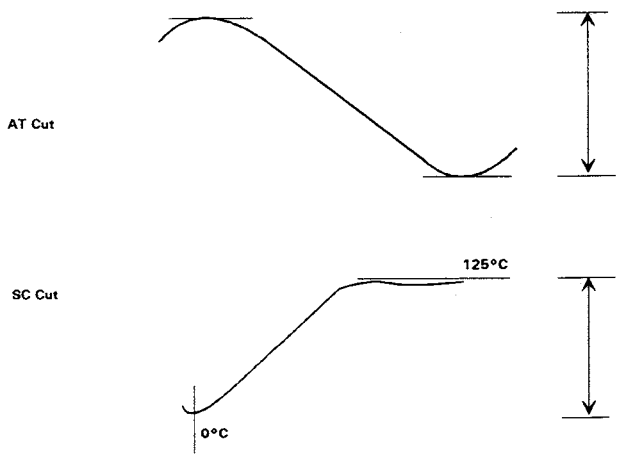
SEALING METHOD	FREQUENCY MHz	QUANTITY	STEP BEFORE	MOUTHLY AGEING 10^{-7}
Soldering	20	2	Burn-in	0,2 to 1,5
	20	2	Thermal cycles	2 to 3
	40	2	Burn-in	1,2 to 1,5
Laser	20	1	Burn-in	6
	120 - AT	1	Burn-in	10
	120 - SC	1	Burn-in	10
resistance welding	20	4	Burn-in	0,5 to 1,8
	120 - AT	4	Burn-in	5
	120 - SC	4	Burn-in	5

TABLE 4

5.2.3./ SHIFT OF FREQUENCY TEMPERATURE RESPONSE AFTER EACH TEST

Table 5 presents shift values in percent.

$$\frac{\Delta F \text{ after each test}}{\Delta F \text{at To}} \times 100$$



SEALING METHOD	F	CUT	CRYSTAL DIAMETER MM	QUANTITY	SECOND LEVEL MOUNTING SIMULATION	SHOCKS AND VIBRATIONS	TEMPERATURE CYCLES	MOISTURE RESISTANCE	FAILURE
Soldering	20	AT	6,5	10	± 1	± 1	± 2	0	0
	40	AT	5	10	± 1	± 2	± 5	0	2 (1leakage) (1broken)
	120	AT	5	3	± 1	± 2	± 5	0	1 (1 broken)
	120	SC	5	3	± 1	± 2	± 5	0	0
Laser	20	AT	5	5	± 2	± 2	± 6	0	0
	40	AT	5	5	± 2	± 2	± 10	0	1 (1 broken)
	120	AT	5	3	± 2	± 3	± 5	0	0
	120	SC	5	3	± 2	± 2	± 3	0	0
Resistance welding	20	AT	6	10	± 2	± 3	± 15	0	1 (leakage)
	120	AT	6	5	± 2	± 4	± 15	0	0
	120	SC	6	5	± 2	± 4	± 10	0	1 (leakage)

TABLE 3

SEALING METHOD	FREQUENCY MHz	CUT	CRYSTAL DIAMETER	QUANTITY	SECOND LEVEL MOUNTED SIMULATION	SHOCKS AND VIBRATION	TEMPERATURE CYCLING
Soldering	20	AT	6,5	10	1	2	6
	40	AT	5	10	1	3	7
	120	AT	5	3	1	5	8
	120	SC	5	3	1	5	8
Laser	20	AT	5	5	1	3	7
	40	AT	5	5	1	3	7
	120	AT	5	3	2	5	9
	120	SC	5	3	1	4	9
Resistance welding	20	AT	6	10	1	2	7
	120	AT	6	5	1	3	10
	120	SC	6	5	1	3	9

TABLE 5

5.4./ DISCUSSION

Comments are given about the electrical performances for each sealing method

5.4.1./ SHIFT OF FREQUENCY

Soldering method

Resonators enclosed in chip carriers provide the best performances under environmental stress. To explain that, the following causes may be invoked :

The sealing temperature, about 320°C, is more important than the temperature of the screening tests and so it minimizes the sensitivity of the resonators.

The use of a transfer substrate could also have been invoked to explain the higher stability of this kind of assembly - Nevertheless the two broken samples make an evidence that it failed to demonstrate its efficiency.

Otherway, we have to notice that the sensitivity of a resonator to the stress also depends on its diameter value.

Laser method

The results of the resonators enclosed in laser sealing packages are not very different from those enclosed in solder sealing packages for the same diameter.

We also have to notice one broken sample.

Resistance welding

No broken pieces are observable - So, the handling system demonstrate its efficiency. However, the resonators show the highest sensitivity to the thermal/mechanical stresses in spite of the use of handling system.

These results are probably due to :

- a still limited optimization of the handling system design
- a lack of stabilization for the packages before use.
- a still limited optimization of the sealing parameters

In fact, we think that an annealing step at a high temperature value (~ 250°C) or an extra cycling step should widely improve the stability versus further cycling tests.

5.4.2./ EVOLUTION OF THE TEMPERATURE VERSUS FREQUENCY BEHAVIOUR

For the three technology ways the shift remains under 10% of initial drift.

Resonators enclosed in electrical sealing package show the most important shift for the reasons explained before.

Otherwise, we notice that the first two sealing methods generate some distorsions on the frequency vs temperature curve like fig. 5.

This kind of anomaly limits the performances of frequency stability when the resonators are used in temperature-compensated oscillators.

On another hand it is an evidence of an excess of stress resulting from the mounting. This result is consistent with the lack of handling system for soldering and laser type packages. It's also consistent with the previously mentionned broken samples during temperature cycling.

This is of critical interest, because the above mentionned distorsions are impredictable and cannot be managed on an industrial line.

5.4.3./ AGEING

The ageing aim is reached for the resonators enclosed in chip carriers and nearly for the resonators enclosed in resistance welding packages (a typical ageing curve is presented figure 8).

Less performing results has been obtained with resonator enclosed in laser sealing packages probably due to the action of pollution of the crystal consequently to the metal evaporation during the sealing process. At this time, we don't have concluded on this way.

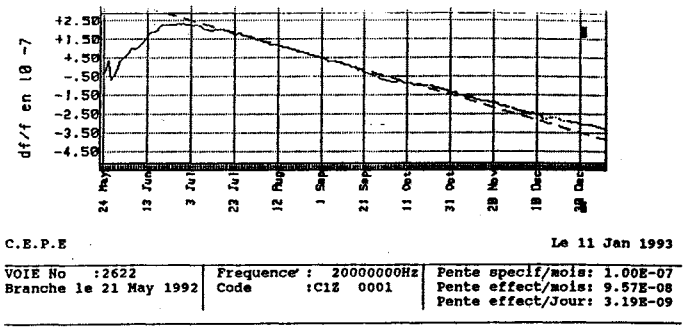


FIGURE 8

VII CONCLUSION

To summarize the results, we can say that, even when quite.interesting electrical performances are reached with the soldering sealing and laser sealing processes, some inconveniences appear for both technologies regarding the industrial application. The main one is the stress evidence on the plate due to the lack of mechanical decoupling between the plate and the substrate which generates casual strong defects on the frequency versus temperature curves and even plate breakages.

At this point, mounting with a handling system should be prefered.

On theother hand, the availability of low cost packages for electrical welding, make it very interesting to use this technology because it is very closed to the today state of this art for the sealing of pin packages which has shown to be highly performing regarding stability. When tested, this sealing technology has shown to be satisfactory on the point of view of ageing and temperature behaviour under typical test file. Performances are still insufficient regarding the frequency shift under environmental tests. However the comparison was not fair because the samples enclosed by soldering or laser means are exposed to a high temperature during the sealing process, which acts as an annealing step. Conversely, electrical welding is performed at ambiant temperature.

In the coming months, test will be performed on electrical welding with an adding annealing step. Regarding the available experience on the pin-package, and on the relaxing procedures for metal handling systems, we think the results should be successfull.

In addition, such equipments as laser soldering furnaces are not yet used in the piezoelectric process. It should be an extra (and important) investment to get them. Conversely, the existing electrical sealing equipment can be converted for the described electrical sealing package.

At this point our opinion about trends is that the package for surface mounted crystal and, further, oscillator will be of the electrical sealing type.

IMPROVING OPERATIONAL AND TECHNICAL PARAMETERS OF QUATZ OSCILLATORS

Prof. Arvid G.Gevorkyan, Dr. Alexander Myasnikov
Sergey P.Kovita

Russian Institute of Radionavigation and Time (RIRT)
2,Rastrelli square,St.-Petersburg, 193124,Russian Federation

Abstract One of the main research directions in RIRT is creation of synchronization systems and devices for space vehicles. In so doing,it was necessary to develop precision quartz oscillators (QO) used as a part of atomic frequency standards as well as self-contained sources of reference frequency. Emphasis was on realizing limiting parameters as regards short-term frequency stability as well as decreasing long-term frequency drifts and reaching a minimum of modulation performance non-linearity. As a result of this work, QO were designed with parameters σ_{1s} ~$3 \cdot 10^{-12}$, mean value of frequency drifts up to $5 \cdot 10^{-11}$ per one day, non-linearity of frequency modulation performance ~ 3 - 5 per cent.

Introduction

Up to the present QO's make up a bulk of main sources of high-stable oscillation for various systems in the fields of navigation, synchronization, time keeping etc.

The Russian Institute of Radionavigation and Time (RIRT) specializes, as regards QO's, in developing precision QO's being used as parts of atomic frequency standards, as well as self-contained high-stable frequency sources.

This paper deals with precision voltage-controlled QO's used in high-accuracy radio systems. Such QO's must meet increased requirements on long-term frequency drift (τ_m > 1 h), short-term frequency stability (τ_m < 100 s),as well as on spectral power density of phase noise. With respect to operational factors, such parameters are of importance as warm-up time, consumed power, weight, noise immunity and stability in destabilizing environment.

1.QO concept and parameters of QO's of the first generation

The main building principles of precision QO's structure were worked out in 70 s - 80 s. Such principles involve as follows: design of QO's package; arrangement for fixing the oven and the temperature bridge with signal generator in QO's package by means of foam encapsulation; "bookstand" design of signal generator; oven design with balance cylinder. Those design and circuitry decisions were of progressive nature and enabled to develop and to product QO's used as self-contained frequency sources as well as parts of atomic frequency standards. Those decisions are being used up to the present time without any important change.

With respect to the operation, RIRT's QO's are divided to three following groups: thermally compensated QO's (TCQO), ovenized QO's and QO's with a crystal integrated with an oven. The ovenized QO's are the largest in number. At present these QO's enable to realize the best accuracy parameters, as well as the potential of precision quartz crystals. Up to the present time nearly 10 types of such QO's are mass-produced; the output of production was as much as 400 in mid-80 s.

All those QO's may be classified as follows: on circuitry type of signal generator; on operation of heat regulator (switching or linear), on number of oven stages (one or two); on frequency adjustment (electromechanical, electronic or digital).

The QO's of the first generation use as an oven a glass Dewar vessel with heat isolating foam encapsulations. A high uniformity of temperature field within the oven is performed by means of metal balance cylinder, on which heater components and temperature sensors are placed. In such a case, the temperature within the area of piezoelement placement is maintained with the accuracy from 0,01 to 0,005°C.

Functional structure of QO's models 298, 305, 329, 339 is given in Fig.1. As is seen, the precision QO's are characterized by high degree of unification (both circuits and design); the circuitry is carefully worked out. These features enable to improve the reliability of these QO's.

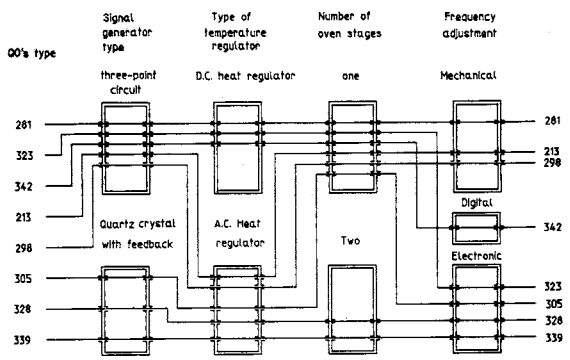


Fig.1. QO's functional structure

An example is 5 MHz-QO 339. The high level of its technical parameters is provided due to specific design with two-stage temperature stabilization in Dewar vessel, as well as by using SC-cut crystal. An external view of QO is shown in Fig.2.

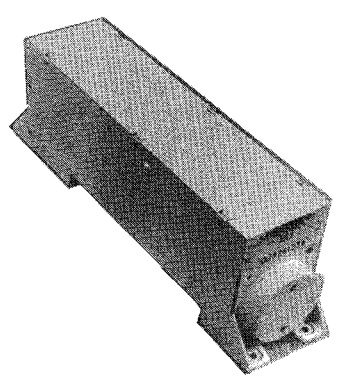


Fig.2.QO 339 for atomic beam frequency standard on board of GLONASS satellite

The possibilities of QO's applications for various types of user equipment are largely expanded due to two independent frequency-controlled inputs with different slope of modulation performance. Specific care was taken for the circuit controlling an oscillator frequency to linearize a rough frequency adjustment performance. The film temperature sensors and heaters with distributed structure are used in the oven, thus providing an accuracy of operational temperature maintaining within the area of piezoelement placement 0,01 °C. The high QO's parameters are confirmed by its real operation as a part of space systems.

The QO 323 for airborne equipment has also a rated frequency of 5 MHz which is controlled by precision AT-cut crystal (Fig.3). High spectral parameters and frequency stability are provided due to the circuitry of signal generator and high-frequency amplifier.

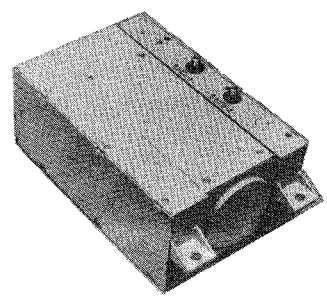


Fig.3. QO 323 for Rb vapour airborne standard

A combination of unconventional temperature regulation system for an oven quick warm-up with high-effective thermoisolation realized as a Dewar vessel provides a high operational availability and a wide range of operational temperatures. In order to update a frequency, there are two inputs for frequency electronic adjustment in QO, which slopes differ by one order of value, thus enabling to compensate irreversible crystal frequency drifts over all the period of QO's operation. This QO can function as a part of airborne atomic frequency standard from switching electrical power sources.

The specific features of QO 323 are as follows: wide range of operational temperatures, high resistance to mechanical disturbances and shocks, built-in buffer amplifier with three outputs having a high degree of output isolation and increased maintainability while using as a part of equipment to be repaired. Technical parameters of QO 339 and QO 323 are given in Tab .1.

The real operation of QO's as parts of radio systems confirms its high technical parameters. For instance, AT-cut QO's provide daily frequency drift at the level of $(1,5 - 6,5)\cdot 10^{-10}$; as for SC-cut QO's, this value is $(1,0 - 3,5)\cdot 10^{-11}$ (after 30 days of continuous operation). A frequency drift over the 1st day doesn't exceed 10^{-8}. Short-term frequency stability (Allan variance) is at the level of $(1,5 - 3,0)\cdot 10^{-12}$.

Table 1

Value of parameter	QO 339	QO 323
1	2	3
1. Rated frequency, MHz	5,0	5,0
2. Fractional fluctuation of operational frequency for the 1st day over the whole t° range	2×10^{-9} after 6 hours of preparation 5×10^{-11} after 90 days	1×10^{-7} after 40 min of preparation
3. Fractional deviation of operational frequency for 30 days over the whole t° range	7×10^{-9} after 5 days in operation	1×10^{-7} after 4 hours in operation
4. Fractional deviation of operational frequency over 1 year of continuous operation	6×10^{-8}	5×10^{-7}
5. Temperature frequency coefficient over environmental temperatures, 1 / °C	3×10^{-11}	3×10^{-10}
6. Standard deviation of frequency random variation (Σ_m = 1 s)	3×10^{-12}	2×10^{-11}
7. Number of output signals	1	3
8. Form of output signal	sinusoidal	sinusoidal
9. Output voltage at 50 Ohm (r.m.s.), V	0,3 - 0,5	0,3 - 0,5
10. Output isolation, dB	-	60
11. Number of control inputs for frequency correction	2	2
12. Limits for frequency updating, Hz	+- 2,5	+- 2,5
13. Slope of fine frequency correction, HZ/V	0,04 - 0,08	0,04 - 0,1
14. Range of operational temperatures	0...+ 50	- 60...+ 60
15. Power source voltage	+ 20; - 18	27
16. Consumed power, W warm-up steady-state	6,5 4,5	33 8
17. Volume, dm^3	1,2	1,5
18. Weight, kg	1,2	1,2

2. Ways to improvement of QO's operational/technical parameters

In order to improve QO's operational/technical parameters, a set of measures is being realized in RIRT which gives grounds to say about the development of precision QO's new generation. Such design/circuitry decisions involve, in the first place, use of metal Dewar vessel with two stems and "active" notches-heaters. This enables to increase a mechanical stability due to excluding console elements, to improve electromagnetic compatability and noise parameters because of shielding nature of metal vessel and physical spacing of critical links through the opposite stems while reducing at the same time the electrical connections. Besides, an improvement of temperature field is achieved due to the symmetry of design, as well as an improvement of hardness to environmental radiation.

Second, a quartz crystal in metal package with ceramic components is purposely designed. It is expected it will enable to improve short-term stability and to reduce warm-up time.

Third, specific circuits are developed for signal generator assemblies as well as for high frequency amplifier, resulting in improvement of frequency temperature stability and QO's noise parameters.

The combination of this decisions enables to reduce a warm-up time up to the achievement of high-stability mode to 5 - 10 min while maintaining an oven temperature with the accuracy of (0,01 - 0,005) °C. Short-term stability is $(1 - 2) \cdot 10^{-12}$. The experimental sets of such QO's are now under the tests.

The disadvantages of ovenized QO's are relatively large dimensions (volume 0,8 - 1,5 dm^3) and large steady-state consumed power due to the large stabilized volume in the Dewar vessel. Such factors for fast warm-up of crystal are required as large initial power and use of heaters with a forced warm up, especially where low environmental temperature takes place. Heaters' placement and choice of its configuration providing the prescribed temperature field would add complexity to design and production of QO's.

Development and production of precision crystals integrated with ovens may be considered as an important achievement in the field of microminiaturization. It is a combination of a quartz crystal itself with elements of temperature stabilizing/thermoisolation system, that is, the crystal body is not only a piezoelement carrier, but is at the same time its oven and contains temperature sensors and executive elements of temperature stabilizing system. Transistors and posistors are usually utilized as elements heating an internal volume of integrated crystal-oven. As a result, there is no need to have an individual oven to place crystal body.

While using this concept, a compact QO 251 was designed, in which customized microcircuits were used containing oscillator, amplifier, thermoregulator and voltage stabilizer. The temperature sensor as well as heating transistors are also contained with the semiconductor chip, along with thermoregulator and measurement bridge. The developed circuit of a signal generator enabled to realize an universal chip without external reactances. The chip breadboard is shown in Fig. 4.

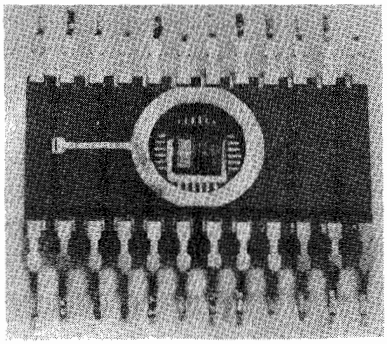


Fig. 4. Chip breadboard

Using a crystal with an internally ovenized quartz wafer of double-rotated cut, as well as a customized circuit for the signal generator, has provided high accuracy parameters within a wide range of environmental temperatures, along with small steady-state power, dimensions and weight (Q0's volume is 13 cm^3)(See Fig.5).

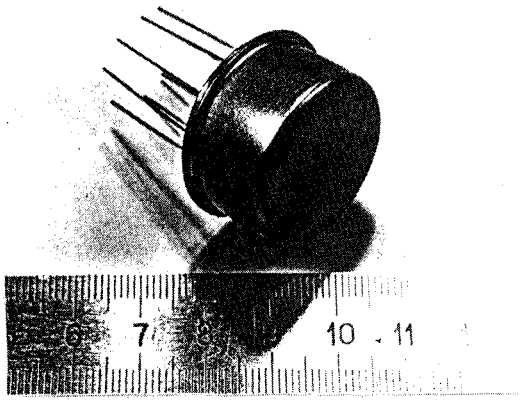


Fig.5. An external view of Q0 251

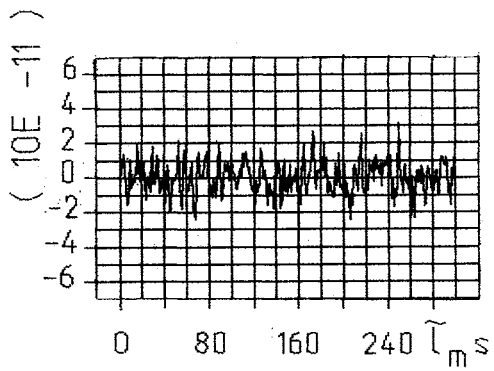


Fig.6. Short-term stability exluding frequency drift

Short-term stability exluding frequency drift is shown in Fig.6.

The main Q0 251 parameters are given in Tab.2.

Q0 251 is at present at the stage of mastering serial production. The works are being conducted at the same time on improving of short-term frequency stability. The obtained results enable to hope that under conditions of serial production the Allan variance will be constantly reproduced at the level of $5 \cdot 10^{-12}$ over time intervals 1 s.

Table 2

Value of parameter	Q0 251
OUTPUT SIGNAL FREQUENCY	5 or 10 MHz
OUTPUT SIGNAL - pulse shape - voltage	 sinusoidal 0,3 V (rms)
FREQUENCY STABILITY - long-term (aging) over 1 day over 1 month over 1 year - short-term (τ_m = 1 s) - within the temperature range from - 60 to + 70°C - within the load range 1 KOhm +-10% - within the power voltage change +_ 10%	 1×10^{-9} 5×10^{-8} 1.5×10^{-7} 1×10^{-11} 1×10^{-8} 2×10^{-8} 1×10^{-8} 5×10^{-9}
FREQUENCY SETTING Warm-up time for frequency setting with an accuracy: 5.10^{-7} 5.10^{-8}	 3 min 5 min
FREQUENCY UPDATING - range	 $+-5.10^{-7}$
REQUIREMENTS TO POWER SOURCE Voltage Consumed power warm-up steady state	 15...27 V 20 W 0,7 W
WEIGHT, g	35
DIMENSIONS, mm	28 (diameter) 20 (height)

As for Q0's of the second generation, its vibrational and mechanical tests have given the value of quartz crystals' g-sensitivity within the range of $(5 - 15) \cdot 10^{-10}$ /g for used systems which fixed piezoelements. In so doing, it was stated that the value of operational frequency is practically independed of the vibrational frequency within the range from 10 to 2000 Hz, that is to say there are no resonances in the fixing system.

Spectral measurements of Q0's output signal enabled to state a dependence of its spectrum on vibrational amplitude/frequency, thus providing to determine the maximum value of mechanical disturbances in order to maintain an equipment normal operation (for example, in the case of atomic frequency standard).

For decreasing the g-sensitivity, the works are now being conducted in RIRT on realizing Q0's with QAS- and BVA-type quartz crystals.

Conclusion

Q0's developed and produced in RIRT totally meet the requirements imposed by various systems. The experience in the field of operation of such devices over more than 20 years gives grounds to say about its high reliability. As a result of this work on improving Q0's operational parameters, the devices of a new generation are developed combining high operational/metrological parameters and small volume.

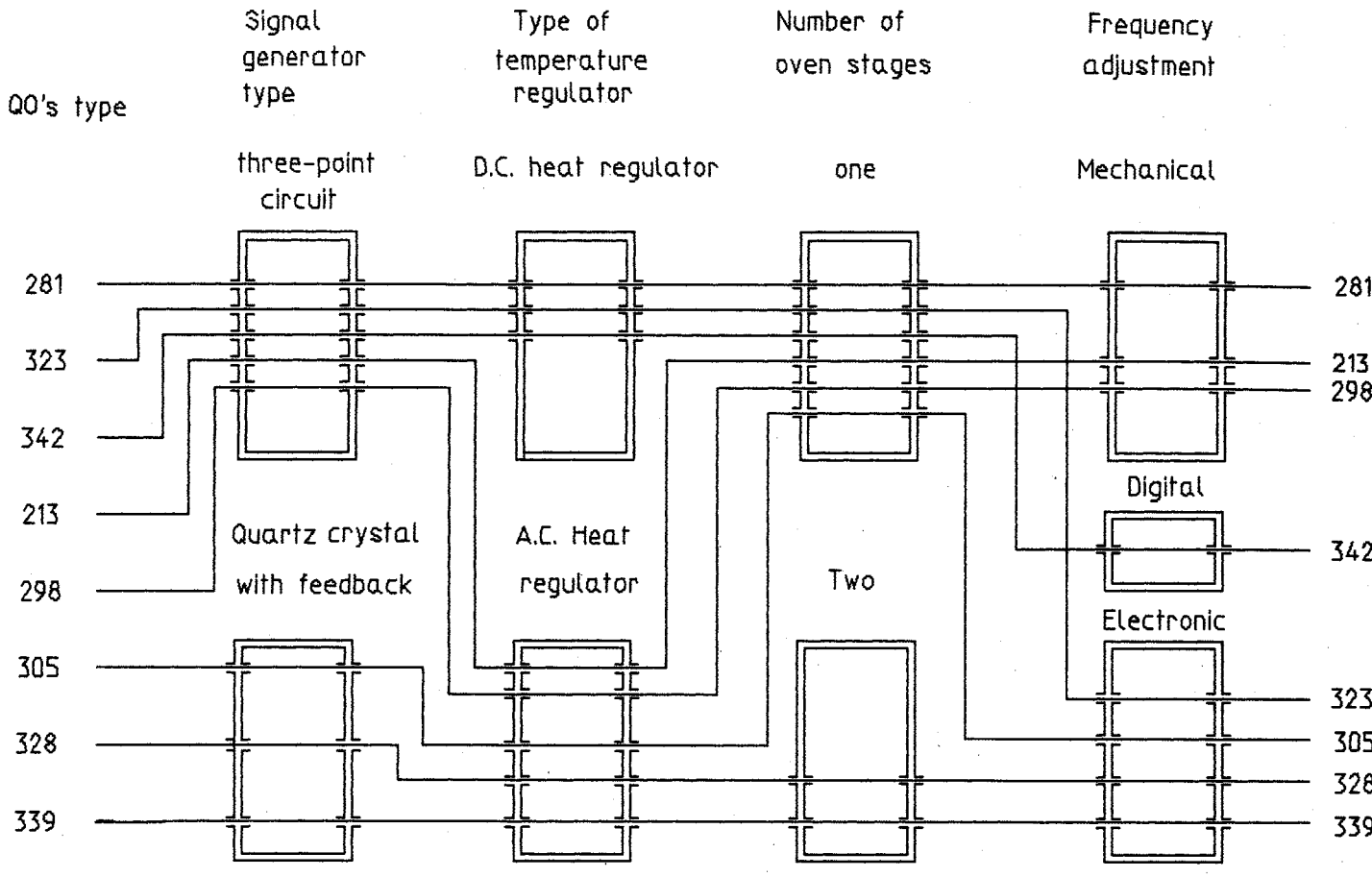


Fig. 1. QO's functional structure

TIME TRANSFER

Chairman: C. Thomas

CHARACTERIZATION OF THE SIGNAL DELAYS IN A GROUND STATION DESIGNED FOR SATELLITE TWO WAY TIME TRANSFER

J.A.Davis and P.R.Pearce

Division of Electrical Science,
National Physical Laboratory,
Queens Road, Teddington, Middlesex, TW11 0LW.

ABSTRACT

A major source of uncertainty in satellite two way time transfer arises in the determination of the delay asymmetry in the ground station instrumentation. In this paper, the results of a detailed examination of the signal delays in a ground station operating with a Hartl MITREX modem are presented. Several different ground station test loop arrangements are studied. The up-link and down-link delays are separately measured and the delay asymmetry determined. Variations of the signal delays in each element of the ground station and satellite simulator are examined. Coaxial multiplexers are used to switch the signal path between different loop arrangements every 15 minutes. This enables variations in the signal delay in each element to be measured with a resolution of 10ps. The dependence of the delay changes on operating parameters are discussed.

keywords
time transfer, satellite communications, microwave measurement, delay measurement.

1. INTRODUCTION

Two way time transfer via geostationary satellites offers the possibility of sub-ns accuracy [1], [2], [3]. The limiting uncertainties arise in the determination of the propagation delay asymmetries of the satellite signals. The asymmetry is defined as the difference between the propagation delay in the two directions. Delay asymmetries occur during the signals propagation through the ground station instrumentation, geostationary satellite, atmosphere and ionosphere. Asymmetries are also produced by the relativistic effects of the earth's rotation [4].

A major source of delay asymmetry occurs in the ground station up-link and down-link instrumentation. Traditionally, the asymmetry has been determined by calibration with a portable station which acts as a transfer standard between participating laboratories [1],[2]. However, this is both costly and time consuming and does not take account of future changes in the instrumentation delays. These may occur due to aging, or to replacement or improvement of the ground station components.

The delay asymmetry within a ground station may be measured using a satellite simulator and microwave mixers, thus eliminating the requirement to use a portable station [5]. In this paper, studies of the ground station propagation delays are analysed. Loop arrangements were formed using elements of the ground station, satellite simulator and microwave mixers. Coaxial multiplexers were used to switch the signal path between a variety of loops which contained some common and some individual elements. This permitted the delay variations within each element of the ground station to be determined. The origin of these changes are examined.

2. COAXIAL MULTIPLEXER METHOD

2.1 NPL ground station

The NPL ground station is shown schematically in figure 1, and is typical of a ground station designed for two way time transfer. An older analogue type MITREX modem was used; the 5 MHz modem frequency reference was supplied by a frequency synthesizer. Adjustment of this frequency enabled the required 1kHz up-link frequency agility to be obtained. The modem's internal 1pps signal was used in a variety of loop tests.

2.2 Experimental arrangement

Two Tastronic TAS023 coaxial multiplexers were used to switch between different test loop arrangement shown in figure 2. The common elements in the loop form signal paths C while the multiplexers switched between paths A and B. The instrumentation contained in loops A and B depended on the test. Typically loop A consisted of only a short length of coaxial cable. Loop B contained either single elements of the ground station or the whole of the up-link or down-link instrumentation, along with microwave mixers to perform frequency changes. Two multiplexers were used to achieve a high degree of electrical isolation between the two paths. The replacement of one multiplexer by a tee piece, introduced unwanted signal reflections which changed the delays.

Data was collected continuously using a PC. 14 minutes of data (840 readings) were averaged to produce each point. Both multiplexers were switched every 15 minutes to produce two data sets. Each data set was interpolated and the difference between the two sets determined (loop B delay - loop A delay). If path A contained a short cable, whose delay is assumed to be constant, then the longer term delay variations within the elements in path B may be determined.

2.3 Addition of delays

Propagation delays were attributed to each of the elements of the ground station instrumentation, these delays were assumed to add linearly. A knowledge of the operating conditions is vital for a meaningful determination of the delay of each element since propagation delays are sensitive to the signal level, carrier to thermal noise ratio (C/N_0), and frequency. The addition of an element to a loop arrangement changes signal characteristics and hence the delays in other elements positioned in a later part of the loop. This effect was most pronounced where frequency changes were involved and in some examples the addition of elements to a loop resulted in a reduction in the total delay.

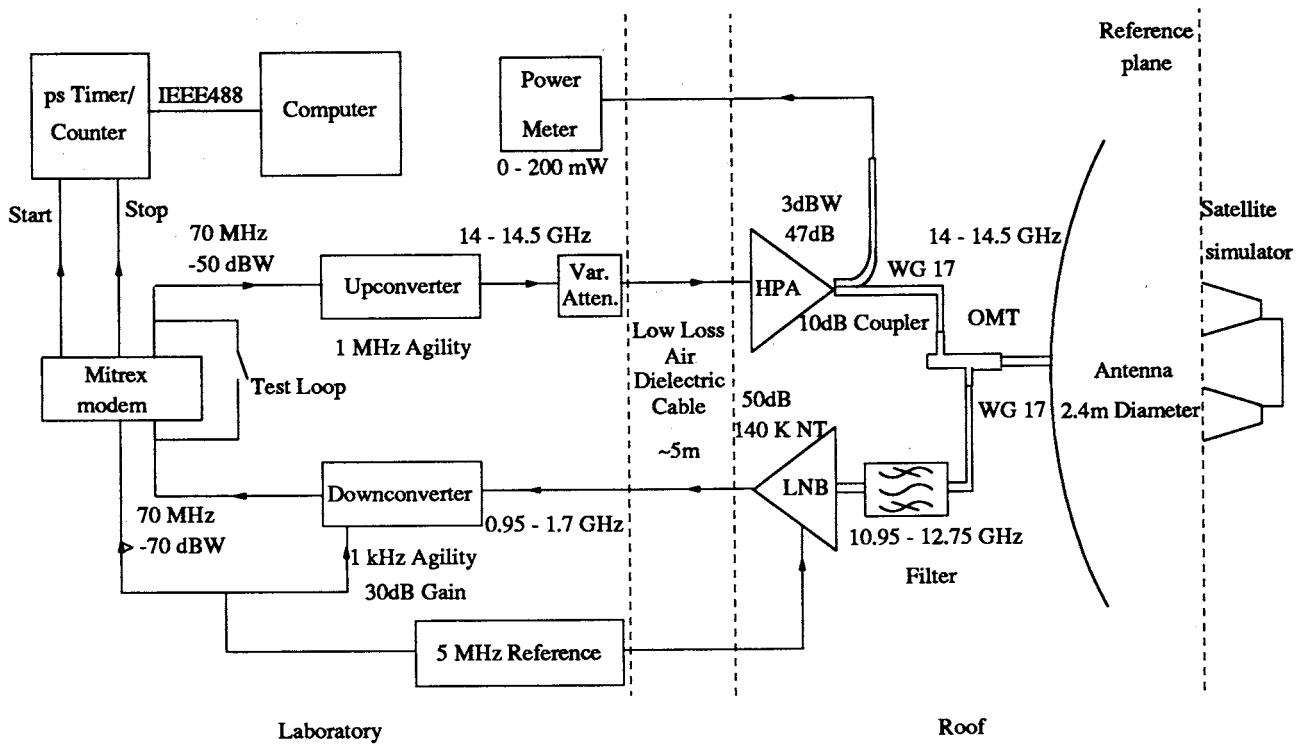


Figure 1 NPL ground station arrangement

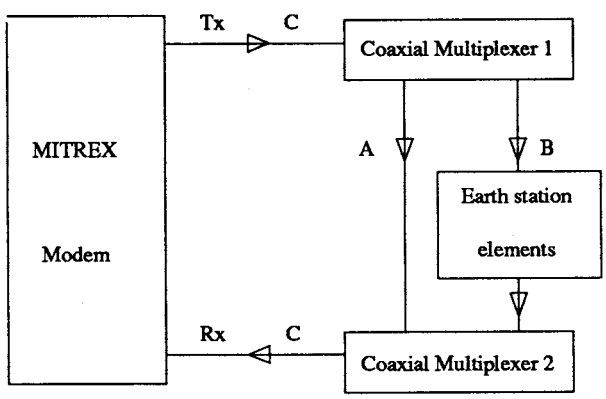


Figure 2 Coaxial multiplexer arrangement

2.4 Modem lock

Where the delays in the two signal paths A and B have a difference of less than 1μs, the modem lock was maintained when switching between loop configurations. Breaking and reforming the modem lock at each switch of the signal path did not change the delay stability. The loop delay was also insensitive when switched between different modem codes.

3. DELAY VARIATIONS OF GROUND STATION COMPONENTS

3.1 Multiplexer delay stability

The delay difference (loop B delay - loop A delay) between the two multiplexer signal paths with no additional elements in the loops, is shown in figure 3. Fluctuations in the measured delay difference set a limit for the resolution with which delay changes in other elements of the ground station may be determined. Delay changes of ±10ps (1σ) were observed during a period of three days. The introduction of the multiplexers does not add significantly to the total delay variations within the ground station instrumentation.

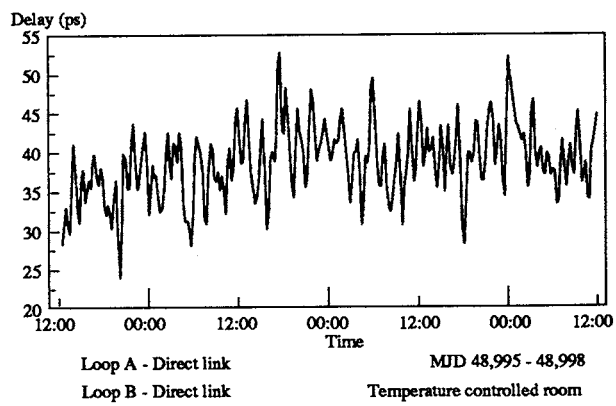


Figure 3 Coaxial multiplexer null test, loop B delay - loop A delay

3.2 Coaxial cable.

The atomic clocks maintaining the national time scale UTC(NPL) are situated in a building approximately 100m from the NPL ground station. A 1pps timing signal is transmitted to the ground station via coaxial cables. Two cables were combined to form a loop between the NPL ground station and timing centre, forming loop B of figure 2. Changes in this cable delay are shown in figure 4. The changes correlated with the outside air temperature. At times T_1, T_2,and T_3, rapid air temperature changes occurred, followed by a slower change in the cable delays. The resulting time constant was attributed to the burial of the cables below the road and buildings, which unfortunately prevented a direct temperature correction being added to the cable delay.

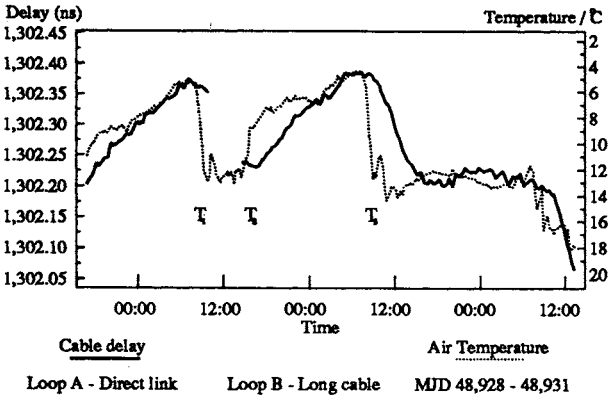


Figure 4 Cable delay.

3.3 Modem

G. de Jong has measured both the total modem loop delay and the delay in the absence of the up-link filter. The modem up-link delay was assumed to occur within the filter and connecting cables, other delay sources, for example the microwave mixer, were neglected [5]. Examples of the delay variation of the modem with filter removed are shown in figure 5. A temperature coefficient of -63ps/°C was determined using a curve fitting technique. After correcting for temperature variations, residual delay changes of ±50ps were observed. A similar temperature co-efficient was measured for the total loop delay. This suggested that the up-link filter delay and hence the modem up-link delay were relatively insensitive to temperature changes.

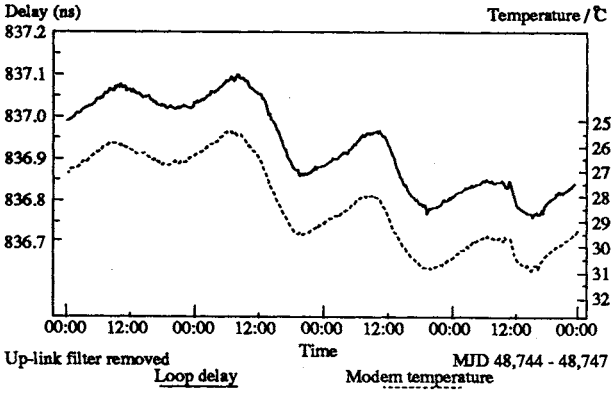


Figure 5 Modem loop delay.

3.4 Mixers

Wide-band Watkins Johnson microwave mixers type M80C and M2AC were used for frequency conversion at Ku and L band. Loop tests were performed in a temperature controlled room, using only the coaxial multiplexers and mixers. Delay changes of ±30ps were observed with both the Ku and L band mixers over a 3 day period.

The temperature of a single mixer was repeatedly cycled between 20°C and 50°C over a six hour period, to determine its temperature coefficient. Typical temperature coefficients of 12ps/°C and 20ps/°C, were obtained for the L and Ku band mixers respectively. The coefficients were dependent on both the mixer frequencies and port configuration. When the mixers were used for frequency up-conversion and down-conversion, negative and positive temperature coefficients were obtained respectively. After correcting for temperature variations, a delay stability of ±20ps was observed with both mixer types, this set a limit for measuring delay variations of other components when Ku and L band mixers were included in the test loop. The loop delay was dependent on both the signal levels and frequencies of the mixer inputs. Typical coefficients of 20ps/dB and 5ps/MHz were observed.

3.5 Amplifiers and frequency converters

The ground station operated with its active elements situated in a variety of environments. The up-converter and high powered amplifier (HPA) operated outdoors at a relatively high signal level, and C/N_0 ratio. The low noise block down-converter (LNB) also operated outdoors with a low C/N_0 ratio (55dBHz), and at an extremely low signal level. The down-converter was situated in a temperature controlled room, again at a low C/N_0 ratio. Consideration was given to these environments when performing the loop tests.

The delay stability of each active element was examined in a temperature controlled room. Examples of the delay variation of the HPA at high C/N_0 are shown. Delay change of less than 70ps was observed for all elements. The temperature of each element was repeatedly cycled to determine any temperature coefficients. The up-converter and down-converter had temperature coefficients of -25ps/°C, and -30ps/°C respectively. In contrast, the delay changes within the LNB and HPA did not correlate with temperature.

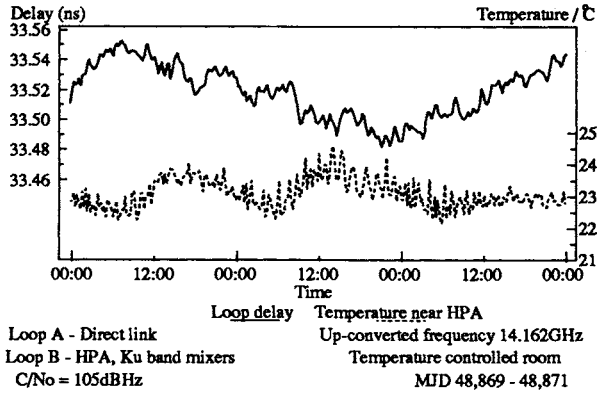


Figure 6 HPA delay.

3.6 Satellite simulator

A satellite simulator consisting of two WG17 20dB gain microwave horns connected to a microwave mixer (Figure 1) was used to receive and re-transmit the ground station signal. The satellite translation frequency was introduced in the mixer. The horns' planes of polarization were arranged to be orthogonal. Polarization isolation enables separate simulator ports to be used to receive and transmit the signal. The two horns were placed as close together as possible to minimise the errors produced by their imperfect alignment in the same reference plane.

3.7 External up-link and down-link delay

The up-link and down-link delays external to the modem occurred between the modem Tx and Rx ports and the plane of the satellite simulator horns. The total ground station delay asymmetry was determined from the sum of the modem and external delay asymmetries. Measurements of the external up-link and down-link delays were made using the test loop shown in figure 7.

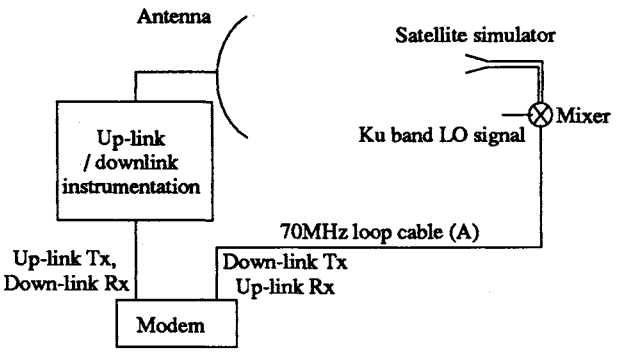


Figure 7 Up-link and down-link delay measurement.

3.8 External up-link delay

Measurements of the external up-link delay stability were made using signals with a high C/N_0 ratio. Loop delay variations of 600ps were observed, which did not correlate with temperature. Studies were made of parts of the up-link, and the delays were measured both outdoors and in a temperature controlled room. Test loops containing the antenna, waveguide feed, and satellite simulator revealed delay changes which correlated with temperature. These changes were attributed to the microwave mixers used in the test. Similar results were obtained with test loop containing only the up-converter and mixer. These results indicated that the non temperature dependent delay changes occurred within the HPA.

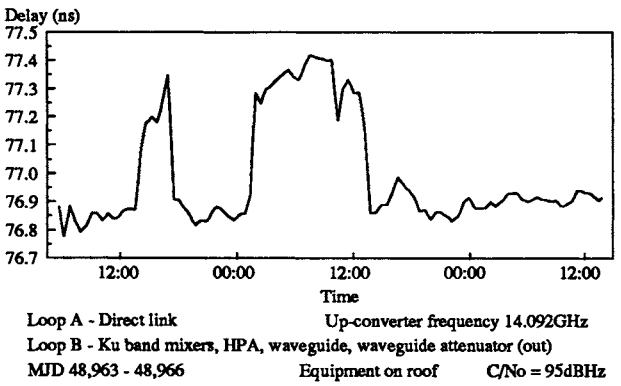


Figure 8 Up-link delay showing near discrete delay changes attributed to the HPA.

Rapid changes in the loop delay of 500ps are shown in figure 8. These 'steps' may occur instantaneously, or over a few minutes. The occurrence of these delay steps depended on the operating conditions of the HPA, possibly in a damp environment. There was a correlation between the delay steps and irregular changes in the gain of the HPA.

The external up-link delay was relatively insensitive to small changes of the direction or plane of polarization of the simulator horns; rotations of either the horn direction, or the plane of polarization by 45° was required to produce a 100ps delay change. Similar rotation of the antenna direction resulted in irregular delay changes of up to 1ns, when the antenna was rotated by half a degree. Possible causes included the greater directionality of the antenna, the presence of the satellite simulator in the antenna's near field, and multi-paths due to signal reflections off local brick walls. A 45° rotation of the antenna plane of polarization resulted in a delay change of 200ps.

3.9 External down-link delay

Delay measurements were made on the whole and part of the external down-link, at a variety of C/N_0 ratios. Examples of the variation of the loop delay at different C/N_0 ratios are shown in figures 9 and 10. This loop contained the down-converter and LNB. Reducing the C/N_0 ratio increased the short term delay fluctuations, while the maximum delay change remained at approximately 150ps. The delay changes correlated with temperature. At low C/N_0 ratios the correlation were masked by the short term delay fluctuations.

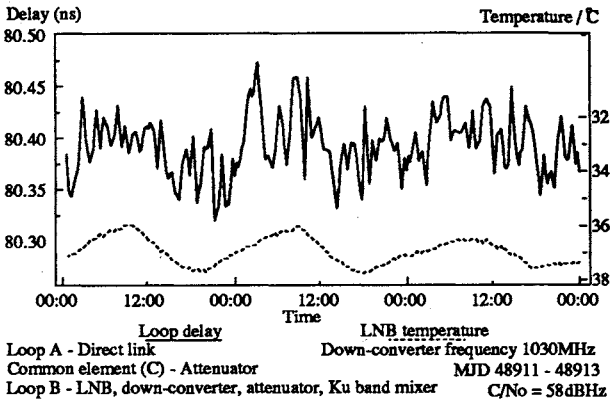


Figure 9 Loop delay measured at a medium C/N_0 ratio.

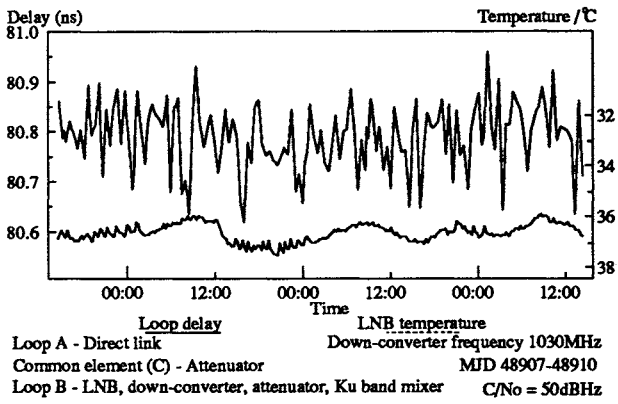


Figure 10 Loop delay measured at a low C/N_0 ratio.

An example of the variation of the external down-link delay is shown in figure 11. Delay steps were observed, which were attributed to delay changes within the LNB.

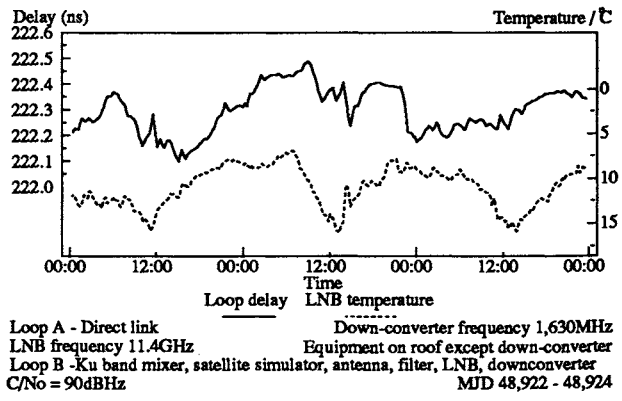


Figure 11 Total external down-link delay measured at a high C/N_0 ratio.

The external down-link delay had a similar dependence to the external up-link delays when either the simulator horns or the antenna were rotated. The delay was less sensitive to the rotation of the antenna.

3.10 External delay asymmetry
Coaxial multiplexers were used to reverse the modem Tx and Rx port configurations, to switch between measuring the external up-link and down-link delays. From these measurements, changes in the external delay asymmetry were determined. Changes in the up-link and down-link delays, delay asymmetry and mixer temperature are shown in figure 12. There was a correlation between both the up-link and down-link delays and the mixer temperature.

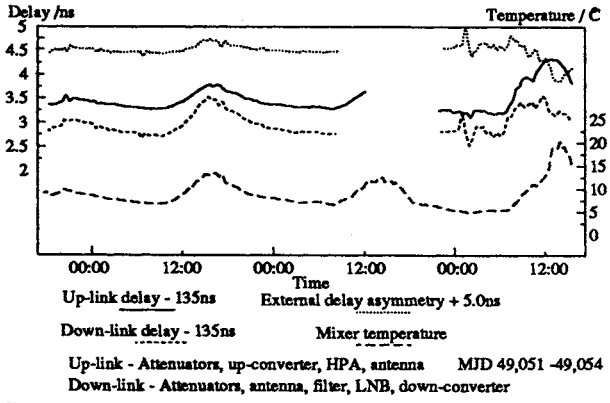


Figure 12 External delay asymmetry.

4. ABSOLUTE DELAY CALIBRATION

4.1 Introduction
Determination of the absolute value of the ground station delay asymmetry is required for satellite two way time transfer. The asymmetry was determined at the reference plane defined by the satellite simulator horns.

4.2 External up-link and down-link delays
The external up-link and down-link delays were determined from the test loops shown in figure 7. The delay of loop cable A was measured separately, as was the delay of the simulator. Separate simulator horns were positioned orthogonally to match the orthogonal transmit and receive polarizations of the antenna.

The external up-link and down-link delays were relatively insensitive to rotation of the plane of polarization of the antenna. A single simulator horn was orientated with its polarization at 45° to the antenna transmit and receive polarizations. The up-link and down-link delays were then measured using the same simulator horn. The signal propagation through the simulator and loop cable A (figure 7) were in opposite directions when measuring the external up-link and down-link delays. These delays largely cancelled in the determination of the external delay asymmetry. Dispersion effects, and errors due to the 45° rotation of the horns remain. This method was an improvement on the previous method [5], where the up-link and total loop delays were measured, resulting in the symmetry of the present system being lost. Errors due to the mis-alignment of the simulator horns were found to be approximately ±200ps. This is minimised due to the close positioning of the horns and the relative insensitivity of the delay asymmetry to horn rotation.

4.3 Mixer delays
The largest uncertainty in the absolute delay calibrations is due to error in determining the delay in the mixers. The delay of a single mixer cannot be directly measured in a loop since at least one frequency up-conversion and one down-conversion, and hence two mixers will be required. Mixer dispersion causes delay changes of up to 2ns occurring within a mixer pair. Adding a mixer pair into a test loop may change the total loop delay by between -1ns and +3ns. A maximum uncertainty of ±1.0ns in the determination of the absolute delay of a single mixer (1σ) has been estimated from the study of mixer delays and dispersion.

4.4 Modem delays
The up-link and down-link component of the modem delay were determined as described previously [5], with respect to the modem Tx and Rx ports. Errors in the determination of the cable's delay within the modem will be greater than those which occurred in the determination of the external delay asymmetry, since the propagation paths are not reversible. The measurement of the up-link mixer delay is the largest error within the determination of the modem delay asymmetry. Values of ±1.0ns and ±2.0ns for uncertainties in the modem up-link delay and delay asymmetry have been estimated.

4.5 Results
The external and modem delay asymmetries (down-link delay - up-link delay) were determined at -1.1ns ±1.4ns, and 510.1ns ±2.0ns, resulting in a total delay asymmetry of 509.0ns ±2.5ns (1σ).

5. FUTURE WORK

The delay stabilities may be improved by controlling the temperature of the instrumentation situated outdoors, and by providing further protection against moisture entering the instrumentation. Delay steps occurred within the HPA and LNB. These components are relatively inexpensive, the comparison of different models may provide instrumentation with improved delay stability.

The measurement of the total delay asymmetry may be used to improve two way time transfers. If the clocks at each location are sufficiently stables and the propagation delay stabilities are limited by the ground station instrumentation, then improved variances of time transfer σ_y may be obtained by correcting the two way time transfers for ground station delay asymmetry changes.

The determination of the absolute value of the ground station delay asymmetry is limited by uncertainties in measuring the delay of the microwave mixers and modem. Microwave mixers and a modem may be transported as a transfer standard between ground stations. By calibrating against these reference components, the uncertainty in determining the total delay asymmetry within a pair of ground stations may be reduced to under a ns. The delay stability of these reference components was limited by measurable temperature coefficients. Unexpected delay changes are not likely to occur during the course of a calibration.

6. CONCLUSIONS

Changes in the delay asymmetry of a ground station operating with a MITREX modem have been determined with a resolution of ±10ps. Variation in the total delay asymmetry of ±1.0ns over a period of 3 days were observed. Discrete delay changes of ±500ps were observed and attributed principally to the HPA. The delay of many of the ground station elements correlated with temperature. Absolute values of the delay asymmetry were measured with an uncertainty of ±2.5ns.

ACKNOWLEDGEMENTS

The authors thank P.B.Coates, J.P.Howes, K.P.Holland, D.R.Allen, B.R Swabey, R.Jefferies, D.J.E.Knight and P.Whibberley for valuable advice given during the course of this work and for the loan of experimental equipment. This work has been supported under the UK National Measurement System Programme for Time and Frequency.

REFERENCES

[1] D.A.Howe, 'Progress towards one-nanosecond two-way time transfer accuracy using Ku-band geostationary satellites,' IEEE Trans. Ultras. Ferro. And Freq. Contrl., vol. UFFC-34, No. 6 pp. 639-646, 1987.

[2] D.A.Howe, 'High accuracy time transfer via geostationary satellites: Preliminary results,' IEEE Trans. Instrum. Meas., vol. 37, pp. 418-423, 1988.

[3] D. Kirchner 'Two way time transfer via communication satellites' IEEE Proc. No. 7, pp 983-990, 1991.

[4] Y. Saburi, M. Yamamoto, and K. Harada, 'High precision time comparisons via satellite and observed discrepancy of synchronization', IEEE Trans. Instrum. Meas., vol IM-25, pp. 473-477, 1976.

[5] G. de Jong, 'Accurate delay calibration of satellite ground stations for two-way time transfer,' Proc. 3rd European Frequency and Time Forum, pp. 1898-203 and Proc. 21st PTTI meeting, 1989, pp 107-116, 1989.

FREQUENCY AND TIME TRANSFER WITH GEODETIC GPS RECEIVERS

P. Baeriswyl, T. Schildknecht, W. Gurtner, G. Beutler

Astronomical Institute, University of Berne
Sidlerstasse 5, CH - 3012 Berne / Switzerland

ABSTRACT

Time transfer using GPS timing receivers with the common view method yields accuracies in the order of 10 nanoseconds over intercontinental distances. Special investigations have demonstrated improved accuracies of a few nanoseconds for time and 10^{-14} for frequency intercomparisions after several days of integration. This is acheived only when using precise orbits, accurate station coordinates and measured ionospheric electron content. We want to proceed even further and present a technique based on principles used in geodetic applications: the full potential of the GPS system is exploited by using all GPS observables (phase, C/A- and P-code) of all available satellites in the interferometric mode. We expect an improvement of one order of magnitude for intercontinental time transfer and a unique method for syntonization measurements over long distances.

Keywords: precise orbits, IGS, geodetic GPS receiver, synchronization, syntonization

1. INTRODUCTION

Today the time transfer using the signals of the Global Positioning System (GPS) is usualy done in the so-called "common-view mode". The main characteristics are as follows:

- specialized GPS timing receivers are used
- the C/A code is used as observable
- many receivers observe only one satellite at a time
- the difference of the synchronization errors with respect to GPS time as estimated by each receiver is analyzed
- ionospheric correction are taken from the broadcast model
- the orbits broadcast by the satellites are used to analyze the signals
- usually the positions of the timing receivers are not accurately known in a well established earth fixed reference frame like e.g. the ITRF (International Terrestrial Reference Frame, see (Refs. 4,5).

It is amazing that despite these characteristics time transfer using GPS in the common view mode could become so popular. Probably this is due to the simplicity of the method. Different authors have demonstrated that the potential of the common view method is by far not yet fully explored (Ref. 1). We may especially emphasize that today, due to the establishment of the International GPS Service for Geodynamics (see section 2), the common view method can easily be improved by

- refering the antenna coordinates of the timing receivers within one centimeter to the ITRF by collocating the timing receiver's site to (one or more) nearby IGS sites (within 2000 km). For this purpose a standard dual-band geodetic receiver may be used, the observations from the nearby IGS site(s) are available for such a project. At maximum one day of observations should be sufficient for such a purpose. The baseline(s) from the timing receiver site to the IGS sites have to be processed by a standard GPS processing package using the accurate orbits from IGS (section 3).
- using the IGS orbits (usually accurate to within 50 cm in the ITRF) instead of the Broadcast Orbits
- using ionosphere content estimates stemming from IGS processing centers (Ref. 11) to account for ionospheric refraction

These improvements are obvious and easy to achieve. It would be a constructive achievement to implement them as soon as possible. Time transfer in the common view mode should then be possible with an accuracy of 1 - 3 ns even over very long distances. In our project we want to proceed even further: we want to use geodetic receivers and geodetic software, in particular we want

- to use all observables available (P-code and phase on both frequencies)
- to upgrade the well-established Bernese GPS Software (Ref. 10) for time transfer purposes in order to make full use of the potential of GPS (use of all satellites and all stations)

In order to achieve this goal geodetic receivers have to be modified (section 4).

2. THE CONTRIBUTION OF IGS

2.1 The Purpose of IGS

The primary goal of the International GPS Geodynamics Service (IGS) is to offer to the scientific coomunity high quality GPS orbits to perform regional or local GPS analysis without further orbit improvement.

The emphasis is on the word "service": IGS orbits are made available on a regular basis through three network centers (Table 1, below) with a delay of a few days. Earth rotation parameters are also estimated by the IGS processing centers. It is planned to provide ionosphere information in the future. Several IGS processing centers produce ionosphere single layer models for test purposes.

2.2 The 1992 IGS Activities

Two events took place in 1992:

(i) the IGS test campaign (21 June 1992 - 23 September 1992)
(ii) Epoch'92 (two weeks campaign centered around August 1, 1992)

The main purpose of the three months campaign was to prove that the scientific community is able to produce high-accuracy orbits on an operational basis: Data (code and phase on both carrier frequencies) were gathered by a core network of about 30 globally distributed stations (equipped with high precision dual-band P-code receivers and near-real time data links to the network centers). 30 sec observational data from the core network are running through operational and regional data centers to end up at three network centers (Table 1) from where they are retrieved by the seven IGS processing centers (Table 1), which committed themselves to daily analyze the data of the core network and to make their results available to the scientific community. More information concerning the structure and the development of IGS may be found in (Ref. 6).

Abrev.	Institution	Center Type
CDDIS	Crustal Dynamics Data Information System	Network
IGN	Institut Géographique National	Network
SIO	Scripps Inst. of Oceonography, USA	Network
UTX	University of Texas at Austion	Processing
CODE	Center for Orbit Det. in Europe	Processing*
GFZ	German Geod. Research Institute	Processing
ESOC	Eurpean Space Operations Center	Processing
JPL	Jet Propulsion Laboratory, USA	Processing
SIO	Scripps Inst. of Oceanography, USA	Processing
EMR	Energy, Mines, Resources, Canada	Processing

Table 1: *Network and Processing Centers of the 1992 IGS Campaign.*
* Collaboration of IfAG (FRG), IGN (France), Federal Office of Topgraphy, and Astronomical Institute, Bern (both Switzerland)

A first analysis of the orbits of the IGS processing centers indicates a consistency on the 0.5 m level for the orbits (Ref. 7) whereas the pole parameters estimated have a quality of about 0.5 milliarcseconds (Refs. 8, 9). As a matter of fact the 1992 IGS test campaign was so successful, that data collection and processing continued even after 23 September.

On 1 November 1992 the IGS PILOT SERVICE was established as an intermediary organization bridging the gap between the 1992 IGS test campaign and the start of the official IGS service. This means that the scientific community now has acces to high quality GPS orbits and earth rotation parameters.

3. TIME TRANSFER BETWEEN CLOCKS OF TWO GEODETIC GPS RECEIVERS

Processing techniques used for geodetic GPS applications are all based on differential approaches in the sense that the differences of simultaneous observations of one satellite from several receivers are introduced. At a first glance this seems to be identical with the common view technique. The difference however is the interferometric use of the observables, the combination of the quasi-simultaneously observed raw measurements from the receivers. It is only the latter which permits the use of the so-called carrier phase observations of the GPS signals. Phase measurements of modern receivers exhibit an astonishingly low noise of a few millimeters and thereby considerably enhance the results not only in geodetic applications but also with respect to time transfer. There is an additional reason for this differencing which resides in the fact that the influence of the satellite clock behaviour is canceled and that errors from the satellite ephemerides are greatly reduced (see discussion below).

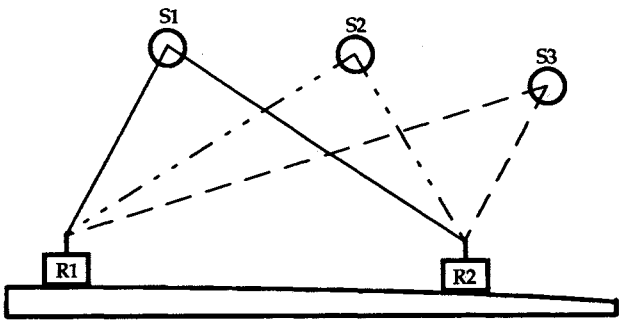


Figure 1 *Scheme to visualise the principle of the single differences.*

The observation equations of the so-called single differences of measurements from two receivers R1 and R2 may be written (somewhat simplified, for a more detailed discussion see Ref 2) as:

For code:

$$\Delta d_k + \left(c + \dot{d}_k\right) \cdot \Delta t_k + \left(\Delta\delta_{ik}\right)_{ion} + \left(\Delta\delta_k\right)_{trop} - c'_{ik} = v_{cik} \qquad (1a)$$

For phase:

$$\Delta d_k + \left(c + \dot{d}_k\right) \cdot \Delta t_k + \left(\Delta\delta_{ik}\right)_{ion} + \left(\Delta\delta_k\right)_{trop} + N_i\lambda_i - p'_{ik} = v_{pik} \qquad (1b)$$

where

l : carrier index (1,2)
k : epoch index
Δd_k : slant range difference between satellite S and the two stations R1 and R2 involved
Δt_k : synchronization error between the two receiver clocks
c : speed of light

$\dot{d}_k$: radial velocity of S with respect to R1 at time t_k

$(\Delta\delta_k)_{ion}$: differential ionospheric range correction

$(\Delta\delta_k)_{trop}$: differential tropospheric range correction

N_l: initial ambiguity parameter

λ_l: wavelength of carrier l

c'_k, p'_k: code resp. phase observation

v_{ck}, v_{pk}: residuals

There is one set of equations (1) for each observed satellite. The measurement rate is usually fixed to one observation per 30 sec but may be as high as one observation per second.

The final accuracy of the receiver synchronization Δt_k is - of course - limited by the errors of all remaining terms in the above equations. We are however in a fortunate position: terms containing Δd_k and d_k (the satellite ephemeris) can be computed from the IGS precise ephemerides (provided that the station coordinates are 'perfectly' known) and the ionospheric refraction $(\Delta\delta_k)_{ion}$ may be modeled using both carrier frequencies (by means of the known frequency dependence of this quantity). There is only one remaining critical quantity, namely the unknown phase ambiguity N_l. It is important to remember that N_l is time independent (as long as the receivers do not lose phase lock).

For demonstration purposes we will analyze equations (1) for code and phase separately (although the only correct way is to process them simultaneously in one adjustment run). The phase part of equations (1) yields a very precise determination of the the evolution of Δt_k (RMS a few mm!) with an unknown (constant) bias. We should emphasize at this point that this already permits high precision syntonization of receiver clocks on short time scale of minutes to hours. The remaining bias term is now estimated by means of all code observations. Thanks to the fact that only one unknown parameter has to be determined the rather large noise of the code measurements (~ 1 m RMS for P-code) scales down with the square root of the number of observations. Figure 2 shows synchronization errors between two receiver clocks on a 41 km baseline estimated from code and phase measurements. The code data (small crosses) showing the larger scatter was used to adjust the curve from the phase data (dots) vertically (the noise of the phase data is not visible in this scale). The strong variation with time is the real synchronization behaviour and stems from the fact that both receivers were equiped with medium quality cristal oscillators.

As with any differntial technique the errors are always, although not strictly, distance dependent. The worst case influence of the orbit part may be estimated with the well proven rule of thumb (Ref. 3):

$$\frac{db}{b} = \frac{dr}{r} \qquad (2)$$

where
b: baseline length
db: error in (one of) the components of the baseline b
r: average distance between satellites and receivers
dr: orbit error

The resulting error in the baseline components is directly related with the error introduced into the synchronization:

$$dt = \frac{db}{c} \qquad (3)$$

where
c: speed of light

Assuming 50 cm orbits, equation (2) and (3) indicate a worst case synchronization error of 0.4 ns on global baselines of 6000 km.

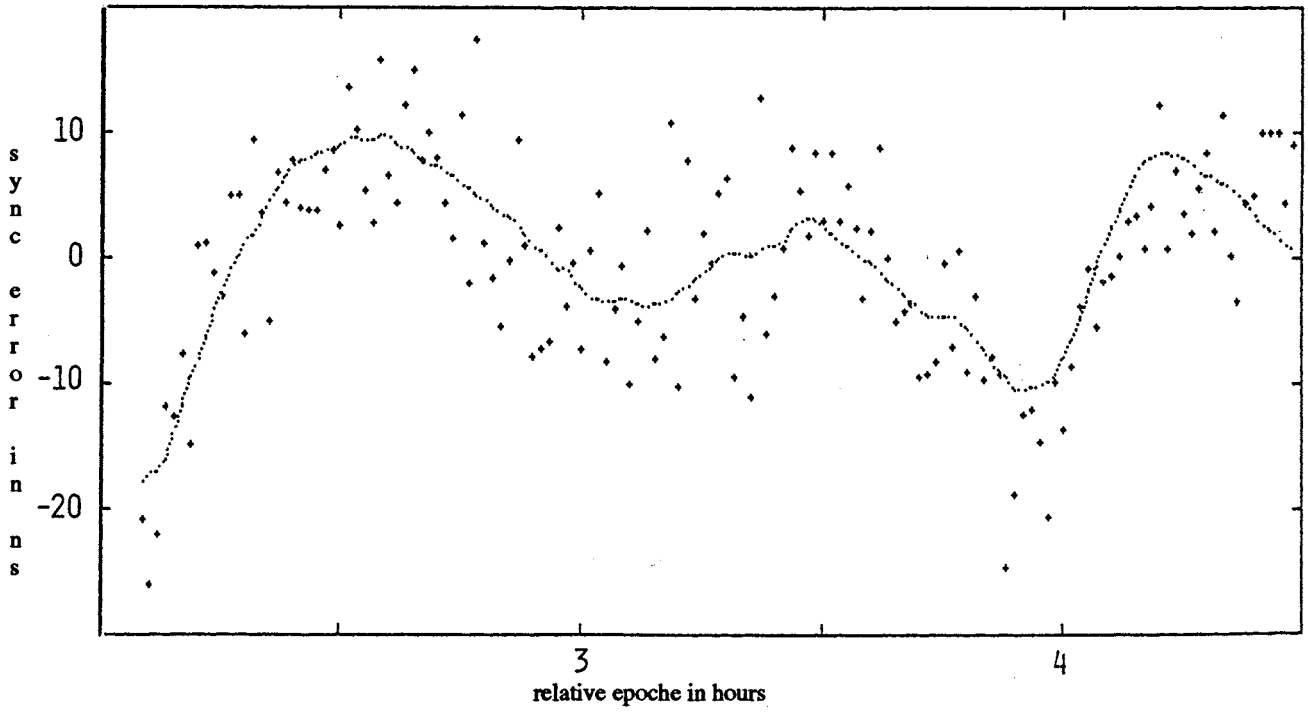


Figure 2: *Receiver clock synchronization error for the baseline La Cumbre -Chaffee (41km): Estimation with code only is drawn with the symbol +, estimation using code and phase is drawn with the dotted line.*

Residual influences from atmospheric refraction (ionosphere, troposphere) can often not be investigated separately. Nevertheless results from global geodetic networks allow to assess the overall accuracy being in the order of a few centimeters and thus well below the nanosecond.

Effects from the often discussed selective availability (SA) compromising the results of the classical common view method are completely eliminated by the differential technique (assuming the use of precise orbits) as long as all participating receivers perform their measurements quasi-simultaneously within 1 millisecond. State of the art receivers guarantee the latter requirement by monitoring their internal clocks permanently.

We should finally not forget to mention that all considerations in this section have to be understood with respect to the internal receiver clock! There is still the problem of getting access to this clock and to compare it with an external reference of interest.

4. THE EXPERIMENTAL NANOSECOND GPS TIME TRANSFER TERMINAL

4.1. The new GPS receiver

The final goal of our time transfer procedure is not the synchronization of some receiver internal clocks but the intercomparision of external time standards. It is therefore necessary to relate the receiver clock to the external clock to be compared. There is theoretically not only one method to achieve this but the most simple (and most reliable) approach is to perfectly syntonize the receiver clock with the reference and to measure their synchronization by means of an output signal (e.g. a 1 PPS signal) from the receiver clock. The receivers must therefore meet some specific requirements:

(i) The receiver internal clock must be the same for code and phase.
(ii) It must be possible to replace the receiver internal oscillator by an external reference (syntonization).
(iii) A 1 PPS output of the receiver clock must exist (synchronization). This output should not be confused with the common 1 PPS GPS time output of certain receivers.
(iv) All internal delays should be as stable as possible.

Unfortunately these requirements do not coincide with the specifications of any of the comercially available receivers. Discussions with several manufacturers resulted in an agreement with Ashtech (Sunnyvale, CA). The company will provide their new Z-12 GPS receiver with the following modifications for this project:

ad (ii) An input is provided to replace the internal frequency base.
ad (iii) A 1 PPS output of the internal clock is implemented. An additional input allows an initial coarse synchronization of the receiver clock.
ad (iv) The temperature stability of the electronics board is improved to keep the internal delays constant.

(The requirement (i) is met by the standard version). The receivers master frequency, a 20 MHz ECL signal, will be replaced (and not phase locked!) by an equivalent signal derived from the external time base. This guarantees the above mentioned 'perfect' syntonization. The 1 PPS output signal always occurs at a leading edge of the 20 MHz signal and is thus strictly related to this fundamental frequency. For technical reasons the output signal will be delayed by 3 to 4 20 MHz cycles with respect to the actual receiver clock. This delay will be known (on the sub-nanosecond level) and constant.
To minimize the power consumption GPS receivers often turn off unused channels (if there are more channels than satellites in view). This may lead to temperature variations of the electronics parts (depending on the number of tracked satellites) and thus to changes in the receiver delays. By changing the receivers power management and operating all channels continuously the internal temperature (and delays) may be stabilized.

4.2. The synchronization of the GPS receiver with the reference clock

In order to tie the receiver clock to the reference clock, let us call it the station clock, we need to measure their synchronization. Figure 3 shows the relation of the different 1 PPS signals from the station clock, the receiver clock and the receiver output. By measuring the delay Δt_{TIC} with a time interval counter we can determine the synchronization Δt_{sync}:

$$(\Delta t_{sync}) = (\Delta t_{TIC}) - (\Delta t_{out})$$

Δt_{out} is a constant delay in the order of 150 to 200 ns which has to be calibrated. If the 1 PPS from the station clock is not far from UTC, Δt_{sync} will have a random value between -1 ms and +1 ms. This is explained by the receivers internal operation: the 1 PPS output is derived from a fundamental 1 kHz signal of the internal clock. The phase of the latter with respect to UTC is defined by the moment when the receiver is powered on (and hence random). The 1 PPS however is synchronized with the 1 kHz cycle lying next to a full GPS second (GPS time is synchronized to UTC, apart from missing leap seconds, to within 1 μs).

As an option the additional synchronization input of the receiver may be used to synchronize the 1 KHz and the 1 PPS with the station clock. The resulting Δt_{sync} would be in the order of 1 μs which in turn simplifies the time interval measurement (shorter time intervals can be measured with higher accuracies).

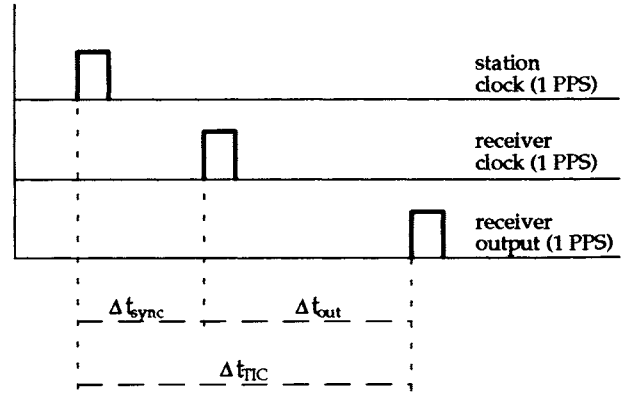


Figure 3: *Synchronization of the receiver and station clocks*

Given the 'perfect' syntonization of the receiver and the station clock, Δt_{sync} must theoretically be determined only once after each 'power on'. Nevertheless we should be prepared to detect slight variations in any of the mentioned delays de-

pending on temperature or other environmental parameters. Some of these delays like Δt_{TIC} can be monitored continuously but others like Δt_{out} can not. This is not problematic as long as we require the latter to be calibratable in the sense that: a) they depend from measurable parameters only (e.g. from temperature), b) their functional dependence from this parameter(s) can be calibrated off-line and is constant over a certain period. As a matter of fact much of the experimental effort is devoted to the calibration and monitoring of the delays, discriminator time walks, etc. .

We are finally able to close the link between the reference clocks of two stations. Their synchronization is simply calculated as the sum of the individual synchronizations between the reference clocks and the GPS receivers as measured at each station (Δt_{sync}) and the synchronization of the two receiver clocks (Δt_j) as determined by the GPS data processing.

4.3. The experimental setup

A functional layout of the experimental time transfer terminal is presented in Figure 4. The diagram is somewhat simplified in the sense that additional outputs and calibration paths have to be foreseen in order to make reference outputs available for the user of the station clock. The following main characteristics can be identified:

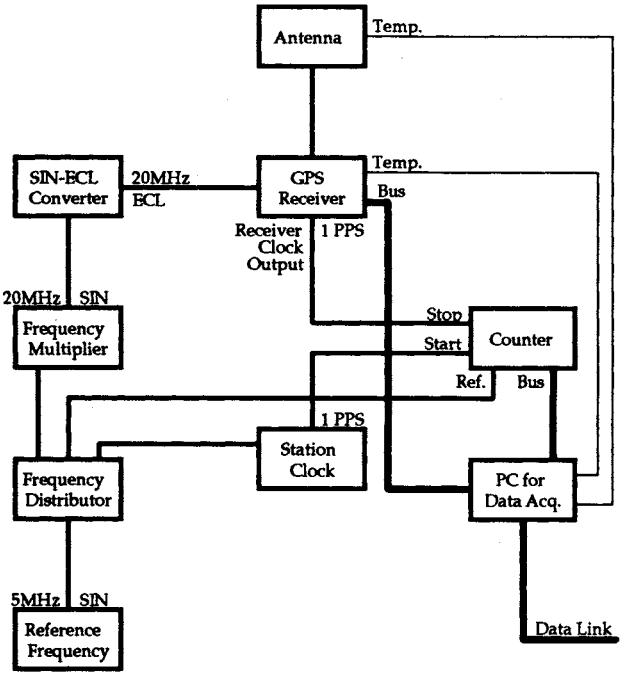


Figure 4: *Functional layout of the experimental GPS time transfer terminal. To use the receivers coarse synchronization option an additional 1 PPS from the station clock has to be provided.*

- Station clock, GPS receiver, and time interval counter are all driven by a common frequency base. Carefull design of the frequency distribution system is needed to maintain a low phase noise.

- The temperature of the receiver electronics and the antenna is monitored and if necessary stabilized.

- Delays from cables and auxiliary electronics are calibrated off-line. Potential dependencies from environmental parameters have to be investigated. As possible countermeasures we may either monitor these parameters and correct the observations using a calibrated functional model or stabilize the environment with respect to the critical parameters.

- System control, data acquisition, calibration and preprocessing is performed by a local personal computer. A data link allows control and data download operations from a remote computer. The time transfer terminal may also autonomously download data to the processing host.

- The GPS and additional synchronization data from two or more stations are evaluated on a central processing facility with a modified version of the Bernese GPS Software.

5. CONCLUSIONS

The presented technique has the capability to perform clock intercomparisions over intercontinental distances with accuracies of at least 1 ns. As a special characteristic of the method we emphasize the possibility to continuously monitor the syntonization of the clocks with a very high resolution and a superior accuracy on short time scales of hours. The fact that the results are not corrupted by the Selective Availability (SA) is an important aspect for our applications.

The improvements with respect to the common-view time transfer are achieved by

- using adapted geodetic dual band P-code receivers
- using precise orbit and ionosphere information as determined by the IGS
- refering the station coordinates within a centimeter to the ITRF
- processing all observables (phase and code on both frequencies) of all available satellites in an interferometric mode using a modified geodetic GPS processing package

Acknowledgments

We would like to thank to the Federal Office of Metrology (Switzerland) for the financial support of this project.

References

[1] W. Lewandowski, C. Thomas: GPS Time Transfer (Proc. of the IEEE, Vol 79, No7, 1991).

[2] Th. Schildknecht, G. Beutler, W. Gurtner, M. Rothacher: Towards Subnanosecond GPS Time Transfer using Geodetic Processing Techniques (Proc. of the 4th European Frequency and Time Forum Neuchâtel, 1990).

[3] I. Bauersima: Radiointerferometrische Satellitenbeobachtungen, (GPS II) (Mitteilungen der Satellitenbeobachtungsstaion Zimmerwald, Nr 10, Astronomisches Institut der Universität Bern, Schweiz, 1983).

[4] C. Boucher, Z. Altamimi: ITRF 89 and other realizations of the IERS Terrestrial Reference System for 1989 (IERS Technical Note No. 6, April 1991, Observatoire de Paris).

[5] C. Boucher, Z. Altamimi: The IERS GPS Terrestrial Reference Frame (IERS Technical Note No. 6, April 1991, Observatoire de Paris).
[6] I.I. Mueller, G. Beutler: The International GPS Service for Geodynamics - Development and Current Status (Proc. of the 6th International Geodetic Symposium on satellite Positioning, Columbus, Ohio, March 1992, pp. 823-835).
[7] G. Beutler: The 1992 Activities of the International GPS Geodynamics Service (IGS) (7th International Symposium, IAG Symposium No. 112, 1992, in press).
[8] D. McCarthy: Accuracy of high-frequency observations of earth orientation (7th International Symposium, IAG Symposium No. 112, 1992, in press).
[9] M. Feissel: IGS'Campaign, Comparison of GPS, SLR, and VLBI Earth Orientation Determinations (IERS Central Bureau, Final Report, November 1992).
[10] M. Rothacher, G.Beutler, W. Gurtner, Th. Schildknecht, U. Wild: Bernese GPS Software, Version 3.4, Documentation 1993.
[11] U. Wild: Ionosphere and Ambiguity Resolution (paper presented at the 1993 Bern IGS Workshop, proceedings in preparation).

TWO-WAY TIME COMPARISONS VIA DFS-KOPERNIKUS 2 BETWEEN FTZ AND PTB

Peter Hetzel* and Armin Söring**

*Physikalisch-Technische Bundesanstalt, Braunschweig
** Forschungs- und Technologiezentrum, DBP Telekom, Darmstadt

ABSTRACT

In September 1992 a two-way time transfer between the FTZ and the PTB via the German domestic satellite DFS-Kopernikus 2 has been started. The earth stations installed at the FTZ and at the PTB are VSAT's (very small aperture terminals) of identical design, each with a 1,8 m dish, operating at Ku-band and equipped with MITREX modems. A short description of the main characteristics of the earth stations is given. The results of the time scale comparisons UTC(FTZ)-UTC(PTB) by means of common-view GPS differ from those by two-way time transfer by 2 ns to 20 ns. The standard deviation of the one-second two-way measurements of UTC(FTZ)-UTC(PTB) was typically 300 ps.

Keywords: two-way satellite time transfer
time scale comparisons

1. INTRODUCTION

During 1991, at the Forschungs- und Technologiezentrum, DBP Telekom (FTZ) in Darmstadt and at the Physikalisch-Technische Bundesanstalt (PTB) in Braunschweig two VSAT earth stations of identical design, each with a 1,8 m dish, were installed for the purpose of performing two-way satellite time transfer at the Ku-band. The first experiments via the German domestic satellite DFS-Kopernikus 2 (28,5° East) were carried out in January 1992 using two digital MITREX modems 2500 A. Because of the insufficient phase stability of this type of modem the results were of limited value only. Therefore the modems have been put back from the digital mode of operation into the original and well-proved analogous mode (Refs. 1,2) by the Institute of Navigation in Stuttgart.

After their modification the modems worked satisfactorily and the two-way experiments could be restarted at the end of September 1992. Unfortunately, troubles with the up-converters first occured in Darmstadt, later in Braunschweig interrupted the measurements again from November 1992 to February 1993. In spite of all these technical difficulties some promising results could be obtained which are reported below.

2. MAIN CHARACTERISTICS OF THE EARTH STATIONS

The earth stations are mounted on the flat roofs of a 8-floor building in Darmstadt (N 49°52', E 8°37') and of a 6-floor building in Braunschweig (N 52°18', E 10°28'), with free sight to all directions. The up- and down-converters in the outdoor units are connected with the modems in the laboratories by cables with a 70 MHz-IF-attenuation of 5 dB at the FTZ and of 11 dB at the PTB, respectively.

Each VSAT system has a 1,8 m offset paraboloid reflector with a gain of 46 dB. The cross polarisation isolation is 35 dB.

The up-converter frequency range is 14 GHz to 14,5 GHz, while the down-converter frequency range is splitted into the two ranges 10,95 GHz to 11,7 GHz and 11,7 GHz to 12,75 GHz by two exchangeable synthesizers. The transmit and receive frequencies can be selected independently with a frequency step size of 1 kHz.

The minimum amplification of the low noise amplifier is 50 dB, its noise temperature at 23°C is less than 140 K. The G/T ratio is (22 + 20 log (f/12,5 GHz)) dB/K.

The output power of the high power amplifier is 38,5 dBm (small signal amplification 50 dB) and the specified EIRP (equivalent isotropic radiated power) of the stations is 53 dBW.

3. PRINCIPLE OF TWO-WAY TIME TRANSFER

A detailed description of the basics of the two-way technique is given in many papers, for example in (Refs. 3,4). The point to point time transfer between FTZ and PTB consists of the following procedure: Both partners measure simultaneously the time

differences between the second pulses referred to their own UTC time scale and the corresponding second pulses referred to the other time scale, which are exchanged as PN-coded time information via the Kopernikus 2 satellite. The time difference between the two time scales is then given by the equation:

$$\begin{aligned} UTC(FTZ)-UTC(PTB) &= 0,5\ (\Delta_F-\Delta_P) \\ &+ 0,5\ (\tau_{FP}-\tau_{PF}) \\ &+ 0,5\ [(\tau_F^{TX} - \tau_F^{RX})-(\tau_P^{TX}-\tau_P^{RX}) \\ &+ \Delta\tau_R \end{aligned} \qquad (1)$$

Δ_F and Δ_P are the counter readings at the FTZ and at the PTB, which are exchanged for the computation of the time scale difference.

The second term at the right-hand side of (1) contains the differences of the propagation delays for both signal directions and of the transponder delays. Since the same transponder is used in both directions and path reciprocity can be assumed, this term cancels out.

The third term at the right-hand side of (1) comprises the differential equipment delays of the modems, of the IF-cables and of the transmit and receive parts of the earth stations. Before the delivery of both stations - when both stations were still collocated at the manufacturer - the differential delays of the transmit and receive parts have been determined (Ref. 5). Because of the same design of both stations the delays are approximately equal. In the receive parts the delays differ less than 0,1 ns and the delay of the PTB transmit part exceeds that of the FTZ transmit part by 2,6 ns.

The differential modem delays have also been measured by collocation of both modems at the FTZ. In the FTZ-modem the delay of the receive part is 19 ns longer and the delay of the transmit part is 206 ns shorter than the corresponding delays in the PTB-modem.

The delays of the IF-cables which connect modems and earth stations have been determined by loop measurements with the MITREX modems and an auxiliary cable (Ref. 6). At the FTZ delays of 453 ns for the IF transmit cable and 455 ns for the IF receive cable have been found. The corresponding values at the PTB are 1120 ns and 1125 ns, respectively.

The third term in equation (1) may be written as

$$0,5\ [(\tau_F^{TX}-\tau_F^{RX}) - (\tau_P^{TX}-\tau_P^{RX})] =$$
$$0,5\ [(\tau_F^{TX}-\tau_P^{TX}) - (\tau_F^{RX}-\tau_P^{RX})]$$

Adding up all differential equipment delays results in a correction value of -112 ns for the receive and transmit parts of both stations.

The last correction which has to be taken into account is the fourth term (Sagnac effect) in equation (2). It is caused by the earth rotation and amounts to 1 ns for the comparison between FTZ and PTB.

Finally the time scale difference UTC(FTZ)-UTC(PTB) is obtained from half of the difference of the counter readings by applying a total correction of -111 ns. All contributions to this correction have been determined with an uncertainty of 1 ns.

4. RESULTS

If possible the two way measurements have been performed twice a week on Monday and Thursday. The transmitting frequency was 14,145 GHz, the receiving frequency 12,645 GHz and the EIRP of both stations 50 dBW.

Figure 1 gives the one-second differences UTC(FTZ)-UTC(PTB). The standard deviation of a sample of 300 one-second measurement values is typically 300 ps and conforms with the jitter expected from the measured carrier-to-noise density ratio (C/N_0) of about 65 to 70 dB/Hz (Ref. 1).

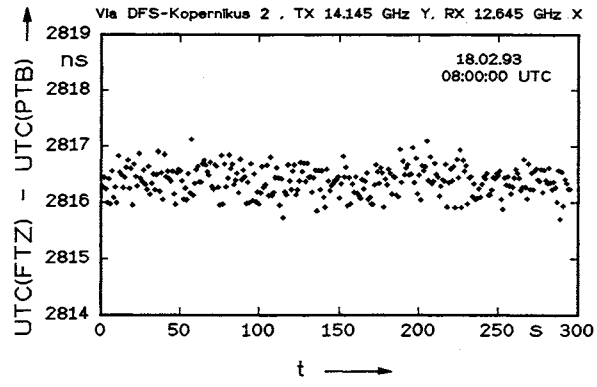


Fig. 1 Plot of the second-to-second time scale differences UTC(FTZ)-UTC(PTB) of Febr. 18, 1993, starting at 08:00:00 UTC.

Figure 2 shows the time scale differences UTC(FTZ)-UTC(PTB) obtained by common-view GPS and by two-way time transfer lasting from September 28, 1992 (MJD 48893) to March 1, 1993 (MJD 49047). As already mentioned, the gap in the two-way data was caused by technical problems with the earth stations. The GPS data between MJD 48964 and 48998 had to be omitted because the FTZ could not follow the updated BIPM tracking schedule of MJD 48964 for common view comparisons by GPS. Furtheron it has to be mentioned that the GPS link FTZ-PTB did not yet have the full possible accuracy: The

GPS receiver of the FTZ was calibrated by only one portable clock comparison with an estimated uncertainty of 20 ns and the GPS antenna coordinates at the FTZ were not yet determined with the highest accuracy available. Therefore it is not clear whether the differences of up to 20 ns between the data UTC(FTZ)-UTC(PTB) via GPS and those via DFS-Kopernikus 2 (see Fig. 3) have to be attributed to the GPS link or the two-way time transfer.

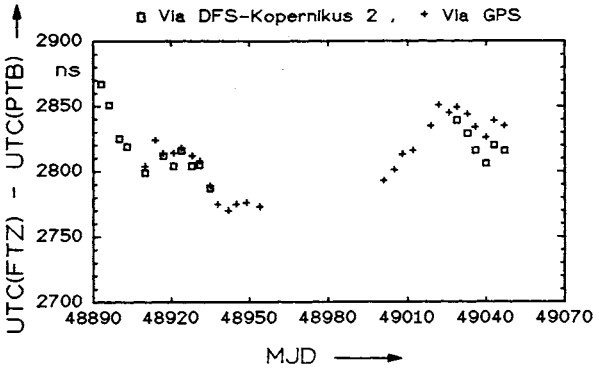


Fig. 2 Time scale difference UTC(FTZ)-UTC(PTB) measured by GPS and by two-way time transfer.
MJD Modified Julian Date.

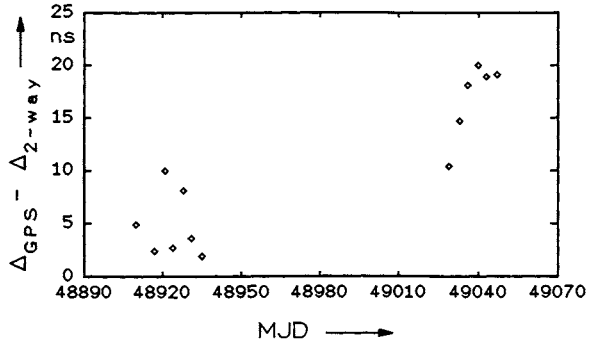


Fig. 3 Difference between Δ_{GPS} = $[UTC(FTZ)-UTC(PTB)]_{GPS}$ and Δ_{2-way} = $[UTC(FTZ)-UTC(PTB)]_{2-way}$ measured by GPS and by two-way time transfer.
MJD Modified Julian Date.

5. CONCLUSIONS

The results confirm the capability of the spread-spectrum two-way technique for high accuracy time transfer. The rapid development of the satellite technique allows the establishment of a system of time links between major time keeping laboratories using small VSAT earth stations. Suitable civilian telecommunication satellites are present for regional and intercontinental links. Efforts in studying and calibrating the equipment delays still have to be done.

ACKNOWLEDGMENT

The authors wish to thank Mr. H. Audrit and Mr. M. Mößle (German Telekom, FTZ Darmstadt) for their support in operating the stations.

REFERENCES

[1] Ph. Hartl et al.: High accuracy global time transfer via geosynchronous telecommunication satellites with MITREX. Z. Flugwiss. Weltraumforsch. 7 (1983), pp. 335 - 342.

[2] H. Ressler and D. Kirchner: Experience with the MITREX-modem. Report INW 8501 (1985), Institut für Nachrichtentechnik und Wellenausbreitung, Technische Universität Graz.

[3] D. Kirchner: Two-way time transfer via communication satellites. Proc. of the IEEE, Vol. 79 (1991), No. 7, S. 983 - 990.

[4] G. de Jong: Prospects of two-way international time scale comparisons. Proc. of the 6th European Time and Frequency Forum (1992), Nordwijk, pp. 69 - 73.

[5] K.-H. Thiel, Institut für Navigation Stuttgart: Private communication of Oct. 9, 1991.

[6] G. de Jong: Accurate delay calibration of satellite ground stations for two-way time transfer. Proc. 3rd European Time and Frequency Forum (1989), Besancon, pp. 198 - 203.

IMPROVED TIME TRANSFER USING A GEOSTATIONARY DIRECT TV SATELLITE

F. Meyer, F. Vernotte, M. Vincent

Observatoire de Besançon,
41 bis Avenue de l'Observatoire
25010 Besançon Cedex

M. Granveaud

LPTF, Observatoire de Paris
61, Avenue de l'Observatoire,
75014 Paris

B. Laporte

CNES Toulouse
18, Avenue E. Belin
31055 Toulouse Cedex

ABSTRACT

In a previous paper (Ref. 1), we have shown that the TV signals transmitted by direct TV satellites may easily be used to perform time transfers at the sub-microsecond level, the main source of uncertainty coming from the satellite position. In this paper we first present a method for improving time transfers ; it consists in estimating the longitude variations of the satellite by one link and in correcting consequently the data of the other links. According to the link used for the calibration, the uncertainty of these corrected time transfers varies between 9 and 50 nanoseconds.

A method has been proposed (Ref. 1) to determine the satellite position by use of the data collected by three stations ; the installation of the (OC) station has allowed the verification of the estimated accuracy of this method. Time transfer between Paris and Grasse has been shown to be achievable with an uncertainty of 15 nanoseconds, and less than 10 nanoseconds between Paris and Besançon.

Keywords : time measurement, time transfer, direct broadcasting satellite.

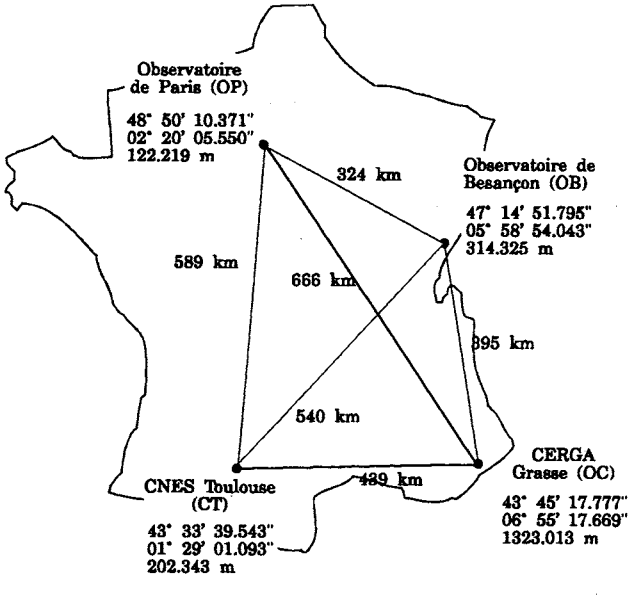


Figure 1 : The 4 stations in use.

1. INTRODUCTION

The principle of time transfers through direct TV satellites has been described in (Refs. 1, 2). The time differences of the two clocks are obtained by timing the arrival of a given pulse at two different locations and by taking into account the co-ordinates of the receiving antennas and of the satellite.

In (Ref. 1), the measurements were provided by three receiving stations located in Besançon (OB), Paris (OP) and Toulouse (CT), observing the TDF2 satellite, located at 18.9° W and transmitting a D2-MAC standard TV signal ; we now use the data from a fourth station recently set up at the Observatoire de la Côte d'Azur in Grasse referred to as OC (figure 1). The Cs clocks of these four stations are moreover linked by GPS allowing the calibration of the results.

In all the results presented in this paper, the clock differences between the different stations have been corrected by mean of these GPS links. This means that the plots presented are in fact a comparison between the time transfers given by the different methods and those provided by GPS.

The lack of precise knowledge of the satellite position is the main source of uncertainty on the time transfers. In order to reduce this uncertainty, it is first possible to remove (e. g. by averaging) the 12 and 24 hour periodic components of the satellite latitude, longitude and radius variations (Refs. 1, 3, 5).

Under the assumption that the main source of error is then the long term drift of the satellite longitude, the known geometry of a given link may be used to deduce the satellite longitude variations from the time measurement variations ; finally the time measurements of the other links may be corrected for these calculated longitude variations.

Another way of improving the time transfers is to explicitly determine the satellite position by using the data from three stations as shown in (Refs. 1, 3) ; the measurements made between two stations of which clock differences are known by GPS give an information on the satellite position. Adding the data from a third station decreases the uncertainty on the satellite longitude and latitude down to a few millidegrees ; time transfer between one of the three stations and a fourth one may then be achieved. These measurements will also be compared with those provided by GPS links.

2. CALIBRATION AND CORRECTION OF THE DIFFERENT LINKS

2.1. The satellite motion

The satellite motion on the geostationary orbit is a combination of the following components, in decreasing order of amplitude (Ref. 5) :

1. the non zero inclination *h* of the orbit causes latitude periodic variations of period one day ; in the same way, the small eccentricity *e* of the orbit causes radius and longitude periodic variations of period one day ; the values of *h* and *e* are controlled to ensure that the amplitude of these variations remains less than the size of the satellite parking box i.e. 0.1° in longitude and latitude ; the amplitude of the radius variations is then less than 10 km ;

2. a non-periodic longitudinal drift is caused by the irregularities of the gravitational potential which induces a longitudinal force on the satellite, making it evolve as on an inclined plane ; the magnitude of this force depends on the location of the satellite along the geostationary orbit ;

3. long period eccentricity and inclination variations (periods varying between 14 and 365 days) have small amplitude and may then be neglected ;

4. the influence of the sun and the moon causes periodic variations of half-day period of the satellite longitude latitude and radius ; these variations are of small amplitude (a few millidegrees in longitude and latitude, a few km in radius) ;

So, the hypothesis we make is that if we remove the short period components (half-daily and daily variations) of the satellite motion, the only significant perturbing element will be the satellite longitudinal drift.

2.2. Calibration of the different links

As it can be seen on figure 2, an accurate estimation of TOA(A) - TOA(B) i. e. of the delay corresponding to the difference of the distances satellite-station A, satellite-station B is the key problem for the determination of UT(A) - UT(B) (TOA(A) stands for the time of arrival of the signal pulse at station A) .

The residual motion of the satellite in its "parking box" is the main source of uncertainty on the time transfer because of the fluctuations of TOA(A) - TOA(B) it generates ; for a given couple of stations (A, B) this uncertainty depends on two main parameters (Ref. 1) :

1. the difference ΔV_{A-B} between the two unit vectors station-satellite (see table I, row 1).

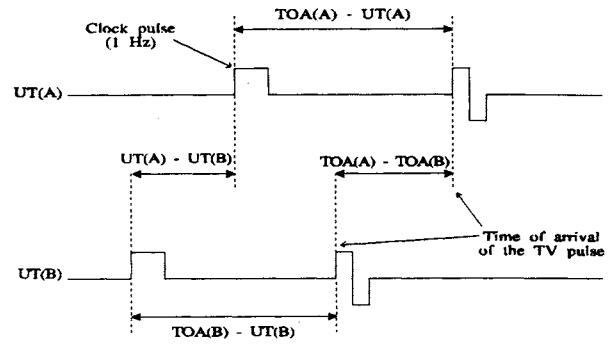


Figure 2 : Principle of the time transfer

2. the vector δP defined by the real position of the satellite with respect to its mean position. As seen above, the satellite is maintained in a cell of ± 0.1° (i. e. ± 73.6 km at the geostationary radius) in longitude and latitude, which implies variations of ± 10 km in radius ; taking into account the longitude φ_0 of the satellite (about 18.9° W for TDF1-2) , we obtain the following vector :

$$\delta P = \begin{pmatrix} 73{,}600 \sin\varphi_0 + 10{,}000 \cos\varphi_0 \\ 73{,}600 \cos\varphi_0 + 10{,}000 \sin\varphi_0 \\ 73{,}600 \end{pmatrix}$$

The uncertainty σ_{A-B} of a raw time transfer is then given by :

$$\sigma_{A-B} = \frac{(\Delta V_{A-B} \cdot \delta P)}{c}$$

where c is the speed of light. The values obtained for the 6 different links (table I, row 2) vary between 2.5 and 5.5 μs. The measured maximal variations (peak to peak) ΔT_{A-B} (row 3) are well within the theoretical limits.

Though there is no way to reduce the contribution of the first parameter, it is possible to manage to reduce the influence of the second one by averaging two measurement values separated by 12 hours. In a first approximation, this will average out the contribution of the daily and half-daily variations of the satellite longitude, latitude and radius. The main source of error is then the longitudinal drift ; in the calculation of the uncertainty in this case, the δP vector will be modified as δP_l with :

$$\delta P_l = \begin{pmatrix} 73{,}600 \sin\varphi_0 \\ 73{,}600 \cos\varphi_0 \\ 0 \end{pmatrix}$$

The values of the uncertainty will then be given by :

$$\tau_{A-B} = \frac{(\Delta V_{A-B} \cdot \delta Pl)}{c}$$

For each of the six links, the τ_{A-B} value is lower than σ_{A-B} (table I, row 4).

			OP-CT	OB-OP	OP-OC	CT-OB	CT-OC	OB-OC
1	ΔV_{A-B} ($* 10^{-3}$ m)	X Y Z	0.03 2.97 -9.14	-2.51 -7.43 3.01	3.70 12.44 -9.33	2.53 4.46 6.14	3.72 9.47 -0.02	1.18 5.00 6.32
2	σ_{A-B} (μs)		± 2.97	± 2.64	± 5.49	± 2.69	± 2.52	± 2.82
3	ΔT_{A-B} (μs)		5.51	3.11	6.30	4.27	3.68	3.76
4	τ_{A-B} (μs)		±0.69	±1.93	±3.18	±1.24	±2.49	±1.25
5	S_{A-B} (μs/arc') S_{A-B}/S_{OP-CT}		+ 0.115 1	- 0.322 - 2.8	+ 0.530 4.6	+ 0.207 1.8	+0.415 3.6	+ 0.208 1.8
6	ΔT_{A-B} (μs) 48996 - 48326		0.663	-1.897	_	1.236	_	_
7	$\Delta\lambda_{est}$		5.76'	5.90'	_	5.98'	_	_
8	ΔT_{A-B} (μs) 48996 -48736		0.43	-1.22	2.00	0.80	1.58	0.78
9	$\Delta\lambda_{est}$		3.72'	3.71'	3.77'	3.87'	3.79'	3.75'

Table I : Geometry and calibration of the different links

Furthermore, assuming that the variations of the averaged data only reflect the satellite longitudinal drift, it is possible to obtain from the geometry of a given link a direct relationship between the time measurement variations and the satellite longitude variations. From this point of view, the τ_{A-B} value can be seen as a coefficient of sensitivity (in terms of time transfer) of the link A-B to the longitude variations of the satellite position. This allows us to deduce the satellite longitude variations from the time measurements variations. The time measurements made for the other links may then be corrected for the effects of these calculated longitude variations. The sensitivity of each link relative to OP-CT is given in table 1, row 5. The negative sign for OB-OP indicates that the time variations for this link are opposite to those of the other links ; this is mainly due to the sign of the longitude difference between OB and OP.

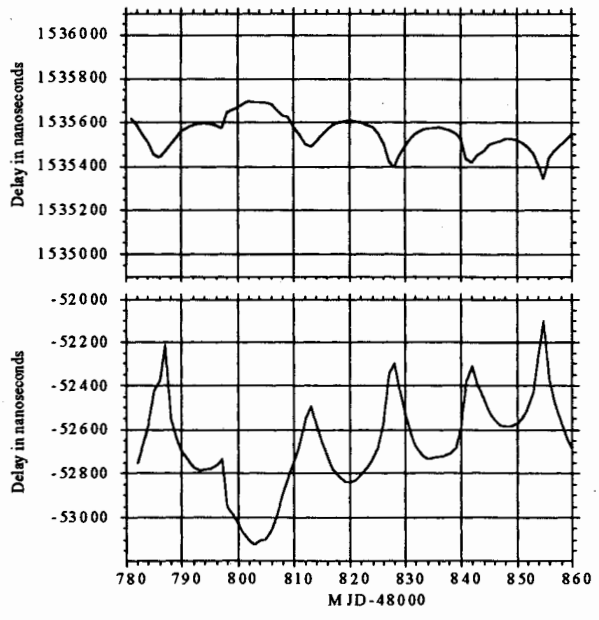


Figure 3 : Averaged timing data :
up : T_{OP-CT} (S_{OP-CT} = 1)
down : T_{OB-OP} (S_{OB-OP} = -2.8)

In table I, row 6 represents the maximum (peak to peak) variations of the averaged data and row 7, the estimation of the corresponding longitude variation, calculated from the τ_{A-B} coefficients. Rows 8 and 9 show the maximum variations since the set up of (OC). In the 2 cases, the values obtained for the longitude variations are in good accordance, showing the validity of the hypothesis.

Figure 3 shows the timing data (T_{A-B}) for the 2 links OP-CT and OB-OP. In these data and the following ones, the clock differences have been corrected by use of the GPS links. Therefore what we observe on these plots is essentially due to the geometric configuration of the satellite with respect to the stations. We have in fact :

$$T_{A-B} = (TOA(A)-UT(A)) - (TOA(B)-UT(B)) + [UT(A)-UT(B)]_{GPS}$$
$$= TOA(A) - TOA(B)$$

The 2 plots of figure 3 are at the same scale, demonstrating the correlation between these data and the satellite longitude variations. The difference of amplitude and the opposite senses of variation correspond to the calculated coefficients (table I, row 5).

2.3. Corrected time transfers

The hypothesis we made is that the τ_{A-B} coefficients actually characterise the sensitivity of the link A-B to the longitude variations of the satellite. The data collected for one link A-B then give the evolution of the longitude of the satellite ; knowing the sensitivity to these variations of another link CD, it is possible to correct the data of this link for the

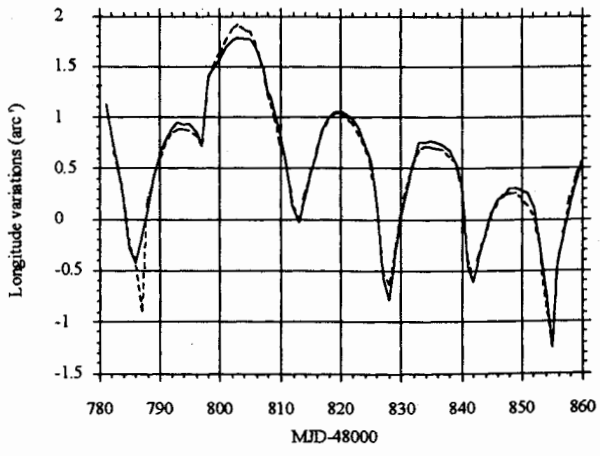


Figure 4 : Longitude variations determined by OP-CT (dash line) and OP-OC (bold line)

longitude variations. Figure 4 shows the longitude variations deduced from the data of OP-CT and OP-OC. The general shape of the curves is made of parts of parabola (Refs. 4, 5) corresponding to the satellite "fall" towards the place of minimum gravitational potential on the geostationary orbit. These arcs are separated by sharp changes due to E-W station-keeping manoeuvres, during which the satellite is pushed back to higher potential.

Since the sensitivity of OP-OC is far greater than the sensitivity of OP-CT, the discrepancies between the two plots can be attributed to the inherent uncertainty of OP-CT data.

Figure 5 shows the data of the link OB-OC corrected after calibration by the link OP-OC (lower curve ; σ = 9 ns) and of the link CT-OB corrected by OP-CT (upper curve ; σ = 35 ns). This last link is much more noisy since its correction was made through a longitude calibration by OP-CT which is the less sensitive link to longitude variations. The calibration made via OP-OC is much better (high sensitivity to longitude variations) giving a good correction of the data.

Thus, the link between OP and OC is the most accurate calibration link, and the least accurate corrected link (see table II). The uncertainty on corrected time transfer between OP and CT varies between 11 and 20 ns according to the link used for calibration.

Some of the peaks appearing on the plot of CT-OB are due (as on figure 4) to the discontinuities occurring during the station-keeping manoeuvres : the hypothesis that the satellite

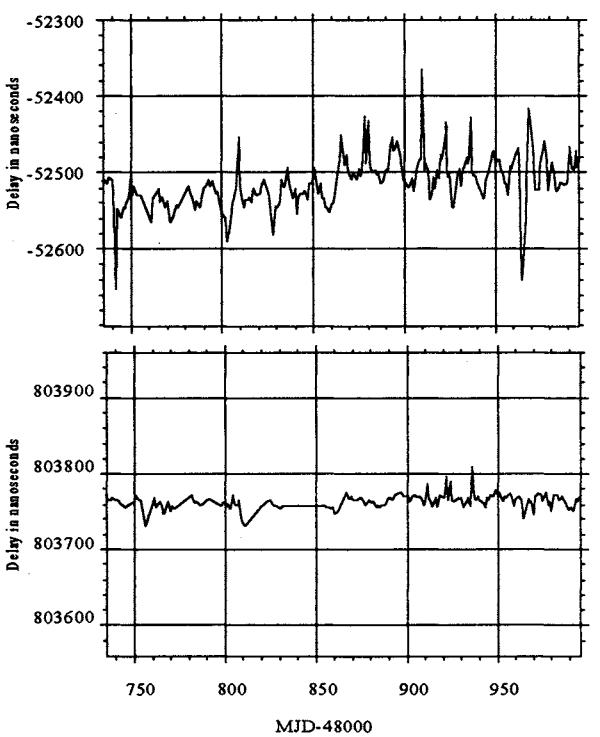


Figure 5 : Corrected time transfer :
up : T_{CT-OB} corrected by OP-CT calibration (σ = 35 ns)
down : T_{OB-OC} corrected by OP-OC calibration (σ = 9 ns)

radius and latitude daily variations can be averaged out is no longer valid (particularly during N-S manoeuvres).

Figure 7 shows the Log-Log plot of the Allan variance for these 2 links (upper and lower curves). They show that a stability better than 10^{-14} can be reached for an averaging time of about 10 days, even for long links if the calibrating couple of stations is favourable. Moreover, for the worst configurations of the experiment a stability of about 10^{-13} can still be obtained.

Calibration link / Corrected link	OP-CT	OB-OP	OP-OC	CT-OB	CT-OC	OB-OC
OP-CT	_	13	12	20	15	11
OB-OP	35	_	11	20	12	24
OP-OC	53	18	_	43	15	23
CT-OB	35	13	17	_	14	27
CT-OC	52	15	11	27	_	27
OB-OC	19	16	9	27	13	_

Table II. Uncertainty for the corrected links (ns)

3. TIME TRANSFER AFTER DETERMINATION OF THE SATELLITE POSITION

3.1. Principle

The principle of this method is described in (Ref. 1) : it consists in solving the geometric problem of finding the set of solutions of the system defined by the differential data obtained from three stations. This gives the equation of a curve (intersection of two hyperboloïds) on which the satellite must be found.

When using the 3 stations OP-OB-CT as ground base triangle, the error generated by the theoretical motion of the satellite along this curve is always less than 10 ns all over France. Nevertheless, if we take into account the different sources of error on the time measurements (such as atmospheric and particularly tropospheric delays, dispersion in the receivers, errors on GPS links...), this value is greatly increased.

It is actually not easy to predict the final uncertainty on such time transfers : for extreme values of the satellite radius and latitude and of the error on the time measurements, the final error could reach values up to 80 ns for the longest links. In the worst cases, uncertainty values of about 50 ns on the link between OC and OB via OP-OB-CT were expected.

3.2. Results

The situation is in fact not so unfavourable as we can see on figure 6. The dispersion is about 15 ns for the link between station OC and the triangle of stations OP-CT-OB (figure 6 upper graph, upper curve) and about 11 ns for the link between station OB and the triangle of stations OP-CT-OC (figure 6 upper graph, lower curve). When averaged over 24 hours, these values fall down to respectively 10 and 7 ns (see figure 6 lower graph).

The difference between the two cases can be explained, as in the previous method, by the fact that the second triangle is a little more extended than the first one. Consequently, this increases the sensitivity of the triangle to each component of

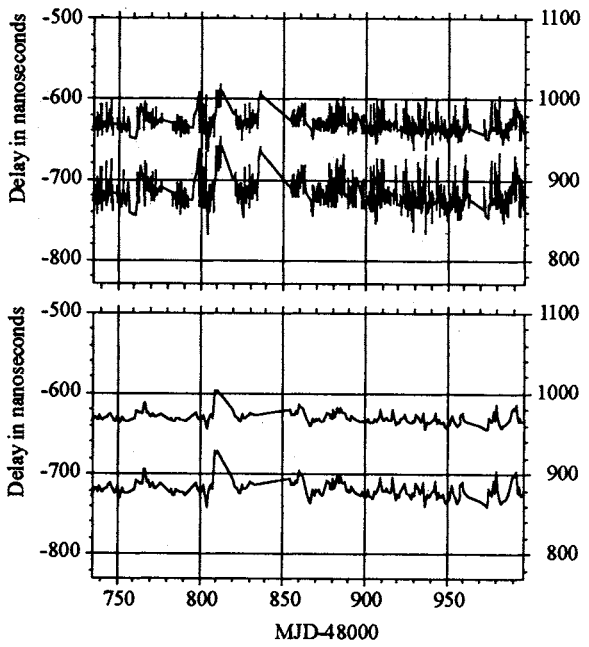


Figure 6 : Time transfer comparisons
$[UT(OB)-UT(OC)]_{GPS}$ - $[UT(OB)-UT(OC)]_{TDF}$
using OP-CT-OB (offset about 880 ns)
$[UT(OB)-UT(OC)]_{GPS}$ - $[UT(OB)-UT(OC)]_{TDF}$
using OP-CT-OC (offset about -630 ns)

upper graph : 6 measurements / day
lower graph : averaged data

the satellite motion. Therefore the determination of the satellite position is more precise and finally the uncertainty on the time transfer is weaker.

The time measurements data used for these plots have been corrected for the receiver delays, that have been determined by the CNES. The remaining offsets observed seem to be mainly due to errors on the antenna co-ordinates.

Figure 7 shows that the stability reached by this method is of the order of a few 10^{-14} for an averaging time of 10 days and is a little better (as seen above) for the triangle OP-CT-OC than for OP-CT-OB.

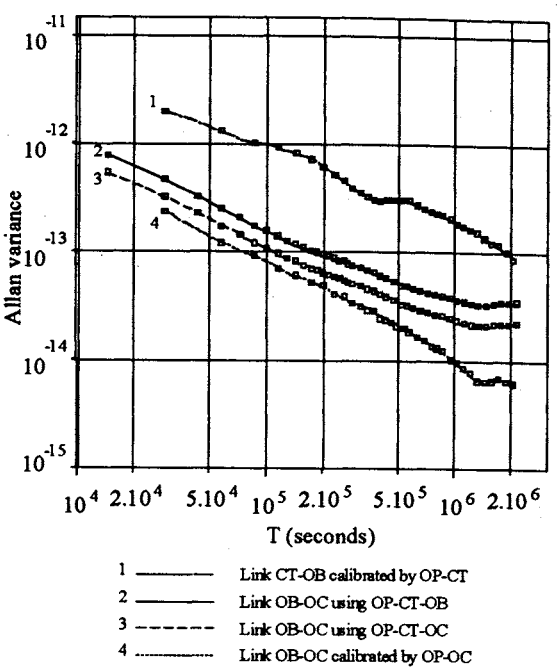


Figure 7 : Log-Log plots of Allan variance versus averaging time

The advantage of using three stations instead of two as in part 2 does not appear clearly because the stations are not distant enough : at this level of precision, the improvement brought by the third station is of the same order than the one obtained by averaging data. The use of a triangle of stations with larger extension in both longitude and latitude would of course give better results.

4. CONCLUSION

Time transfers via geostationary direct TV satellites are largely improved by adding information on the satellite position. Two ways have been presented : the first one uses daily averages of the measurements and corrections for the longitude drift of the satellite ; the second one takes into account the satellite position as known through the time measurements by 2 or 3 stations. Both method can lead at best, to agreement with GPS at the level of 10 nanoseconds (1σ).

More precise information would be obtained by the use of an extended ground basis, for example Kourou (Guyana) - France. The longitude modelization and the satellite position could be highly improved.

5. ACKNOWLEDGEMENTS

This work is supported by the Bureau National de Metrologie. The authors wish to thank A. Vincent for the study and the realisation of the receivers, J.F. Dutrey (CNES), R. Helmlinger (LPTF), P. Grudler (Observatoire de la Côte d'Azur) and B. Mougin (Observatoire de Besançon) for their helpful work of collecting and filing data.

6. REFERENCES

1. F. Meyer, M. Granveaud, M. Brunet, F. Vernotte, M. Vincent - Time transfer using a geostationary direct TV satellite, Proceedings of the Fifth EFTF Besançon 12-14 Mars 1991, pp 82-88.

2. J. A. Davis, P. R. Pearce, M. Hutchins, M. Sagin - Analytical techniques for improved time transfer using common view of direct broadcasting satellites. Proceedings of the Fifth EFTF, Besançon, 1992, pp. 74-81.

3. F. Meyer, M. Granveaud, B. Laporte, F. Vernotte, M. Vincent - Synchronisation d'horloges atomiques par TDF1 et TDF2 : résultats et perspectives, Bulletin du Bureau National de Métrologie, 90, Octobre 1992, pp. 15-22.

4. J. A. Davis, P. R. Pearce - Analytical techniques for improved time transfer using common view of direct broadcasting satellites. Proceedings of the Sixth EFTF, Noordwijk, 1992, pp. 195-199.

5. Cours de technologie spatiale : "Le mouvement du véhicule spatial en orbite", CNES, 18 Avenue Edouard Belin, 31055 Toulouse Cedex, 1980, pp. 583-610.

TWO-WAY SATELLITE TIME TRANSFER BETWEEN USNO AND TUG USING THE INTELSAT SATELLITE AT 307° EAST.

D. Kirchner
Technical University Graz, Austria

H. Reßler and R. Robnik
Space Research Institute, Graz, Austria

W. Klepczynski, P. Wheeler, W. Powell and A. Davis
U.S. Naval Observatory, Washington, D.C.

L. Veenstra
International Telecommunications Satellite Organisation
Washington, D.C.

ABSTRACT

During the second half of 1992 two-way time transfer test measurements were carried out between the U.S. Naval Observatory and the Technical University Graz, Austria, via the INTELSAT satellite at 307° E. Information is provided on the stations, the satellite and the link performance. Two-way results and a comparison with GPS results are presented and the problem of delay calibration is addressed.

Keywords: Two-way satellite time transfer, transatlantic link, MITREX-Modem, GPS time transfer.

1. INTRODUCTION

Two-way time transfer via communication satellites using pn-modulated signals as generated by MITREX modems provides highest precision and has the potential of high accuracy (Refs. 1, 2). In 1983 the first experiment using MITREX modems was carried out between the United States of America and Germany via an INTELSAT satellite (Ref. 3). Since that time no transatlantic measurements using this technique were performed mainly because of the advent of the Global Positioning System (GPS) and its potential for accurate time transfer (Ref. 4). Due to the growing availability of small and inexpensive satellite earth stations, so-called very small aperture terminals (VSAT's), and of Ku-band (14/11 GHz and 14/12 GHz) satellite capacity two-way satellite time transfer links between North America and Europe can now easily be established (Refs. 5, 6).

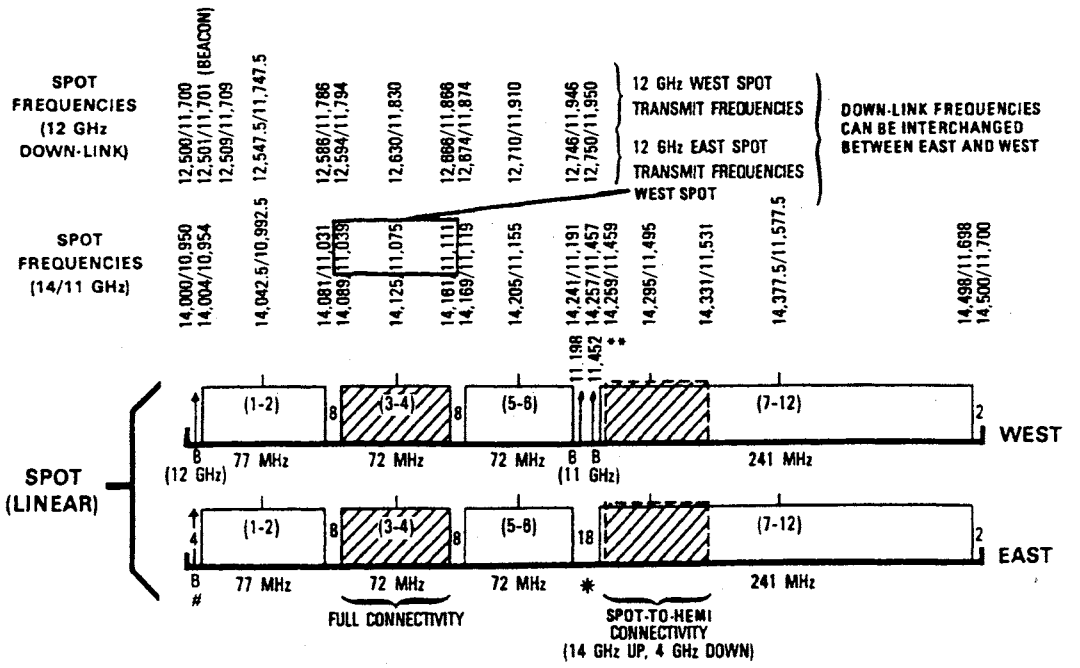
In June 1992 line-up tests were carried out for a two-way link between the U.S. Naval Observatory (USNO) and the Technical University Graz (TUG) using the INTELSAT satellite at 307° East followed by time transfer measurements for test purposes. The low elevation encountered at TUG was of special interest. Measurements on a regular basis will begin in the near future and will include additional North American and European stations with the aim of setting up a two-way time transfer network according to the Recommendation S6(1989) of the Comité Consultatif pour la Definition de la Seconde (Ref. 7).

2. EARTH STATIONS

The earth stations used in the experiment are of different types. The USNO station employs an antenna with a rather large diameter (4.6 m) and is of the expensive conventional design providing full frequency agility. The TUG station is of the inexpensive VSAT type providing frequency agility only in steps of 1 MHz, but using additional up- and down-conversion stages between the MITREX-modem and the transceiver employing synthesizers with 1 kHz step size any assigned satellite frequency within a given band can be used and the requirements of the MITREX-modem are met (Ref. 8). The main characteristics and the positions of the stations - both collocated with the time keeping stations - and the elevation angles for the INTELSAT satellite at 307° E are given in Table 1.

	USNO	TUG
Antenna Dia. (m)	4.6	1.8
G/T (dB/K)	31.6	22
Max. EIRP (dBW)	58	52
Used EIRP (dBW)	55	49
Position	77° 04' W 38° 55' N	15° 30' E 47° 04' N
Elevation	38.5°	5.8°

Table 1 Main characteristics and locations of the earth stations.



NOTES:
1. See Tables 1 and 2 for further information.
2. *On Spot 11 GHz down-link there is a 250 MHz gap between (5-6) and (7-12)(11,200-11,450 MHz).
3. The 11 GHz or 12 GHz spot frequencies may be selected independently for the East or West.
4. **The polarization of the 11 GHz beacons is Right Hand Circular (RHCP).
5. #On Spot 12 GHz down-link there is a 9 MHz guardband before (1-2) to accommodate the beacon.

Figure 1 INTELSAT VA (IBS) spot beam transponder layout (from Ref. 9).

The USNO station is approved as INTELSAT Standard E2 earth station and the TUG station as INTELSAT Standard G earth station. Line-up tests were performed prior to the spread spectrum time transfer tests which began in July 1992.

3. SATELLITE LINK

The INTELSAT VA (F-13) satellite located at 307°E (Ref. 9) carries most of the Europe-North American IBS (INTELSAT Business Service) traffic and provides great flexibility concerning the configuration of the transponders. The full-connectivity transponder configuration which was available during the experiment and was later on discontinued provides the most flexible environment for time transfer links. During this type of transponder operation it is possible for each site to monitor its own signal together with the signal from the remote site. The spot beam transponder layout is given in Fig. 1.

The tests were scheduled on frequencies allocated to the occasional use IBS service (USNO: 14090.575/11795.575 MHz, TUG: 14090.575/12595.575 MHz). The actual receive frequency differed from the nominal one by about 13 kHz, again demonstrating the need for receive frequency agility (1 kHz step size) of the stations or of frequency agile modems (Ref. 8).

The K-band West spot beam covers the United States and southern Canada and the K-band East spot beam covers most of western Eu-

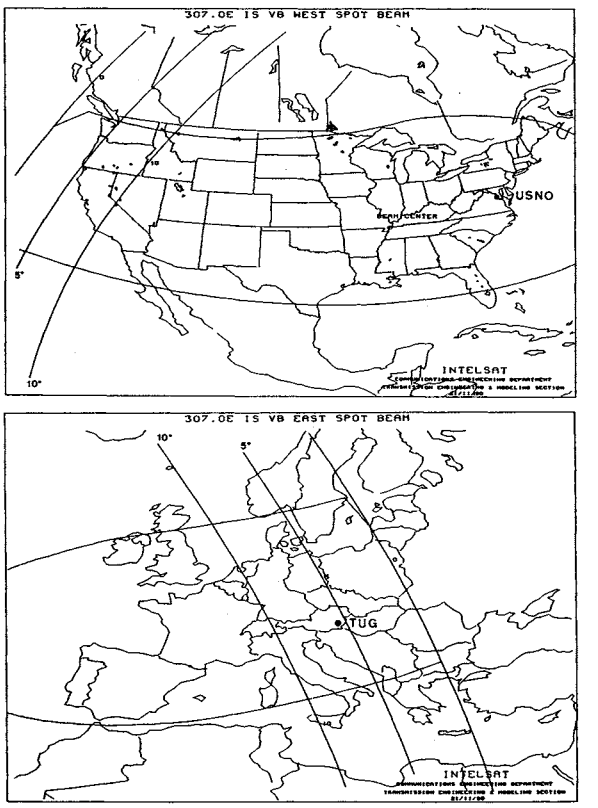


Figure 2 INTELSAT VA (IBS) West and East spot coverage from 307° East (from Ref. 6).

rope (see Fig. 2). The elevation angles from Europe are relative low but even at an elevation of lower than 6° as encountered at TUG successful operation is possible. Fig. 3 is a plot of typical traffic in a band of 50 MHz centered at the TUG receive frequency.

To achieve a jitter (standard deviation) of the received signal below 1 ns (see Section 4), the link should provide a carrier-to-noise power density ratio (C/N_o) of at least 54 dBHz. The result of the link budget calculations is given in Table 2.

Transmit E/S	USNO	TUG	USNO	TUG
Receive E/S	USNO	TUG	TUG	USNO
C/N_o (dBHz)	63	55	55	63

Table 2 Result of the link budget calculations for ranging and time transfer.

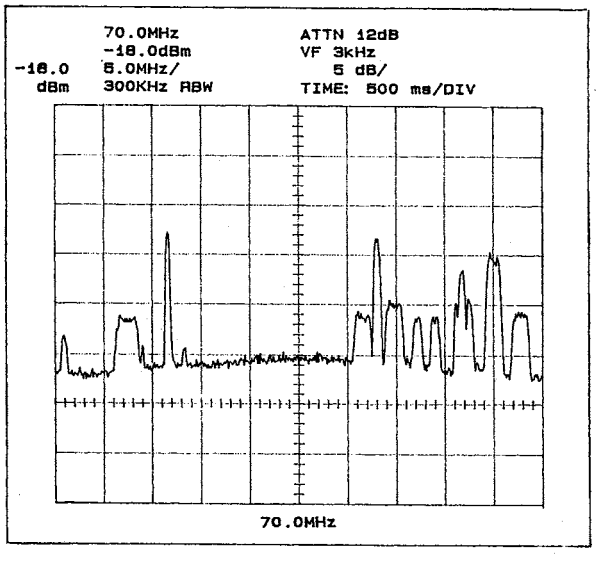


Figure 3 Plot of typical traffic via INTELSAT VA (IBS) as recorded at TUG (50 MHz scan centered at TUG receive frequency).

4. MEASUREMENTS AND RESULTS

After the initial line-up tests time transfer measurements were carried out between USNO and TUG for test purposes at irregular intervals beginning on July 16 and ending on October 30. The measurements were done on 17 days in several successive blocks of about 5 minutes duration (1 measurement per second) and the total of 99 measurement blocks were performed. In addition to the two-way time transfer measurements GPS time transfer measurements employing the common-view method were carried out.

In the following two-way and GPS results are given without a preceding presentation of the basics of both methods which can be found elsewhere (Refs. 1, 10, 2). Problems concerning the comparison of the results obtained by both methods are discussed in (Ref. 11). For each block of measurements a linear regression through the differences of the corresponding one-second measurements of both stations (the measurement data were exchanged via INTERNET) and the expected value of this regression referred to the midpoint were computed. Fig. 4 gives the standard deviations of the residuals of these regressions (on the average 730 ps) and Fig. 5 the Modified Allan Variances (Mod.$\sigma_y(\tau)$) which are a more appropriate measure to assess the performance of a time and/or frequency transfer link (Ref. 12).

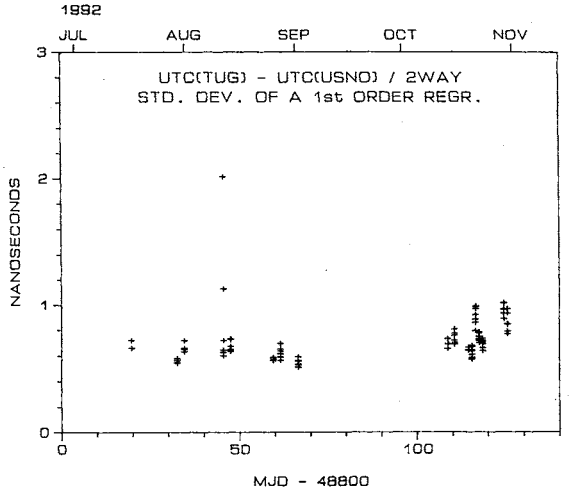


Figure 4 Standard deviations of the two-way time transfer measurements.

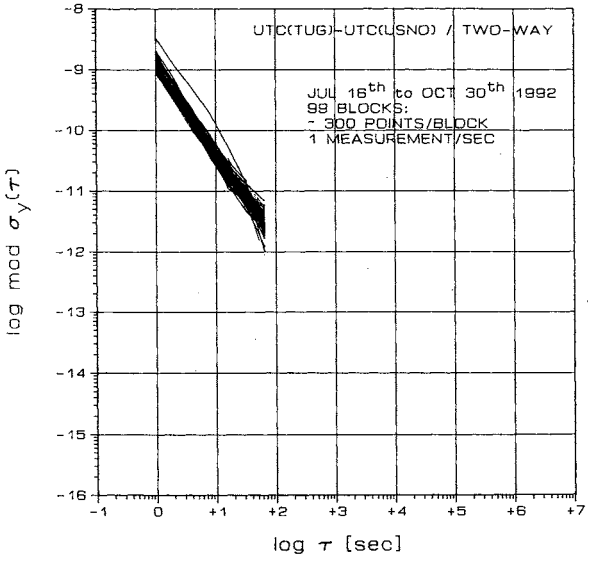


Figure 5 Modified Allan variances of the two-way time transfer measurements.

For white-noise PM the slope in a Mod.$\sigma_y(\tau)$ diagram is -3/2 and dividing Mod.$\sigma_y(\tau=1s)$ by $\sqrt{3}$ one obtains the standard deviation. The noise levels in Fig. 5 depend on the actual C/N_o available during the measurements and the average slope indicates a noise process which is obviously of white-noise PM type. Looking to the separate Mod.$\sigma_y(\tau)$ plots of the single blocks - not included here - one can sometimes find

small deviations from a straight line indicating periodic phenomena which can also be seen by plotting the residuals and which have to be investigated further.

In addition to the computation of the differences of the corresponding one-second measurements of both stations (see above) also quadratic regressions through the individual one-second measurements of both stations have been computed and the expected values of these regressions referred to the midpoint were used to calculate the differences. Using this approach the amount of data which has to be exchanged in a future operational system can be reduced drastically. The differences between the results obtained by both methods have a mean value of 4 ps, a standard deviation of 80 ps and a maximum and minimum deviation of about 200 ps and -200 ps, respectively. Similar results have been obtained in earlier experiments (Ref. 13).

The differences UTC(TUG) - UTC(USNO) obtained by two-way and GPS are depicted in Figure 6. The difference of about 550 ns between the results obtained by both methods is due to the fact that no corrections apart from that for the differences between the reference pulses of the measurements and the transmitted pulses have been applied. Fig. 7 gives the difference between UTC(TUG) - UTC(USNO) obtained by two-way and GPS after applying all available two-way corrections (the one applied in Figure 6, the respective differences between UTC(Lab) and the reference pulse, and the Sagnac effect of -207.8 ns). The remaining difference of about 350 ns is due to the unknown delays in the earth stations and the path non-reciprocity in the satellite because of the use of different spot beams (Ref. 14). It should be noted that several changes in the earth station equipment were made before October 16, this caused the overall peak-to-peak deviation to change from about 20 ns to about 10 ns after October 16.

5. CONCLUSION

A transatlantic two-way time transfer test link has been established employing pn-coded signals generated by MITREX-modems and exchanged via the INTELSAT VA (IBS) satellite at 307° E using a satellite terminal of VSAT type at the European side.

For the measurement blocks of 5 minutes duration the one-second measurement data show white-noise PM behaviour with an average level of about $1.3 \times 10^{-9}\ \tau^{-3/2}$ (τ in seconds) and therefore using conventional cesium frequency standards the clock noise is encountered for averaging times in the range of minutes. Under the assumption that this white-noise PM behaviour will persist for longer integration times - an assumption which has to be proved by future experiments studying the stability of two-way satellite links - remote frequency standards could be compared to the 10^{-15} level employing integration times of some hours.

No problems were caused by the low elevation at Graz compared with higher elevations apart from the expected larger attenuation during rainstorms. Due to the decreasing C/N_0 during such events at rare occasions the link could not be established at all.

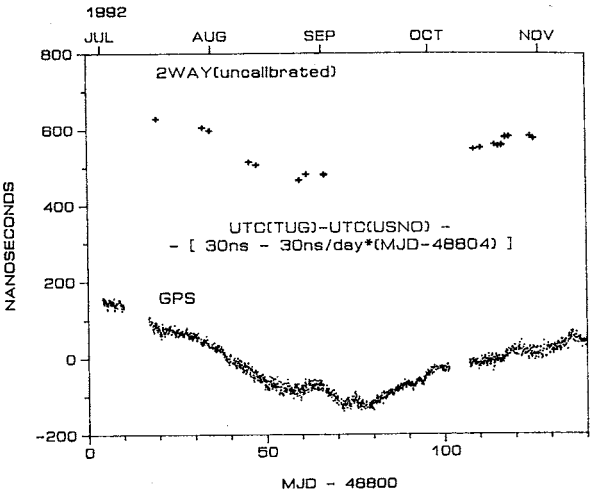


Figure 6 UTC(TUG) - UTC(USNO) computed from two-way measurement blocks (corrected for the differences between the reference pulses of the measurements and the transmitted pulses) and single GPS common-view tracks after removal of the mean rate difference between UTC(TUG) and UTC(USNO).

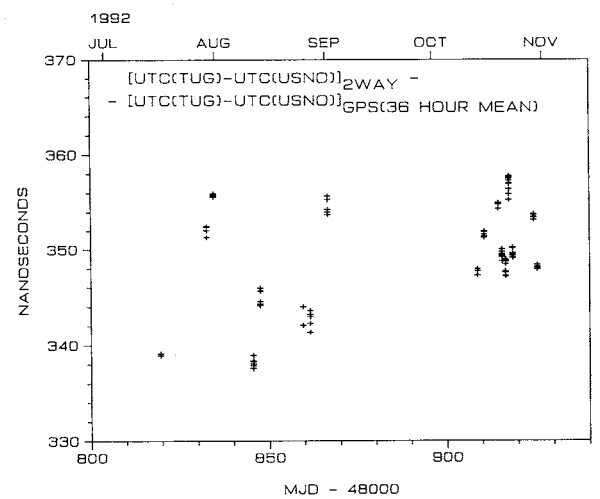


Figure 7 Difference between UTC(TUG) - UTC(USNO) measured by two-way and GPS (GPS data smoothed over 36 h) after applying all corrections apart from that for the earth station delays and satellite non-reciprocity.

Because different spot beams are used at both sites the signal delay in the satellite is different for both signal directions and therefore the link cannot be calibrated by transportation of an earth station or by local measurements. Fortunately for frequency comparisons only the stability of the link is of importance and therefore no absolute calibration is necessary. To calibrate the link other high

accurate time transfer methods could be used.

In a very near future additional stations in America and Europe will be ready and therefore plans for test operations including these stations are being developed. Unfortunately the full-connectivity transponder configuration will not be available in foreseeable future and therefore employing the INTELSAT satellite at 307° E for transatlantic and regional links stations will have to switch between different frequencies because different transponders will have to be used for transatlantic and regional operation.

ACKNOWLEDGEMENTS

The assistance of the COMSAT Corp. in obtaining INTELSAT satellite transmission time is gratefully appreciated and contributed to the success of these preliminary experiments. The help of W. Schladowsky (ÖPTV) in administrative and technical matters is gratefully acknowledged. The work was supported by the Austrian Academy of Sciences and the Jubilee Fund of the Austrian National Bank.

REFERENCES

[1] D. Kirchner, "Two-way satellite time transfer via communication satellites", Proc. IEEE, vol. 79, pp. 983-990, 1991.

[2] P. Hartl et al., "MITREX 2500, a modem for microwave time and ranging experiments via telecommunication satellites", Inst. of Navigation, Univ. Stuttgart, 1985.

[3] Ph. Hartl, L. Veenstra, N. Gieschen, K.M. Müssener, W. Schäfer, C.M. Wende, W. Klepczynski, H.H. Nau, and R. Stoiber, "Spread spectrum time transfer experiment via Intelsat", in Proc. 15th Annual PTTI Meeting pp. 331-356, 1983.

[4] W. Lewandowski and C. Thomas, "GPS time transfer", Proc. IEEE, vol. 79, pp. 991-1000, 1991.

[5] L.B. Veenstra and D.W. Hanson, "Two-way satellite time transfers between and within North America and Europe", in Proc. 42nd Annual Symp. on Frequency Control, pp. 179-185, 1988.

[6] L.B. Veenstra, "International two-way satellite time transfer using INTELSAT space segment and small earth stations", in Proc. 22nd Annual PTTI Meeting, pp. 383-400, 1990.

[7] Comité Consultatif pour la Définition de la Seconde (CCDS), Rapport de la 11e session, Recommandation S6(1989), Paris: BIPM, 1989.

[8] D. Kirchner, "Design considerations for Ku-band two-way satellite time transfer terminals", in Proc. 4th European Frequency and Time Forum, pp. 631-637, 1990.

[9] INTELSAT Earth Station Standards (IESS). Document IESS-407 (Rev.4), 1990.

[10] D.W. Hanson, "Fundamentals of two-way time transfer by satellite", in Proc. 43rd Annual Symp. on Frequency Control, pp. 174-178, 1989.

[11] D. Kirchner et al., "Comparison of two-way satellite time transfer and GPS common-view time transfer between OCA and TUG", in Proc. 23rd Annual PTTI Meeting, pp. 71-88, 1991.

[12] D.W. Allan, "Time and frequency metrology: Current status and future considerations", in Proc. 5th European Frequency and Time Forum, pp. 1-9 , 1991.

[13] D. Kirchner et al., "A two-way time transfer experiment via ECS-1 using the MITREX Modem", IEEE Trans. Instr. Meas., vol. 37, pp. 414-417, 1988.

[14] G. de Jong, "Prospects of two-way international time scale comparisons", in Proc. 6th European Frequency and Time Forum, pp. 69-73 , 1992.

REAL-TIME RESTITUTION OF GPS TIME

Claudine THOMAS

BIPM, Pavillon de Breteuil, F-92312 Sèvres Cedex

ABSTRACT/RESUME

The Global Positioning System is an outstanding tool for the dissemination of time. Equipped with C/A-code GPS time receivers, civil users may access the time scale GPS time, disseminated by the GPS satellites, and then the reference time scales UTC(USNO) and UTC. At a given instant, the restitution of GPS time through the satellite constellation presents a peak-to-peak discrepancy of several tens of nanoseconds without SA and several hundreds of nanoseconds with SA. This paper investigates diverse solutions for improving the real-time restitution of GPS time, with implementation of SA, to the level obtained without SA. One proposal relies upon the statistical analysis of the noise brought about by SA. It shows that averaging on numerous simultaneous observations of Block II satellites may help. If such data is not available, a second solution proposes the application of a Kalman filter to GPS timing data. This allows local access to GPS time through an estimation based on past data and updated at each new observation.

Keywords: Time, GPS, Selective Availability, Kalman filter.

1. INTRODUCTION

The NAVSTAR Global Positioning System, GPS, is a military satellite navigation system based on satellite ranging using on-board atomic clocks. In its final stage it will provide position, velocity and time instantaneously and continuously anywhere on or above the Earth (Ref. 1). To determine a position in three space dimensions and time, a user must receive signals simultaneously from four satellites. However, if a user has a known position it is necessary only to track one satellite in order to determine time. In this way, GPS can be used as a time distribution system. The time scale GPS time distributed by the GPS satellites is a continuous time, maintained in agreement with UTC(USNO) to within 100ns (Refs. 2,3).

Observational data delivered by GPS time receivers are time differences between the 1pps (1 pulse-per-second) of the user's local clock and GPS time, as deduced from the received signal of a given GPS satellite.
Such timing-data may be used for time transfer purposes by the common-view method (Ref. 4). The principal feature of this method is that satellite clock error contributes nothing, as GPS time disappears in the difference. This is of utmost interest during implementation of Selective Availability (SA) (Ref. 5). The ultimate accuracy of the common-view mode is a few nanoseconds (Ref. 6).
The GPS timing data may also be exploited simply for access to the time scale GPS time and then eventually to reference time scales such as UTC(USNO) and UTC. The precision of access to GPS time depends on local conditions of observation, mainly on the quality of the receiver antenna coordinates. Generally, the realization of GPS time through the constellation shows a peak-to-peak discrepancy, at a given instant, of up to 30ns. However this performance is largely degraded by implementation of Selective Availability.

Selective Availability is an intentional degradation of GPS signals, implemented on Block II satellites only, which aims to deny unauthorized users access to the full accuracy of instantaneous positioning and distribution of time. It has been implemented since November 1991 as a dither-noise added to on-board clocks. As a consequence, the real-time restitution of GPS time from Block II satellites suffers peak-to-peak discrepancies up to several hundreds of nanoseconds as shown in Section 2 of this paper.

In this paper solutions are proposed for real-time processing of local GPS timing data with the aim of improving the restitution of GPS time, in the case of SA, to a level comparable with that obtained without SA.
A first solution is detailed in Section 3. It relies upon statistical analysis of the noise brought about by SA. It appears that for an averaging time of a few hundreds of seconds, the SA noise is white and corresponds to an uncertainty (1 sigma) of about 100ns. It follows that an average of data from numerous Block II satellites, observed simultaneously, greatly decreases the impact of SA, though it does not protect against its systematic effects in the long term.
In the case where such data is not available, a second solution is investigated. This is presented in Section 4 and relies upon Kalman filter theory. It provides a real-time access to GPS time from filtered past data and from present observations. Assuming a linear model of evolution of the system, the Kalman process gives an optimal estimation, in the least-squares sense, to the degree that the actual model noise and observation noise are white and Gaussian. The theoretical considerations concerning the Kalman filter, and the application of the Kalman formalism to timing-data, are detailed in Ref. 7. Section 4 of this paper only recalls the method used and presents experimental and numerical results.

2. REAL-TIME ACCESS TO GPS TIME FROM THE OBSERVATION OF ONE GPS SATELLITE

The observation of one given GPS satellite, via channel i of a multi-channel C/A code GPS time receiver, leads to the recording of the quantity

$$z_i(t) = [\text{Local Clock - GPS time (SV)}]_i(t), \qquad (2.1)$$

generally obtained through short-term data processing over a time interval of duration τ_o, centred on date t.

Quantitative information on the precision of real-time access to GPS time at date t, may be obtained by comparison of the real-time observation $z_i(t)$ with a deferred-time estimation of

$$x(t) = [\text{Local Clock - GPS time}](t). \qquad (2.2)$$

This deferred-time estimation is efficiently achieved through the application of a smoothing technique on a set of data from Block I satellites not affected by SA.

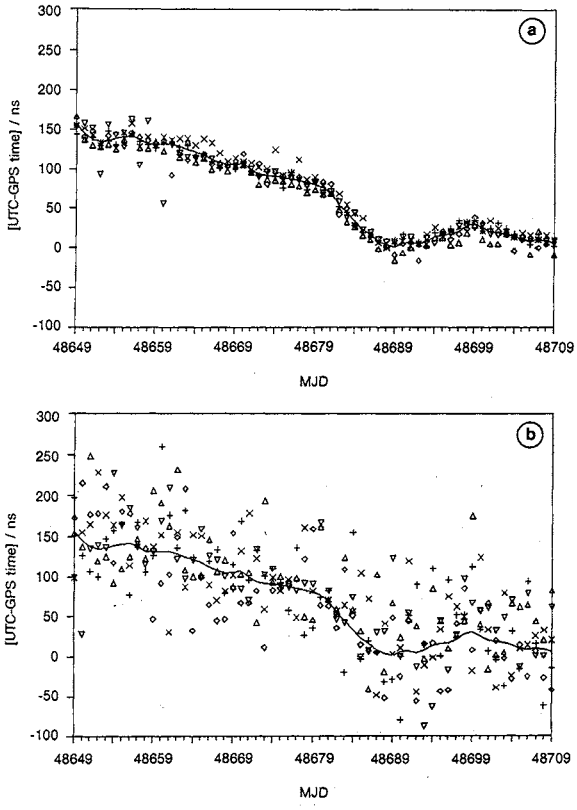


Fig. 1. Restitution of [UTC - GPS time] in deferred-time as published in the BIPM Circular T.
Measured ionospheric delays are used to correct raw GPS data.

1a. Data from Block I tracks and corresponding smoothed values (——).
1b. Data from Block II tracks and smoothed values of Fig. 1a.

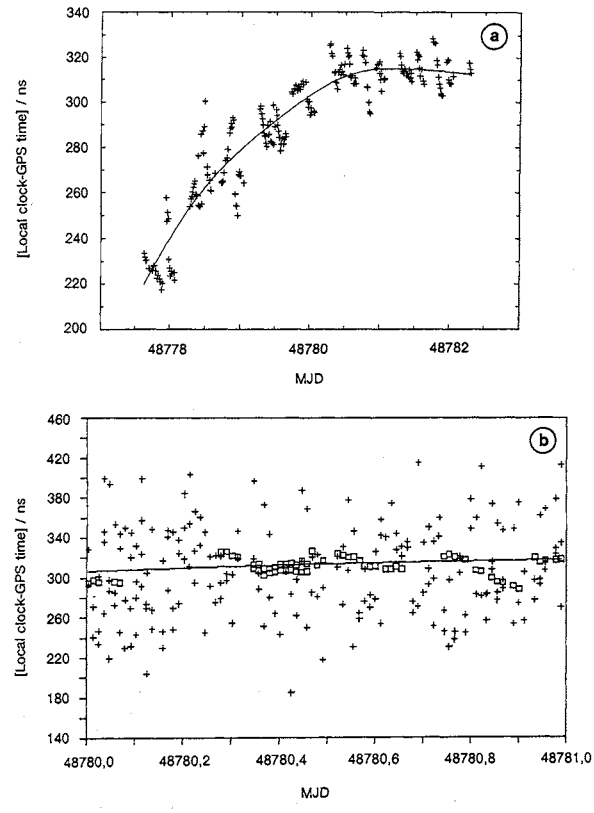


Fig. 2. Data [Local Clock - GPS time] taken at the BIPM between 4 and 9 June 1992.

2a. + + + Block I satellite data.
—— Smoothed values obtained in deferred-time from Block I satellite data.
2b. □ □ □ Block I satellite data.
+ + + Block II satellite data.
—— Smoothed values of Fig. 2a.
Fig. 2b. concerns one day (7 June 1992).

Such estimates are regularly made at the BIPM which publishes values of [UTC - GPS time] in its monthly bulletin, *Circular T.* An example is shown in Fig. 1a for the two-month period February-March 1992, with data following the international GPS common-view schedule.

The precision of GPS time restitution can be expressed in terms of the root mean square, r, of the residuals to the smoothed values; here r = 7,3ns. This value, deduced from Block I satellite data only, reflects the impact of systematic errors due to conditions of observation such as:

i. residual errors in local antenna coordinates, in general very small for national timing centers contributing to TAI (Ref. 8),
ii. errors in estimating the ionospheric delay of GPS signals, here negligible as measured ionospheric delays are used (Ref. 9),
iii. errors in ephemeride parameters broadcast by the satellite (Ref. 10),
iv. other minor error sources such as multipath effects and the sensitivity of the local receiver to external temperature (Ref. 11).

In Fig. 1b are reported the smoothed values [UTC - GPS time] of Fig. 1a, together with raw data obtained from a selection of Block II satellites. The impact of SA may be expressed quantitatively as r equal to 48ns, which is much larger than the value 7,3ns obtained from Block I.

Another example is given in Figs. 2a and 2b. This corresponds to GPS data taken at the BIPM and used for the numerical study detailed in Section 4 of this paper. A four-channel C/A-code time receiver was especially scheduled for a 5-day period (4 to 9 June 1992) with all available Block I and II satellites: four different satellites were scheduled every 16 minutes, each satellite being observed for a 13 min track length. The local reference is a commercial caesium clock installed at the BIPM.

Raw data $z_i(t)$ taken from Block I satellites only are shown on Fig. 2a, together with a smoothed curve. Here no ionospheric measurement is used and r is equal to 9,8ns. Fig. 2b corresponds to Block I and Block II data for 1 day (7 June 1992) when r becomes equal to 52ns. When no Block I satellite is observed, for instance near 5h UTC on 7 June 1992, the real-time access to GPS time is degraded to the level of 100ns.

It follows that with reasonable conditions of observation, the real-time reading of the quantity [Local Clock - GPS time] presents discrepancies up to 30 ns from Block I satellites and up to 150ns from Block II satellites.

3. REAL-TIME ACCESS TO GPS TIME FROM THE SIMULTANEOUS OBSERVATION OF A NUMBER OF BLOCK II SATELLITES

If real-time access to GPS time is degraded to the level of 150ns from the observation of one single Block II satellite, one can imagine that further information may come from the simultaneous tracking of several Block II satellites. To develop a procedure based on this idea, it is first necessary to have a good knowledge of the statistical properties of the SA noise.

For this purpose, timing data from different Block II satellites have been analysed at the BIPM. They correspond to 15s data (τ_0 = 15s) sequentially recorded with a one-channel C/A-code receiver at the BIPM on 10, 11 and 12 February 1993. Five different Block II satellites (SV 16, 19, 20, 26, 27) and one Block I satellite (SV 12) were observed for several hours (from rising up to going down).

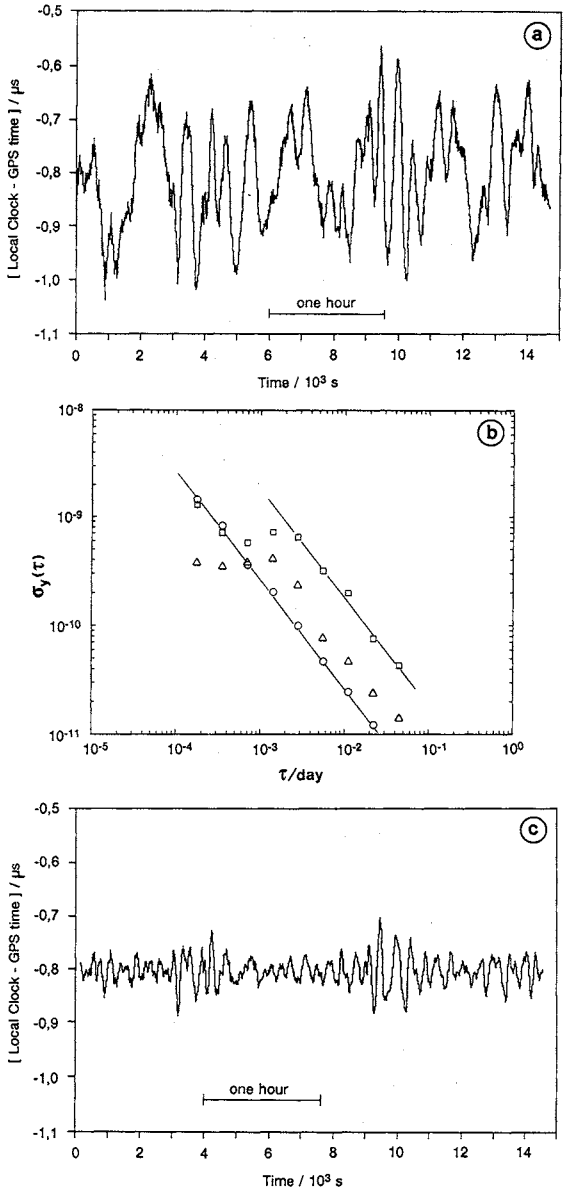


Fig. 3. Study of observational data from satellite 26, taken for more than 4 consecutive hours.

3a. Raw 15s measurements.
3b. Allan deviation of:
- □ *15s raw measurements from satellite 26.*
- ○ *15s raw measurements from satellite 12.*
- △ *15s smoothed measurements from satellite 26.*

3c. Measurements from satellite 26, smoothed for short-term and long-term white noise.

Figure 3a shows a plot of 15s raw data recorded for satellite 26 over more than 4 consecutive hours. The Allan deviation computed with this timing data is reported in Fig. 3b together with the Allan deviation values obtained with 15s data from SV 12.

Timing data from SV 12 shows white phase noise, the level of which is about 15s, for all averaging times. It corresponds to observational noise, as detailed in Section 2. The same noise is apparent on Block II data for the very short term, up to τ = 30s. In the longer term, the SA degradation brings additional noise, characterized by the typical Allan deviation 'bump', which corresponds to sinusoidal variations of timing data with a half-period of order 150s. For larger τ (240s to about 2000s), the predominant noise is white phase noise, corresponding to a level of about 85ns.

White phase noise can be reduced by using simple averages. Figure 3c shows the timing data from SV 26, after smoothing out of both of the short-term and long-term white phase noises. The corresponding Allan deviation values are added in Fig. 3b: they show the efficiency of the smoothing procedure. Figure 3c clearly indicates the presence of residual sinusoidal variations with periods of order some hundreds of seconds.

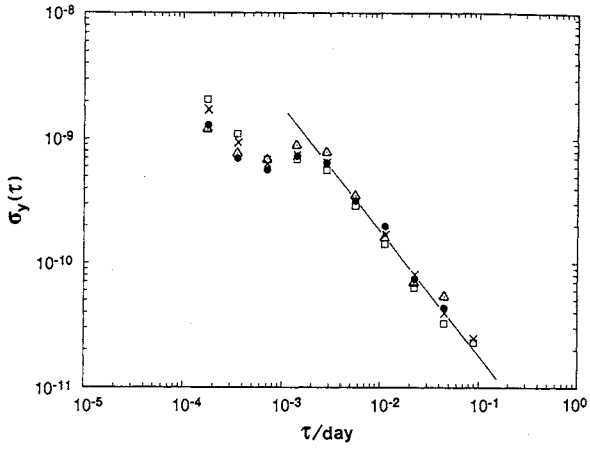


Fig. 4. Allan deviation of 15s observational data from 4 Block II satellites:

- ● *Satellite 26.*
- □ *Satellite 19.*
- × *Satellite 27.*
- △ *Satellite 16.*

The same analysis have been performed with observational data from different Block II satellites. The corresponding Allan deviations are reported in Fig. 4 for four of them. This demonstrates that the SA noise has the same statistical features in each of the four examples, though they were not tracked simultaneously.

This conclusion may have very important consequences. One can easily imagine that, due to SA, all Block II satellites are affected by white phase noise of comparable level over averaging times of order 1000s. To reduce this SA effect, it is sufficient to observe simultaneously a number of Block II satellites over such averaging times and to compute a mean of the individual satellite results. For instance, observation of ten satellites with a multi-channel receiver for a 1000s track-length reduces the level of the SA white noise to about 13ns. This is comparable to usual observational noise. In addition, the result of the mean over these 10 results may be obtained immediately after recording the simultaneous tracks, thus providing access, in real-time, to GPS time.

However, the preceeding analysis does not prove that SA does not introduce other effects over much longer averaging times. For instance, Fig. 5a presents a complete set of timing data taken at the BIPM over a 26 hour-period and composed of Block I and II tracks, following the international GPS common-view schedule, and short-term data for satellites 27 and 20 observed over several consecutive hours. None of the previous noise can be detected on satellite 20 data. Moereover, the corresponding Allan deviations, in Fig. 5b, are similar to those from SV 12. Nevertheless, a bias gradually appears and reaches 150ns within several hours. The SA degradation can thus be very different in its effects.

The use of an average over a number of simultaneous and sufficiently long observations cannot solve all the problems. In fact, it can be applied only if long term biases are correctly detected and controlled. This leads, of course, to longer observational times and thus delays access to GPS time.

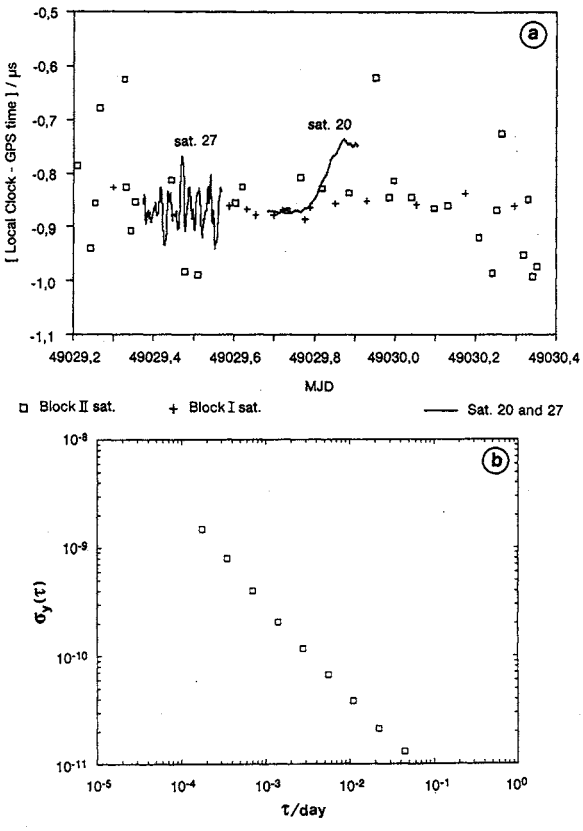


Fig. 5. Timing data taken at the BIPM over a 26 hour-period:

5a. + + + Common-view tracks from Block I satellites.
□□□ Common-view tracks from Block II satellites.
——— Short-term measurements from satellites 20 and 27.
5b. Allan deviation of 15s measurements from satellite 20.

4. REAL-TIME ACCESS TO GPS TIME FROM THE IMPLEMENTATION OF KALMAN FILTERING ON PAST AND PRESENT OBSERVATIONS OF GPS SATELLITES

The procedure proposed in Section 3 may be difficult to apply simply because the observer does not have at his disposal a sufficient number of Block II satellite data. It is then convenient to couple the small number of current observational data with understanding of the system behaviour obtained through the analysis of past data. A tool well-adapted to such a goal is the Kalman prediction technique.

4.1 Theoretical considerations

Theoretical considerations about the Kalman filter, as well as how it can be adapted to timing observations are detailed elsewhere (Ref. 7).

Consider the experiment already evoked in Section 2, where four different GPS satellites are scheduled each 16 minutes for observation at the BIPM. Observational data z(t), affected by the measurement noise r(t), are taken following a sequence of discrete and equally-spaced dates t, with τ of order 1000s as the basic time step.

We study here the dynamic system corresponding to the observations which are taken. This is represented by the quantity $x(t_1+\tau)$ at date $t_1+\tau$, already defined in (2.2):

$$x(t_1+\tau) = [\text{Local Clock - GPS time}](t_1+\tau). \qquad (4.1)$$

The quantity x has evolved with time since t_0, the date of the first observation. Its model of evolution is linear and written as:

$$x(t_1+\tau)/\tau = x(t_1)/\tau + y_0 + v, \qquad (4.2)$$

where y_0 is the slowly varying frequency of the local clock relative to GPS time and v is the white frequency noise (WFN) which affects y_0 for τ of order 1000s. The WFN level is characterized by the Allan deviation, $\sigma_y^2(\tau)$, computed with time data obtained by comparison of the local clock with some external reference.

The Kalman prediction technique provides an estimate, written as $x(t_1+\tau/t_1+\tau)$ (and read as 'estimation of x at date $t_1+\tau$, knowing all observations since date origin') of the quantity $x(t_1+\tau)$, together with the confidence on this estimate $g(t_1+\tau/t_1+\tau)$.

The Kalman filter is linear and optimal in the least-squares sense if the model noise and the measurement noise are white and Gaussian, all deterministic effects being accounted for in the evolution model and in the observational procedure. It performs a filtering of past data, according to the model used and the level of the noise involved. It thus provides a prediction for the next time interval. It only runs forward and is a tool well-adapted for real-time estimation.

In addition, the Kalman filter is recursive. It provides an estimate of the quantity x(t) at date $t_1+\tau$, $x(t_1+\tau/t_1+\tau)$, and gives the confidence on this estimate, $g(t_1+\tau/t_1+\tau)$, knowing:
- the initial conditions, deduced from the first observation performed at t_0,

$$x(t_0/t_0) = z(t_0) \text{ and } g^2(t_0/t_0) = r^2(t_0), \qquad (4.3)$$

- the previous estimate, $x(t_1/t_1)$, of x(t) at date t_1,
- the confidence, $g(t_1/t_1)$, on the estimate of x(t) at date t_1,
- and the observation, $z(t_1+\tau)$, at date $t_1+\tau$, characterized by the observational noise, $r(t_1+\tau)$.

The corresponding equations are written:

$$x(t_1+\tau/t_1+\tau)=[x(t_1/t_1) + y_0\tau][1 - k(t_1+\tau)] + k(t_1+\tau)z(t_1+\tau), \qquad (4.4)$$

$$g^2(t_1+\tau/t_1+\tau) = [g^2(t_1/t_1) + \sigma_y^2(\tau)\tau^2][1 - k(t_1+\tau)] \qquad (4.5)$$

where $k(t_1+\tau)$ is the Kalman gain

$$k(t_1+\tau)=[g^2(t_1/t_1)+\sigma_y^2(\tau)\tau^2]/[r^2(t_1+\tau)+g^2(t_1/t_1)+\sigma_y^2(\tau)\tau^2]. \qquad (4.6)$$

These equations include four parameters, τ, y_0, $\sigma_y^2(\tau)$ and $r^2(t_1+\tau)$, directly derived from the conditions of the experiment.

The Kalman gain, $k(t_1+\tau)$, is the key feature of the Kalman procedure. It is a positive number between 0 and 1. Equation 4.4 indicates that it directly specifies the fraction of the observation taken at a given date in the estimate of x for that date. For example, if the observation $z(t_1+\tau)$ is very bad, the level of the observational noise, $r(t_1+\tau)$, is rather large and $k(t_1+\tau)$ is close to zero. The estimate $x(t_1+\tau/t_1+\tau)$ thus does not take into account the observation $z(t_1+\tau)$ and relies mainly on the knowledge of the past. The confidence on this estimate, given in Eq. 4.5, is the confidence obtained at the previous step, increased with the local clock noise.

In the case of a quasi-stationary filter, the Kalman gain is nearly constant and independent of the date. The expression of $x(t_1+\tau/t_1+\tau)$ thus represents an exponential weighted average of past and present observations. The time constant of this exponential filter is inversely proportional to k and is written as:

$$T = \tau/k. \qquad (4.7)$$

The value of k thus indicates the memory extension of the filter.

4.2 Application

To apply the procedure detailed in Section 4.1, it is first necessary to provide values for the parameters involved.

The local clock installed at the BIPM (Sèvres, France) has been compared over a 6-month period with the master clock UTC(OP) of Paris Observatory (Paris, France) using the common-view method with more than 30 daily common views. The BIPM clock presents a frequency relative to UTC(OP) of about 25 ns/day when averaged over 10-day

intervals. This value is probably much larger than the frequency of GPS time relative to UTC(OP) for the same averaging time. Thus the value taken for y_0 is:

$$y_0 = 25\text{ns/d}. \quad (4.8)$$

The value for τ=1000s of the Allan deviation of the BIPM clock relative to UTC(OP) is:

$$\sigma_y(\tau) = 2\text{x}10^{-13}, \quad (4.9)$$

and corresponds to the WFN regime. The accumulated error for τ=1000s is then:

$$\sigma_y(\tau)\tau = 0{,}2\text{ns}. \quad (4.10)$$

At a given date t_1, we have at our disposal four individual measurements issued from the four channels of the GPS receiver:

$$z_i(t_1) = [\text{Local Clock - GPS time (SV)}]_i(t_1),\ i=1 \text{ to } 4, \quad (4.11)$$

directly obtained as a linear fit result over 780s of short-term data.

The noise affecting $z_i(t_1)$ can be characterized by the variance:

$$\sigma_i^2(t_1) = s_i^2(t_1) + s^2, \quad (4.12)$$

obtained as the quadratic sum of the two quantities, $s_i(t_1)$ and s where,

i. $s_i(t_1)$ is the RMS characteristic of the 13-minute linear fit applied by the receiver software on 15-second measurements. It is the single available information about SA in real-time and can be as large as several tens of nanoseconds for Block II satellite.

ii. s represents the level of uncertainty due to systematic effects in local conditions of observation as described in Section 2, it is chosen equal to 10ns.

The mean observation at a given date is thus the weighted average of the four simultaneous observations, with weights chosen as reciprocal to σ_i^2. The corresponding value of r is of order 15ns to 30ns.

4.3 Results

In order to test the real-time restitution of GPS time obtained by the method described in Section 4.1, the Kalman estimations are compared with the smoothed Block I data shown in Fig. 2a, which corresponds to the application, in deferred-time, of a Vondrak smoothing with a cut-off period of about two days (Ref. 12). This is a reasonable choice made specifically for this paper. Another smoothing technique could equally have been chosen.

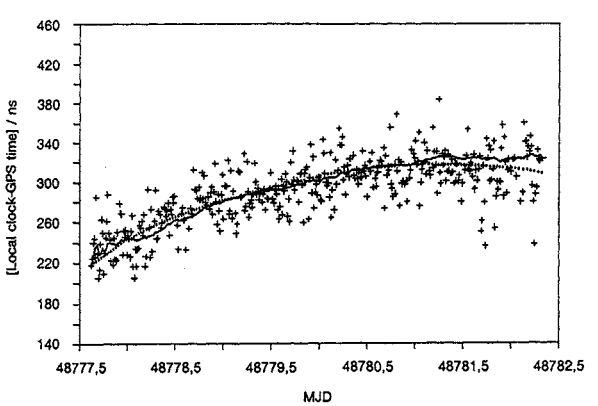


Fig. 6. Kalman estimation of the quantity [Local Clock - GPS time]:

+ + + Averaged observations.
——— Kalman estimation.
- - - - Smoothed Block I data.

The result of the Kalman estimation of the quantity [Local Clock - GPS time] is presented in Fig. 6, together with the averaged measurements z(t) obtained at each step. The Block I smoothing of Fig. 2a has been added to Fig. 6 and shows that the Kalman estimation is always very close to the Block I data. On the 5-day experiment, the mean difference between the estimations is of order 5ns. The Block I smoothing corresponds to a rather strong smoothing of data while the Kalman estimation presents short-term fluctuations with an amplitude of order 2ns.

During the 5 day experiment, one can distinguish three different phases:

i. For the first 12 hours, the difference between the estimations oscillates rapidly. This corresponds to a transient regime. The Kalman estimation here is excessively influenced by the initial conditions derived from the first observation. After running for several hours, the Kalman estimation becomes independent of the initial conditions and reaches a quasi-stationary state.

ii. For the following 3 days, the Kalman estimation is always smaller than the smoothed Block I values. This is because the value of y_0 is too small: it does not sufficiently increase the part of the estimation arising from the model of evolution between two dates (see Eq. 4.4).

iii. For the last part of the period, the Kalman estimation is always larger than the smoothed Block I data. The value of y_0 is too large: it would have been necessary to update y_0 or to add a frequency drift in the model equation.

From these observations, one can conclude that the real-time Kalman estimation does not diverge from that expected in deferred-time from Block I observations only. The discrepancy never exceeds 14 ns with the parameter values and the degree of smoothing of Block I data chosen here. However, it must be stressed that efficient implementation of the Kalman technique supposes a deep knowledge of the physical phenomena involved so as to provide realistic values for the different parameters involved. These parameters may have long-term variations and thus must be updated from time to time.

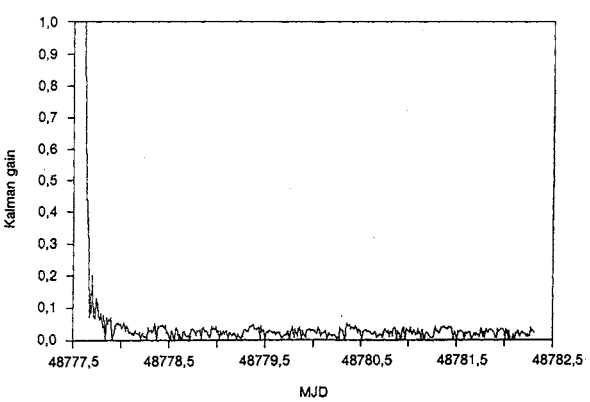


Fig. 7. Kalman gain of the estimation of the quantity [Local Clock - GPS time].

A curve showing the variation of the Kalman gain k is shown in Fig. 7. At the first step, k is equal to 1 as the estimation relies entirely on the observation chosen as the initial condition. During the transient phase, k rapidly drops below 0,06. This means that the new observation never accounts for more than 6% of the update estimation. In Fig. 7, k is sometimes equal to zero: this corresponds to steps for which no observation is available.

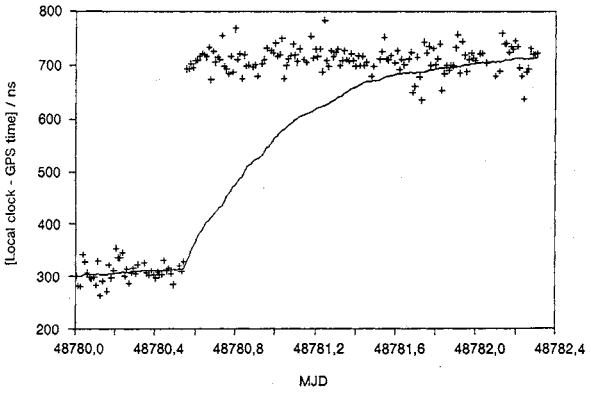


Fig. 8. Response of the Kalman estimation to a 400ns-time step introduced intentionally in raw observational data.

To examine the effect of the time constant of the filter, a time step of 400ns was intentionally introduced in the raw data $z_1(t)$. The Kalman estimation has no information about this intentional step in its model and so follows the observations as they arrive. The result is shown in Fig. 8. The filter keeps a memory of past data and responds according to the function $[1-\exp(-t/T_c)]$, where T_c is the time constant of the filter. The analysis of the response curve in Fig. 8 gives T_c of order 9 hours.

The mean value of the Kalman gain k is of order 0,03 which gives, according to (4.7), a value for T of order 8h40m. The estimated values of T_c and T are then in complete agreement.

5. CONCLUSIONS

The precision of the real-time restitution of GPS time from the occasional observation of one single Block II satellite is degraded at the level of several hundreds of nanoseconds because of SA. This effect can be reduced if many simultaneous observations of Block II satellites are continuously performed through the use of a multi-channel GPS time receiver.

A simple average of simultaneous observations of about 10 satellites, integrated over a sufficient long time, can partly smooth out the white phase noise brought about by SA. Nevertheless this method can be used only if the statistical features of SA noise are known. It supposes the analysis of short-term data with checks to ensure that the dominant noise is white phase noise, and detection of any long-term bias. This can delay the restitution of GPS time by several hours.

If the above conditions are not entirely fullfilled, an efficient solution is to couple the real-time observations with knowledge of the past through the application of Kalman filtering. To implement this technique, it is necessary to suppose a linear evolution of the time difference between the local reference clock and GPS time, and to provide realistic values to describe the systematic changes and random noise level involved in the estimation process. The Kalman filter is then able to determine how a new observation should influence the updated estimation.

Acknowledgements

The author is grateful to her colleagues from the Time Section of the BIPM, especially to J. Azoubib, W. Lewandowski and G. Petit, for continuous discussions.

The experimental part of this work was done at the BIPM thanks to the loan from USNO (Washington DC, USA) of a commercial caesium clock. The staff of the Time Section of the BIPM is grateful to USNO for its generosity.

References

1. Lewandowski W., Thomas C., GPS Time Transfer, *IEEE Special Issue on Time*, 1991, **79**, 991-1000.

2. *NATO Standardization Agreement (STANAG) 4294*, Arinc Research Corporation, 2551 Riva Road, Annapolis, MD 21401, USA, Publication 3659-01-01-4296, 1990.

3. Series 4 of USNO.

4. Allan D.W., Weiss M.A., Accurate Time and Frequency Transfer during Common-View of a GPS satellite, *Proc. 34th Annual Symposium on Frequency Control*, 1980, 334-346.

5. Allan D.W., Granveaud M., Klepczynski W.J., Lewandowski W., GPS Time Transfer with Implementation of Selective Availability, *Proc. 22nd Annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting*, 1990, 145-156.

6. Lewandowski W., Petit G., Thomas C., Accuracy of GPS Time Transfer Verified by the Closure around the World, *Proc. 23nd Annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting*, 1991, 331-339.

7. Thomas C., Real-Time Restitution of GPS time through a Kalman Estimation, *Metrologia*, 1993, accepted.

8. Lewandowski W., High Accuracy Ground-Antenna Coordinates for GPS Time Transfer, *Proc. IAG Symposium G2-Permanent Satellite Tracking Networks for Geodesy and Geodynamics*, 1991, in press.

9. Lewandowski W., Petit G., Thomas C., The Use of Precise Ephemerides, Ionospheric Data and Corrected Antenna Coordinates in a long-distance GPS Time Transfer, *Proc. 22nd Annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting*, 1990, 547-558.

10. Lewandowski W., Petit G., Thomas C., Precise GPS Ephemerides from DMA and NGS Tested by Time Transfer, *Proc. 23nd Annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting*, 1991, 59-70.

11. Lewandowski W., Tourde R., Sensitivity to External Temperature of some GPS Time Receivers, *Proc. 22nd Annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting, 1990*, 307-316.

12. Vondrak J., A Contribution to the Problem of Smoothing Observational Data, *Bull. Astron. Inst. Czechoslovakia*, 1969, 20, 349-355.

GLONASS COMMON-VIEW TIME TRANSFER

W. Lewandowski, P. Moussay
Bureau International des Poids et Mesures
Pavillon de Breteuil
92312 Sèvres Cedex, France

G. T. Cherenkov, N. B. Koshelyaevsky, S. B. Pushkin
Institute of Metrology for Time and Space
Russian National Time & Frequency Service
Mendeleevo, Moscow region, 141570, Russia

ABSTRACT

One challenge facing the international time metrology community is use of the Russian Global Navigation Satellite System, GLONASS, for the common-view time transfer. This paper describes a possible organization for the GLONASS international common-view schedule and data collecting and treatment.

Keywords: GLONASS, GPS, common-view, track, schedule.

1. INTRODUCTION

The Russian Global Navigation Satellite System, GLONASS, is used on an experimental basis by only a few laboratories (Refs. 1,2), although its space segment is at about the same level of development as that of the GPS. This is because commercial automatic GLONASS time receivers are not available. This paper shows that techniques for data collection and treatment developed for GPS common-view time transfer may be used without change for GLONASS common-view time transfer.

GLONASS common-view time transfer may be illustrated by an experiment set up jointly by the Bureau International des Poids et Mesures (BIPM) and the Institute of Metrology for Time and Space (VNIIFTRI). At each site, at Sèvres in France and at Mendeleevo in Russia, GLONASS receivers of A-724 type are connected to the local atomic clocks. At the same time, at both sites, GPS receivers of NIST type are connected to the same clocks. The GLONASS and GPS receivers are differentially calibrated. As GPS common-view time transfer between Sèvres and Mendeleevo is accomplished at a level of a few nanoseconds in precision and accuracy, it gives an excellent reference with which to evaluate the performance of GLONASS common-view time transfer.

2. GLONASS CONSTELLATION - SOME CHARACTERISTICS

The Russian Global Navigation Satellite System, GLONASS, was inaugurated in 1982. The full constellation of twenty four satellites, according to latest information, should be completed by the end of 1995. In March 1993, fifteen satellites were operational.

GLONASS satellites, equipped with caesium clocks, are located on near-circular orbits with periods of about 11 h 15 min at a height of 19100 km and an inclination of 64,8°. Three orbital plans are separated by 120°. Within each plane, satellites are separated by 45° and between the planes there is a phase displacement of ±30° (Ref. 3). One feature of this configuration is of utmost interest for the organization of GLONASS common-view time transfer: satellites of the same plane separated by 45° appear in precisely the same direction on the successive days with a daily advance of 4,07 min.

3. POSSIBLE ORGANIZATION OF GLONASS COMMON-VIEW TIME TRANSFER

It follows that with a full constellation of twenty four satellites a GLONASS common-view tracking schedule could be arranged to follow precisely the same pattern as is used in the GPS common-view tracking schedule. Only one change is required: with the GPS schedule one observes the same satellite in the same direction on the successive days, less 4 min, with the GLONASS schedule a different satellite. As in satellite common-view time transfer satellite clock errors vanish in the calculation of time differences (Fig. 1), daily observations of a different GLONASS satellite in the same direction do not affect the principal conditions of observation which are atmospheric refraction and errors in antenna coordinates.

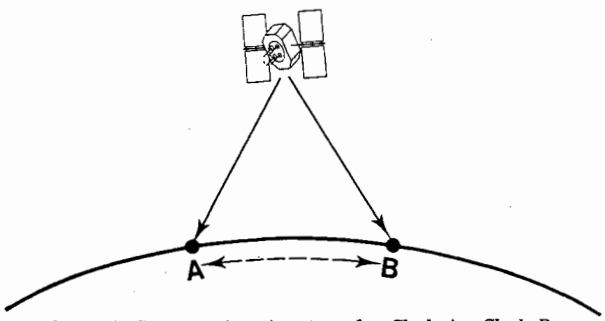


FIGURE 1. *Common-view time transfer: Clock A - Clock B = [Clock A - satellite time] - [Clock B - satellite time].*

A separate issue is the length of the GLONASS track and the treatment of short-period data. The choice of a 13 min track for the GPS common-view technique was determined by the 12,5 min duration of the navigational message. As the GLONASS navigational message lasts 2,5 min the choice of a 3 min track appears appropriate. If the same arrangements were made for observation of single tracks as are now made for GPS observations, this could accelerate implementation of the GLONASS common-view technique. GLONASS common-view tracks collected in this way could be treated by the techniques already developed for GPS.

The main obstacle preventing practical implementation of this scheme is the lack of commercial automatic GLONASS time receivers. During the experiment on GLONASS common-view time transfer described here, suggested solutions could not be fully applied as the receivers were operated manually.

4. AN EXPERIMENT OF GLONASS COMMON-VIEW TIME TRANSFER

An experiment on GLONASS common-view time transfer, based on two Russian navigation receivers was carried out in the period from April 2 to June 23, 1992. To evaluate the quality of the GLONASS common-view time transfer, an independent GPS common-view link was organized at the same time.

4.1. The set up of the experiment

The two remote atomic clocks used in the experiment are separated by about 2800 km. On each site a GLONASS receiver and a GPS receiver are connected to a single clock, an Hydrogen-maser at Mendeleevo and a caesium standard at Sèvres (Fig. 2). The C/A Code receivers for each satellite system are of the same type, AOA TTR6 at Mendeleevo and AOA TTR5 at Sèvres, for GPS, and A-724 at both locations, for GLONASS.

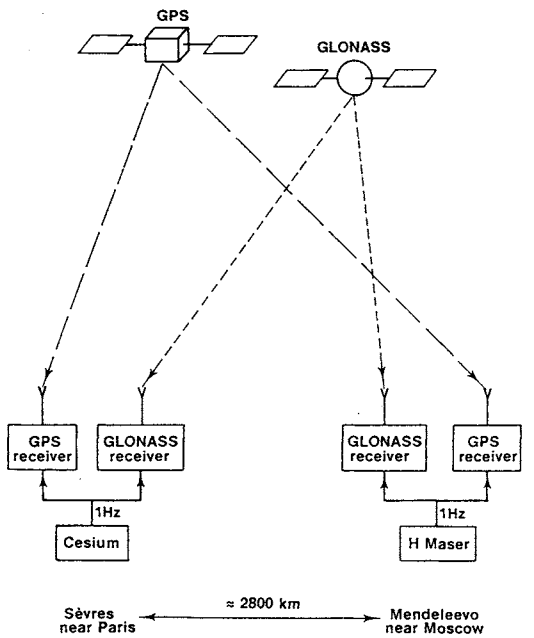


FIGURE 2. *The experiment configuration.*

In common-view time transfer, two remote stations receive signals from the same satellite at the same time and exchange the data to compare their clocks (Ref. 4)). Over distances of up to a few thousands of kilometres the impact of errors, such as poor estimation of ionospheric delay or broadcast ephemerides, is diminished during common-view time transfer. By using the same type of receiver, as in our case, the consistency of the comparison is further improved as possible software errors are removed by differencing.

4.2. GLONASS common-view link

The need to operate the A-724 navigation receiver manually strongly influences the organization of the GLONASS common-view link. This receiver has no interface to allow the connection of an external micro-computer so two operators, one at Mendeleevo and one at Sèvres, have to read the receiver screens simultaneously and write the observations on paper. This limits the length of track to 3 minutes and the number of common-view observations to about 5 per day. In addition, as the GLONASS constellation is not yet complete, the schedule applied here does not follow pattern described in paragraph 3.

Because the receivers were operated manually, the short-term GLONASS observations also do not follow the GPS scheme. These observations were taken every 15 seconds. A typical GLONASS track is given in Figure 3. A linear fit of the short-term data was used in post-processing to deduce the time difference between the satellite and laboratory clocks for the middle of the 3 min track. The slopes of the linear fit do not usually exceed a few tens of picoseconds per seconds, and the corresponding standard deviations are of order a few nanoseconds. These standard deviations are particularly small and can possibly be explained by the fact that the observations displayed on the screen are not raw values, but come from a smoothing done by the receiver. No information is available on the nature of this smoothing.

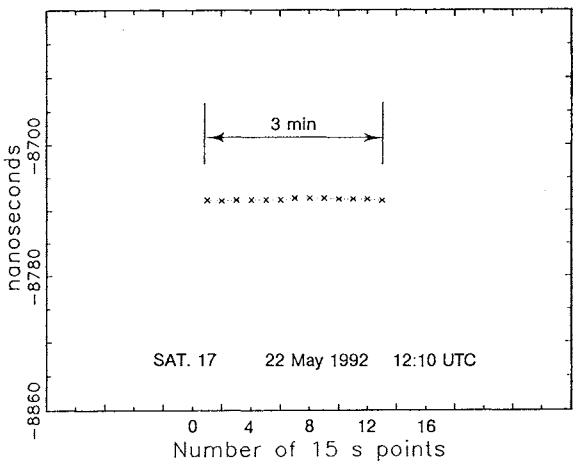


FIGURE 3. *A typical GLONASS track. Short-term GLONASS observations [Local Cs Clock - GLONASS time] taken manually every 15 seconds at the BIPM, for a 3-minute track of Sat. 17 on 22 May 1992.*

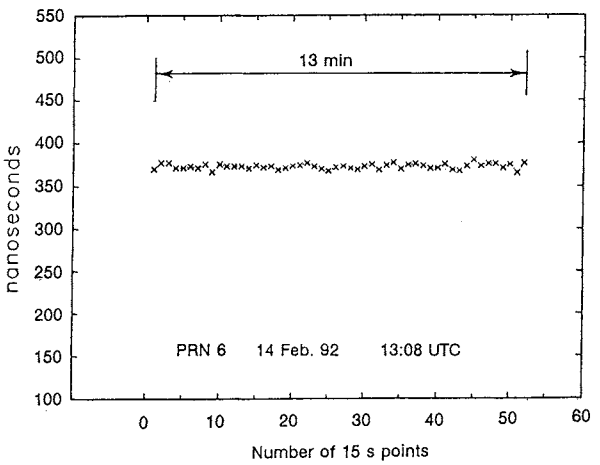


FIGURE 4. *A typical GPS track of a Block I satellite. Short-term GPS data [Local Cs Clock - GPS time] taken every 15 seconds at the BIPM, for a 13-minute track of PRN 6 on 14 February 1992.*

The A-724 GLONASS navigation receiver does not correct its observations for tropospheric delay. The ionospheric delay is estimated by a fixed model which does not depend on external parameters.

The coordinates of the GLONASS antennas are expressed in the SGS 85 reference frame with an estimated uncertainty of 5 m. They were obtained at each site by averaging a series of navigation solutions. The uncertainties of GLONASS ground-antenna coordinates can have an impact on the accuracy of the common-view link of a few tens of nanoseconds.

The GLONASS receivers were compared side-by-side for several days before one was shipped to the BIPM. Their differential delays are thought to be known to within a few nanoseconds.

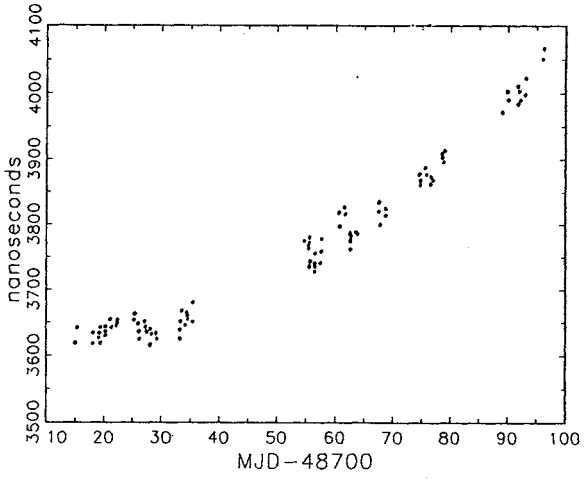


FIGURE 5. *[BIPM Cs Clock - VNIIFTRI H-maser] by GLONASS common-view.*

The GLONASS common views are presented in Figure 5. The numerous gaps in the data, due to interruption of observations during weekends and vacations, preclude any sophisticated statistical analysis and smoothing.

4.3. GPS common-view link

The two GPS receivers used are fully automatic units especially designed for common-view time transfer. A typical GPS track is shown by Figure 4. The receiver processes short-term measurements, smoothing them over a period of 15 s through use of a second degree fit (Refs. 5,6). A linear fit of the 15-second points is used to deduce the time difference between the satellite and laboratory clocks for the mid-point of the 13 min track. The slope of the linear fit for Block I satellites is usually no bigger than several picoseconds per seconds and the corresponding standard deviation ranges from 3 ns to 20 ns depending on the multipath effects.

The GPS time receivers, unlike these for GLONASS, correct observations for tropospheric delay on the basis of a model which is a function of the elevation of the GPS satellite and of the altitude of the observation site. Ionospheric delay is estimated from a model using broadcast parameters.

The coordinates of the GPS antennas are expressed in the ITRF88 reference frame and were provided by the BIPM method of differential positioning (Refs. 7,8) with respect to the closest ITRF sites. The estimated uncertainty of coordinates at Mendeleevo, obtained through a link with the ITRF laser site in Graz (Austria), is 70 cm, and that at Sèvres is 30 cm (link with the ITRF laser site in Grasse, France). The uncertainties of the GPS ground antenna coordinates can have an impact on the accuracy of this GPS common-view link of up to a few nanoseconds.

The GPS receivers were differentially calibrated with an uncertainty of 2 ns.

The GPS common-view time transfer was realised from the about 35 tracks of Block I and Block II satellites available daily according to the BIPM international tracking schedule No. 18 (Fig. 6). During this experiment the Block II satellites were subject to Selective Availability, so that strict common views were required (Ref. 9). A Vondrak smoothing (Ref. 10), which acts as a low-pass filter with a cut-off period of about 0,5 day, was performed on the raw common-view values. For this experiment, the smoothed values were interpolated for the times of occurrence of GLONASS common views. The precision of this GPS link, estimated from the residuals of the smoothed values, is about 4 ns.

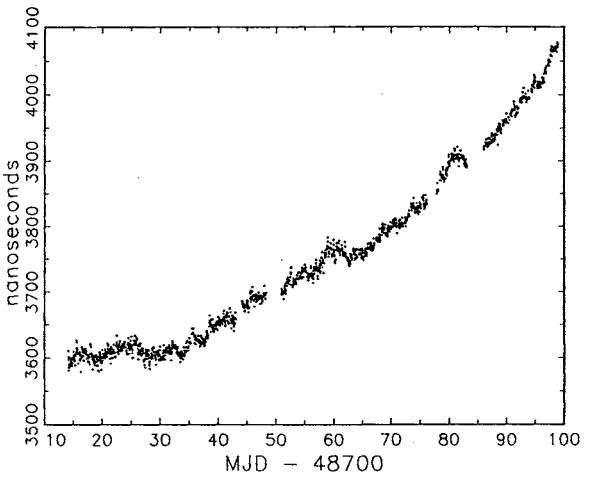


FIGURE 6. *[BIPM Cs Clock - VNIIFTRI H-maser] by GPS common-view.*

4.4. Comparison of GLONASS and GPS time transfers

As the GPS common-view link between Mendeleevo and Sèvres is realised with a precision and accuracy of a few nanoseconds it serves as an excellent reference for estimation of the performance of the GLONASS common-view link. GLONASS raw common-view values are compared to the GPS smoothed and interpolated values in Figure 7. Note that the GPS and GLONASS results differ by a fairly constant bias with peak-to-peak discrepancy of about 40 ns. The mean of these differences over the duration of the experiment is 32 ns. The root mean square of the residuals to the mean, which is taken as an estimation of the confidence of the mean, is 13 ns.

The bias of 32 ns between the GLONASS common views and the GPS common-views is partially due to an approximate calibration of the GLONASS equipment and partially to the large error in GLONASS ground antenna coordinates. The noise affecting the GLONASS common views is also partially due to coordinate error, to the absence of a tropospheric correction and to an imprecise estimate of ionospheric correction. Table I shows a comparison of typical GLONASS raw data with GPS smoothed data. In the last column the 32 ns bias has been removed. If we compare the last column of Table I with the differences between GPS raw data and GPS smoothed data, given in Table II, we find that the noise affecting raw GLONASS common views is not much greater than that affecting raw GPS common views.

To illustrate this, consider Figures 8 and 9 which present, respectively, comparisons of GPS smoothed common views with GPS raw common views, and GPS smoothed common views with GPS common views affected by an, artificially introduced, coordinate error of 14 m. A constant bias and noise similar to that affecting raw GLONASS common views can be seen.

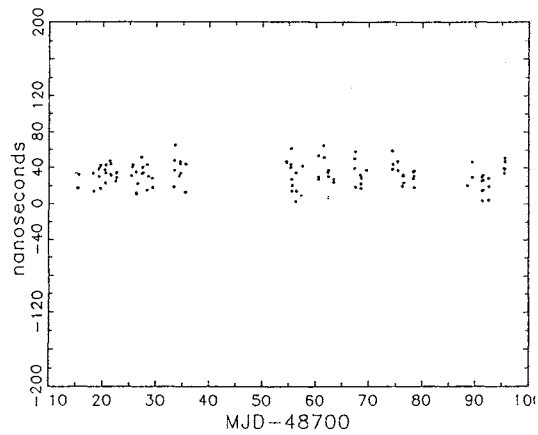


FIGURE 7. *[BIPM Cs Clock - VNIIFTRI H-maser] by raw GLONASS minus [BIPM Cs Clock - VNIIFTRI H-maser] by smoothed GPS.*

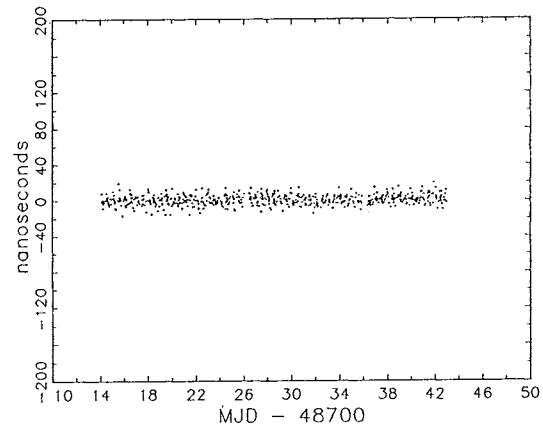


FIGURE 8. *[BIPM Cs Clock - VNIIFTRI H-maser] by raw GPS minus [BIPM Cs Clock - VNIIFTRI H-maser] by smoothed GPS.*

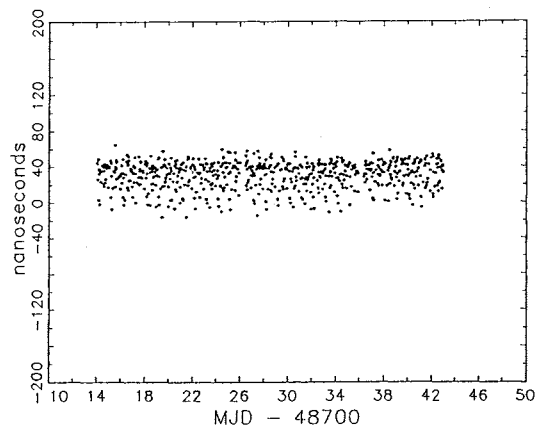


FIGURE 9. *[BIPM Cs Clock - VNIIFTRI H-maser] by raw GPS affected by artificially introduced 14 m antenna coordinate error minus [BIPM Cs Clock - VNIIFTRI H-maser] by smoothed GPS.*

Date April 1992	Start Time UTC h	m	SV GLO	GLO -GPS smoothed ns	GLO -GPS smoothed -32ns ns
3	7	40	7	16.8	-15.2
3	9	40	19	17.0	-15.0
3	14	20	22	31.8	0.2
6	6	48	2	33.5	1.5
6	7	40	3	13.6	-18.4
7	6	40	3	37.9	5.9
7	7	10	3	29.7	-2.3
7	11	10	17	40.2	8.2
7	13	15	15	16.6	-15.4
7	14	10	19	42.3	10.3
8	9	40	17	37.2	5.2
8	11	10	17	22.7	-9.3
8	13	10	19	33.6	1.6
8	14	10	20	42.7	10.7
9	6	40	5	46.8	14.8
9	11	10	19	43.8	11.8
9	12	10	19	31.9	-0.1
9	14	10	20	30.6	-1.4
10	9	40	19	24.5	-7.5
10	13	40	22	33.8	1.8
10	14	10	22	28.5	-3.5
13	9	40	22	30.6	-1.4

TABLE I. *A sample of [BIPM Cs Clock - VNIIFTRI H-maser] by raw GLONASS minus [BIPM Cs Clock - VNIIFTRI H-maser] by smoothed GPS.*

Date April 1992	Start Time UTC h	m	PRN	GPS raw - GPS smoothed ns
11	15	36	12	-2.8
11	16	4	3	-4.3
11	17	24	3	-2.2
11	19	32	3	0.1
11	20	4	23	-2.3
11	21	56	11	3.6
11	22	28	23	2.4
12	1	56	14	-0.1
12	3	16	15	-3.9
12	3	48	13	-1.9
12	4	36	14	-1.5
12	4	52	18	-1.9
12	5	24	24	-5.9
12	7	32	6	-7.2
12	9	24	6	5.5
12	11	0	12	-9.3
12	11	16	2	7.4
12	12	4	13	14.0
12	12	20	12	-1.7
12	13	40	12	3.2
12	13	56	13	-0.7
12	14	28	20	-0.9
12	15	0	24	6.1
12	15	32	12	0.9
12	16	0	3	-9.3

TABLE II. *A sample of [BIPM Cs Clock - VNIIFTRI H maser] by raw GPS minus [BIPM Cs Clock - VNIIFTRI H maser] by smoothed GPS.*

5. CONCLUSION

The GLONASS common-view time transfer could be organized following the pattern developed for GPS common-view time transfer. This would allow the use of the techniques of data treatment already developed for GPS.

The results of the experiment reported here demonstrate that, even with GLONASS navigation receivers not designed specifically for timing purposes, the GLONASS common-view time transfer can be realised with precision close to that found in the early stages of GPS common-view time comparisons. More accurate determinations of GLONASS antenna coordinates in the SGS 85 reference frame would significantly improve GLONASS common-view time transfer.

Manual operation of the GLONASS receivers was possible only during short period of this experiment. The development of automatic GLONASS receivers dedicated especially for time transfer is an urgent challenge.

Acknowledgements

The authors are pleased to express their gratitude to Prof. P. Daly of the University of Leeds, for his encouragement and advice.

The experimental part of this work was done at the BIPM thanks to the loan from USNO (Washington DC, USA) of a commercial caesium clock. The staff of the Time Section of the BIPM is grateful to USNO for its generosity.

References

1. P. Daly, G. T. Cherenkov, N. B. Koshelyaevsky, S. B. Pushkin, Satellite time transfer between UTC(USNO) and UTC(SU) using Navstar GPS and GLONASS, *Proc. 4th Institute of Navigation Technical Meeting*, pp. 199-206, September 1991.

2. P. Daly, N.B. Koshelyaevsky, W. Lewandowski, G. Petit and C. Thomas, Comparison of GLONASS and GPS time transfers between two West European time laboratories and VNIIFTRI, *Proc. 23rd PTTI meeting*, pp. 341-350, 1991.

3. P. Daly, Aspects of the Soviet Union's GLONASS Satellite Navigation System, Journal of Navigation, Vol. 41, No. 2, pp. 186-197, May 1988.

4. D.W. Allan and M.A. Weiss, Accurate time and frequency transfer during common-view of a GPS satellite, *Proc. 34th Ann. Symp. on Frequency Control*, pp. 334-346, May 1980.

5. W. Lewandowski, C. Thomas, GPS time transfer, *Proc. of the IEEE, Special Issue on Time and Frequency*, pp. 991-1000, July 1991.

6. W. Lewandowski, G. Petit, C. Thomas, The Need for GPS Standardization, *Proc. 23rd PTTI meeting*, pp. 1-9, 1991.

7. B. Guinot, W. Lewandowski, Improvement of the GPS Time Comparisons by Simultaneous Relative Positioning of the Receiver Antennas, *Bulletin Géodésique*, 63, pp. 371-386, 1989.

8. W. Lewandowski, High Accuracy Ground-Antenna Coordinates for GPS Time Transfer, in *Proc. IAG Symposium G2-Permanent Satellite Tracking Networks for Geodesy and Geodynamics*, Vienna, August 1991, in press.

9. D.W. Allan, M. Granveaud, W.J. Klepczynski and W. Lewandowski, GPS time transfer with implementation of Selective Availability, *Proc. 22nd PTTI meeting*, pp. 145-156, 1990.

10. J. Vondrak, A Contribution to the Problem of Smoothing Observational Data, *Bull. Astron. Inst. Czechoslovakia*, 20, pp. 349-355, 1969.

QUARTZ CRYSTALS 2

Chairman: B. Neubig

PHYSICAL LIMITATION ON THE QUALITY FACTOR OF QUARTZ RESONATORS. *QUARTZ OSCILLATOR AT 4.2 K*

A. El Habti, A. Zarka* ,F. Lardet-Vieudrin and F. Bastien

Laboratoire de Physique et Métrologie des Oscillateurs du CNRS
associé à l'Université de Franche-Comté - Besançon
32, Avenue de l'Observatoire 25044 Besançon cedex FRANCE

* Laboratoire de Minéralogie-Cristallographie
Université PARIS VI et PARIS VII
Tour 16 - 4, Place Jussieu - 75252 Paris cedex 05 FRANCE

Abstract

The measurement of the resonator quality factor as a function of temperature for different overtones gives a method to distinguish between different causes of acoustic wave damping. Several quartz resonators with different surface finishing have been measured and compared. Futhermore the different causes of damping are theoretically analyzed. The conclusions derived both from a theoretical analysis and from the measurements. Therefore at T = 4.2 K for good resonators three residual causes can be considered. It seems possible to reach, using a careful chemico-mechanical polishing, a Q value near to the intrinsic crystal quartz limit at 4.2 K that is to say about 100 million.
In order to use such large quality coefficient it is necessary to minimize the temperature sensitivity of resonator. For this purpose, we have realized an oscillator working at very low temperature. Below 15 K , the frequency variation as a function of the temperature is linear and small (about 0.2ppm/K). Furthermore this variation is weakly depending on the particular cut under consideration at least for one axis rotation cut with θ between 33° and 45°.

KEY-WORDS : Quartz Resonator ; Quartz Oscillator ; Very low Temperature ; Wave Attenuation.

1 Introduction

First, this work has been done to gain a better understanding of the limitations of quartz resonators. Quartz resonators make use of the energy transfer between elastic energy and mechanical energy. The acoustic wave attenuation is an important parameter for the resonator. Therefore the study of Q as a function of temperature may provide some insight on the acoustic attenuation mechanism in quartz. The acoustic attenuation, for frequencies in the 10 MHz range is very weak. The direct measurement of absorption in few millimeter of quartz is not possible but with a resonator the attenuation is measured for millions of oscillations. The determination of Q for T from 1.5 K to 10 K allows us to distinguish between the effect of acoustic wave absorption by the quartz crystal itself and other attenuation causes. Futhermore, the use of different overtones provides a means for modifying the energy trapping, therefore the mechanical damping due to mounting structures can be reduced.
Point defects, density of dislocations and the presence of scratches under the electrodes are evaluated by x-ray examination.
In a second part, we present an oscillator working from room temperature to 1.4 K. Using this oscillator we study the sensitivity to the temperature for one axis rotation cut with θ between 33° and 45°.

2 - Resonator quality factor at low temperature

The quality factor Q has been measured in two ways : first by the evaluation of the peak width around the resonance, and second by the study of the resonator free oscillations. This last method, already described in [1], permits almost automatic measurement given Q as a function of temperature. In fact, this method gives only the loaded Q. This is not a problem for two reasons. First, considerations of loaded Q or unloaded Q lead, for our study, to the same conclusions ; and second, for high overtones, the motional resistance is high enough, even at low temperature, to give an unloaded Q very close to the loaded Q.
The experimental device has been already described [1]. It permits measurements to be made from room temperature down to 1.5 K. For the present discussion, we present results between 20 K and 1.5 K only.
The results obtained for higher temperatures, have already been discussed in a previous paper [1]. The $1/Q = f(T)$ curve shows a fast decrease as T decreases from 20 K and reaches a plateau with a level depending on the resonator (see fig. 1). If the plateau value is subtracted from the curve, we get a straight line for $\mathrm{Ln}(1/Q) = f(\mathrm{Ln}\ T)$ (see fig. 2). The slope of this line gives the relation $1/Q \propto T^4$. For very low temperatures the dependence on T^4 is that given by the Landau-Romer theory [2, 3] for acoustic wave absorption by phonon-phonon interaction in the crystal. This fact and other considerations discussed in a previous paper [1] show that the contribution proportional to T^4 is due to the intrinsic absorption of the crystal itself. Therefore for $1.5\ \mathrm{K} < T < 15\ \mathrm{K}$ we can take the relation $1/Q = \alpha + \beta T^4$.
We propose several possible origins for the constant term :
1 - mechanical damping due to mounting structures.
2 - wave diffraction on point defects.
3 - wave diffraction on dislocations.
4 - wave diffraction on scratches on the resonator surface.
5 - wave diffraction on the surface irregularities (polishing).
6 - losses due to stress gradients into the resonator.
7 - losses due to modified superficial layers (amorphous layers) or absorption in metallic electrodes.
For type 1 losses, the energy trapping provides a fast decrease of damping as the overtone order increases. We previously showed that for a diameter of 15 mm and a 7th overtone with a frequency f = 11.8 MHz we obtained negligible limitation losses for this phenomena at 4.2 K. The ultimate limitation, that is to say the attenuation due to a perfect quartz crystal, corresponds to Q = 100 million at 4.2 K. By comparison, we consider as a negligible limitation an attenuation which produces a contribution for 1/Q smaller than 10^{-8}.

The losses due to type 2, 3, 4 or 5 sources are linked to acoustic wave diffraction with different obstacles. In order to evaluate the relative importance of these causes we have performed computations and experiments. In paragraph 3 we give the computation of such losses.
The measurements are performed with resonators with different surface roughnesses. The surface roughness was evaluated by an optical interferometric method. In addition, the resonators were examined by x-radiation.

3 - Evaluation of magnitude for Q limiting factors below 10 K

As mentioned in the previous paragraphs, we now evaluate the magnitude of diffraction and the magnitude of the influence of a stress gradient.

3.1 - Acoustic wave diffraction via different obstacles

We have evaluated two cases.
a. We take a reflecting surface and we evaluate the diffraction at an infinite distance.
b. We study the near diffraction on spherical or cylindrical defects with an infinite rigidity.
None of the previous cases is rigorously the case of a real device for several reasons. In a real case, the diffraction is not at an infinite distance, the material is not isotropic, the waves are not pressure waves and the resonator is not a volume limited by two infinite planes. Nevertheless the following calculations are precise enough for a classification of the relative magnitude of the different causes of wave diffraction.

3.1.1 - Diffraction by a phase grating

The diffraction by a phase grating represents diffraction related to a reflection on the irregular surface of quartz. For the pressure wave approximation in an isotropic media, the displacement and stress continuity leads to a description of the diffraction according to Huygens' principle.
Using the theory of diffraction it is possible to calculate the intensity of zero order diffraction for a diffraction structure with characteristic dimensions smaller than λ (wavelength). As a simple model we choose a surface given a phase shift which present the following function of ζ (coordinates in the emitting plane)

$$exp\ \{ja\ sin\ (2\pi\frac{\zeta}{\zeta_o})\}$$

where ζ_o is the grating step and a is a constant corresponding to the phase shift maximum value.
In a first approximation see [4], we obtain for the ratio of the intensity of the diffracted wave to the intensity of the incident wave

$$1 - J_o^2\,(a).$$

Therefore, if Q_{dif} is the contribution of this kind of diffraction for Q, we have

$$Q_{dif} = 1/(1 - J_0^2\,(a))$$

For a small value of a = $2\pi e/\lambda$ (e being the surface defect characteristic dimension) we can approximate $J_o(a) \simeq 1\text{-}0.25\ a^2$

$$\text{Finally,} \qquad 1/Q \simeq 0.5\ a^2 = 2\pi^2 \left(\frac{e}{\lambda}\right)^2$$

With $\lambda = 400$ µm (f ≃ 10 MHz) we obtain

e(µm)	10^{-3}	$5\ 10^{-3}$	10^{-2}	$5\ 10^{-2}$
Q_{dif}	$8\ 10^9$	$3\ 10^8$	$8\ 10^7$	$3\ 10^6$

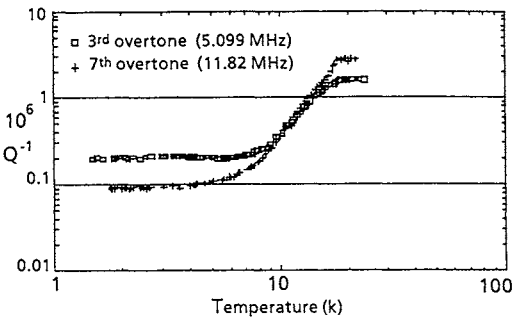


Fig. 1 : *1/Q as a function of temperature for a mechanically-polished quartz crystal; (SC cut, 15 mm diameter)*

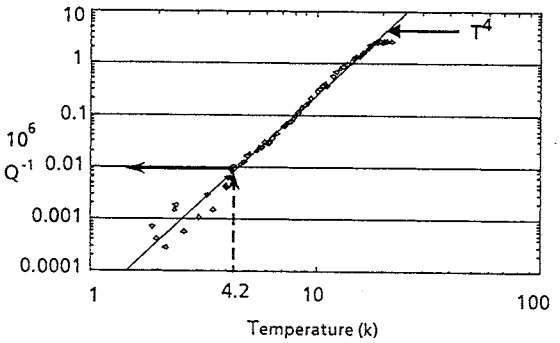


Fig. 2 : *7th overtone for the quartz used for Fig. 1*
◇ : *Value of the low-temperature plateau has been substracted from the experimental results.*
— : *T^4 variation*

We note that the ratio $e/\lambda = 2{,}5\times10^{-5}$ is much smaller than the equivalent ratio required to obtain specular reflexion on a surface. In fact, in our case we must consider a large number of reflexions. The previous calculation is obviously very approximate. Nevertheless, it shows that polishing is important ; but good optical polishing seems to be sufficient.

3.1.2 - Diffraction on a rigid cylinder

For a cylinder with infinite rigidity, the total diffracted power is given by Morse [5].

$$\Pi_c = \frac{6\pi^5\, a^4\, I}{\lambda^3} \qquad if\ \lambda \gg 2\pi a$$

I being the incident intensity.
We consider a cylinder with a radius a, and a length equal to the unit of length.
For the case where $\lambda >> 2\pi a$ and if a_c is the ratio of diffracted power to the incident power we obtain

$$a_c = \frac{\Pi_c}{2aI} = 3\pi^5 \left(\frac{a}{\lambda}\right)^3$$

For the 7th overtone we can set $Q = 7\pi/a_c$

$$\text{Therefore} \qquad Q = \frac{7}{3\pi^4}\left(\frac{\lambda}{a}\right)^3 \frac{1}{L}$$

For a cylindrical object like a scratch, and for λ = 400 µm we chose a = 10 µm and L = 10^{-2} m, we then obtain $Q = 1.6 \times 10^5$. The effect is large. Obviously, the present value is overestimated because a scratch cannot be taken to be a rigid cylinder. Nevertheless, this result underlines the importance of scratches.

For dislocations, let us considere a dislocation density of 10^3 cm^{-2} for a resonator with a surface of about 1 cm^2 and a thickness of about 1 mm. With L=1 m and with a=100 Å we obtain Q=1.6 10^{12}.
The conclusion of this evaluation shows that it is very important to avoid scratches, and to obtain a good optical polish. Moreover, it seems that the effect of dislocations is negligible.

3.2 - Q modification due to thermal stress

We first evaluate the order of a magnitude of stress due to a difference in thermal expansion between quartz and the metal electrode. We consider a slice of quartz with two metallic layers [6].
A_i : cross sectional area of the slice
σ_i : stress
E_i : Young's modulus
h_i : thickness
α_i : coefficient of linear expansion
L : length of slice
i=Q for quartz and i=M for metal
T : temperature increase
By applying the following rule that the extension of the quartz+the compression of the metal = the difference in "free" lengths, we obtain the following equation :

$$\frac{\sigma_Q L}{E_Q} + \frac{\sigma_M L}{E_M} = \left(\alpha_M - \alpha_Q \right) L T$$

By applying Newton's law of equal action and reaction

we obtain $\sigma_Q A_Q = \sigma_M A_M$

with $A_M \ll A_Q$ and $\frac{h_Q}{h_M} = \frac{A_Q}{A_M}$

finally we get $\sigma_Q \simeq \frac{h_M}{h_Q} E_M (\alpha_M - \alpha_Q) T$

With $(\alpha_M - \alpha_Q) T \simeq 3.10^{-3}$, $E_M = 8.10^{10}$ N/m^2, $\frac{h_M}{h_Q} \simeq 3.10^{-4}$

we obtains $\sigma_Q \simeq 70$ KPa
Let us assume that this variation of stress induces a variation of $\Delta\rho$ in the specific mass ρ of quartz, and a variation of ΔV for the sound velocity V.
According to the value of the reflection coefficient a on the boundary of two layers having ρ_1 and ρ_0 for their specific mass respectively, and V_1 and V_0 as the respective sound velocities, we obtain,

$$a = \left[\frac{\rho_1 V_1 - \rho_o V_o}{\rho_1 V_1 + \rho_o V_o} \right]^2$$

for our case,

$$a \simeq \frac{1}{4} \left[\frac{\Delta\rho}{\rho} + \frac{\Delta V}{V} \right]^2$$

Therefore we can assume that $a \leq \left(\frac{\Delta f}{f} \right)^2$

$\Delta f/f$ is the variation of frequency f when the quartz is submitted to a stress.
In order to obtain an order of magnitude for this variation of frequency we have submitted a resonator to static pressure during a frequency measurement. The result is a shift of about 0.2 Hz/KPa. Using this value with a stress of 70 KPa, and with the relation $Q = n\,\pi/a$ we obtain for 7th overtone ; Q = 1.1.10^{13}.
With Δf = 14 Hz for 70 KPa
This result value with the large overestimation of attenuation shows that the effect of the thermal gradient remains rather unimportant in all cases.

3.3 - Evaluation of the effect of a non crystalline surface layer

The intrinsic attenuation coefficient in quartz at 4.2K can be evaluated from the maximum value of Q at 4.2K (i.e. 100 millions). The value obtained α_Q = 10^{-4} m^{-1} is much lower than the attenuation coefficient α_A = 20m^{-1} in amorphous quartz at 4K [7]. Therefore as long as the absorption is considered, one millimeter of crystalline quartz corresponds to a thickness of $h = \alpha_Q/\alpha_A$ m of amorphous quartz. As a result we find that a layer of 50Å of amorphous quartz begins to play a role for the acoustic wave attenuation. This thickness can reasonably be considered to be caused by mechanical polishing.
For example, J.R Vig et al [8] find, using high energy electron diffraction (80 Kev), a disturbed layer of about 40 Å.

3.4 - Possible role of acoustic attenuation in the metal of the electrodes

It is well known that wave attenuation in a metal does not vanish as the temperature decreases toward 0 K due to the presence of free electrons. According to the literature [9] the attenuation in silver at 4.2 K for a frequency of 150 MHz is of the order of 5 db/cm. This value corresponds to an attenuation which is much greater in metal than in quartz.
The exact determination of the attenuation requires taking into account the zero stress on the outside face. This fact reduces the attenuation inside the thin metallic layer. We have not performed this computation, but in our case the effect of the metal cannot be distinguished from the effect of absorption due to the disturbed surface layer of quartz. Therefore both effects can be considered as surface layer absorption. The separation of different effects requires the use of resonators with electrodes not directly deposited on the quartz.

4 - Experiments on Q limiting factors for resonators below 10 K

The previous discussion shows that the most probable causes for the limitations of quartz resonators at low temperatures are wave diffraction on the surface irregularities (cause 5) and losses due to modified surface layers on the quartz (cause 7) if the attenuation due to the mounting structure has been eliminated.
In order to verify this hypothesis, we have performed several experiments. We obtain the quality factor Q as a function of temperature for several overtones for different resonators. Specifically, we investigated the region 1.5K<T<15K in order to obtain results on the origin of Q limitations at low temperatures.
The most probable cause of these limitations being the surface quality, we have performed several tests on resonators with different surface finishing. The surface was studied by an optical interferometric method.
Futhermore, the general quality of the resonator has been studied by x-ray diffraction.
Surface quality has been classified into 3 categories :
1 - good mechanical polishing (referred to as M) obtained using 0.2 µm grain abrasive ;
2 - chemical polishing (referred to as C) obtained by ammonium bifluoride etching of the quartz polished as in 1 ;
3 - mechanical polishing (referred to as D) obtained using 2 µm grain abrasive.
The x-rays examination shows that the different types of quartz used are either good or very good. The dislocation density is of the order of 10^2 cm^{-2}. The existing scratches are also visible on the X-rays photographs.

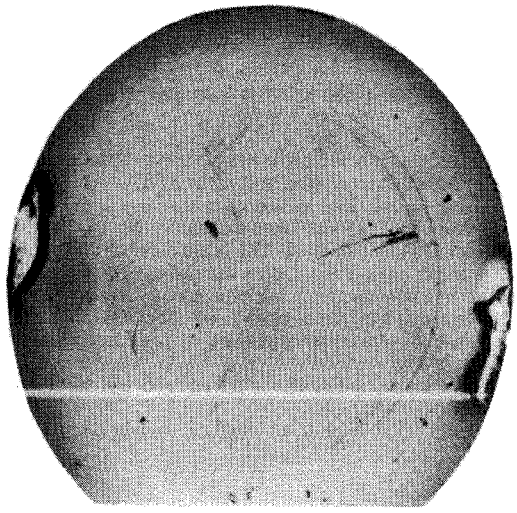


Fig. 3: *X-ray image of quartz "D2", mechanical polishing ; [diameter : 15 mm]*
Quality factor for the 3rd overtone at 4.2K: $Q_3 = 7.10^6$
Quality factor for the 7th overtone at 4.2K : $Q_7 = 14\ 10^6$
Low dislocations density = 10^2 cm^{-2}

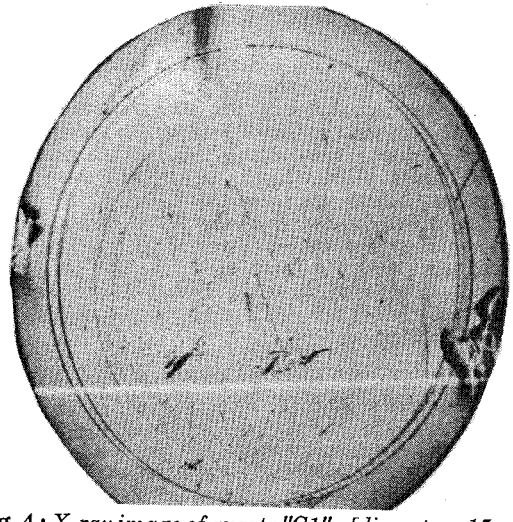


Fig. 4 : *X-ray image of quartz "C1" ; [diameter : 15 mm]*
$Q_3 = 3.6\ 10^6$; $Q_7 = 3.4\ 10^6$;
Low dislocation density ≃ 10^2 cm^{-2}

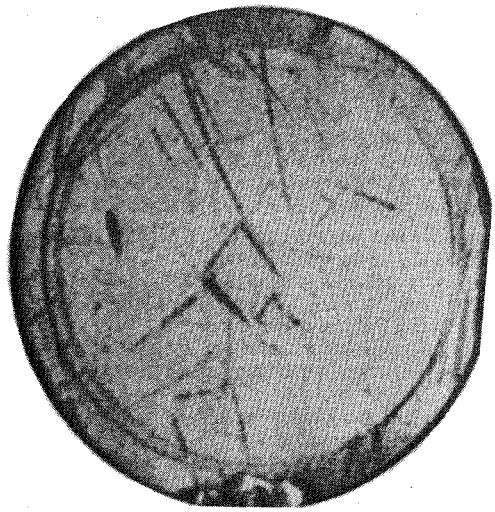


Fig. 5 : *X-ray image of quartz "D4"; [diameter : 15 mm]*
$Q_3 = 3.10^6$; $Q_7 = 1.6\ 10^6$;Scratches are visible

The figure shows the case of a very good quartz resonator (fig. 3), a good quartz resonator (fig. 4), and a quartz resonators with scratches (fig. 5).
The quality coefficient difference between quartz resonators C1 and D2 does not correlate with a difference in dislocation density.
The poor quality coefficient of the resonators D4 does correlate with the existence of scratches.
In order to show the different results, we have used a synoptic diagram (fig. 6) on which we have a point for each resonator in the following coordinates. First, Q at the liquid helium temperature (4.2 K) for the 3rd overtone (f=5.1 MHz), and second, Q at the same temperature for the 7th overtone (f=11.80 MHz). The diagram also gives, for each point, the surface finish category (M, C or D).
On this diagram we can distinguished two zones. Zone 1 corresponds to the best resonators. All the quartz resonators are in category M. Zone 2 corresponds to poor quality oscillators. In this category all the quartz resonators but one are from category C or D. The only quartz resonator from category M shows a large number of scratches.
We note that the chemically-polished quartz resonators have been submitted to extensive large etching. This etching induces rather large surface irregularities. The previous results show the importance of polishing, but the dipersion results indicate that the limiting factor is not only the surface polishing. The most probable contribution, according to previous evaluations, is the absorption due to the surface layer, especially the layer disturbed by mechanical polishing.
Another possible limitation may be the contribution of damping of the structure. This cause decreases rapidly as the overtones increase because the energy is trapped more efficiently. Therefore it is useful to choose high overtones for low temperature experiments.

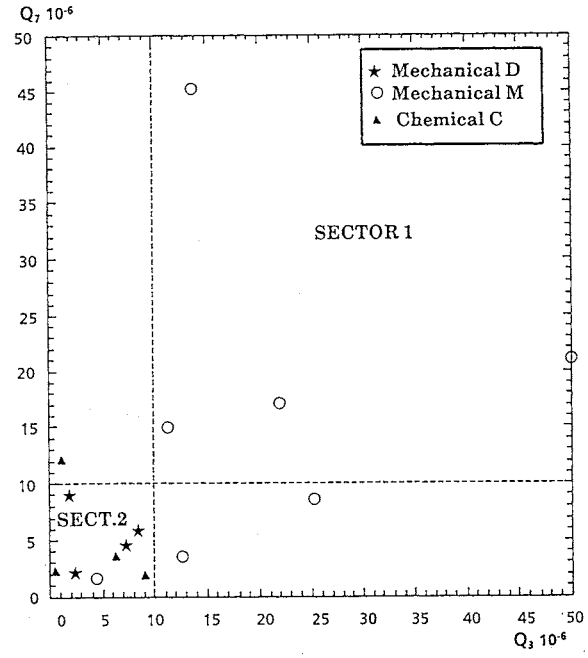


Fig. 6 : *Q_3 = quality factor for the 3rd overtone at 4.2 K; Q_7 = quality factor for the 7th overtone at 4.2 K*
Each point corresponds to an resonator.
Sector 1 : best resonators region
Sector 2 : poor quality resonators region

5 - Very low temperature oscillators

The realization of resonators with a very high quality coefficient Q at low temperature allows us to reduce $1/f$ noise because this kind of noise decreases rapidly as Q increases. (see ref. [10]). Nevertheless, the noise due to temperature fluctuation remains. Consequently the carrying out of an oscillator working at low temperature is not only interesting in itself but also allows us to study the problem of **temperature compensated** cuts at 4.2 K. The quartz oscillator is built with a field effect transistor with low temperature characteristic not provided by the manufacturer. Therefore the study of transistor BF 966 characteristics at 77 K and 4.2 K allows us to realize an oscillator working at least from 1.4 K to room temperature except around 55 K if the resonator is built with a quartz crystal containing Na+ impurities [fig. 7].
We have measured the oscillator frequency f as a function of temperature. The experimental values are compared with the result given by a computer programme developed by J. J. Gagnepain and G. Theobald.
The relative variation of frequency is given by

$$\frac{\Delta f}{f} = \frac{1}{2}\frac{\Delta C}{C} - \frac{1}{2}\frac{\Delta \rho}{\rho} - \frac{\Delta h}{h}$$

Where C is the stiffness coefficient
ρ is the specific mass
h is the quartz thickness

Using a development in Taylor's series up to the third order we can write

$$\frac{\Delta C}{C} = \frac{C(T) - C(To)}{C(To)} = \sum_{i=1}^{3} a_c^{(i)} (T - To)^i$$

$$\frac{\Delta \rho}{\rho} = \sum_{i=1}^{3} a_\rho^{(i)} (T - To)^i \text{ and } \frac{\Delta h}{h} = \sum_{i=1}^{3} a_h^{(i)} (T - To)^i$$

We obtain the relative frequency variation as

$$\frac{\Delta f}{f} = \frac{f(T) - f(To)}{f(To)} = \sum_{i=1}^{3} a_f^{(i)} (T - To)^i$$

An initial computation has been performed taking for a_e, a_ρ and a_h the value obtains by Bechmann, Ballato et al [11,12] between 77 K and 443 K. We find a good fitting with experiment at T>200 K but a large discrepancy at T<50 K see fig. 8.

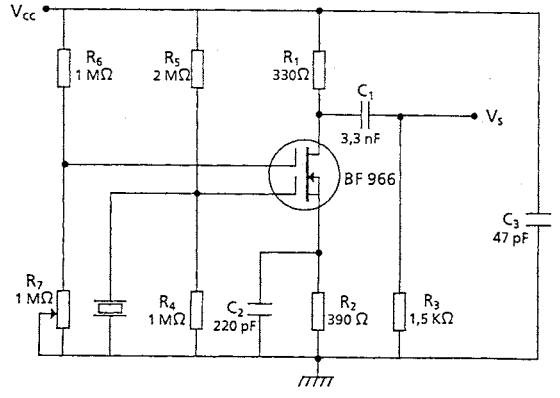


Fig. 7 : *Quartz oscillator circuit for 1.5K<T<300 K*

Another computation with a_c obtained for 4 K<T<50K by A. Smagin and B. G. Milstein [13] and with thermal expansion coefficients one million times smaller than those used at room temperature, we obtain a curve with a behavior near the experimental one at low temperature but not identical. (see fig. 9).
The experimental behavior between 4 K and 10 K presents a linear frequency variation. The slope of the straight line is about 0.2 ppm/K for Δf/f as a function of the temperature. The table I gives the results for the different "one axis rotation cut" of quartz

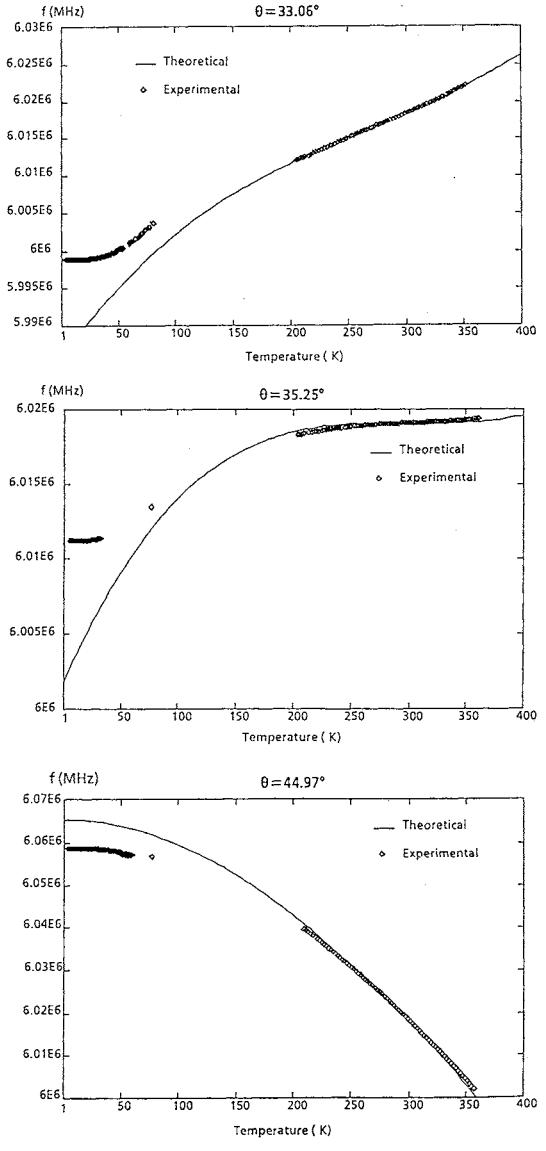


Fig. 8 : *Frequency variation as a function of temperature for a "one axis rotation quartz". See text for the coefficient values used in the theory.*
a) θ = 33.06° ; b) θ = 35.25° ;c) θ = 44.97°

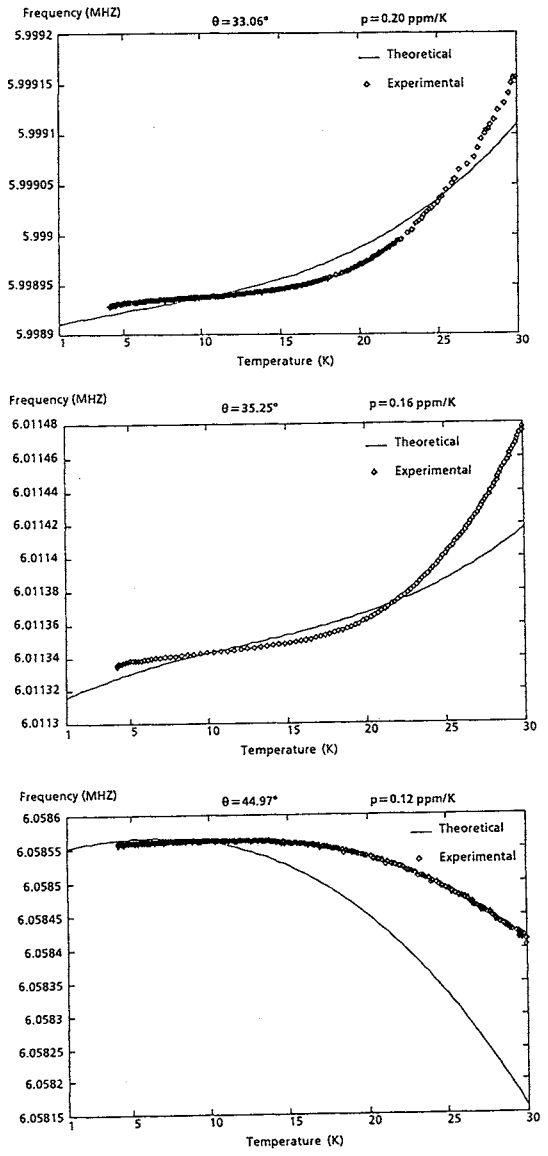


Fig. 9 : *Frequency variation as a function of temperature for a "one axis rotation quartz". See text for the coefficient values used in the theory.*
a) θ = 33.06° ; b) θ = 35.25° ; c) θ = 44.97°

θ	33.06°	35.25°	44.97°
Slope (Δf/f) f ≃ 6MHz	0.2 10-6	0.17 10-6	0.12 10-6

The slope is small but greater than the better compensated temperature cut at room temperature. The determination at low temperature of better compensated temperature cut requires further investigation.

6 - Conclusion

The most important result of this work is the interpretation of the origin of the low temperature plateau for the 1/Q = f(T) curve. Probable causes which prevent the attainment of 100 million for the quality coefficient Q at 4.2 K are

- the presence of an amorphous layer on the quartz due to mechanical polishing. This absorption can be reduced by using a careful chemical-mechanical polishing method.
- the presence of metal on quartz. This absorption can be suppressed by separation of electrodes and quartz.
- in some cases there exist a mechanical damping at least for the lower overtones. This damping can be reduced by use of the higher overtones or alternatively by the choice of a larger diameter for the resonator.

Therefore it seems possible to reach the intrinsic limit of the quartz crystal itself. Thus one would obtain a Q of 100 million at T = 4.2 K for f = 11.8 MHz (i.e. Q.f = 1.2 10^{15} Hz)

REFERENCES

1. A. El Habti, F. Bastien
Low temperature limitation on the quality factor of quartz resonators.
46th Annual symposium on frequency control, Hershey, Pennsylvania USA may 1992.
[Submitted at IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control].

2. L. Landau, G. Rumer
Physik Z. Sowjet Union, 11, 18, (1937).

3. P. G Klemens
"Effect of Thermal and Phonon Processes on Ultrasonic Attenuation" in Physical Acoustics edited by Warren P. Mason Vol III, part B, Academic Press New-York, (1965), pp 201-234.

4. A. El Habti, A. Zarka and F. Bastien
"Physical limitation on the quality factor of quartz resonators" to be published in J.A.S.A.

5. P. M Morse and K. U Ingard
Theoretical Acoustics, MC Graw Hill, New York, (1968), p 404.

6. E. J Hearn
Mechanics of Materials, Pergamon press, (1976), p 26.

7. S. Hunklinger W. Arnold
"Ultrasonic Properties of Glasses at Low Temperatures" in Physical Acoustics edited by Warren P. Mason and R. N Thurston Vol XII, Academic Press New-York, (1976).

8. J. R. Vig, C. F Cook Jr, K. Schwidtal, J. W Lebus and E. Hafner
"Surfaces Studies for Quartz Resonators" 28 th annual symposium on frequency control (1974) proceedings p 96 - 108.

9. R. E Macfarlane and J. A Rayne
Ultrasonic Attenuation in the Noble Metals, Physical Review, 162, 532-543 (1967).

10. J. J Gagnepain, J. Ubersfeld, G. Goujon and P. Handel Proc. 35th Ann. Freq. Cont. Symp. pp. 476-483, 1981.

11. R. Bechmann, A. D. Ballato, and J. J. Lukasek
Proc. IRE 50, pp. 1812-1822, 1963.

12 P. C. Y. Lee, Y. K. Yong
J. Appl. Phys. 56, 5, 1984.

13. A. Smagin and B. G. Milstein
Kristallografya, 19, 4, 1974.

CORRELATION BETWEEN RESONATORS PERFORMANCES AND RAW MATERIAL QUALITY - A REVIEW

J.-P. AUBRY, J.-P. MICHEL, M. ZAFRANY, O. CAMBON

THOMSON / CEPE
44, av. de la Glacière - BP 165
95105 ARGENTEUIL - FRANCE

SUMMARY

High performances for ovenized oscillators and high yields for filters are closely dependant on the crystal raw material. This paper reviews the most important kinds of defects in the raw material and the way they are measured (dislocations, impurities, inclusions, OH content). Then it will try to give mechanisms to show the way this defects may act on the electrical characteristics of resonators, oscillators and filters regarding phase noise, noise floor, G-sensitivity, retrace, temperature stability, ageing, radiation sensitivity, intermodulation, spurious responses and process ability. It is also given some evidences of the forseen limits of the conventionnal physical characterisation procedures when applied to higher quality materials.

1./ INTRODUCTION

The progress in order to achieve ever higher performances for crystal resonators and oscillators have essentially been performed on technology and crystal design. Resonator shape, mounting, enclosure, annealing, cleaning, plating, plate making have been the subject of a lot of works widely reported in the scientific litterature. On the other hand the raw material manufacturers have tried to improve the quality of the material they deliver thinking there may be a correlation between the physical characteristics of the material and the performances of the resonator. Actually some correlations have been demonstrated: one of the most well known is the evolution of the mechanical quality factor of the resonator with the infra-red absorption coefficient of quartz. Also well known is the dependancy of the irradiation behaviour with the impurities content in the raw material. But, in fact, regarding the numerous very high level characteristics which are required for high performance resonators and oscillators, very few exact correspondances have been precisely established with the material and with the elaborating conditions.
The difficulty to make such a correlation comes from the fact that a good correlation requires a statistical treatment - ie - a high number of equal high quality devices. This is not an easy point: it means that measuring procedures must be accurate enough, that test oscillators must be of the most performing type, that resonator processing need the high reproducibility of an industrial process, that the design of the resonator does not cancel the material evidences - ie that technological effects on the resonator performances have previously been widely studied and mastered - and, last but not least, that the used material is sufficiently defined or characterized. This last point is an actual to-day limitation because, if raw quartz manufacturers are able to deliver sometimes very high performances materials, they do not seem to be able to ensure a sufficient reproducibility for a very high level of quality material. In addition, taking as-coming quartz production and sorting is not realistic because the block is no longer big enough after characterization to ensure the manufacturing of a statistically significant number of resonators. Even so, block characterization is not sufficient because the block itself is not homogeneous enough and the remaining part of a block after characterization cannot be guaranteed to be of the same level of quality as the actually tested part and nobody can be sure that a casual dislocation or inclusion may not stand in the middle of the resonator supposed to be made of a defect free material.
At that point, one can only suppose that a block or a batch which will demonstrate after characterization a low level of defects, should be statistically better than else. But connecting this to the resonators and oscillators performances will need to perform a test matrix with widely dimensionned samples expecting that the raw material batch is large enough. The driving of such test matrix is very costly because it obviously adresses the best resonator and oscillator manufacturing processes. It needs to be strongly optimized using for instance Taguchi methods.

2./ CRYSTAL DEFECTS

2.1./ DISLOCATIONS

The dislocations comes from the shift of an atomic plane on another. There can be two types of dislocations: edge dislocation for which the Burger vector is directed at right angle of the dislocation line

and the screw dislocation for which the Burger vector is colinear to the dislocation line. Many works have concluded that actual dislocations were essentially of the edge type [1],[2],[3] in the ratio of approximately 85%. The same references have observed that the dislocations have always the same direction and they stay within a solid cone of 25 degrees off the Z axis. It has also been demonstrated that the slight inflexion off Z axis tends to ensure gathering of the dislocations during their propagation towards Z face; this is responsible for the formation of the commonly observable hillocks on the Z surface of as-grown crystals. This hillocks themselves seem to be the result of the spiral growing of crystal as described by [13] and called references. One can suppose that the spiral growing mechanism tends to reject dislocations on the spiral edge. The origin of a dislocation in high quality crystal is mainly a pre-existing dislocation in the seed: during crystal growth along the Z direction, the dislocation - for which Z is a preferred direction - grows too. Consequently, when a dislocation is pre-existing in the seed, it will generate a dislocation in the crystal. Another cause of dislocation generation is the presence of inclusions at the growing front: the inclusion may initiate a dislocation which will pursue its growing along the Z axis. This last cause of dislocation is more a problem for medium quality raw material than for high quality one which is quite inclusion free. Nevertheless, as manufacturers make their seeds from previously grown crystal and more and more rarely from natural crystal (for cost saving reasons), there is a not negligible risk for a constant increase of the dislocation density in synthetic quartz. Dislocation may also come from unsteady growing conditions often associated with high growing rates: when there is no concentration gradient in the region near the growing front, the oversaturated solution cristallizes in sites presenting a minimum of energy for the crystalline lattice. In case of more dynamic conditions in the growing area, there is a flow of solution around the growing area. This flow may casualy generate local concentration gradients. So, crystallization may prefer to occure at the site of a concentration peak instead of at the optimum site in the cristalline lattice from the crystal energy point of view.

It is generally assumed that moving a dislocation in a high quality crystal requires too high an energy to be achieved in the normal using conditions of the crystal. Nevertheless Griggs and Franck [4] gave a theory (also reported by [5]) which assumed that an increased concentration of $[OH]^-$ may ease the gliding of atomic plans on each others and lead to a higher probability of occurence of dislocations. The lowering of lattice energy with $[OH]^-$ concentration is a commonly observed behaviour of quartz homeotype crystals. It has been demonstrated for instance that the growth starting of berlinite is more tolerant to seed defects when the growing conditions facilitate a reasonnably high level of water inclusion [6]. Good results have been obtained by [7] when trying to grow berlinite crystal from multi seeds positionned side to side because the growing process developped was highly generating of water introduction. This tolerance has also been experienced to make feasible an heterogeneous epitaxy: single $GaPO_4$ crystals have been grown on berlinite seeds [8]. Furthermore, [9] made the assumption that there could be a bi-univoque relation between $[OH]^-$ concentration and the dislocation density given by the following formula:

$$\log N_D = 5 +/- 0.48 + 2.5 \log a$$

where N_D is the dislocation density per cm^2, a is the infra-red absorption at 3500 cm^{-1}. The order of magnitude issued from that formula is acceptable but, in fact, it should not be considered as an illustration of the above theory because a reversed mechanism can be invoked as an explanation: the dislocations may be attractive sites for $[OH]^-$, which may be the cause of increased $[OH]^-$ concentration in a highly dislocated material.

The measurement of dislocations density is made by X-ray topography. The sample must be chosen keeping in mind that the dislocations are oriented toward the Z direction: a Z cut will give a studded aspect due to right angle interceipt of the dislocations; an AT or SC cut will give an idea of the importance of the perturbation caused by the dislocation on an actual resonator; dislocations are figured by dark lines whose length depends on the thickness of the plate; Y cuts or X cuts through the seed lead to visualize the whole extent of each dislocation (fig. 1); It is particularly interesting to make connections between seed defects, or inclusions and dislocation generation.

Another way to visualize dislocations has been described by [5] : it is a light scattering tomography method which has the merit to be less expensive than X-ray topography.

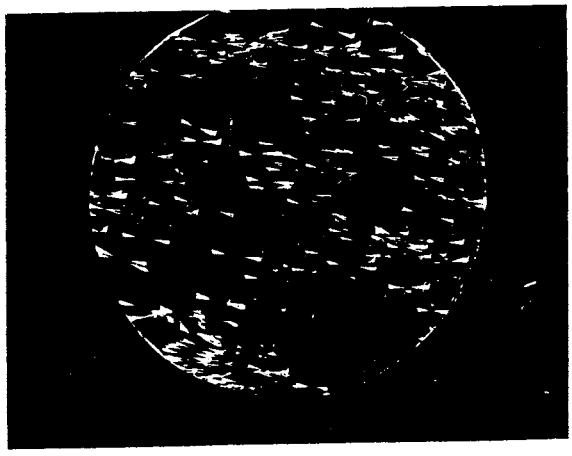


FIGURE 1 : DISLOCATIONS OBSERVED BY X-RAY TOPOGRAPHY [45]

2.2./ ETCH CHANNEL DENSITY

The etch channel density is a raw material characterization parameter which is quantified by the IEC norm [12]. It is not properly speaking a defect but an evidence of defect appearing after a chemical etching of the crystal: after etching by fluorhydric acid the crystal shows preferred quasi-ponctual etched spots called etch pits. This etch pits look like conic holes prolongated inside the crystal by tunnels called etch channels.
It has been verified that these channels are the surface interceipt of dislocations [1],[2],[10],[11],[3]. Taking this into account, it can be assumed that etch channels have the same origin as dislocations. It must be underlined that, if each etch pits can be considered as caused by a dislocation, each dislocation, conversely, does not necessarily produce an etch channel under chemical etching. One of the main counter-example comes from swept material: sweeping has been proved to be an efficient factor to reduce etch channel density. But all authors agree to write that the energy involved during sweeping is far from sufficient to cause a dislocation motion. X-ray topography performed before and after sweeping confirmed this assertion. The mechanism should be else and is described as follow: during sweeping, impuritiy cations are extracted from their original sites. These cations are collected in the dislocations. A dislocation is called to be decorated when impurities are trapped in it. These impurities in the dislocation make it chemically inactive under etching.
The observation of etch channels is performed on an AT cut plate submitted to a standard test procedure [12], under binocular (fig.2). Quantification is made by counting the number of pits per surface unit.

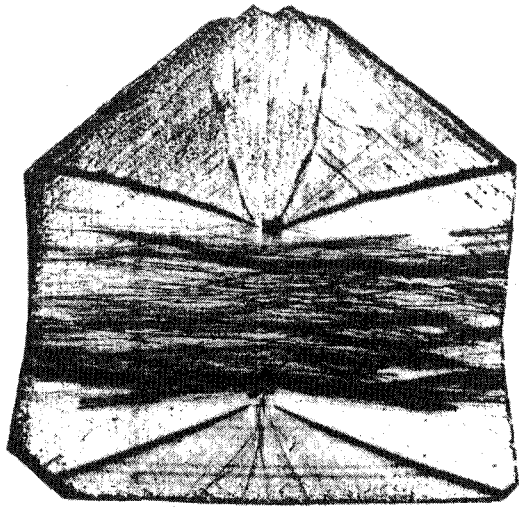


FIGURE 2 : VIEW OF ETCH CHANNELS ON A AT PLATE

2.3./ INCLUSIONS

Inclusions are solid or liquid inserts in the crystal. Regarding piezoelectric application the solid inclusions seem to be the most dommageable. Conversely, liquid inclusions are more a problem for optical application of quartz. Analysis of the solid inclusions have often lead to invoke the presence of acmite. Acmite is an iron/sodium silicate $NaFeSi_2O_6$ involving Fe(III). Iron is suspected to come either from the autoclave walls and bearers or from the nutrient. More recent works [14] tend to demonstrate that, under special conditions, Al containing inclusions may be grown. In any case, solid inclusions are small foreign crystallites grown in the crystal. The mechanism could be the following [14]: taking into account that there is a given level of impurities in the nutrient, there is a conflicting situation about the way these impurities may be inserted in the lattice: under some local conditions in the growing front region, they are incorporated in a distributed way under their ionic form; in other conditions, impurities develop themselves - under a pure or a complexed form - as separate single crystallites deposited on the growing front. As a comparison, a similar behaviour has been observed when trying to grow berlinite on a quartz seed: in that case, the small crystallites are made of berlinite. Conflict between cristallytes growing and insertion in the lattice is mainly directed by the local conditions resulting from nutrient flow on the growing front. Another mechanism has also been invoked: the walls of the autoclave being made of steel, it occurs an acmite layer generation on its surface. This layer acts fortunately as a protection for the wall toward any excess corrosion. But one may suspect some particules of acmite to break away from the walls and to be carried by the convective nutrient flow toward the growing front. Expecting an improvement in inclusion densities, several authors have tested autoclaves with noble metal liners. In fact, no improvements have still been reported [15] .
Inclusions are observed by immersion of the whole block in a liquid of optical index closed to the one of quartz, namely 1.55. The block is illuminated along the Y axis. Observation is made at right angle either on X- or Z-face. Inclusions appear as white dots inside the crystal. Micron size particules can be detected. For a better observation, the IEC norm [12] suggests a 30 to 40 magnification. Due to the matching of optical index, the faces of the block do not need to be polished.
Some authors have also detected a light blue coloration of the inner cristal when illuminated with an intense light. This phenomenon has been associated with Tyndall effect due to diffusion on very small inclusions. It is more probably caused by liquid inclusions.

2.4./ IMPURITIES

The main impurities encountered in a quartz crystal are Al, Fe, Li, H, Na. As substantial improvements in impurity concentration have been made by growing from well refined nutrients [15],[16], it can be assumed that impurities essentially come from nutrient. Impurities can be inserted in the crystal on

two sites: they can substitute to an Si atom in the lattice. This mechanism is more probable for Al whose size and affinity make it particularly suitable to Si replacement. Trivalent Fe is also a good candidate for Si substitution. Inserting a trivalent impurity on the site of a 4-valent Si generates a lack of electrical charge. So the lack needs to be compensated by a monovalent ion. This kind of ion may be either Li^+, Na^+ or H^+. The smallest ones may be situated directly on the Si site (Li^+, H^+), when a bigger one as Na^+ will be aside of the normally Si site [17]. Another preferred place for impurities is the breaking line of the dislocations which manages space for foreign atoms location. During crystal growth, when impurities are present around the growing front of the crystal, there is a conflict between three insertion mechanisms depending on the local lattice energy, and SiO_2 concentration gradient: the substitution on a Si site [18], the location in a dislocation and, as previously described, the growing of a discrete inclusion.
Excepted for inclusions, once inserted in the crystal, impurities need quite small energy to be moved from one site to another. Impurity displacement can be achieved by irradiating of the crystal or by sweeping: it has been shown above that sweeping makes chemically inert the outside interceipt of previously decorated dislocations by moving impurities.
Al, Fe, Li, H, Na are the main impurities. As impurities come from the nutrient, other impurities may be present in the growing solution but high levels of concentration are less probable.
The analysis of the impurities can be performed by various methods: chemical analysis gives a precision of a few ppm, inductively coupled plasma or atomic absorption lead to a better precision of a few hundredth of ppm. Nuclear magnetic resonance can be useful to detect the presence and location of Li, Si and Al.

2.5./ OH^- CONTENT

When speaking of impurities a special mention must be made concerning H^+: in the lattice, hydrogen is quite always linked to oxygen creating OH^-. This is of particular interest because OH^- is easily estimated by infra-red absorption. A simple correspondance has been established between infra-red absorption at 3500 cm^{-1} and the mechanical quality factor of the material. This correspondance is quite good as long as OH^- can be considered as the main impurity whose effect on the mechanical Q-value exceeds widely all others. For very low OH^- contents the correspondance is no more so accurate as it will be shown further.

2.6./ REQUIREMENT OF PERFORMANCES FOR RAW MATERIAL AND COMMERCIALLY AVAILABLE MATERIALS

Different ranges of performances have been defined by the IEC norm [12] concerning mass production of quartz. The requirement of crystal devices industry for high performance resonators and oscillators looks for higher characteristics of the raw material: Infra-red absorption at 3500 cm^{-1} should be better than 0.018, dislocation density better than 10 cm^{-2}, impurity content better than 1ppm for each species Al included and inclusion density below 5 cm^{-3} for particule sizes bigger than 30 microns. To-day more approaching specifications are obtained from french manufcturers. Nevertheless a good reproducibility for that range of specification is still a challlenge.
American open-market manufacturers are limited at a dislocation density of 30 cm^{-2}, and a few ppm of impurities (even 10 ppm regarding Na). Japanese production is more dedicated to high mass-production of medium specification crystal devices. Russian manufacturers seem to be able to deliver quite high level of specification material but there is no available information about their ability to manage a good economically acceptable reproducibility.

3./ RAW QUALITY MATERIAL AND RESONATOR RERFORMANCES

The performance requirements for material come from the need to improve the characteristics of the resonators and oscillators. The target specifications have been defined considering some dependancies between each of the critical requirements for resonators characteristics and the raw material performances. Some of them may be quantified but numerous others are provided by reasonning on the resonator behaviour regarding each type of material defect. This links will be reviewed.

3.1./ PHASE NOISE

A low close to the carrier noise level for an oscillator is tightly connected to a high mechanical Q-value for the resonator: the frequency for which the noise floor is equal to $1/f^3$ noise is given by $Df = f/2Q$. So, the higher is the Q-value, the closer to carrier frequency the floor level is reached. Q-value is more or less affected by each of the material defects above listed. The historical main cause of Q-value limitation is the OH^- content. The IEC norm [12] makes a correlation between infra-red absorption in the material due to OH^- and the mechanical Q-value. An approximative formula is: $10^6/Q = 0.114 + 7.47\ a - 0.45\ a^2$ with $a = 1/t\ Log(T_{3800}/T_{3500})$ where T_{XXXX} is the infra-red transmission at XXXX cm^{-1} and t is the thickness of the sample in cm. This formula is quite good for high values of a; it becomes unaccurate when a comes below 0.02: as it has been mentionned above, OH^- is produced by the presence of H^+ impurities; as long as H^+ is the major cause of impurity in the crystal it can be considered - regarding the effect - as the only cause of damping. H^+ being introduced in the crystal for the need of charge balance, other monovalent ions are also candidates such as Li^+, Na^+. In fact, each of these

three ions may be responsible for damping. Others of minor importance may also be invoked. The mechanisms of damping may be of different kinds: firstly, an impurity inserted in the lattice must be responsible of a local strong elastic constant variation in a region closed to it, due to the local variation of the cohesion energy. During lattice vibration, it will be the spot of a local maximum of vibrating energy (this hypothesis has been invoked - although in a quite different formulation - for explanation of Q variation at temperatures below 4°K by [19]) involving non linear effects. In that hypothesis, a slight spread of energy over the whole spectrum of frequency should be responsible for a lack of energy at the observed frequency and, consequently, for a decrease of the Q-value. According to [19], this lack of energy should have an order of magnitude of 10^{-7} for a reasonnable content of impurity. Another explanation which has been proposed as a damping mechanism is a vibration coupling effect between lattice and impurities. It has often be noticed that important Q fluctuations may occur when varying the temperature of the resonator between 10°K and ambient. Different peaks of mechanical absorption have been observed; a main one is attributed to Na at 53°K. It has also been demonstrated that other impurities such as Li may also be responsible for absorption peaks at other temperatures. [20] mentionned that Na causes an extra peak at 335°K and Li a peak at 305°K. Both peaks are of particular interest because they are closed to the normal operating range of temperature for the resonator. A more complete observation has been made by [17]: he made a difference between the various locations and bondings of the impurity in the lattice. He established that different temperature peaks may be generated depending on the species and locations involved. Firstly, one must keep in mind that monovalent impurities located in the lattice are the result of Al insertion on a normally Si site and that monovalent ions may be located either at the Si site, closely connected to Al, or at a dislocation place. The corresponding absorption peaks have been attributed as follow: 305°K for Li in a dislocation, 335°K for Na in a dislocation, 53°K and 135°K for Na at Al proximity; no absorption peaks have been observed connected to OH in a dislocation, nevertheless the dislocations act as storage sites for a coming move of OH toward Al sites. It has also been observed no evidence of a peak due to Li at Al proximity. The proposed explanation is that Li being of small size is located on the same site as Al. Conversely, Na cannot stand on Al site and is deported aside. The different peaks observed are not equal regarding their sharpness. Especially the 135°K, 305°K and 335°K are quite flat: there is still a remaining contribution at ambient temperature. This contribution is responsible for an extra damping factor in the temperature range of use of the resonator.

In summary, to improve Q-value it is necessary to avoid: 1/ Al sites 2/ dislocations 3/ Na impurities 4/ Li in dislocations; Li located on a Al site does not seem to be an inconvenient: it may give the way to a solution even when Al or dislocations are unavoidable.

Dislocations can also be more directly responsible of a Q attenuation. Inclusions can also be included in the same reasonning: as it has been above mentionned for impurity, these defects are the spot of local variations of the cohesion energy of the lattice. This is firstly the cause of a spread of frequency responses directly responsible for noise increase and secondly the energy spent by the spreading is missing for the main frequency response, leading to a Q decrease. The local variations of the cohesion energy must induce a local variation of the elastic constants. If the dislocation density is quite high, it must have a consequence on the macroscopic elastic constants. This phenomenon has been observed by [31] for a dislocation density of $10^3/cm^2$. Is it possible to conclude on an actual effect of the dislocations ? May not it be the damping effect of any decoration in the dislocation ? May be more conclusive are the observations made by [25] using stroboscopic X-ray topography: they stated that, when a dislocation lies under the electrode of a resonator, it generates a deviation of the acoustic wave propagation. This deviation has been observed to be responsible for the appearing of coupled acoustic modes clearly visible when stroboscopy is tuned on other frequencies. It has been concluded on a possible mechanism for Q decrease [26].

3.2./ G-SENSITIVITY

Lot of work has been done in the recent past years in order to determine the causes for G-sensitivity of the resonators. Previously the G-sensitivity effect coming from the raw material, which is in the 10^{-9} range, has been canceled by technological more important causes of frequency fluctuations under vibrations. Nowadays this order of magnitude is no more negligible regarding the target specifications for oscillator stability. It has first been demonstrated that the G-sensitivity is due to an unsymmetrical stress field applied to the resonator plate which is exposed to an acceleration. The causes of an unsymmetrical stress field have been investigated. Among them the practical main one seems to be the quasi unavoidable non perfectly symmetrical disposal of the suspension ribbons and sealing on the plate. Pursuing in that direction a lot of work has been performed in order to theoretically estimate the G-sensitivity contribution of a given mispositionning of the suspensions. On another way one tried to define the condition of a G-sensitivity minimization taking the mispositionning as a data. Practical laws have been established [22] which are in good agreement with the actual behaviour. Another way to reduce the effect of the mispositionnig of the suspensions is to make a mechanical shielding between the suspension and the vibrating area. Here is the origin of the autosuspended resonator (QAS) [27] where the vibrating area is linked to an outer quartz ring by quartz bridges. The efficiency of this structure on an

industrial production has been proved: the G-sensitivity has evoluated from a mean value of 10^{-9} to a value of $3\ 10^{-10}$ [28]. Regarding our subject the most interesting observation has been made by [21]: he exposed that the actual origin of the G-sensitivity of a resonator was not exactly the existence of an unsymmetrical stress field but the presence of a difference of location between the center of the stress field and the center peak of the mode shape. In other words, even when the stress field may be symmetrical on the plate, if the mode shape is not itself centered there will be an increase of G-sensitivity. Pursuing on the same mind he demonstrated that it was possible to compensate the G-sensitivity of a resonator by a properly selectively loading of a part of the electrode off the center. Another demonstration of the same assertion has been given by [23] who reported the feasibility of a G-sensitivity reduction by selectively driving different parts of the electrode off the center. Meanwhile observations of several resonators by stroboscopic X-ray topography [24] gave evidence that dislocations lying under the electrode of the resonator may be responsible for a mode shape deformation because vibrating energy is guided outside the electrode area by the dislocation. Then it is deducible that a dislocation in the vibrating area will lead to a poor G-sensitivity performance [5]. Nobody recently related it, but it may be suspected that the presence of an inclusion in the vibrating area, especially when it is off the center, may cause a similar increase of the G-sensitivity of the resonator.

3.3./ RADIATION SENSITIVITY

The crystal resonator centre frequency have been demonstrated to be sensitive to radiation in the range of 10^{-12} per rad. This point is a particular inconvenience for space application: the spatial norms suppose that the resonator may integrate a total $2\ 10^{6}$ rad dose during its operating life. Fortunately, it seems that, due to shielding, the resonator in its actual location in the satellite will not be exposed to more than a few tousends of rad [29]. Various manifestations may occur in the resonator when exposed to irradiation: a darkening of the resonator is visible for high doses of irradiation. It recovers after annealing. Frequency is shifted downwards or upwards, Q-value is modified, the frequency versus temperature curve is modified. The characteristics recover partially or totally with time or annealing.

The radiation sensitivity has been attributed for long at a move of positively charged impurities. It has been the origin of the development of the sweeping technics to improve the resonator hardness: during sweeping, under an intense electrical field, positively charged ions are extracted from the crystal. Sweeping under an air atmosphere introduces, for the purpose of charge balance, H^+ under an OH^- form. In order to introduce a more accurate explanation, one must remember that the monovalent impurities are present in the crystal in order to compensate for the lack of charges due to the presence of an Al on a Si site and that, according to their location in the crystal lattice, they are responsible for mechanical absorption peaks at various low temperature. Observing the evolutions of the peaks during irradiation [20] has developped an interesting theory which has been completed by [17] by the same approach, and more or less confirmed by [30] by an infra-red absorption approach. The resulting mechanism is the following: during irradiation, monovalent impurities such as Na^+ and Li^+ which stand on an Al replaced Si site are moved towards dislocations. Consequently the mechanical absorption peak due to Na moves from 53°K to 335°K and a peak due to Li appears at 305°K. The proximity of the ambient temperature range make this peaks efficient to modify the electrical characteristics in the temperature range of use, namely Q value and resonant frequency. The frequency versus temperature curve is differently impacted whether lower temperature or higher temperature are considered; consequently the inversion point is modified. Meanwhile, the extracted monovalent ions must be replaced on the Al site. In case of a high concentration of H^+ or OH^- around, the extracted ions may be replaced by OH^-; in case of a lower concentration, only part of the Al site will be filled with OH^-. The remaining are filled with holes. This holes are responsible for the darkening of the crystal. Annealing may partially reverse the mechanism or, more often, simply replace the holes by OH^-.

Consequently, in order to improve the radiation behaviour of the resonators, the raw material must be improved in the following way: 1/ avoid Al 2/ avoid Na and Li 3/ when an exceeding level of this impurities is remaining, sweeping must be performed. Then, one can discuss whether sweeping must be performed in an air atmosphere or in vacuum. Sweeping in air leads to replace monovalent ions by OH^-. At first sight, it should generate a decrease of Q-value. But sweeping has been demonstrated to remove impurities from any interstitial site in the material and consequently to remove impurities from lattice defects. As monovalent impurities are responsible for an extra mechanical absorption when located in lattice defects, their extraction will tend to increase Q. Among this two conflicting mechanisms the most important effect on Q must be considered. In the particular case of Li located in a lattice defect, [20] has observed a simultaneous increase of infra-red absorption and of Q-value after air sweeping. On the other hand, vacuum sweeping could appear to be more promizing. Nevertheless [24] has observed under X-ray topography that vacuum sweeping was responsible for stress in the sweeping anode region.

3.4./ AGEING

The ageing of the crystal oscillators has constantly been the subject of improvements. Industrial performances are around 10^{-9} per month at 5 Mhz

after one or two months of stabilization in operating quiescent conditions. This nowadays performance has been reached by successive successful studies on the resonator technology and the manufacturing process. Such parameters as the mounting stress, the plating stress, the cleanlyness of the plate, the outgassing of the whole assembly have been demonstrated to be very influent on ageing characteristics. Heating treatments at different steps of the process are efficient to stabilize more or less completely those parameters.
Further improvements will have to allow even better industrial performances, i.e. lower ageing per month towards a few 10^{-10}, as well as a shorter stabilization cycle time with a target of half a month. In that range of performances, the influence of the material itself must be considered. At first sight, whatever may be the defect density in the crystal, it will have no effect on the ageing of the resonator provided nothing changes with time. Up to now, practically no pertinent correlations have been made between crystal defects and the contribution of their evolution to ageing. Let us first consider dislocations: dislocations are defects of the lattice. However lattice arrangement involves high levels of energy: it does not seem to be any chance for a dislocation motion except in very hard conditions far from using ones (high temperatures closed to Curie phase transition point associated with a high electrical field [32]). Nevertheless, [5], on its review on dislocations effects, invoked a possible effect when dislocations are present simultaneously with a high OH concentration because OH facilitates the sliding of atomic planes. However he remained quite suspicious about that hypothesis. On the same year, [33] made a review on the causes of the ageing of resonators. He gave no credit to the thesis of dislocation motions at ambient temperature, but no experimental results on the subject are available. Practically no mentions of inclusion effect on ageing are available, but one can extrapolate the reasonning made about dislocations and, consequently, inclusions should not contribute to ageing. More suspectable are the impurities, but [33] underlined that no experimental work has been carried about a possible correlation between impurity concentration or type and ageing. However impurities are quite easily moveable under weakly stringent conditions. The trap of impurities in unsteady conditions during growing is a possible occuring: this has been demonstrated in a very particular case by [25]: when vacuum a crystal block is exposed to vacuum sweeping, there is an evidence of stress at the anode side probably due to a hole defect concentration. The block recovers from that unsteady situation by annealing. [30] (and others previously) demonstrated that a quite low temperature annealing (250°C) is sufficient to eliminate holes - holes having been generated by irradiation in this particular case. This results lead to the possible supposition that: 1/ crystal under using may be in slightly unsteady conditions regarding the impurity distribution 2/ the recovering energy is low enough to be given by the normal conditions of use of the resonator in its environmental range, so that impurities motion may be suspected regarding ageing contribution.

3.5./ RETRACE

When turned off then on again, an ovenized oscillator does not instantaneously recovers it working frequency. The main cause is the temperature decrease of the resonator, and the design of the thermostat has to be optimized in order to achieve an as fast as possible temperature recovery. However, even when temperature has recovered, the frequency fails to recover its initial value in the 10^{-9} range of accuracy. Technological causes may be the reason such as a change in the stress field applied to the plate by the mounting. But the material itself may also be suspected. For a beginning of explanation, the experiments and theories performed by [34] must be put in mind. When a DC polarization is applied to a doubly rotated crystal, there is a change in resonant frequency. This change is due to the electro-elastic effect quantified by the third-order electro-elastic coefficients. However, when the DC polarization is maintained, the frequency shift is gradually reduced and even vanishes on a logarithmic decrease. This phenomenon has been attributed to the motion of electrically charged impurities along the applied DC electrical field. It has been applied as a method to investigate the concentration and type of impurities considering the diffusion constant of the various species. The decrease of the frequency shift has been shown to be activated by temperature. At ambient temperature it is a possible cause of a quite slow frequency shift keeping on over days. To avoid the effect of DC polarization, oscillators are designed with a serial capacitor connected with the resonator. However this solution does not lead to avoid the transient DC electrical current when the oscillator is turned on. This may give an explanation for the retrace shift. In addition this may be the starting point of an impurity unsteady distribution at the first turning on of the oscillator leading to a contribution to initial ageing.

3.6./ TEMPERATURE STABILITY

Temperature stability for an ovenized oscillator depends on the oven design and its temperature stability (about a few tousendth of degrees) and on the accuracy of the positionning of the working point of the oscillator at the inverting temperature of the crystal. The tuning at the inverting point lies on the accuracy of the cut angle value (sometimes 15'). So there is an interest for a highly reproducible frequency versus temperature curve. There is also an interest for a good stability of a given resonator regarding the frequency versus temperature curve. Due to the above mentionned evolution of the elastic constants around crystal defects, there should be a contribution to frequency versus temperature curve.

In addition, the mechanical absorption peaks due to impurities in dislocations are responsible for frequency variations more important in the lower temperature region leading to a frequency versus temperature curve rotation.

3.7./ NOISE FLOOR

The noise floor of an oscillator is improved by increasing the driving level of the resonator. Towards higher values the driving level of the resonator is limited by the anisochronism effect appearing about 0.3 mW for an 10 Mhz SC 3rd overtone resonator. This phenomenon is a consequence of a non linear behaviour of the crystal which is normally described by the third order elastic coefficients of the material [36]. No works have been reported having been done on the changes of the third order coefficients with the quality of the raw material. However, it has been asserted above that crystal defects may be responsible for local elastic constants changes. On those particular regions, it may result on a local accumulation of mechanical vibrating energy, sollicitating the third order coefficients for driving levels below the therotetical level of anisochronism occurence in a resonator made of a defect free crystal. This hypothesis has been explained in other words by [19] speaking of vibration spreading around inclusions and dislocations.

3.8./ INTERMODULATION

The intermodulation characteristics is of particular interest for dual-mode oscillators and for filters. This phenomenon is responsible for the generation of new frequency responses when the crystal is driven by a two tones signal. For the dual mode oscillator, the crystal is worked simultaneously on the two different frequencies dedicated respectively to the metrologic reference and to the thermal sensor [35]. The metrological frequency is corrected at each temperature by the information of temperature provided by the second mode oscillator acting as a thermal sensor. The two frequencies lie approximately in a three ratio, according to the design options. Extra frequencies would be responsible of an enrichment of the noise spectrum of the oscillator. Regarding filters, the input spectrum of the filter is, by definition, composed of various frequencies among which the filter has to make a selection. The generation of extra frequency responses at the filter is obviously an inconvenience. There are many technological reason for a degraded intermodulation behaviour such as mounting, cleaning, ... But finally, the performances are limited by the material and are governed by the fourth-order elastic constants [37] [38]. The same reasonning which has been applied to the driving level may be applied to suppose that the crystal defects, especially inclusions and dislocations, may be responsible for a change in the fourth order elastic coefficients with degrading consequences regarding the intermodulation performances.

3.9./ SPURIOUS RESPONSES

Spurious responses are the consequence of anharmonic responses of the resonator. They are responsible for spectrum enrichment of an oscillator frequency and they are causes of rejection limitations for filters. The main causes of spurii generation is the shape of the plate and of the electrode. When focusing on the first spurious response, which is the closest to the resonant frequency of the resonator, it appear to essentially be generated by any dissymetry affecting the resonator, leading to a dissymetry of the mode shape. The causes of dissymetry may be uncontroled thickness or contouring of the resonator, misalignment of upper and lower electrode, dissymetry of plating, dissymetry of mounting, ... Actually another cause of dissymetry must be considered: it has been assumed above that the dislocations may generate a deformation of the mode shapes, sufficient enough to have consequencies on G-sensitivity performances of the resonators. The same dislocations may be supposed to act as contributors for increasing the closest spurious response. However the spurious frequency responses seem actually to be to-day more widely limited by the previously described technological limitations than by the material defects.

3.10./ PROCESS ABILITY

Process ability is also a very important matter regarding especially filters manufacturing. Such applications as mobile telecommunication nowadays require higher frequencies (70 to 200 MHz) [39]. Now, this kind of filter has been at a manufacturing step for 2 years, then yield is a main factor of interest. According to the wide bandwidths needed, the resonator have to be worked at the fundamental frequency. Consequently, the plate thickness is very low (8 microns at 200 MHz). It is achieved by ionic or chemical etching and the etching amplitude is high (about 40 microns). So any defect in the material leading to a locally modified etching rate may cause strong surface defects on the finished plate.
As for as ionic etching, a cause of plate rejection comes from small hillocks appearing on the plate. Some of this defects have been observed to come from inclusions arised at the surface level during etching and then acting as a mask for adding etching. Concerning chemical etching, a lot of papers have already been published [40-44] relating etch pits and etch channels occurence during etching by fluorhydric acid or ammonium difluoride. By the way, it has been stated above that it is the most common way to visualize dislocations for their density measurement. Nowadays, new solvents are being used with the purpose to avoid etch pits and etch channels. These solvents have been developped in

order to perform chemical etching in closed to equilibrium conditions. The result is a widely lower sensitivity towards etch pits and etch channels generation. These defects may still occur for wide amplitude etchings (40 microns and more) on poor quality materials. However these technics are efficient enough to make practical chemical etching feasible but yields will still have to earn from a lowering of dislocations and inclusions density.

4./ CONCLUSION

The quartz raw material has for long been suspected to give a limitation to resonators and oscillators performances. Some evidences have been given experimentally, but many correlations still need to be experimentally demonstrated. However, the to-day better knowledge of the kind of defects in the raw quartz and of some of their acting mechanisms, allow to deduce some hypothesis about a correlation. According to that reasonning, each of the major performances of the resonators for oscillator applications as for as filter applications have been demonstrated to possibly be affected by the raw material characteristics. In addition, it has been introduced the opinion that standard material characterization procedures - such as infra-red absorption for mechanical Q-value estimation - would not be any more appliable for very high quality raw material.

5./ BIBLIOGRAPHY

[1] Lang, A.R. and Miuscov, V.F., Journal of applied Physics, 38, 1967, p.2477
[2] MacLaren,A.C., Osborne, C.F. and Saunders, L.A., Phys. Stat. Sol. 4, 1971, p.235
[3] Johnson, G.R. and Irvine, R.A., Proceedings of the 41st Annual Frequency Control Symposium, 1987, p.175
[4] Griggs, Franck, J.R. Astr. Soc. v. 14 pp. 19-31, 1967
[5] Murray, Proceedings of the 45th Annual Frequency Control Symposium, 1991, p.2
[6] Philippot, Ibanez, Goiffon, Buisson, Arnaud, Capelle, Zarka, Toudic, Schwartzel, Détaint, 4th European Frequency and Time Forum, 1990, p.577
[7] Chai, Buehler, Flynn, US patent n. 4 324 773, 1982
[8] Philippot, and al., 6th European Frequency and Time Forum, 1992
[9] Brice, Review of Modern Physics, v. 57, n°1, pp.105-146, 1985
[10] Iwasaki, Journal of Crystal Growth, 39, 1977, p.291
[11] Barns, Freeland, Kolb, Laudise, Patel, Journal of Crystal Growth, 43, 1978, p.676
[12] norme IEC 49 218
[13] Cambon, thesis, 1990, Montpellier, France
[14] private conversation with Pr. Demazeau, LCS, Bordeaux, France
[15] Armington, Balascio, Proceedings of the 39th Annual Frequency Control Symposium, 1985
[16] private report, SICN, Annecy, France
[17] Martin, Proceedings of the 43rd Annual Frequency Control Symposium, 1989, p.526
[18] Lias, Grudensky, Kolb, Laudise, Journal of Crystal Growth, 18, 1973, pp.1-6
[19] El Habti, thesis, 1992, Besancon, France
[20] Armington, Balascio, Proceedings of the 42nd Annual Frequency Control Symposium, 1988, p.162
[21] EerNisse, Proceedings of the 43rd Annual Frequency Control Symposium, 1989, p.388
[22] Janiaud, thesis, 1978, Besançon, France
[23] Smythe, Horton, Proceedings of the 44th Annual Frequency Control Symposium, 1990,p.437
[24] Zarka, Proceedings of the 41st Annual Frequency Control Symposium, 1987, p.236
[25] Zarka, Capelle, Detaint, Schwartzel, Proceedings of the 42nd Annual Frequency Control Symposium, 1988, p.85
[26] Capelle, Zarka, 43rd Annual Frequency Control Symposium, 1989, p.470
[27] Aubry, Craveur, Deyzac, Fragneau, Janiaud, 1st European Frequency and time Forum, 1987, p.195
[28] Aubry, Canzian, Candelier, 6th European Frequency and time Forum, 1992
[29] private meeting at CNES, Toulouse, France
[30] Zecchini, Lipson, Proceedings of the 43rd Annual Frequency Control Symposium, 1989, p.490
[31] James, Proceedings of the 42nd Annual Frequency Control Symposium, 1988, p.146
[32] Bachheimer, Dolino,Proceedings of the 43rd Annual Frequency Control Symposium, 1989, p.517
[33] Vig, Proceedings of the 45th Annual Frequency Control Symposium, 1991, p.77
[34] Brendel, Gagnepain, Aubry, Proceedings of the 40th Annual Frequency Control Symposium, 1986, p.121
[35] Kusters, Fischer, Leach, Proceedings of the 32nd Annual Frequency Control Symposium, 1978, p.389
[36] Gagnepain, thesis, 1972, Besançon, France
[37] Tiersten, Proceedings of the 28th Annual Frequency Control Symposium, 1974, p.1
[38] Smythe, Proceedings of the 28th Annual Frequency Control Symposium, 1974, p.5
[39] Michel, Zafrany, Soubirous, Mouton, Billon, Berté, Schwartz, Lechopier, Le Corre, this forum
[40] Brauer, Muller, Crystal Res. and Technol.,1984, p.101-104
[41] Jones, Proceedings of the 40th Annual Frequency Control Symposium, 1987
[42] Tellier, Surface Technology, 21, 1984, p.83-89
[43] Brandmayr, Vig, Proceedings of the 39th Annual Frequency Control Symposium, 1985
[44] Brandmayr, Vig, Proceedings of the 40th Annual Frequency Control Symposium, 1986
[45] Bignon, Brottier, 4th European Frequency and Time Forum, 1990, p.609

QUARTZ RESONATORS FOR MARS PLANETARY MISSIONS

Michel E. ROULET, Jean-Luc JOSSET

CSEM, Swiss Center for Electronics and Microtechnology, Inc.

Maladière 71, CH-2007 Neuchâtel (Switzerland)

ABSTRACT

Control and Data Management Systems implemented for planetary probes or automatic vehicles will require real time clocks (RTC) capable of being operated in extreme environment. In order to save power energy, mass and volume, there is a considerable interest to assess the possibility to obtain a RTC, based on a miniature quartz resonator and associated integrated electronics, which can be used in a broad temperature range. This would enable to relax or even avoid thermal control requirements.

This paper describes a preliminary study of a 2.1 MHz RTC based on the ZTY-cut quartz micromachined resonator developed by CSEM under contract for ROLEX for wrist watches. Devices manufactured by ROLEX and encapsulated in its own cold-weld holders have been characterized in the framework of a contract for the European Space Agency (ESA). Results show that this quartz resonator is a good candidate for developing a high-performance MARS spatial RTC.

1. INTRODUCTION

The exploration of MARS was initiated by a planetary mission in 1976-1977 when the Viking probes landed to study the chemical and physical environment of the planet.

Although the probes provided interesting measurements, their recent analysis and results have generated surprising open questions. In the interest of the international scientific community for a Martian mission, ESA has defined a project called "MARSNET" which consists of placing three fixed stations on MARS at different latitudes and longitudes. The scientific study of MARS will evolve in the course of the next decade. A cartography and an environmental study by the MARS OBSERVER satellite of NASA will be made and the exploration of the ground with a rover will be performed.

Planetary missions have a number of criticalities. A critical issue is of course the generation of energy in a harsh environment while solar radiation is reduced because of the distance. Viking probes used radio-isotope thermoelectric generator (RTG). Because of their poor figure of merit regarding yield of electrical energy production, mass, volume and radiation pollution, new missions will avoid to rely on RTG. Low-power consumption of electrical as well as thermal energy is consequently becoming a technological driver for the design of probes.

ESA has awarded to CSEM a project in the framework of the Technological Research Programme (TRP) to explore the possibilities and the technological criticalities of using low-power electronic technologies in extreme environment [1]. This paper is focused on one part of the project, namely the critical assessment of a novel industrial miniature quartz resonator. It might be used as RTC on each probe of MARSNET in particular for the synchronization of seismological measurements. The different RTC can be linked to a master clock placed on the orbiter satellite used as relay satellite.

2. MARTIAN ENVIRONMENT

As a result of the successful Viking missions, there are some data available about the temperature for two different local sites on MARS (Viking lander 1 and Viking lander 2). From these data and thermal mode calculations, it is possible to summarize the temperature variation on MARS versus latitude and time.

For a middle latitude, the temperature variation can be 80° between the daily maximum and minimum. The maximum temperature occurs about 1 hour after noon and the minimum temperature occurs at dawn. This is true during all the Martian year. Furthermore due to the seasonal effect, the temperature could rise up to about 10°C in summer, but could be as low as -130°C in winter. This is shown on Figure 1. It has to be noted that the lower limit of the temperature range is below that of the military range (-55,+125°C). In addition for a typical mission of one Martian year, corresponding to 1.88 years on earth, the number of thermal cycles is 670. This corresponds to a harsh thermal environment. Beside low temperature and rapid thermal cycles, other environmental conditions are not severe (atmospheric pressure of typically 1mb, very weak magnetic fields, low ionizing radiation level).

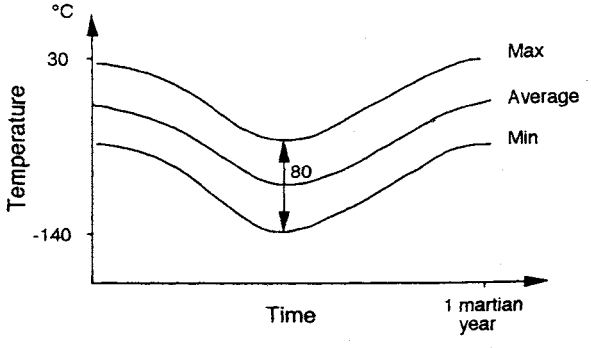


Figure 1 Temperature variations on planet MARS

3. ZTY-CUT QUARTZ RESONATOR

3.1 BASIC CHARACTERISTICS

A few years ago CSEM has developed a high-performance quartz resonator for the Swiss watch-making company ROLEX. The resonator was brought by themselves to their own severe technical requirements. The need for more accurate electronic watches has indeed prompted the development of miniature quartz crystals with better frequency-temperature behavior and lower ageing rates. Among the many efforts devoted to this goal, a new quartz crystal cut has been found at CSEM by applying computerized variation techniques. This so-called ZTY-cut, which makes use of coupling between two elastic modes, has been originally applied to rectangular resonators. This novel resonator has been patented [2] and described in several publications [3] and [4]. The shape of the resonator is illustrated on Figure 2.

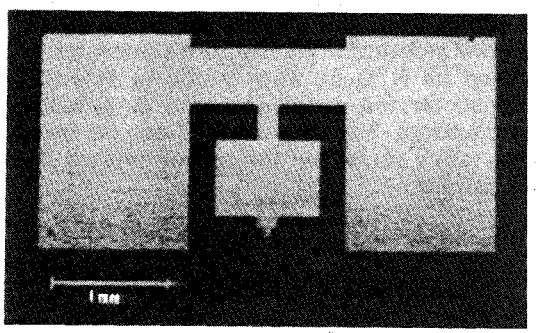


Figure 2 Micrograph of the ZTY-cut quartz crystal resonator

Typical cubic characteristics $\Delta f/f$ versus temperature are shown on Figure 3. This third-order frequency curve is in contrast with the parabolic frequency-temperature curves of the well-known AT- and SC-cut crystals used extensively in various fields. It is important to note that the design of the ZTY-cut quartz resonator can be adjusted to meet specific requirements since the resonator characteristics can be calculated by applying computerized variation techniques. More precisely :

i) the temperature range can be adjusted by varying the temperature of the inflection point;

ii) the slope around the inflection point can also be adjusted by slightly changing the cut angle of the crystal.

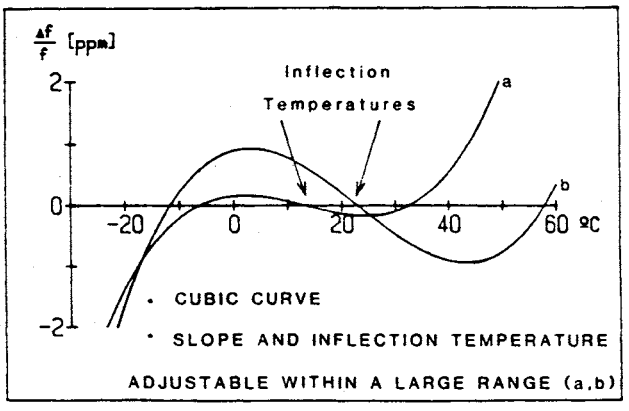


Figure 3 Typical frequency response versus temperature of a quartz resonator designed for room temperature operation

These possibilities of optimizing the design can also be applied to any desired low-temperature range in order to satisfy extreme environment conditions while minimizing temperature effects on the quartz resonator performance.

The typical characteristics achieved for a resonator designed for room-temperature operation are summarized in the following table (Figure 4).

• Frequency (unadjusted)	: 2.1 MHz ± 3 KHz
• Static capacitance	: 2.3 pF
• Motional capacitance	: 0.5 fF
• Quality factor	: > 0.5×10^6
• Low temperature-dependence	: 0 to 40°C < ± 2 ppm
(worst case)	: 20 to 60°C < ± 5ppm
• Ageing (first year, at 28°C)	: 0.5 ppm
• Shock resistance, 1 x 5000g/0.2 ms, 3 axes	: < ± 1 ppm
• Resistance to static accelerations (typical)	: 15×10^3 g
• Static acceleration sensitivity	: 3 to 4 x 10^{-10}g
• Thermal transients sensitivity	≈ 0.1 to 0.2 µs/°C

Figure 4 Typical characteristics of ZTY-cut quartz resonator

3.2 LOW-TEMPERATURE TESTING

3.2.1 DESCRIPTION OF DEVICES UNDER TEST

Several ZTY quartz resonators manufactured and encapsulated by ROLEX have been measured. Their design for a nominal frequency of 2.1 MHz has been made for room-temperature operation, but they have been tested at a temperature as low as liquid nitrogen temperature (-196°C) in order to have a margin for the assessment of their performance with respect to Martian temperature range. The encapsulation has been developed by ROLEX and is based on a proprietary design relying on a so-called "button" packaging. Covar feed-throughs and coldwelding are used. Dimensions are 6 mm of diameter and 2 mm of height. The weight is only 0.3 g. This packaging is large enough to accommodate, if required, an integrated oscillator. In this paper most of measurements performed with the resonator have been carried out at low temperature. The oscillator has not been placed in the "button" packaging. A set of four different devices has been tested.

3.2.2 EXPERIMENTAL TEST SET-UP

A dedicated test set-up has been realized in order to simulate harsh environmental conditions, in particular rapid thermal cycles. Accelerated thermal cycles have to be performed in order to be representative of the 670 cycles corresponding to one Martian year which is the typical duration of a mission. Actually 1000 cycles have been carried out in 28 hours with the set-up

described in Figure 5. One cycle duration is 100 seconds. The encapsulated resonator is directly plunged into liquid nitrogen and subsequently lifted up and heated in front of an infrared lamp. The set-up is programmable and automatically operated. The resonator is not operated during thermal stresses.

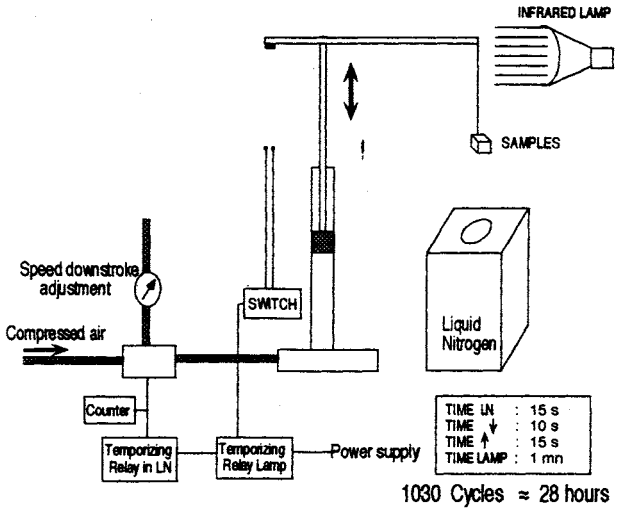


Figure 5 Rapid thermal cycle set-up

Two types of frequency measurements have been carried out with state-of-the-art facilities:

i) off-line measurements at room temperature after an increasing number of thermal cycles;

ii) in-line measurements at fixed temperatures reasonably spread between -196°C and room temperature, the temperature being measured with a calibrated thermo-couple.

For the sake of completeness, a limited number of measurements have also been performed at +165°C with a different test set-up.

The experimental set-up comprises a frequency synthetizer, an oscilloscope, a measuring bridge, a Dewar with liquid nitrogen, resistor heating elements and a calibrated thermocouple.

The process to determine the quartz resonator frequency and its quality factor consists of balancing a measuring bridge made of two fixed and one adjustable resistors. The fourth impedance is given by the quartz resonator itself. At a given temperature, the balancing of the bridge is done with the frequency synthetizer and the adjustable resistor. This measurement yields a good approximation of the series resistance and of the resonance frequency of the crystal, providing that its figure of merit M satisfies the following condition :

$$M = Q \frac{C_1}{C_0} >> 1$$

where :
C1 = motional capacitance of the quartz resonator
$C_1 \approx 0.5$ fF
C0 = static capacitance of the quartz resonator
$C_0 \approx 1$pF
Q = Quality factor of the quartz resonator
$Q \approx 500'000$

Since the figure of merit is about 250, this approach is a very good approximation for determining the frequency and the quality factor of the quartz resonator.
The quartz frequency is measured by tuning the frequency of the synthetizer in order to balance the measuring bridge.
This set-up enables therefore to test the resonator and its encapsulation as a function of temperature and has also the capability to perform thermal shocks as well as accelerated temperature cycles.

3.2.3 TEST RESULTS

Figure 6 shows the frequency-temperature characteristic for temperatures ranging from -196 to +165°C. The 4th-order fit based on experimental data is given in the figure. It is worth noting that the correlation coefficient amounts to 0.99987 and that the low value of the 4th-order coefficient indicates that the curve is almost a cubic.

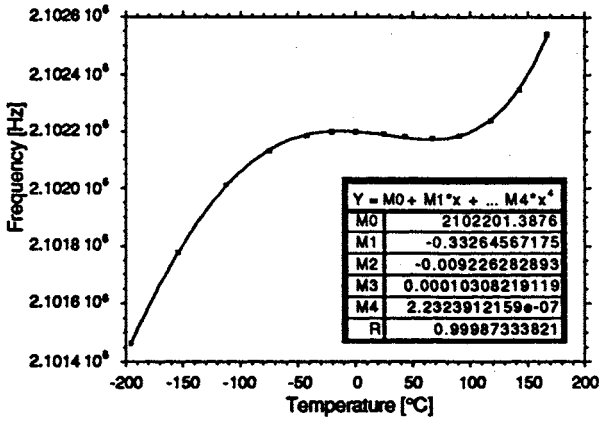


Figure 6 Frequency-temperature characteristic before thermal cycling

The first conclusion to draw is that the theoretical behaviour is verified at low temperature for a design made for room temperature operation. This finding demonstrates that the resonator can be used in an extended temperature range and validates the model applied as well as the computation techniques used.

The relative residual differences of frequencies between the 4th-order fit and measurement data are displayed on Figure 7. For temperatures ranging between -196 and +100°C the differences are smaller than +/- 2 ppm. If required, these differences might even be electronically corrected for a specific device.

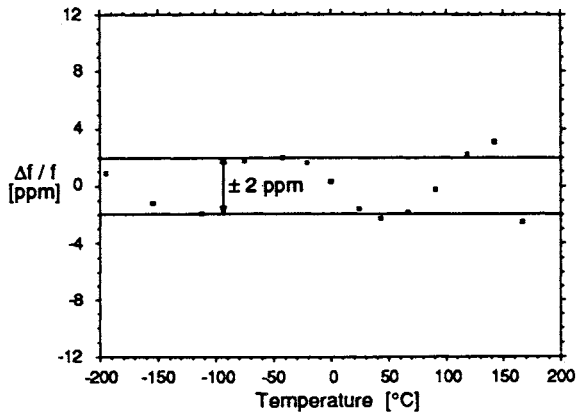


Figure 7 Relative residual differences of frequencies between 4th-order fit and measurement data as a function of temperature

Consistent results have been obtained with the four devices which have been measured in the same temperature range. However, it is necessary to make the 4th-order fit on an individual basis since the nominal frequency differs slightly from one device to the other. Furthermore, results are well reproducible, differences of the order of 1 ppm are given by temperature uncertainties which are measured within +/- 0.1°C.

The quartz resonator encapsulated with the "button" packaging is robust, since the same frequency-temperature characteristic is obtained after 1000 thermal cycles between -196°C and room temperature. Figure 8 shows a typical result obtained with the same device as that of Figure 6. The corresponding 4th-order fit has a correlation factor of 0.99995.

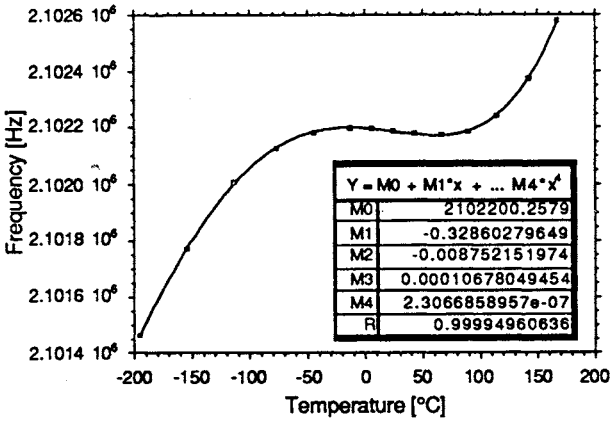


Figure 8 Frequency-temperature characteristic after 1030 cycles between -196°C and room temperature

The issue of ageing is also of paramount importance for long-term applications in space. As already presented in Figure 4, the ageing is small. Figure 9 shows curves measured every 3 months over a period of 5 years for 10 quartz resonators in continuous operation at room temperature. It appears that the nominal frequency increases by less than 0.4 ppm. Actually more than the 50 % of ageing is observed during the first 3 to 6 months and then a very slight increase still occurs (asymptotic behaviour). Similar results are obtained [5] when quartz resonators, stored at room temperature, are not in continuous operation. The ageing phenomenon is supposed to be due to relaxation effects at the quartz-crystal electrode interface and not to any degradation of the piezoelectric material produced by the excitation energy.

The qualify factor of ZTY-cut quartz resonators (typically larger than 0.5 x 10^6) does not vary markedly for temperatures ranging from 196 to 165°C. It is worthwhile to note that no parasitic mode has been observed in that temperature range.

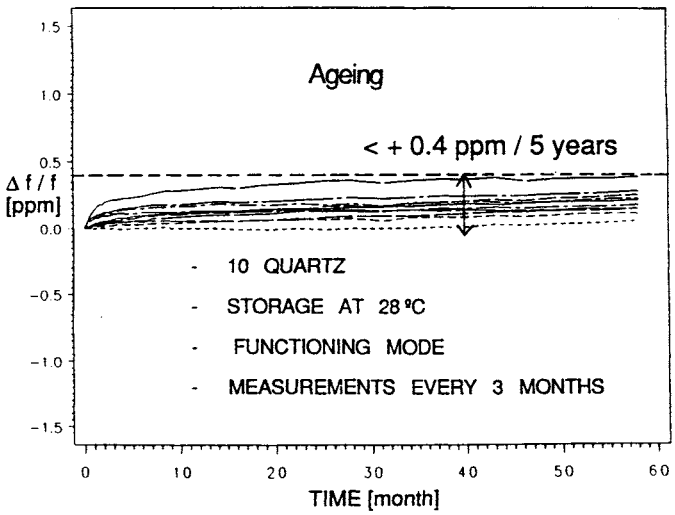


Figure 9 Ageing of ZTY-cut quartz resonators

3.2.4 DISCUSSION

The results presented in section 3.2.3 demonstrate that the ZTY-cut quartz resonator is suitable for low-temperature operation. Figure 10 shows that the design made for room-temperature operation enables to generate a 2.1 MHz frequency signal within +/- 60 ppm for temperatures ranging between -130 and +20°C or within +/- 60 ppm for a reduced temperature range of -100 to +20°C. On the curve of Figure 10, at proximity of the inflection point, a frequency variation of +/- 10 ppm corresponds to a temperature variation of 150°C. Since as explained in section 3.1, the design of the quartz resonator can be modified in order to shift the inflection point at any temperature, the frequency-temperature curve can be shifted towards low temperatures. A design optimized for Martian temperature range, would enable a +/- 10 ppm frequency variation to be obtained. Alternatively, if the quartz resonator is designed for room-temperature operation, an electronic temperature correction based on a look-up table might be implemented with a temperature sensor in order to reduce frequency variations at low temperatures.

Additional testing are being prepared in order to perform 1000 thermal cycles over 3 months, while the quartz resonator is in operation.

Preliminary complementary measurements have shown that the quartz resonator can be operated with an integrated 1.5 V CMOS circuit at a temperature as low as -100°C. This IC which comprises an integrated

oscillator and a frequency divider chain has been bonded to the quartz resonator in the packaging.

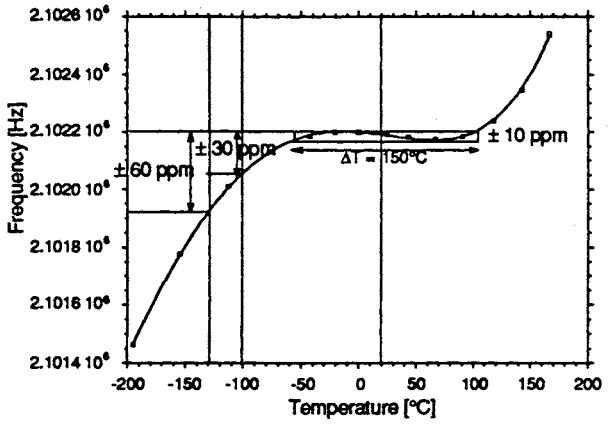


Figure 10 Frequency variations for different temperature ranges of the ZTY-cut quartz resonator designed for room-temperature operation

4. CONCLUSION

This study has assessed the possibility of realizing a 2.1 MHz RTC for planetary missions with the ZTY-cut quartz micromachined resonator developed for wrist watches. This resonator has been designed by CSEM for temperatures ranging from
-10 to 50°C. Devices fabricated and encapsulated with a proprietary design by ROLEX have been characterized in harsh environment representative of missions on MARS. Measurements of the frequency response as a function of temperature have been performed in a broad temperature range (-196, + 165°C). In addition hard thermal cycle stresses have been applied, since for long time duration missions on MARS the temperature range is between -130 and 20°C. The resonators have been subjected to 1030 cycles from -196°C to room temperature corresponding to more than one Martian year. The measurements show that the resonator behaves at low temperature as expected. The 3rd-order theoretical characteristic is obtained without any interference with parasitic modes. A 4th-order fit has been calculated for assessing the resonator performance. Furthermore, the resonator and its encapsulation have sustained quite satisfactorily the severe thermal stresses which have been applied.
In order to reduce frequency variations at low temperatures a significant improvement can be achieved if the quartz resonator design is made for low-temperature instead of room-temperature operation.

In conclusion, this study shows that the ZTY-cut quartz resonator, which has already successfully flown in 1991 as RTC in the G-21 experiment of the Technology Demonstration Programme of ESA [6] is a good candidate for developing a high-performance MARS spatial Real Time Clock of low-power consumption (μw range), low mass (< 1g), and small volume (<0.2 cm^3) and low ageing.
The results obtained demonstrate the capability of this device to be used in extreme environment and should justify further development work aiming at qualifying a RTC for extreme space environment.

ACKNOWLEDGEMENTS

The authors would like to acknowledge valuable discussions with J. Hermann, and the company ROLEX, Geneva, for having made available the devices which have been tested.

REFERENCES

[1] M. Roulet and J.-L. Josset
"Electronic Payloads for Martian Missions : A Preliminary Assessment of Technologies"
Proceedings of 1st International Symposium "Missions, Technologies and Design of Planetary Mobile Vehicles",
Toulouse, Sept. 28-30, 1992
Editor : CNES

[2] Patents : US 4,313,071 / US 4,450,378

[3] J. Hermann
"A Novel Miniature ZT-Cut Resonator"
Proc. of Frequency Control Symposium, Philadelphia, May 29-31, 1985

[4] J.-J. Gagnepain
"Recent Advances in Subminiature/Thin Membrane Resonators
Proc. of 1985 Ultrasonics Symposium, p. 291

[5] Private communications of ROLEX

[6] J. Hermann
"Advanced Real Time Clock for the G-21 TDP Experiment"
CSEM Technical Report No 464 for ESA, October 1991.

Measurement of crystal parameters using a network analyzer in direct impedance mode

H.Matthey
technical responsible
Vibro-Meter SA office Bienne
Ch.de la Scierie 55 CH-2504 Bienne, Switzerland

ABSTRACT

A solution using Network Analyzer HP4195 is described. Some new ways are used to get the dynamic parameters in about 5 seconds in a temperature controled system with sufficient accuracy. The hardware adaptation to Saunders two chamber equipment is described. Obtained results will be discussed.

Keywords: Crystal-parameters, network-analyzer, temperature, hardware

1) INTRODUCTION:

This study was made with the HP 4195 Network/Spectrum analyzer and the impedance test fixture. The measurement is based on the reflection method [1]. The network analyzer has a built-in equivalent circuit analysis, but the results are too much dependent on the settings to be used in an automatic mode. Knowing the equivalent parameters one can adjust the span to get the right value on the display. The crystal is connected between ground and the hot point. Measurements can be made directly with the SMD fixture, or for connecting to Saunders test chambers, we have used 50 Ohm coaxcables. Our way shows some specific solutions to compute the crystal parameters. The internal compensation capability of the network analyzer is used to compensate the line inductivity and capacity even if needed, the crystal wires and case capacity. This capability is well appreciated mainly up to 500 MHz.

2) HARDWARE

One coax-relay switches from one to the other chamber and one coax-relay at each chamber switches between the two rows. It is easy to use the analyzer as a capacitance meter for equalizing the connections and lines up to 500 MHz. For the connection to the test wheel, we have replaced one π by an approximated 50Ω adapted feedthrough and the other by a brass cylinder to ground.The best solutionfor two row systems is to place the ground at the top and the bottom, the 50Ω connections are in the center. The network analyzer has TTL-outputs which can be used as interface between IEEE-bus for driving the relays. One of these TTL-output was also used to switch off the system at the end of the measurement cycle.

3) CALIBRATION AND SETUP

The frequency range from 100 kHz to 500 MHz has been divided in 11 frequency spans in a logarithmic scale. The multiplying factor is about 2.1544. There is an overlap of about 10% included and the last step is extended to 500 MHz. These ranges have been calibrated for open-,short and loaded-output and stored on the floppy disc. Unfortunately it is not possible to load by software these calibrations. The compensation of the test-fixture or the cable to the test wheels has to be done after new start or after change of test fixture, the calibration data has to be loaded from the floppy. Making the compensation, the seen "open" connection will be taken in consideration. If there is an empty crystal base the stray capacity of the base , and by the "short" measurement the length of the wire making the short-circuit will be compensated. This is of importance at higher frequencies .

4) SOFTWARE

The first step is to find the reference frequency, the reference dynamic capacity and the static capacity. Last one will be measured only in the first step. These values are stored in a real array. The used resolutions for later frequency will be 0.1ppm, for the equivalent serie resistance (ESR) 0.1 Ohm and for the dynamic capacity (C1) in 100ppm, stored in a integer array. For the first and the second

temperature step reference frequency and a span of about 300 ppm is used to find fm (f for minimum impedance). For the other the center frequency are extrapolated from the preceding records with a narrow span. So the first sweep starts at about minus 500 ppm from the nominal frequency and stops at nominal +200 ppm with low resolution. The measured frequency for min impedance is used for the next sweep with narrow span and high resolution. Then the frequency for minimum impedance is found, the value of impedance and the phase are measured. For the motional capacitance, we took following steps. The line cursor is put at ESR*sqrt(2). Using the Q-value from the built-in program, we get

$$C1 = 1 / (2*\pi*fm*ESR*Q) \qquad (1)$$

For the static capacitance, start with the frequency for minimum ESR fm, stop at +500 ppm. Ask frequency at max impedance, parallel resonance fp.

$$Co = C1*fm / (2*(fp-fm)) \qquad (2)$$

The stored resonance frequency is corrected for the phase shift and the Co effect by using the approximation given by Roger J. Williamson [3]:

$$fr = fm *(1+ (\tan(-\theta e)/(2*Q)) \qquad (3)$$

with

$$\theta e = \text{Phase(at fm)} + \arctan(ESR/XCo) \qquad (4)$$

It is also possible to select the oscillator output level for adjusting crystal power. The power level (P) at the crystal is:

$$UOUT = 1/2\ Uosc*\ ESR/(ESR+50) \qquad (5)$$

$$P = (UOUT)^2 / ESR \qquad (6)$$

5) CAUTION AND TIPS

At exactly 10 MHz some switches of filters are operated by the network analyzer and the sweep-oscillator has a small jump. If the slope of the measured crystal between serie- and parallel resonance is at exactly 10 MHz, then a step appears in the curve. This can be eliminated by using a narrow span around 10 MHz. In the dialog between the computer and the analyzer care must be taken by the reaction time of the analyzer specially for the service request signal. The delay between instructions must be sufficient for the signal to disappear. In the programming example [2], the trigger routine has to be corrected in the following way:

ON INTR 7,*Mask* GOTO..

The routine doesn't work if Mask is omitted. It is, instead of using the Q information, better to take the upper and lower frequency of the bandwith. If any frequency is negative (-99999), it means that this point is out of range. So center frequency or span can be corrected.

6) RESULTS

The serie resonance frequency will be shifted up if using the sweep mode up and down if sweep mode is down. The amount of shift is sweep-time dependent and quite constant and if needed can be compensated by the program. Table 1 gives some comparison between our measurements and measurements made by an other manufacturer . The left column are obtained with our network analyzer with a 60 cm long coaxcable.
Comparing the results we discover a deviation for the C1 measurements at 16 MHz. The measured Q value by the bandwith method at sqrt(2)*ESR was higher with the coaxcable than direct on the SMD-fixture of the 41951A,which correlate with the CI-meter 150 and the π 444. Table 2 gives the comparison for C1 of the three test methods. A clear explanation is not found. The 155 MHz crystals where compared with the value of the manufacturer. The correlation is, keeping in mind the short measuring time (5s), fairly good.

Analysis of the parameters:

Series-resonance frequency:

comparing 25 peaces of 5.075 MHz between this work and CI-meter 150 as reference, give this figure:

frequency deviation	
mean value	-6.52 ppm
stand. dev.	0.552 ppm

with correction of + 6.52 ppm each frequency:

frequency deviation	
mean value	-0.459 ppm
stand. dev.	0.552 ppm

For Co measurement:
16 MHz crystals:

this method mean val.	1.962 pF
measured at 18 MHz	2.024 pF
delta this method-"18 MHz"	0.062 pF
stddev.	0.070 pF

C1 method (1)
Analyzing table 1 we find:
Value from this work compared with CI-meter 150 and π with HP-4194, where C1 is computed from the phase slope at zero phase (graphic method):

CI 150	mean dev.	0.3 %
	stand. dev.	4.0 %
π-4194	mean dev.	3.1 %
	stand dev.	2.1 %
CI-150 to π-HP4194:		
	mean dev.	-1.8 %
	stand. dev.	2.3 %

The frequency offset (figure 1) is due to the sweeping speed and direction. It is dependent on the frequency and the quality factor of the measured crystals. It can easily be taken into account. The quickest way to find the offset is to take the difference of the frequency found by up and down sweeping the crystal. This is supported by the standard deviations of the offsets (figure 2).

Figure 3 gives the average measuring time per part of a run for each temperature steps. The first step is always longer, because there is no possibility for extrapolation of the frequency and we have adopted one sweep more to find the frequency. All other steps are between 4 and 5 sec per unit.

7) CONCLUSION

The network analyzer is competitive in speed with other systems and permits an easy use, by non specialists, up to 500 MHz. In addition we get the slope of C1 (figure 4) in function of the temperature, and Co is also measured in the frequency range where the resonator is used.

8) ACKNOWLEDGEMENTS

The author would like to thanks Dr. W Zingg from Micro-Crystal Grenchen for the comparison measurement of the Xtal data.

9) REFERENCES

[1] M. HONDA The Impedance Measurement Handbook A Guide to Measurement Technology and Techniques Yokogawa-Hewlett-Packard Ltd.

[2] HP 4195A Network/Spectrum Analyzer operation manual

[3] An Improved Method of Measuring Quartz Crystal Parameters, Roger J. Williamson IEEE Transaction on ultrasonic, ferroelectrics and and frequency control Vol. UFFC-34 No.6 November 1987

Table 1 : Measurements this work, by u-Crystal with CI150 and HP4194

	value this work S&A HP 4195 with 60 cm coax			value from u-Crystal CI 150			HP 4194 π-424		
#	F (MHz)	ESR	C1 fF	F (MHz)	ESR	C1 fF	F (MHz)	ESR	C1 fF
1	4.0958888	44.8	1.63	4.0959089	47.0	1.56	4.0958890	44.3	1.64
2	4.0959034	46.3	1.57	4.0959247	49.0	1.50			
3	4.0958481	43.8	1.62	4.0958694	42.0	1.58	4.0958512	44.3	1.69
4	4.0958449	78.6	1.58	4.0958629	77.0	1.54			
5	5.0750879	93.6	1.42	5.0751297	106.0	1.42	5.0750845	97.0	1.52
6	5.0750512	76.3	1.28	5.0751011	73.0	1.28			
7	5.0750555	108.6	1.41	5.0751039	77.0	1.40			
8	5.0752629	57.9	1.37	5.0753200	62.0	1.33			
9	5.0751294	84.7	1.31	5.0751789	92.0	1.31	5.0751243	89.3	1.33
10	7.9999796	39.0	2.34	8.0000200	38.0	2.32	7.9999780	38.5	2.41
11	8.0000192	51.8	2.31	8.0000530	47.0	2.31	8.0000180	55.8	
12	8.0000155	33.0	2.38	8.0000576	32.0	2.33	8.0000165	33.5	2.40
13	8.0000178	39.0	2.36	8.0000554	39.0	2.34			
14	15.9970474	16.0	8.135	15.9970265	15.0	8.37	15.9970670		8.38
15	15.9968911	12.0	8.073	15.9968640	12.0	8.31	15.9969050		8.35
16	15.9971447	8.9	8.074	15.9971201	9.0	8.29	15.9971550		8.33
17	15.9970589	7.7	8.006	15.9970373	8.0	8.23	15.9970750		8.28
18	15.9970975	9.6	7.709	15.9970802	10.0	7.93	15.9971060		7.97
19	19.9964662	13.4	9.180	19.9962934	13.0	9.64	19.9964800		9.68
20	19.9962459	10.7	9.466	19.9960525	10.0	9.96	19.9962440		10.00
21	19.9962008	12.9	9.520	19.9959945	12.0	10.01	19.9962000		10.00
22	19.9956558	10.5	9.570	19.9954657	10.0	10.06	19.9956650		10.00
23	19.9958173	10.8	9.585	19.9956864	12.0	10.04	19.9958300		10.10
24	65.9915250	9.7	12.634				65.9917000	8.2	12.35
25	65.9972144	26.4	10.925				65.9974500	24.6	11.08
				Toyokom					
26	155.501150	41.3	0.392	155.5016850	41.3	0.38			
27	155.501200	40.3	0.410	155.5017850	40.3	0.40			
28	155.500550	37.2	0.416	155.5009350	37.2	0.41			
29	155.501450	73.3	0.389	155.5021350	75.0	0.34			
30	155.501950	42.3	0.390	155.5024350	42.2	0.38			

Table 2 : Analysis of C1 between Micro-Crystal (MC) S&A and SMD

8 MHz crystals

#	MC	S&A	SMD	MC /S&A	MC /SMD	SMD /S&A
1	8.37	8.135	8.312	1.029	1.007	1.022
2	8.31	8.073	8.239	1.029	1.009	1.021
3	8.29	8.074	8.281	1.027	1.001	1.026
4	8.23	8.006	8.209	1.028	1.003	1.025
5	7.93	7.709	7.878	1.029	1.007	1.022
			mean	1.028	1.005	1.023
			std.dev	0.001	0.003	0.002

19 MHz crystals

#	MC	S&A	SMD	MC /S&A	MC /SMD	SMD /S&A
1	9.64	9.180	9.535	1.050	1.011	1.039
2	9.96	9.466	9.565	1.052	1.041	1.010
3	10.01	9.520	9.689	1.051	1.033	1.018
4	10.06	9.570	9.903	1.051	1.016	1.035
5	10.04	9.585	9.679	1.047	1.037	1.010
			mean	1.050	1.028	1.022
			std.dev	0.002	0.012	0.012

S&A: on Saunders with 60 cm coax cable
SMD: on SMD fixture of HP 41951A

Figure 1: Frequency measurements
average value of offsets

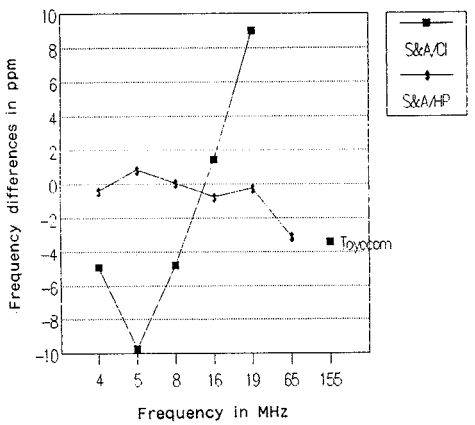


Figure 2: Frequency measurements
standard deviation of offset

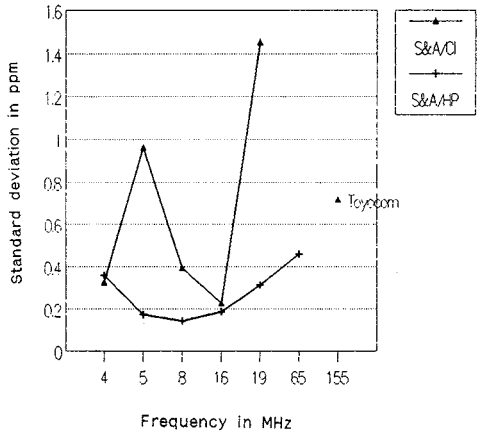


Figure 3 average measurement times per part

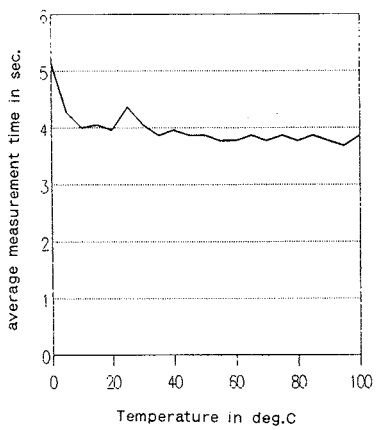
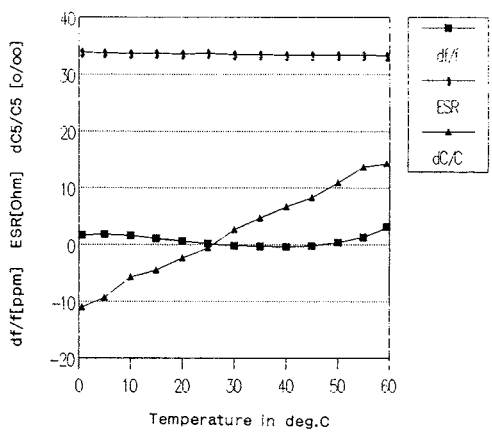


Figure 4: 155.5 MHz crystal



SMALL FREQUENCY/TEMPERATURE PERTURBATIONS IN RESONATORS FOR CRYSTAL OSCILLATORS: THEIR EFFECTS ON SYSTEM APPLICATIONS AND HOW TO MEASURE THEM

P.E. Morley, J. Dowsett and F. Ross

C-MAC Quartz Crystals Ltd
Edinburgh Way, Harlow, Essex, England

ABSTRACT

Recent oscillator requirements have necessitated severe scrutiny of quartz crystals for the presence of small frequency/temperature perturbations. In the past, continuous bandrun methods were used for measurement of frequency/temperature performance of individual crystals, although they were often rather crude and would not have detected very small frequency jumps. More recently, with the introduction of computerised methods of measurement, the continuous measurement has been largely replaced by methods which measure multiple crystals at discrete temperatures. In this type of measurement, there is clearly a compromise between the measurement time and the temperature interval. In this paper a measurement method is described which is essentially pseudo-continuous, together with the presentation of typical results. This technique can be used to screen crystals for optimum oscillator performance.

Keywords: Quartz, resonator, oscillator, frequency, temperature, GPS.

1. INTRODUCTION

The frequency/temperature performance requirements of some recent designs of temperature compensated crystal oscillators (TXCOs) have necessitated severe scrutiny of testing methods for such devices. One of these requirements is the detection and elimination of rapid frequency changes over very small temperature ranges. Under most circumstances, this type of effect is not a problem, but in some system applications, particularly where the oscillator is used as a phase reference in GPS (Global Positioning System) receivers, some system degradation can result.

To manufacture an oscillator with a minimum of such frequency transitions at a reasonable yield, it is necessary to design and manufacture crystals which are devoid of small band-breaks, and, for this purpose, a measurement system has been developed to screen crystals for this parameter. This paper summarises the problems which are caused by these small perturbations and outlines a method for measurement of crystals for characterization of the effect.

2. OSCILLATOR PERFORMANCE REQUIREMENTS

Specifications for TCXOs have been becoming progressively tighter since their development in the late 1960s. The use of integrated circuit methods for temperature compensation has facilitated the design of oscillators with tolerances as low as ±0.5 ppm over wide temperature ranges [1]. A typical oscillator plot is shown in Fig. 1. Since, in this type of device, the compensation function is at most a 5th order polynomial, the crystal frequency/temperature curve must be quite smooth in order to achieve this level of specification.

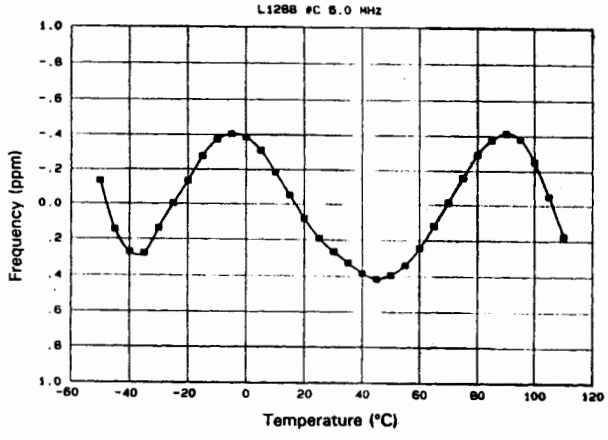


Fig. 1. Typical TCXO Performance Using 6.0 MHz Fundamental Crystal

In recent years, several applications have arisen, particularly for systems such as GPS, which require a temperature compensated oscillator which has constraints applied not only to the frequency itself, but also to the first and second derivatives of frequency versus temperature. In a typical application, the oscillator is configured as TCVCXO (temperature compensated voltage controlled crystal oscillator), and is used as the local phase reference oscillator, phase locked to the primary GPS frequency of 1575.42 MHz. This coherent phase reference is used to synchronise data received from the GPS satellites, and so must remain stable. Although sophisticated second or third order phase lock loop filtering techniques can be used to retain this coherent condition, sudden frequency jumps

in the local oscillator, however small, cannot be tolerated by such a system, and result in either the complete loss of lock or in sudden loss of synchronisation between the two systems [2]. The latter can result in the satellite data being garbled until the receipt of the preamble from the next subframe of data. Digital compensation methods are particularly unsuitable for this type of application because of the discrete nature of a typical compensation function.

The initial measurement problem is that of characterizing the oscillator. This is exemplified in Figs. 2 and 3, which

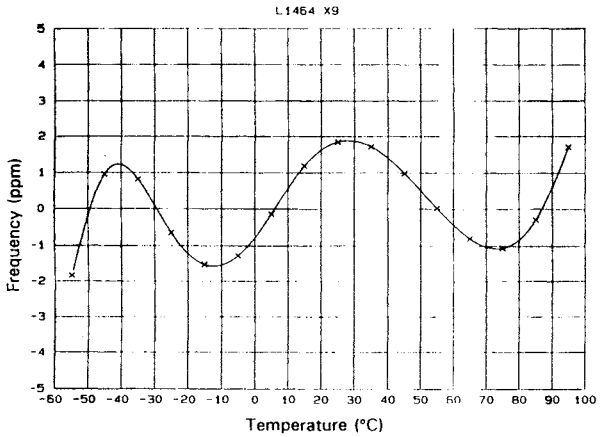


Fig. 2. Oscillator Characteristic Using 40.92 MHz Third Overtone Crystal

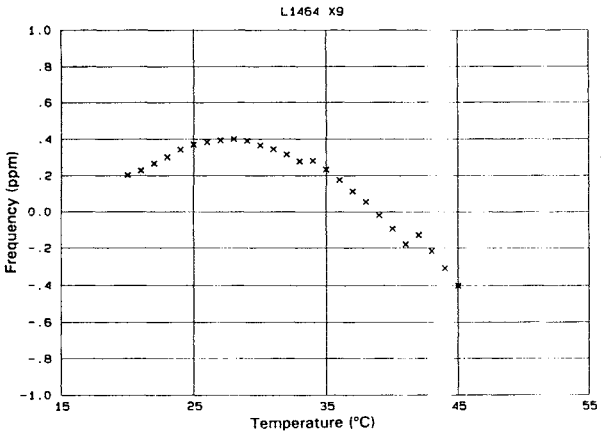


Fig. 3. Same Oscillator Showing Detail of Frequency/Temperature Perturbations

show a reject unit. The device uses a 40.92 MHz third overtone crystal, and the frequency specification is ±2 ppm over the temperature range -55°C to +95°C. Clearly, a temperature step of 10°C does not resolve the problem, but, as shown in Fig. 3, there is a frequency perturbation at about 42°C which is resolved by testing with a 1°C step. The picture can be further enhanced by using an HP5372A Frequency and Time Interval Analyzer or similar instrument, which monitors the oscillator output while the temperature is ramped continuously. In the plot of Fig. 4 (in which the oscillator output frequency has been divided by 40), the

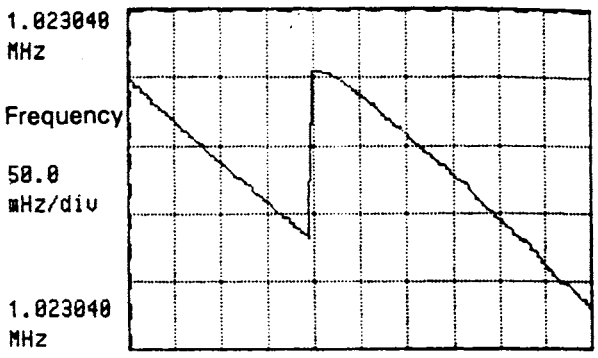


Fig. 4. More Detailed Illustration of Higher Temperature Anomalous Effect

resolution is about 0.02°C, which illustrates the narrow temperature range over which the perturbation occurs.

The next problem is to characterize the oscillator crystal in order to screen out units which exhibit this phenomenon.

3. CONVENTIONAL FREQUENCY/TEMPERATURE MEASUREMENT METHODS

3.1 Continuous Methods

Many systems have been developed in the crystal industry for continuous characterization of crystal frequency and equivalent series resistance (ESR) with respect to temperature. The usual method is to cool the crystal in a copper or aluminium block by some crude method (such as immersion in dry ice or the application of a liquid CO_2 jet onto the surface of the block), and then for heat to be applied by a variety of controlled means. The crystal is connected to an oscillator circuit which applies a similar load and drive condition to that in which it will eventually be used. The temperature, frequency and activity are continuously monitored and an analogue plot is obtained directly from this information. Such analogue systems have been usefully employed for analysis of anomalous crystal behaviour [3].

Whilst this method is able to identify fairly small band-breaks, it does have a number of significant limitations. Firstly, it only lends itself to measurement of single crystals, and is therefore very slow and difficult to automate. Secondly, it is difficult to estimate the exact crystal temperature as it is being ramped, so even if data were recorded for later analysis, interpretation would be difficult. Also, band-breaks are difficult to identify when the frequency/temperature slope is high. However, such methods are useful for checking the proper functioning of more sophisticated systems.

This method is now only recommended for the detection of the presence of trapped water vapour or other undesirable volatiles in the crystal package. The CECC Generic Specification for Crystal Units [4] has such a method defined in Test B.

3.2 Discrete Temperature Methods

Commercially available systems, such as the range manufactured by Saunders and Associates, Scotsdale, Arizona, now perform the vast bulk of measurements of crystal parameters over temperature. In such systems [5], a batch of crystals is placed in a test fixture which is contained in a temperature-controlled chamber. A measurement network is situated outside the chamber, such that impedance measurements are not affected by the temperature dependence of any of the network elements. The crystals are measured sequentially by a multiplexing arrangement: in the case of Saunders equipment, the multiplexing is in the form of open contacts; other systems use mechanical relays for this purpose. The transmission characteristics of the crystal in the network are analyzed to determine the impedance characteristics of the resonator. A computer system controls the equipment, setting the chamber temperature to each prescribed test temperature, and then measuring all crystals at each temperature. Typically, the temperature step is set to 5°C to 20°C, depending on the application. The data are then analyzed at the end of the run as required by the specification.

The advantages of such systems are obvious: they are fully automatic, measuring many crystals simultaneously, and the temperature setting can be very accurate. Data are stored on disc and can be manipulated as required while the system is measuring the next batch of crystals. Also, the stabilisation time can be optimized for the particular crystal package to reduce the thermal hysteresis of the system.

The main problem caused by the discrete temperature measurement method is that sharp band-breaks can be completely missed, particularly if the temperature step is 10°C or greater. Another problem is one of contact integrity, which tends to limit the actual instrument resolution so that smaller temperature steps only detect noise in the measurement method.

An outline of discrete temperature methods is also described in Test A of the relevant section of the CECC specification [4], and is used primarily to define accurately the general form of the crystal frequency and resistance behaviour with temperature without the advantages of an almost continuous plot.

4. DEVELOPMENT OF PSEUDO-CONTINUOUS METHOD FOR CRYSTAL MEASUREMENT

The method that has been developed for screening oscillator crystals for small frequency/temperature perturbations is a slight modification of the conventional discrete temperature method. The system schematic is shown in Fig. 5. The crystal is measured as per IEC-444 [6], such that the zero phase frequency f_r is continuously monitored while the temperature is changed.

The system is fully loaded with crystals and then the chamber temperature is set to 5°C below the specified temperature range. The first crystal is then moved into the measurement position, and it remains in this position while the temperature is increased steadily through the

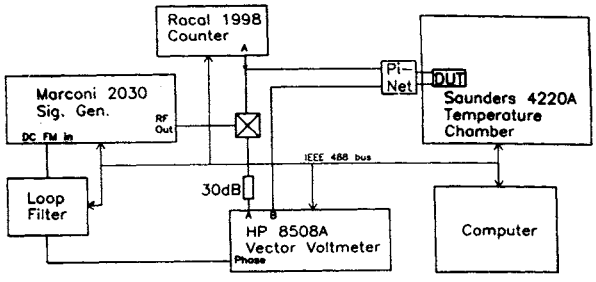


Fig. 5. Schematic of Test System for Pseudo-Continuous Test

specified range to 5°C above the operating range. The rate of change of temperature can be adjusted as required, but is typically set between 5°C and 10°C per minute. The zero phase frequency is measured with a high-resolution frequency counter at a rate of about 10 readings per °C, and the ESR and temperature are similarly monitored. The data are stored onto a network disk for later retrieval. The next crystal is then moved into the measurement position and the temperature is ramped downward while measurements are recorded. This is repeated until all crystals have been measured.

The data are now applied to a curve fitting routine. This calculates the least squares fit of frequency versus temperature to a fifth order polynomial, and the least squares fit of resistance data to a linear function. The residuals can then be plotted, resulting in an easily interpreted demonstration of anomalies in the crystal characteristics.

This test avoids the problem of making and breaking contacts to the resonator, and gives results which closely resemble continuous plots. On the negative side, the treatment of the temperature chamber is somewhat severe, so there may be some long-term consequences of this. Also the use of coolant gases and electric power for heating is much greater than for the standard test. The thermal hysteresis of both system and resonators prevents this method from being used for accurate determination of the general form of the crystal frequency/temperature curve. It does, however, provide an excellent automatic alternative to the continuous method.

5. RESULTS

Typical outputs obtained from this system are shown in Figs. 6, 7 and 8. Figure 6 shows a response which is almost free from perturbations. Figure 7 shows a crystal curve with a number of fairly smooth band-break phenomena while Fig. 8 gives an illustration of a discontinuous curve.

The system has also been used to characterize a number of crystal units still further by repeated measurements of the same unit at progressively increasing drive levels. This is shown in Figs. 9 and 10, in which a number of test results are superimposed on the same graphs. The vertical axes are intended only to illustrate the scale of frequency and ESR variations and do not represent discrete frequency changes with drive level. Figure 9 shows examples of non-linearly coupled modes which

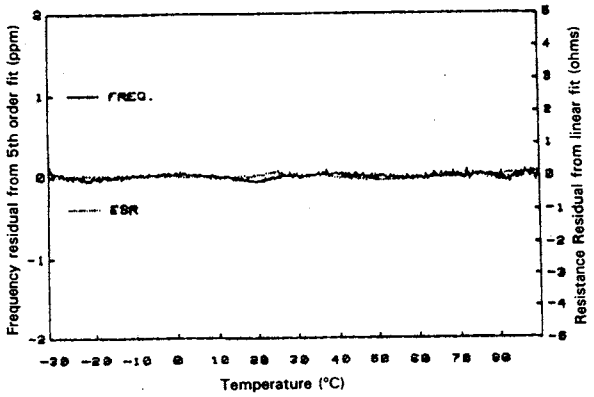


Fig. 6. Crystal Curve Showing no Significant Frequency Perturbations

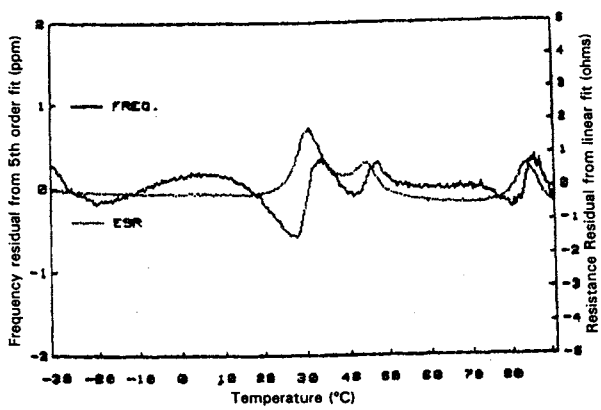


Fig. 7. Crystal Curve With Smooth Bandbreak Characteristics

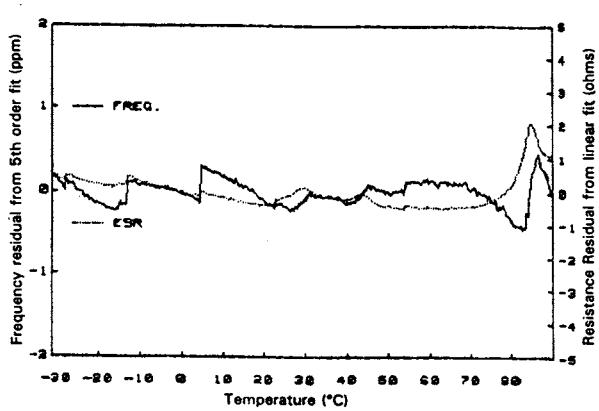


Fig. 8. Crystal Curve with Discontinuous Frequency Perturbations

become more evident at higher drive levels [7,8]. Figure 10 gives an example of a crystal which exhibits one mode which is linearly coupled to the main resonance, and therefore perturbs the frequency versus temperature curve in a way which is largely independent of drive level, and also exhibits a mode which has non-linear coupling. Clearly, in both of these sets of curves, it can be seen that the level of noise in the measurement is most severe at the lowest drive level.

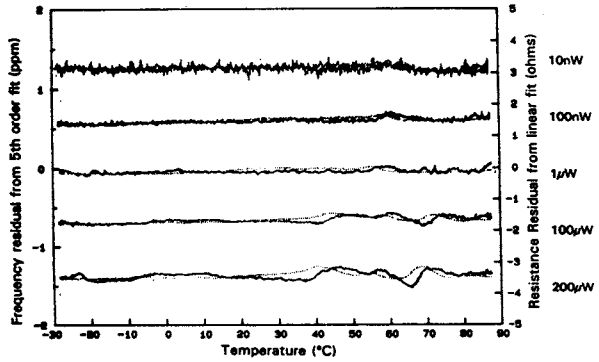


Fig. 9. Example of the Effect of Drive Level on Bandbreaks

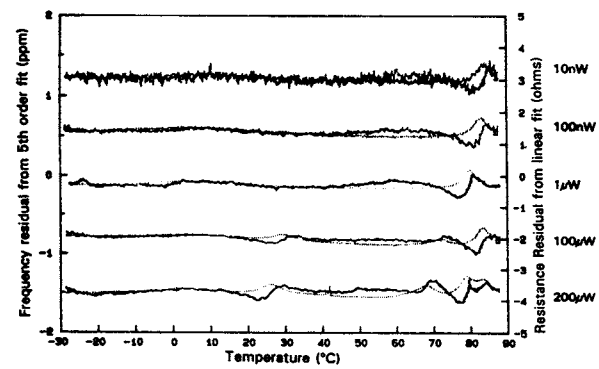


Fig. 10. Example of the Effect of Drive Level on Bandbreaks

6. CONCLUSIONS

As a result of increasing severity of crystal specifications as required by temperature compensated oscillators, a technique has been developed which analyzes crystal impedance characteristics for anomalous behaviour when a continuous temperature ramp is applied. The system used is a standard Saunders 4220A test chamber for adjusting the temperature, and the crystal measurement is based on the method described in IEC-444 [6].

This method is able to resolve band-breaks where the frequency variation is less than one part in 10^7, and can detect anomalies which span a temperature range of less than 1°C, although the exact temperature at which anomalies occur can only be estimated to within ±5°C. The test is already being applied to crystals for internal requirements, and for the full characterization of some crystals for space applications.

The pseudo-continuous test was discussed in a meeting of the CECC Working Group 17 on 5th March 1993, and is proposed as an amendment to the CECC Generic Specification for Quartz Crystal Units [4], to be known as Test C under Section 4.7.3. However, because of the effect of thermal gradients on crystal frequency, this test can be used as an alternative to Test B, but should not be combined with Test A, where a high accuracy is required to define the full frequency/temperature characteristics.

7. ACKNOWLEDGEMENTS

The authors would like to thank Gary Waggitt for his help in assimilating vast quantities of crystal data from which the above results were selected.

8. REFERENCES

1. An Improved Method of Temperature Compensation of Crystal Oscillators, J.S. Wilson, 37th Annual Symposium on Frequency Control, 1983, pp. 442 to 453.

2. Multipath Sensitivity and Carrier Slip Tolerance of an Integrated Doppler DGPS Navigation Algorithm, Proc. IEEE Position Location and Navigation Symposium, March 1990, pp. 638-644.

3. Results of Temperature Slewing Quartz Crystals for Anomalous Responses, M.B. Bloch, M.P. Meirs and A. Strauss, 32nd Annual Symposium on Frequency Control, 1978, pp. 344-353.

4. CECC 68000, Generic Specification: Quartz Crystal Units, Issue 1, 1989, Section 4.7.3: Frequency and Resonance Resistance as a Function of Temperature.

5. General Xtal Test Methods Related to Oscillators and Transmission Instruments, J. Saunders, 1981 Quartz Crystal Conference, pp. 210-228.

6. IEC Publication 444-1, Measurement of quartz crystal unit parameters by zero phase technique in a π-network, 2nd Edition, 1986.

7. Activity Dips in AT-Cut Crystals, A.F.B. Wood and A. Seed, 21st Annual Frequency Control Symposium, 1967, pp. 420-435.

8. On Activity Dips of AT Crystals at High Levels of Drive, C. Franx, 21st Annual Frequency Control Symposium, 1967, pp. 436-454.

TIME SCALES AND INSTRUMENTATION

Chairman: A. Bauch

PRACTICAL STANDARDS FOR PM AND AM NOISE AT 5, 10, and 100 MHz

F. L. Walls
Time and Frequency Division
National Institute of Standards and Technology
Boulder, CO 80303

Abstract

This paper describes a practical implementation of secondary standards for phase modulation noise (PM) and amplitude modulation noise (AM) at carrier frequencies of 5 MHz, 10 MHz, and 100 MHz. In each case the standard simultaneously produces output signals at approximately + 15 and + 17 dBm. The PM and AM noise between the two signals can be set to a calibrated level, which is constant to better than 0.2 dB for Fourier frequency offsets from nearly dc to 5% of the carrier, or set to a minimum value. The calibrated level of PM and AM noise is used to evaluate the accuracy versus Fourier frequency. The accuracy of determining the calibrated PM and AM noise in the standards is ± 0.14 dB for both PM and AM noise at carrier frequencies of 5, 10, and 100 MHz. The temperature coefficients of the calibrated PM and AM noise are less than 0.02 dB/K. The minimum differential noise is used to evaluate the noise floor versus Fourier frequency of PM and AM measurement systems. When the different noise is set to a minimum, the PM noise S_ϕ(10 kHz) between the two signals (PM noise floor of the standard) is less than -190 dB relative to 1 rad^2/Hz at 5, 10, and 100 MHz. The noise floor for AM measurements depends on the configuration. Using two AM detectors in a differential configuration, we have calibrated AM noise detectors with a noise floor below -180 dBc/Hz at all the test carrier frequencies. These noise standards can also be used as the calibration standard for determining the conversion sensitivity of the detector and amplifier gains for PM/AM measurements.

1. INTRODUCTION

In the past it has been difficult to evaluate the accuracy and noise floor of phase modulation (PM) and amplitude modulation (AM) noise measurement equipment because no artifact standards were available. Comparisons of measurement systems in different laboratories using commercially available oscillators as transfer standards were typically limited to a repeatability of roughly ± 3 dB due to the temporal variability of the oscillator noise. This paper describes a practical implementation of portable secondary standards and associated measurement techniques for evaluating the noise floor and accuracy versus Fourier frequency of phase modulation (PM) and amplitude modulation (AM) noise measurement systems [1,2]. In each case the standard simultaneously produces output signals at approximately + 15 and + 17 dBm.

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

The PM and AM noise between the two signals can be set to a calibrated level, which is constant to within ± 0.2 dB for Fourier frequency offsets from 1 Hz to 5 % of the carrier, or set to a minimum value for determining the noise floor. Evaluations of these new PM/AM noise standards at 5, 10, and 100 MHz yield an accuracy of 0.14 dB, a temperature coefficient less than 0.02 dB/K, and a stability better than 0.4 dB over 1 y. The differential PM noise between the two output signals is typically less than -190 dBc/Hz for Fourier offsets of 10 kHz, which is much less than any presently available source. These features make these standards powerful tools for evaluating the performance of PM and AM measurement systems.

These new PM/AM noise standards can also be used to calibrate the conversion sensitivity of the PM/AM detector for PM or AM measurements. This greatly simplifies the measurements since detector sensitivity and amplifier gains versus Fourier frequency are determined simultaneously with an accuracy of better than ± 0.5 dB.

2. DESIGN OF PM/AM NOISE STANDARD

Figure 1 shows a simplified block diagram of the PM/AM noise standard model PMAM 115. A frequency source with very low PM and AM noise is regulated in amplitude and divided into the reference and signal outputs using a passive power splitter or directional coupler. (Active power splitters generally add more PM noise than do passive splitters.) A switch turns the carrier signal "on" and "off." The amplitude and phase of these outputs track one another with great fidelity. The residual differential PM noise between the two outputs is typically much smaller than the resolution of available measurement systems [3-5]. At 10 MHz, for example, the differential phase noise between the two channels, $1/2S_\phi$(10 kHz) is -197 dBc/Hz. This configuration is used for measuring the noise floor of PM measurement systems.

A broadband power combiner is used to add passband-limited Gaussian noise to a carrier (see Fig. 1). The level of added noise is roughly 40 dB above the noise floor of most measurement systems and 60 dB above the residual noise between the two signals. This calibrated noise is typically constant in magnitude to ± 0.2 dB for Fourier frequencies from dc to 5% of the carrier frequency at 5, 10, and 100 MHz. This configuration is used to evaluate the accuracy of PM/AM noise measuring equipment as a function of Fourier frequency. A switch can be used to change the noise level by inserting known

attenuators. This results in a spectral density of PM noise, $S_\phi(f)$ given by

$$S_\phi(f) = \frac{PSDV_N(\nu_0 - f) + PSDV_N(\nu_0 + f)}{2V_0^2}, \quad (1)$$

where ν_0 is the carrier frequency, V_0 is the amplitude of the carrier, and $PSDV_N(\nu_0 \pm f) \equiv V_N^2(\nu_0 \pm f)$ per hertz is the power spectral density of voltage noise at frequency $\nu_0 \pm f$. The level of $S_\phi(f)$ is constant from dc to approximately the half-bandwidth of the filter, assuming that the added noise is constant in amplitude from ν_0 - f to ν_0 + f. A switch connects either a 50 Ω termination or the output of the filtered noise source to the summing device. The summing device linearly adds the noise from either the 50 Ω termination or the filtered noise source with the signal to create a modulated output. Since there is no phase coherence between the signal and the noise, the resulting modulated output has precisely equal AM and PM noise with the condition given by Eq. (2) (assuming negligible amplitude compression in the power summer, which is easily attained) for the same ranges in analysis frequencies about the carrier [3-5]. These noise spectra can be made very nearly constant over a very wide temperature range by stabilizing the carrier and the noise separately using traditional approaches. The level of the added noise is such that

$$\int_0^\infty S_\phi(f) < 0.01. \quad (2)$$

This ensures that the compression of the measured $S_\phi(f)$ is smaller than 0.04 dB [3-5]. The residual differential PM noise between the two signals when the noise source is off is generally much smaller than the PM noise of the source. The ratio of the differential PM noise to the source PM noise can approach -100 dB at low Fourier frequencies, degrading to approximately -16 dB at f = $\nu_0/10$ [5]. This insensitivity to source noise improves the accuracy of the PM standard, especially at low Fourier frequencies where the PM noise of the source becomes important.

Measuring $PSDV_N$ and V_0^2 separately calibrates $S_\phi(f)$ at the modulation output relative to the reference signal, as specified by Eq. (1). Turning the filtered noise off allows good resolution and accuracy in the measurement of V_0^2 by eliminating or greatly reducing V_N. Turning the noise on and the carrier off optimizes the resolution and dynamic range for measuring $PSDV_N$. Usually the attenuation of the noise signal is set to a minimum for this measurement. The components are chosen to minimize the voltage standing-wave ratio (VSWR) so that changing the state of the various switches does not significantly alter the impedance or phase of the output to the measurement system. The attenuators are calibrated using a source routed through the external port.

Implicit in these discussions is the assumption that the phase detector in the PM noise measurement system under test has sufficient discrimination against AM noise and that the amplitude detector in the AM noise measurement system has sufficient discrimination against PM noise. These assumptions are not excessively restrictive since many devices to be calibrated also have similar PM and AM noise [5]. Discrimination of 15 dB reduces the unwanted effects below 0.14 dB. This discrimination is easily met by virtually all AM and PM noise measurement techniques in use today [4,5] (typical levels of discrimination are 25 dB for PM and even higher for AM measurement systems.)

When the AM noise is measured on a single channel, the differential technique applied to $S_\phi(f)$ does not help to extend the range. The AM noise of the signal contains both the original AM noise of the source after the leveling circuit and the calibrated AM noise. Thus, it might be expected that the spectrum of $S_a(f)$ at the output would rise much faster than the PM noise as f approaches 0. Fortunately, most sources have much lower AM noise than PM noise close to the carrier, so this potential problem is usually eliminated.

When the AM noise is measured using two detectors in a differential mode, the AM noise of the source cancels to a high degree. Using these AM standards with two detectors and a two-channel spectrum analyzer we have verified AM noise floors below -180 dBc/Hz for carrier frequencies of 5, 10, and 100 MHz.

3. CALIBRATION CONSIDERATIONS

3.1 Calibration of PM Noise in PM/AM Noise Standard

If a scanning narrow-band receiver is used to measure the relative value of V_0^2 and $PSDV_N$ on the modulated output, only the equivalent noise bandwidth BW of the receiver, and its accuracy for relative measurements from V_0^2 to $(PSDV_N)BW$ must be calibrated to determine $S_\phi(f)$. The form of Eq. (1) is such that small errors in the alignment of the noise passband filter resulting in odd-order variations of $V_N^2(\nu)$ about ν_0 are averaged away. The only limit in setting $S_\phi(f)$ is that $S_\phi(f)$ should be high enough that it is far above the noise floor of the measurement system and yet small enough to satisfy Eq. (2) (15 dB margin reduces the errors to approximately 0.14 dB). The smaller the bandwidth for $S_\phi(f)$, the higher the noise can be set without violating Eq. (2). Typically $S_\phi(f)$ is constant for $0 < f < \nu_0/10$ and the integrated phase noise is in the range from 10^{-2} to 10^{-4} rad^2. We must also check that all of the amplifiers used in the evaluation are operated in the linear gain region. The error in the calibration for the circuit in Fig. 1 is typically 0.05 dB due to random errors and ± 0.05 dB due to uncertainties in the linearity of the receiver, 0.04 dB due to uncertainties in determining the bandwidth, yielding an overall root-sum-of-squares (RSS) uncertainty of ± 0.14 dB. Errors in a prototype PM/AM noise standard at 10.6 GHz are about ± 0.5 dB for 3 kHz < f< 500 MHz and < 1.5 dB for 500 MHz < f < 1 GHz.

3.2 Determining Accuracry and Noise Floor in "2 oscillator" type PM Noise Measurement Systems

Figure 2 shows the preferred method for determining the accuracy and noise floor of the standard "2 oscillator" type PM noise measurement system using the new

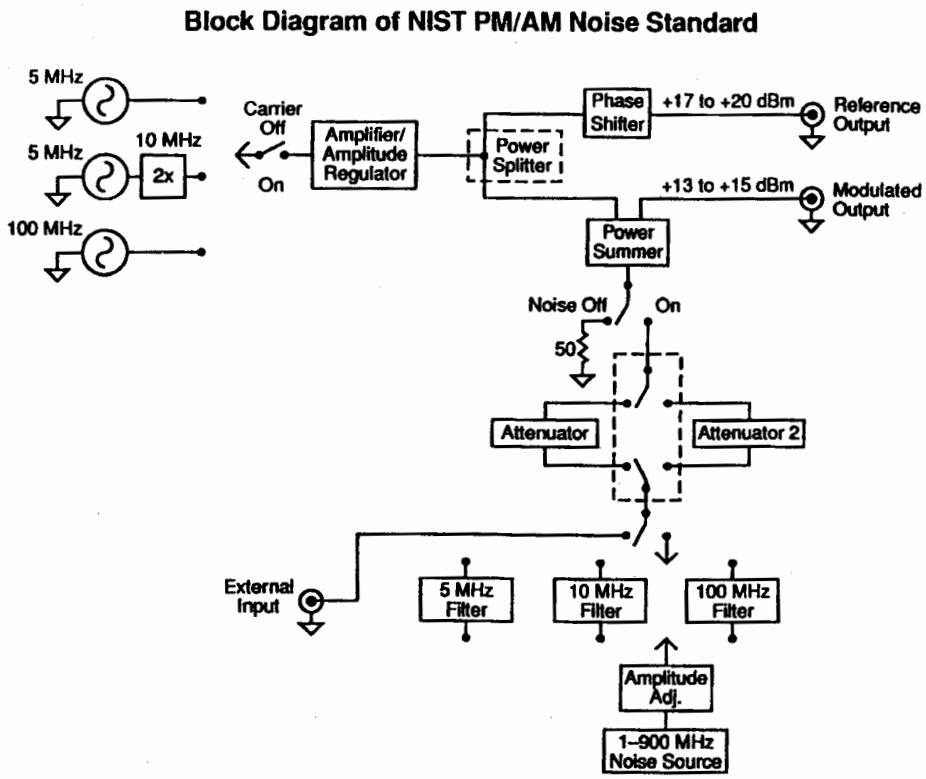


Fig. 1 Block diagram of Calibration Standard PMAM 115 for PM and AM noise.

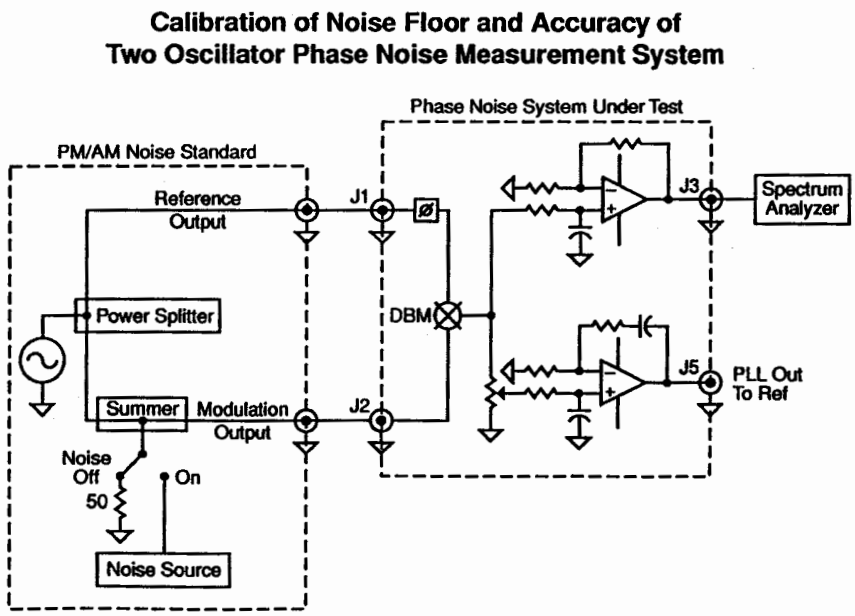


Fig. 2 Preferred calibration method for two oscillator type phase noise measurement systems.

PM/AM noise standard [3-5]. As explained above, the PM noise standard produces signals of about + 15 and +17 dBm. In this configuration there are no phase-lock-loop (PLL) effects to bias the low frequency measurements and the PM noise of the driving source cancels to a very high degree. To implement the measurements, the phase noise measurement system be able to adjust the phase between the two signals at the mixer to 90° so that the output of the mixer is close to zero V. Not all phase noise measurement systems have this provision, and some may require the use of the phase shifter in the phase noise standard. Once the phase shift has been set to 0, the mixer sensitivity is measured using an external source set to the same output power and impedance as that produced by the signal from one output of the PM standard. It is very important to use the same cable and to terminate the unused signal port with a 50 Ω termination. The beat between the two sources is measured to determine the slope at the zero crossing in V/rad. The cable is removed from the external source and reconnected to the signal port of the PM standard for further measurements.

The noise floor of the measurement system is determined with the noise modulation off and the mixer sensitivity determined above. Since the noise floor between the two signals is typically less than -193 dBc/Hz, the PM noise measured in this configuration is generally due only to the measurement system.

Next the PM noise is measured with the noise modulation on. Differences between the measurement of $S_\phi(f)$ and the calibrated PM noise determined using the method in Section 2 represent errors in the PM noise measurement system under test. The contribution of the noise floor of the PM noise measurement generally needs to be considered only near the carrier, (that is, as f approaches zero Hz) [5]. Errors are usually traceable to differences in rf drive and/or output impedance between the internal source and the external source used to determine k_d. The most common mistake is using a different cable for determining k_d and in measuring PM noise.

Figure 3 shows another method in which a second oscillator is used to determine the accuracy of PM measurement systems. The effects of the PLL on the measurements for small f can be important in this configuration. The PM noise of both oscillators and the calibrated PM noise are detected in the measurement. This approach can also lead to accurate evaluations of the measurement system errors versus Fourier frequency if the oscillator PM noise is small enough [5]. Our PM/AM noise standards accommodate this approach by providing oscillators with very low residual AM and PM noise and the ability to turn the calibrated PM noise off to check the noise floor of the measurement. As above, k_d is measured by allowing the two oscillators to beat. The noise floor of the entire setup is measured with the calibrated PM noise off and then with the calibrated PM noise on. From these two measurements we can determine the contributions of the uncancelled PM noise of the oscillators and the measurement system and their contribution to the measurement of the PM noise standard. Unlike the measurement technique outlined in Fig. 2, it is not generally possible to verify the noise floor of the measurement system because the phase noise of the reference oscillators always contributes to the measured noise. Also the accuracy cannot be carried out at small values of f since the PM noise of the reference oscillators dominates over the added noise. This problem can be alleviated somewhat by increasing the added noise and decreasing the bandwidth to satisfy Eq. (2). The accuracy of determining k_d is typically better than in the configuration of Fig. 2 because no substitution oscillator is required.

3.3 Determining Accuracry and Noise Floor of "Single Oscillator" Type PM Noise Measurement Systems

The PM/AM noise standard can also calibrate the accuracy of single-oscillator (frequency discriminator) phase noise measurement systems by the configuration shown in Fig. 4. The PM noise of both the oscillator and the single-oscillator measurement system is detected in the measurement. This approach can lead to accurate measurements if the oscillator PM noise is small enough [5]. The procedure is to measure the residual PM noise in the measurement with the calibrated PM noise off and then to measure the PM noise with the calibrated PM noise on. Unlike the technique of Fig. 2, it is not generally possible to verify the noise floor of the measurement system for these calibration methods because the phase noise of the reference oscillator always contributes to the measured noise. Another limitation is that the measurement cannot be carried out at low values of f since the PM noise of the reference oscillator generally dominates the added noise. This problem can be alleviated somewhat by increasing the added noise and decreasing the bandwidth to satisfy Eq. (2). There are no PLL effects to bias the measurements at low values of f. The noise floor of single oscillator systems is generally inferior to that of 2 oscillator systems [4].

3.4 Measuring PM Noise Added by Amplifiers, Frequency Multiplier, and other Components.

The phase noise added by amplifiers frequency multipliers and various components can be determined using a configuration similar to Fig. 2. The element to be tested is inserted in one or both of the signal paths from the PM/AM noise standard to the PM measuring system. PM measurements with the calibrated noise modulation on are used to calibrate the system sensitivity for converting phase fluctuations to voltage fluctuations versus Fourier frequency. PM measurements with the noise modulation off yield the PM noise (noise floor of the system plus the added noise of the element under test) relative to the calibrated PM noise. With this approach there is no need to measure k_d. Errors due the gain variations with frequency in the amplifiers and/or the spectrum analyzer cancel to a high degree. Accuracy is very good for this method since there is no change in sources or cables and there are no PLL effects to take into account.

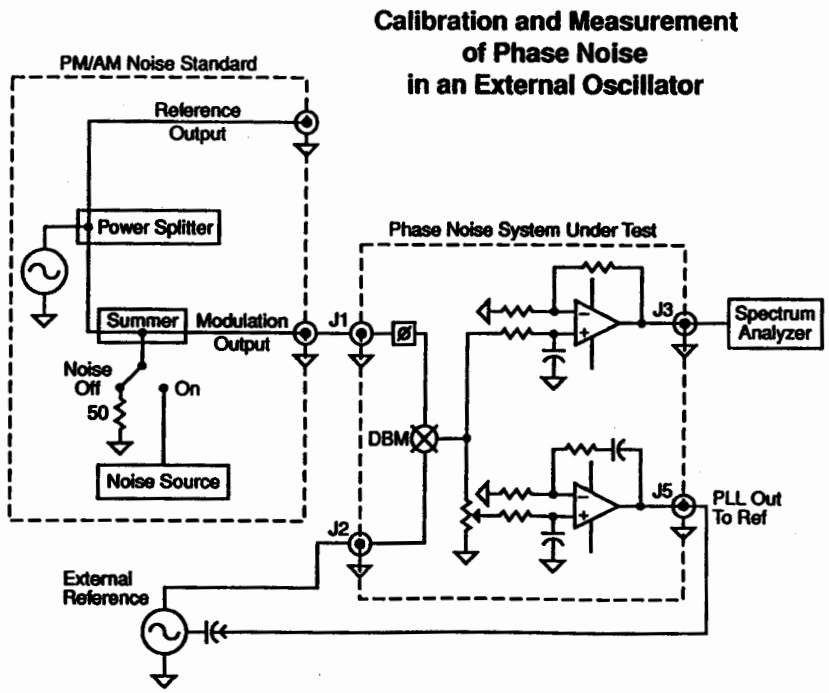


Fig. 3 Calibration method for two oscillator phase noise measurement systems.

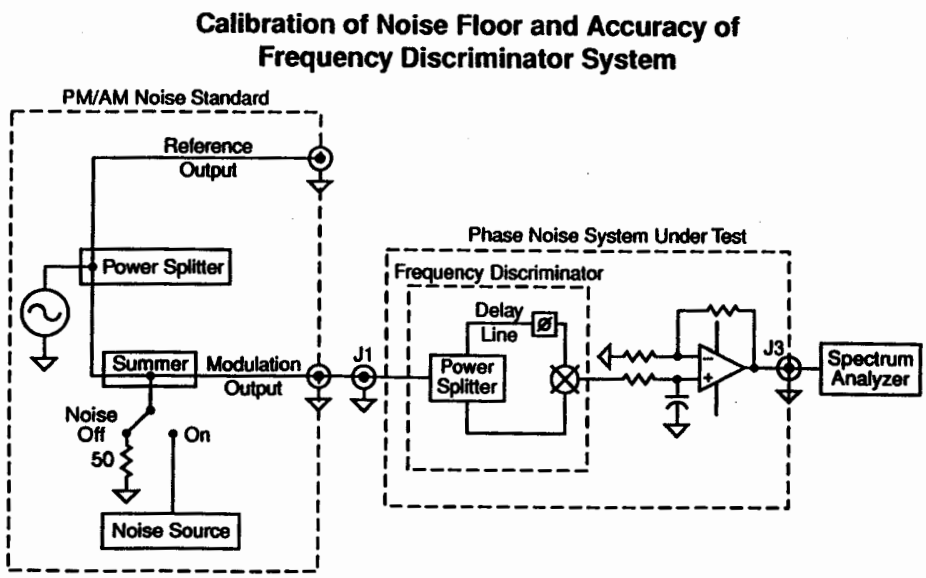


Fig. 4 Single oscillator phase noise calibration method using a frequency discriminator (see [3,4]).

3.5 Measuring PM noise in Oscillators

The configuration of Fig. 3 can be used to measure the phase noise of an external oscillator. PM measurements with the calibrated noise modulation on are used to calibrate the system sensitivity for converting phase fluctuations to voltage fluctuations versus Fourier frequency. PM measurements with the noise modulation off yield the PM noise (noise floor of the system plus the added noise of the oscillator under test) relative to the calibrated PM noise. Unbiased estimates of the PM of individual sources can be obtained using 3 oscillators in the 3 different pair comparisons [5]. Cross correlation which eliminate many of the problems with this type of measurement is discussed in [6,7].

3.5 Calibration of AM Noise in PM/AM Noise Standard

The AM noise in the standard is determined using the procedure described in section 3.1 and Eq. (1), since in the PM/AM noise standard,

$$S_a(f) = S_\phi(f) = \frac{PSDV_N(\nu_0 - f) + PSDV_N(\nu_0 + f)}{2V_0^2}. \quad (3)$$

3.6 Determining Accuracry and Noise Floor of AM Noise Measurement Equipment

Amplitude noise can be measured by several different methods. Each method is characterized by a detector that converts the amplitude of the rf carrier signal to a dc voltage. A working definition of $S_a(f)$ is

$$S_a(f) = \frac{\delta V^2(f)}{V_o^2} \frac{1}{BW}, \quad (4)$$

where $\delta V^2(f)$ is the mean squared fluctuation of the carrier rf amplitude measured at Fourier frequency f from the average frequency of the carrier in a noise bandwidth BW, and V_0 is the average amplitude of the carrier [3]. The most uncertain part of the process is determining the sensitivity of the detector $k_a(f)$ for converting the fractional changes in the carrier amplitude to a dc voltage fluctuation, since $k_a(f)$ is a function of V_0, f, and the carrier frequency ν_0, and in determining the variations of amplifier gain G(f) with Fourier frequency.

3.6.1 Calibration of Simple AM Noise Measurement Systems

The signal input of the AM noise measurement system to be measured is connected to the modulated output of the PM/AM noise standard as shown in Fig. 5. The spectral density of AM noise is calculated as

$$S_a(f) = \frac{V_n^2(f)}{BW}\left(\frac{1}{((k_a(f)G(f))^2}\right), \quad (5)$$

where $V_n^2(f)/BW$ is the PSD of voltage noise measured by the spectrum analyzer.

When the noise modulation is on, the measured noise computed with the aid of Eq. (5) should agree with the results of the calibrated level as given by Eq. (3). The differences between the measured results and that determined from Eq. (4) yield the error of the AM measurement system under test as a function of Fourier frequency. This measurement can also be used to determine $k_a(f)G(f)$ except at very low values of f where the AM noise of the source might mask the added noise. This problem can be alleviated somewhat by increasing the level of the added noise and decreasing the bandwidth to satisfy Eq. (2). The variation of $k_a(f)G(f)$ with f is extremely difficult to determine by traditional methods since few sources can be accurately amplitude-modulated over a broad frequency range. This issue becomes even more serious as the carrier frequency and the offset frequency from the carrier increase. One method for determining k_a is shown in Fig. 6.

3.6.2 Determining the Noise Floor of an AM Noise Measuring Systems

The noise floor for AM measurements can be estimated by turning the noise modulation off, measuring $V_n(f)$, and computing $S_a(f)$ using Eq. (5). The noise floor determined in this manner includes the true noise floor of the AM noise measurement system being calibrated plus the AM noise of the internal source in the PM/AM noise standard. In most cases, the noise floor of the PM/AM noise standard is lower than the detectors being tested.

A differential measurement technique, shown in Fig. 7, can be used to evaluate AM noise in detectors even below that of the source. The dc outputs of the two detectors are preferably adjusted to be the same within approximately 1% using attenuators on the input of the detectors. The gain following the detectors may also be adjusted; however, the cancellation of the residual AM noise in the source may not be as complete as when input attenuators are used. The outputs are then subtracted in a conventional device such as a differential amplifier (see Fig. 7). Some spectrum analyzers also can process the difference between the two input signals. With either of these methods, the AM noise of the internal source is substantially reduced, perhaps as much as 20-40 dB. (At very high Fourier frequencies it may be necessary to adjust the phase slightly on the incoming signal to one of the AM detectors to optimize the cancellation.) The noise floor measured in this manner is that of both detectors (and their associated amplifiers). If they were identical, we would subtract 3 dB from the measured result to obtain the noise floor of a single unit. [4,5,8]

5. RESULTS

The first PM/AM noise standards operate at 5, 10, and 100 MHz. The noise source is flat within ± 0.2 dB over the bandwidth equal to 5 % of the carrier frequency and stable with temperature to 0.02 dB/K. The differential noise added to the carrier ν_0 is exceptionally flat due to

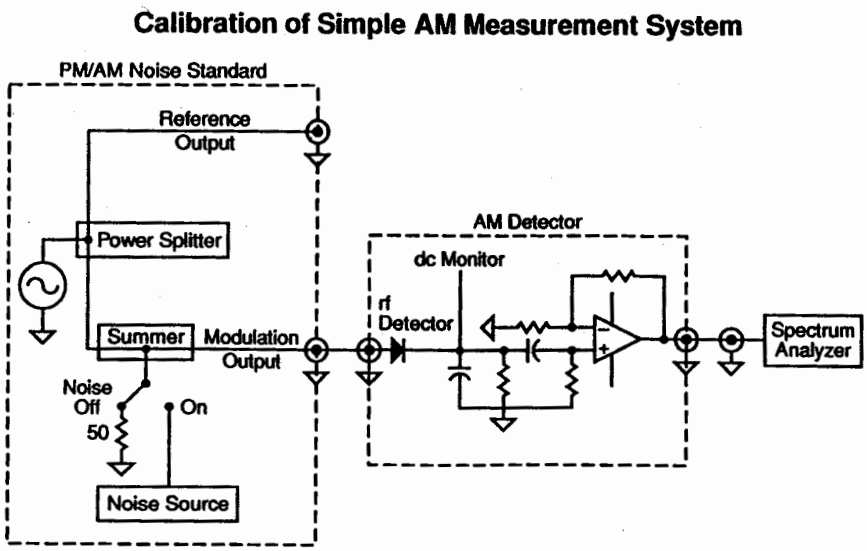


Fig. 5 Simple calibration method for AM noise measurement systems.

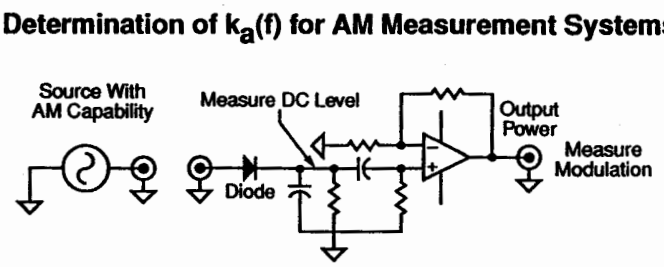


Fig. 6 One method of determining the sensitivity for converting small changes in amplitude to small changes in voltage.

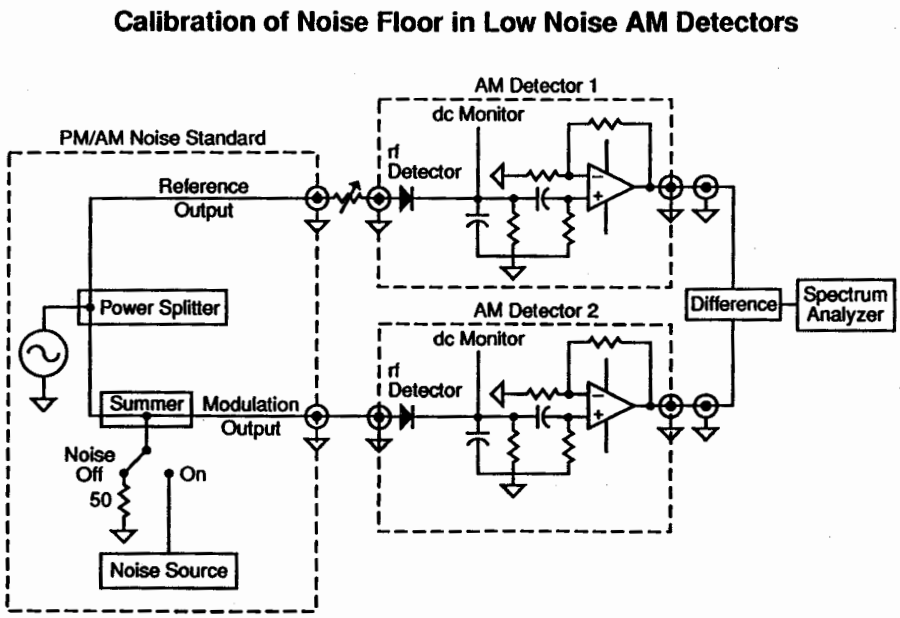


Fig. 7 Differential AM calibration method to determine the noise floor of AM detectors and measurement systems.

the symmetry of Eq. (1). The results obtained at 100 MHz are shown in Fig. 8.

The PM noise of the 100 MHz reference oscillator of the PM/AM standard is shown in curve A of Fig. 9. The added differential PM/AM noise between the reference and modulated outputs at 100 MHz is shown in curve B. The noise floor of the PM/AM standard at 100 MHz using the method of Fig. 2 with the noise modulation off is shown in curve C. The cancellation of the 100 MHz reference oscillator phase noise by using the method of Fig. 2 (the difference between curves A and B) reaches approximately 85 dB at 1 Hz. The large suppression of the reference oscillator PM noise makes it possible to use the PM/AM noise standard over a much larger range of Fourier frequencies about the carrier than any technique that is sensitive to the PM noise of the reference such as that shown in Figs. 3 and 4. At 100 MHz, the difference is having an accurate PM standard for Fourier frequency offsets from 1 Hz to 10 MHz versus 300 Hz to 10 MHz. The peaks at 60, 120, and 180 Hz in curve B are due to spurious signals originating from the power lines.

Curve A of Fig. 10 shows the PM noise in the PM/AM noise standard at 100 MHz determined using a high-dynamic-range scanning receiver in the calibration mode described in Section 2 above. The noise is constant to $\pm$ 0.05 dB for Fourier frequencies from dc to 5 MHz and gradually decreases 0.3 dB from 5 to 10 MHz. Curve B of Fig. 10 shows the measurement of the PM noise level using a separate and independently calibrated phase noise measurement system PNMS [4,5,9]. Curve C of Fig. 10 shows the AM noise determined using the simple measurement technique of section 3.6. The divergence of curve C from curves A and B for Fourier frequencies below 100 Hz is easily shown, by the AM noise floor measurements, to be due to noise in the diode detector used for these measurements. The fluctuations in curve B of Fig. 10 versus frequency are primarily due to the finite number of samples taken to estimate the spectral density [10]. The differences between curves A and B of Fig. 10 are consistent with the estimated errors in the internal phase modulator used in the calibration of the PNMS. Table 1 summarizes the PM/AM calibrated noise and noise floor at 5, 10, and 100 MHz. The independent data from the PNMS confirm the calibration of the PM/AM standard to $\pm$ 0.5 dB.

6. CONCLUSION

A series of new portable PM/AM noise standards has been developed, tested, and used to evaluate the noise floor and accuracy of PM and AM noise ($S_\phi(f)$ and $S_a(f)$) measurement equipment. The differential PM noise in the PM noise standard used to test the noise floor is typically much less than the noise floor of present measurement systems. The PM noise in the standard was evaluated for Fourier frequencies from 1 Hz up to 10% of the carrier frequency to an accuracy of $\pm$ 0.14 dB at 5, 10, and 100 MHz. The temperature coefficient of the PM and AM noise is less than 0.02 dB/K. These calibrations were independently verified for all frequencies to $\pm$ 0.5 dB using the Phase Noise Measurement System [4,5,9]. The design of the standard is such that the fractional amplitude noise $S_a(f)$ is equal to $S_\phi(f)$ over a very wide range of Fourier frequencies. The new secondary standard has maintained the same level of PM and AM noise to within $\pm$ 0.4 dB over the past year [1,2].

These PM/AM noise standards can also be used as the calibration aid for both PM and AM measurements of sources, amplifiers, frequency multipliers and other devices. Their use greatly simplify the measurement process because the sensitivity of the detector and the relative gain of the post amplifier versus Fourier frequency are determined simultaneously.

7. ACKNOWLEDGEMENTS

The author thanks F. Ascarrunz, C. Nelson, and P. Pond for help in construction. This work was supported in part by the calibration coordination group of JCTG/GMT-JLC and the CECOM Center for Space Systems TSSRF, Ft. Monmouth, New Jersey.

8. REFERENCES

[1] Adapted from F. L. Walls, "A Secondary Standard for PM and AM Noise at 5, 10, and 100 MHz," CPEM 1992, Proc. IEEE Trans on I&M, 1992.

[2] F. L. Walls, "Calibration System for Determining the Accuracy of Phase Modulation and Amplitude Modulation Noise Measurement Systems," U. S. Patent No. 5,172,064, issued 15 December 1992.

[3] "Characterization of Clocks and Oscillators," Eds. D. B. Sullivan, D. W. Allan, D. A. Howe, and F. L. Walls, NIST Tech Note 1337, 1990.

[4] F. L. Walls, A. J. D. Clements, C. M. Felton, M. A. Lombardi and M. D. Vanek, "Extending the Range and Accuracy of Phase Noise Measurements," Proc. of 42nd Annual Symposium on Frequency Control, Baltimore, MD, pp. 279-283, June 1-4, 1988.

[5] F. L. Walls, C. M. Felton, A. J. D. Clements, and T. D. Martin, "Accuracy Model for Phase Noise Measurements," Proc. of 21st Annual Precise Time and Time Interval Planning Meeting, Redondo Beach, CA, pp. 295-310, Nov. 30-Dec. 1, 1990; F. L. Walls, A. J. D. Clements, C. M. Felton, and T. D. Martin, "Precision Phase Noise Metrology," Proc. of National Conference of Standards Laboratories, pp. 257-275, August, 1991.

[6] W. F. Walls, "Cross-Correlation Phase Noise Measurements", 1992 IEEE Annual Sumposium on Frequency Control, pp. 257-261.

[7] F. L. Walls "Reducing Errors, Complexity and Measurement Time in PM Noise Measurements", 1993 IEEE Freq. Cont. Symp.,1992.

[8] L. M. Nelson and F. L. Walls, "Relationships between AM and PM Noise in Selected rf Oscillators" 1993 IEEE Freq. Cont. Symp..

[9] F. L. Walls, "Method and Apparatus for Wide Band Phase Modulation." U. S. Patent No. 4,968,908 issued 6 Nov. 1990, F. L. Walls, "Wide-band Phase Modulation." U. S. Patent No. 5,101,506 issued 31 Mar. 1992.

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

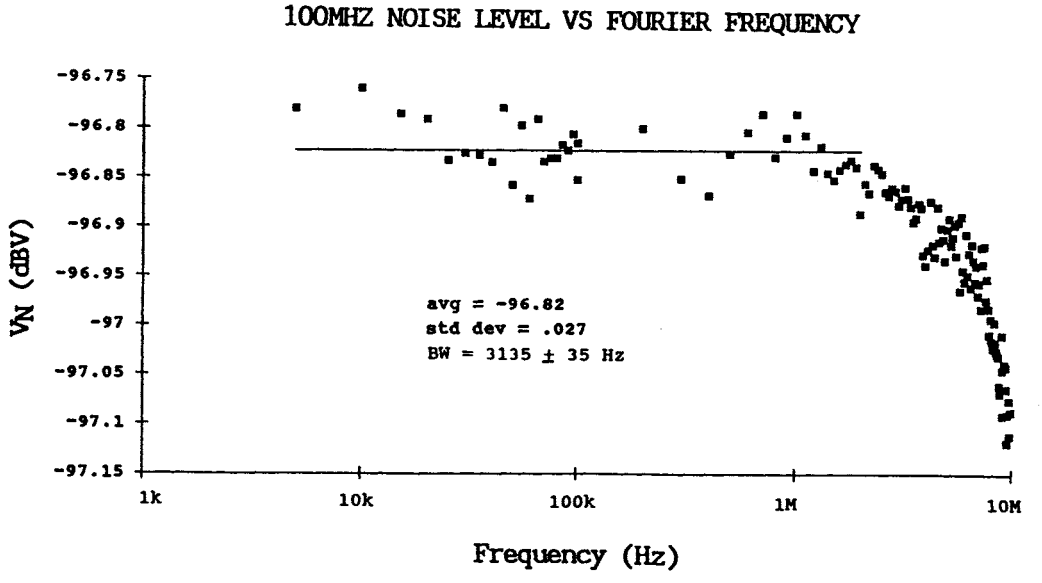


Fig. 8 Calibration of V_N added to the carrier V_0 versus Fourier frequency f at 100 MHz.

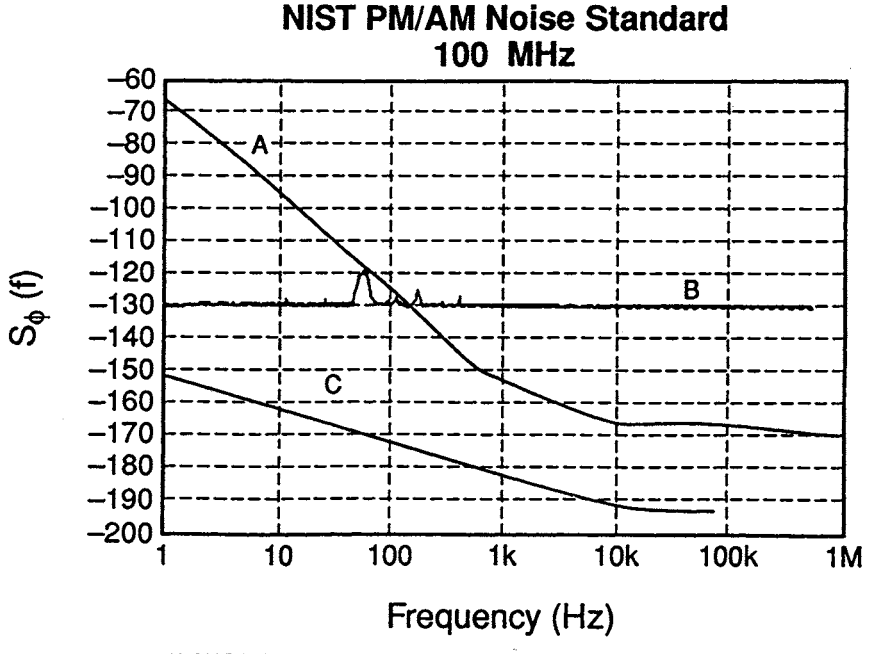


Fig. 9 Measurement results for PM/AM Standard model PMAM 115 at a carrier frequency of 100 MHz. Curve A shows the PM noise of the 100 MHz reference oscillator. Curve B shows the calibrated level of PM and AM noise. Curve C shows the typical noise floor for PM noise in the PM/AM Noise Standard at 100 MHz. Somewhat better results are obtained at 5 and 10 MHz.

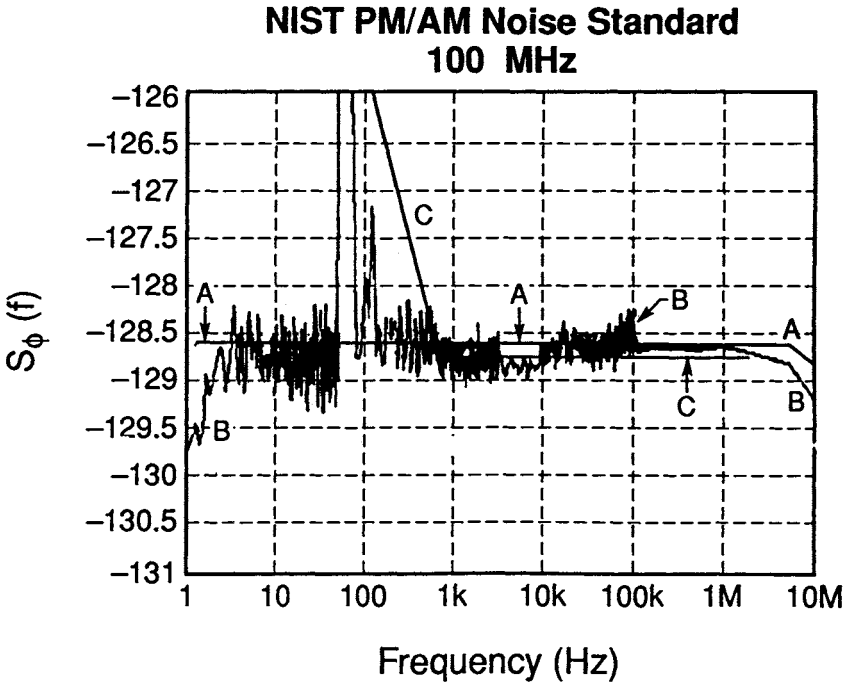


Fig. 10 Measurement results for PM/AM Standard model PMAM 115 at a carrier frequency of 100 MHz. Curve A shows the PM noise level of the standard determined using the technique of Section 3. Curve B shows the measurements of PM noise using the totally independent NIST Phase Noise Measurement System (see [4,5,9]). Curve C shows the level of AM noise of the standard at 100 MHz using the method shown in Fig. 5. The peaks at 60 and 120 Hz are due to the power line. Similar results are obtained at 5 and 10 MHz.

TABLE 1. SYSTEM NOISE FLOOR AND PHASE NOISE OF SOURCES

SOURCE FREQUENCY	FOURIER FREQUENCY							
	NOMINAL SOURCE PM NOISE/CHANNEL (dBc/Hz)							
	1 Hz	10 Hz	100 Hz	1 kHz	10 kHz	100 kHz	1 MHz	10 MHz
5 MHz	-121	-151	-163	-171	-174	-174	-174	
10 MHz	-115	-145	-157	-165	-168	-168	-168	
100 MHz	-70	-100	-130	-156	-170	-170	-173	-173
MAXIMUM RESIDUAL PM NOISE BETWEEN CHANNELS (dBc/Hz)								
5 MHz	-162	-172	-182	-190	-194	≤-175	≤-175	
10 MHz	-161	-176	-183	-191	-197	≤-175	≤-175	
100 MHz	-152	-162	-172	-182	-193	≤-175	≤-175	≤-175
DIFFERENTIAL PM/AM NOISE LEVEL (dBc/Hz)								
5 MHz	-128.3	-128.3	-128.3	-128.3	-128.3	-128.3		
10 MHz	-129.1	-129.1	-129.1	-129.1	-129.1	-129.1	-129.2	
100 MHz	-128.9	-128.9	-128.9	-128.9	-128.9	-128.9	-128.9	-129.2

SIFTING THROUGH NINE YEARS OF NIST CLOCK DATA WITH TA2*

Marc A. Weiss
Time and Frequency Division
NIST

Thomas P. Weissert
LiteroPhysics

Abstract

We have extended the new TA2 post-process time scale at NIST beyond our previous reports to include all of the period from January of 1984 to the present. Derived from the ensemble of clocks at NIST, this time scale includes the benefits of several recent refinements to the algorithm. By iteratively running the algorithm on the ensemble clock data and characterizing anomalous behavior in the dominant individual clocks of the ensemble between iterations, we obtain an optimized scale which benefits from the informed anticipation of that anomalous behavior and demonstrates an overall decrease in scale disruption. Herein we discuss: changes to the TA2 algorithm that we made while processing the eight-year run, our method of characterizing anomalous behavior in individual clocks, the way unanticipated anomalous behavior is dealt with by the algorithm, and our resulting nine-year time scale.

keywords: time scale, clock ensemble, post-process, stochastic noise, deterministic errors

1. INTRODUCTION

We report here our development of a nine-year time scale derived from the measurements among the best atomic clocks at NIST over the years from 1984 through 1992, using the TA2 algorithm [1,2,3]. In this paper we refer to both our algorithm and the time scale resulting from the application of that algorithm to the NIST clocks as "TA2;" whenever contextual clarity demands it, we differentiate between the TA2 algorithm and the NIST TA2 time scale. The algorithm takes data from an ensemble of clocks and computes a time scale in the form of time and frequency offsets of each clock from ensemble time and frequency.

Because the primary purpose of a post-processed time scale at NIST is to carry forward the evaluation of a primary frequency reference, TA2 was designed to optimize frequency stability in both short and long term. Whereas AT1, the real-time scale at NIST [4,5,6], shares these goals with the post-processed TA2 scale, it must be administered differently due to its greater vulnerability. In real-time, the most recent ensemble data must be synthesized into the time scale immediately, leaving little or no opportunity for the anticipation of anomalous behavior. Freed from the constraints of maintaining a real time scale, we are able to run the TA2 algorithm over a fixed period of data, characterize most of the existent anomalous behavior, and correct for it in subsequent iterations of the algorithm.

TA2 is the name for the process of running the AT2 algorithm both forwards and backwards in time to determine frequency estimates of clocks in each direction, then running a final third pass forward in time to generate ensemble time using the frequency estimates from both the forward and backward passes. The AT2 algorithm is very similar to the AT1 algorithm, but it uses a simple Kalman filter technique to estimate a confidence on the frequency estimate of clocks. This confidence facilitates both automatic frequency step detection and combining the forward and backward passes to obtain a smoothed ensemble frequency.

In addition to our primary goal of optimum stability, we designed the TA2 time scale at NIST to facilitate: the addition and/or removal of clocks from the scale with minimal perturbations to the scale; the detection of steps in either the time or frequency offsets of an individual clock from the scale; and the general ease of use in running the scale. While much of the design of TA2 has been reported before, some corrections and improvements have been made to the algorithm in the process of running the scale over nine years of data.

Using the behavior of the real-time scale AT1 as a guide, we examined all major anomalous features occurring in the ensemble data during the nine years of measurements among the best clocks at NIST. When we encountered frequency steps or drift [7,8,9], we characterized this behavior and inserted anticipatory information into the run-time command file for the scale, and so prevented these effects from pulling the frequency of the scale. We made these characterizations carefully with the conscious intent of maintaining the free-running independence of TA2; thus we say that TA2 has not been steered to any clock in the ensemble or to an external reference. We discuss our method of inserting clock characterizations into the time scale in some detail.

We analyze the performance of the scale by making comparisons both with the AT1 time scale and with time scales and clocks external to NIST. We show that TA2 is comparable in performance to other NIST scales and so

*Contribution of the U.S. government. Not subject to copyright.

provides redundancy and additional diagnostic abilities for characterizing clocks.

2. ANOMALOUS BEHAVIOR CORRECTION

TA2 has the ability to detect frequency steps automatically and correct for them by removing the offending clock from the ensemble until the scale can learn the new frequency. When running the scale in the forward direction a clock is removed chronologically after the frequency step, while in the backward direction, the clock is removed chronologically before the step. Thus the automatic frequency step algorithm keeps the clock in the ensemble at full weight in only one direction in the neighborhood of a frequency step. Whereas this limits the pulling of the ensemble frequency by the clock which has taken a frequency step, it reduces the stability of the ensemble, since the clock is no longer contributing. When we ran TA2 over the nine years, if a high weight clock took a frequency step we perefered to characterize that step to allow the clock to maintain full weight in the ensemble in both forward and backward directions.

As with other algorithms [10,11,12,13], TA2 incorporates a definition of the ensemble time offset from the common clock using a weighted average of the measured offsets of all the clocks from the common clock. TA2 computes this weighted average beginning with a prediction of the time offset of each clock from ensemble time. Because this prediction represents our expectation for each clock's nominal time offset from the scale, it serves as a natural point of entry for our foreknowledge of anomalous behavior. Just as anomalous behavior breaks from our expectations for a clock, we direct the algorithm to break from its prediction, which is otherwise based solely on a clock's history.

We specifically consider two types of anomalous behavior: frequency steps and frequency drift. We characterize these effects by studying the differential performance of clocks in the ensemble. If a clock has low weight and therefore little correlation with the scale, then we can run the algorithm and characterize that clock's behavior against the ensemble itself. But for higher-weight clocks, the value of frequency steps or drift relative to the ensemble are biased by that clock's correlation with the ensemble, and so better characterizations are obtained by comparing the clock under question with other clocks in the ensemble. When we see a frequency step between two clocks, we need to determine which clock took the step. To do this, we can compare each with a third clock and use majority voting. It also helps to look at more than three, and to consider the possibility of an environmental effect that could perturb several clocks simultaneously at the same site.

We emphasize the significant difference between adjusting a clock's prediction to reflect pre-characterized behavior, and steering the ensemble frequency to an external reference. If a clock in the ensemble takes a frequency step, the ensemble will be pulled off in frequency by that clock because previous frequency estimates contribute to current updates. Thus, it is the prediction of frequency based on previous history which can pull the ensemble frequency if the clock frequency has shifted unpredictably. Since we run the scale after this has occurred, we may characterize this clock's frequency step by comparing the frequency of this clock to other clocks in the ensemble. If we see a consistent value for the frrequency step which is considerable larger than the stochastic fluctuations of frequency for this clock, we use that value as the characterization of the clock's frequency step. We insert this value as a shift in the prediction of frequency for this clock against the ensemble at the reference time when the clock stepped. The ensemble frequency is not steered by this effort to any clock. We must assume only that one or more clocks in the ensemble are stable enough in frequency that our characterization of a frequency step is significant.

The process is similar for frequency drift, but in practice the characterization requires more care. We must see a consistent drift in one clock as compared to several clocks. If we accept this value as the frequency drift of the one clock, we are assuming that the other clocks are neither drifting deterministically nor walking off randomly together. As we shall see, it is possible for clocks to move together. Once we have characterized a frequency drift, we adjust the prediction of a clock to account for it much as with a frequency step, but now the frequency prediction is adjusted continuously over many measurement cycles. Others have studied using maximum likelihood estimation of drift [19]. This requires the assumption that all clocks have a constant drift during the period used for the maximum liklihood estimation.

3. RESULTS

Figure 1 shows the fractional frequency offset of AT1-TA2. Overall the two scales nominally remain within $1 \cdot 10^{-13}$ of each other. There seem to be significant periods of frequency drift between the two scales reaching a magnitude of perhaps 10^{-16}/d for extended periods. Most of this drift appeared when we introduced a robust technique for time steps [17]. This new method tapers clock weights quadratically as prediction error increases from three to four standard deviations, instead of completely removing a clock whose prediction error exceeds three standard deviations. Whereas AT1 completely removes a clock whose prediction error exceeds three standard deviations, it is not surprising that changing this technique increased differences between the two scales. The magnitude of the change reflects the non-linear way in which the algorithms respond to small differences.

We emphasize that we can not know which method produces less drift in an absolute sense. This drift of 10^{-16}/d is much smaller than effects we see in comparing the two scales against the best available external references. We add the technique to TA2 since it should theoretically improve the scale.

There are occasionally more rapid excursions which existed even without the robust time step technique. These are probably due to using somewhat different clock ensembles and the differences in the algorithms. In TA2 when a clock

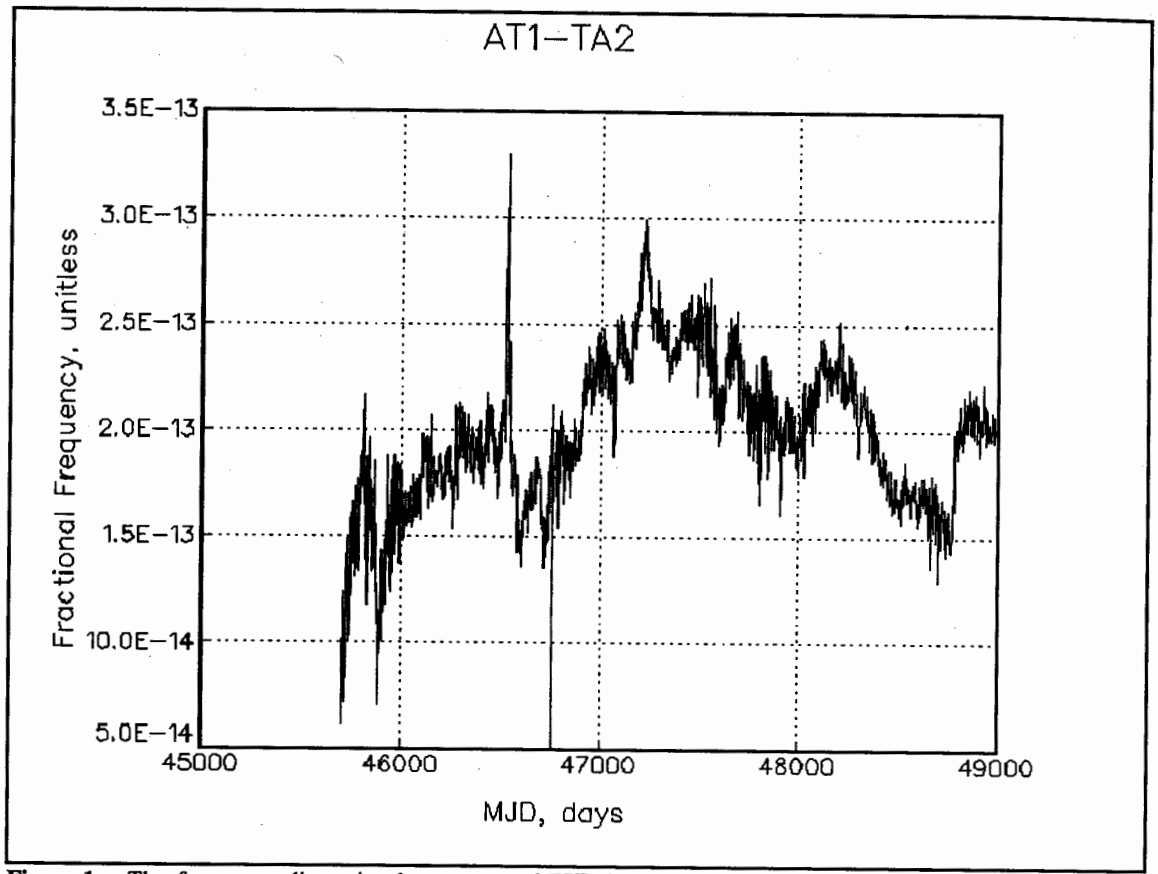


Figure 1: The frequency dispersion between two NIST time scales, the real-time AT1 scale minus the post-processed TA2 scale.

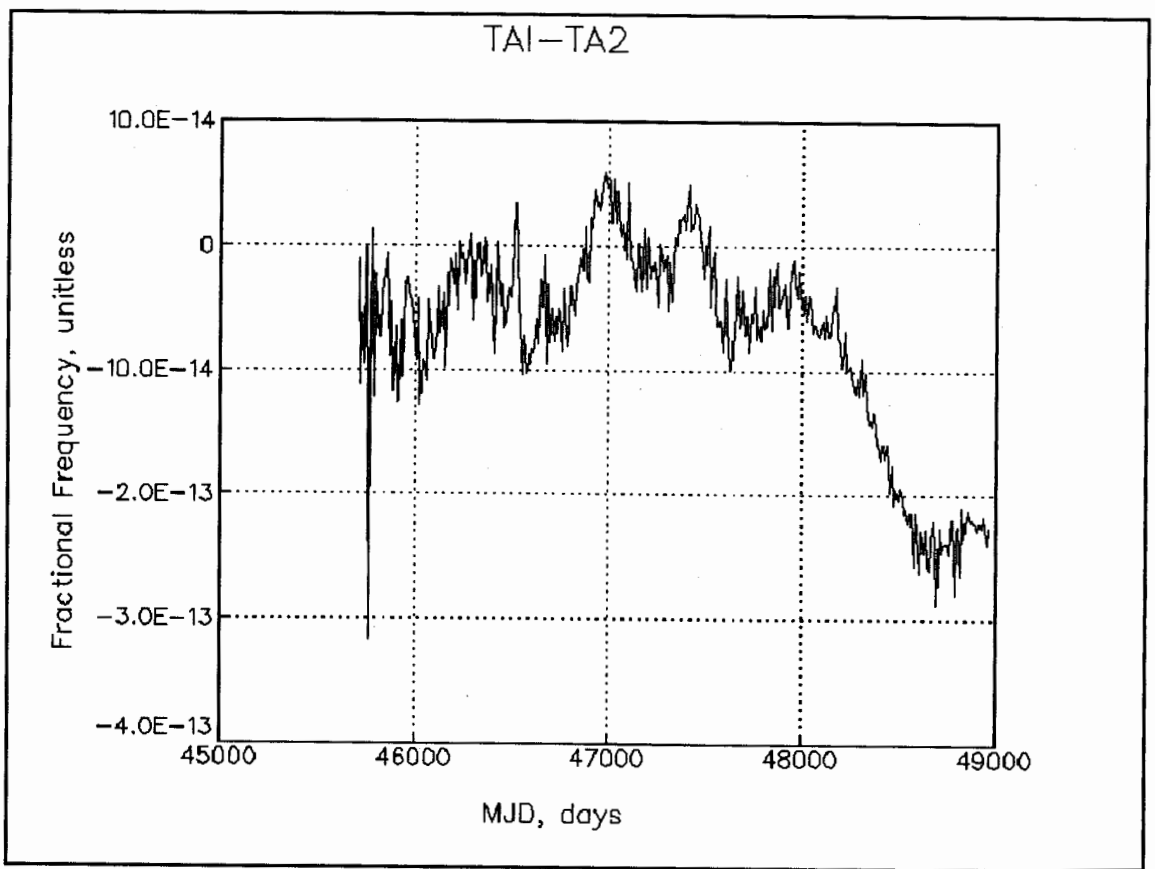


Figure 2: The frequency dispersion of International Atomic Time minus the NIST TA2 time scale.

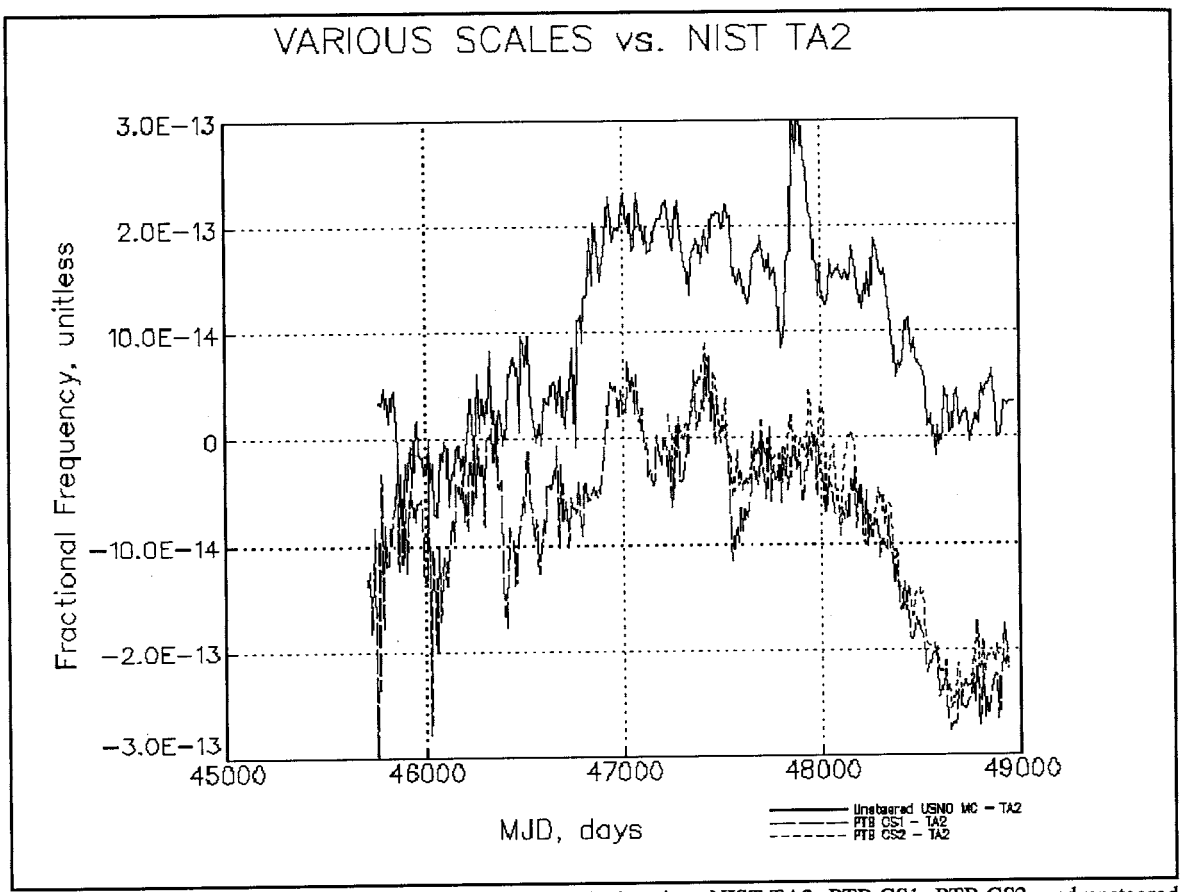


Figure 4: the frequency dispersion of three independent clocks minus NIST TA2, PTB CS1, PTB CS2, and unsteered USNO Master Clock.

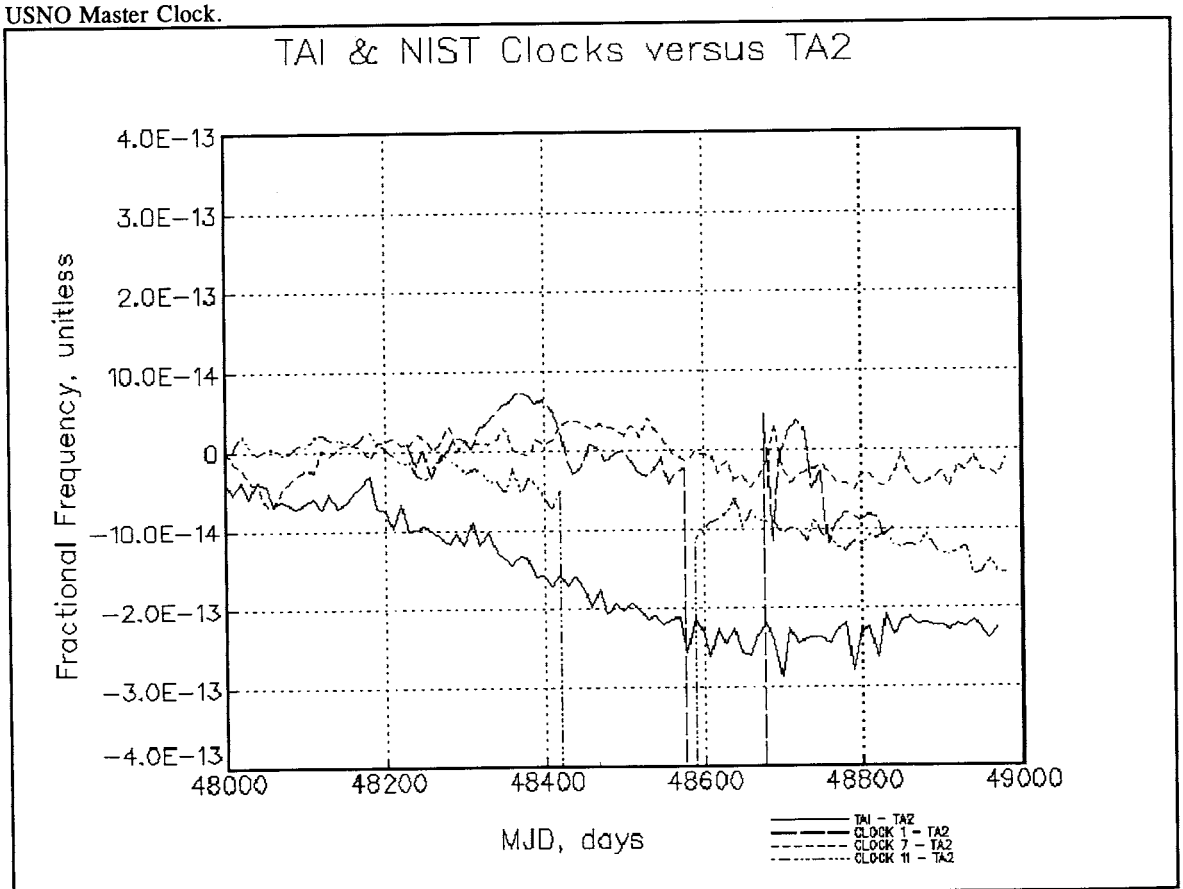


Figure 3: TAI and three clocks at NIST with a similar frequency drift during the period of significant drift for TAI-TA2.

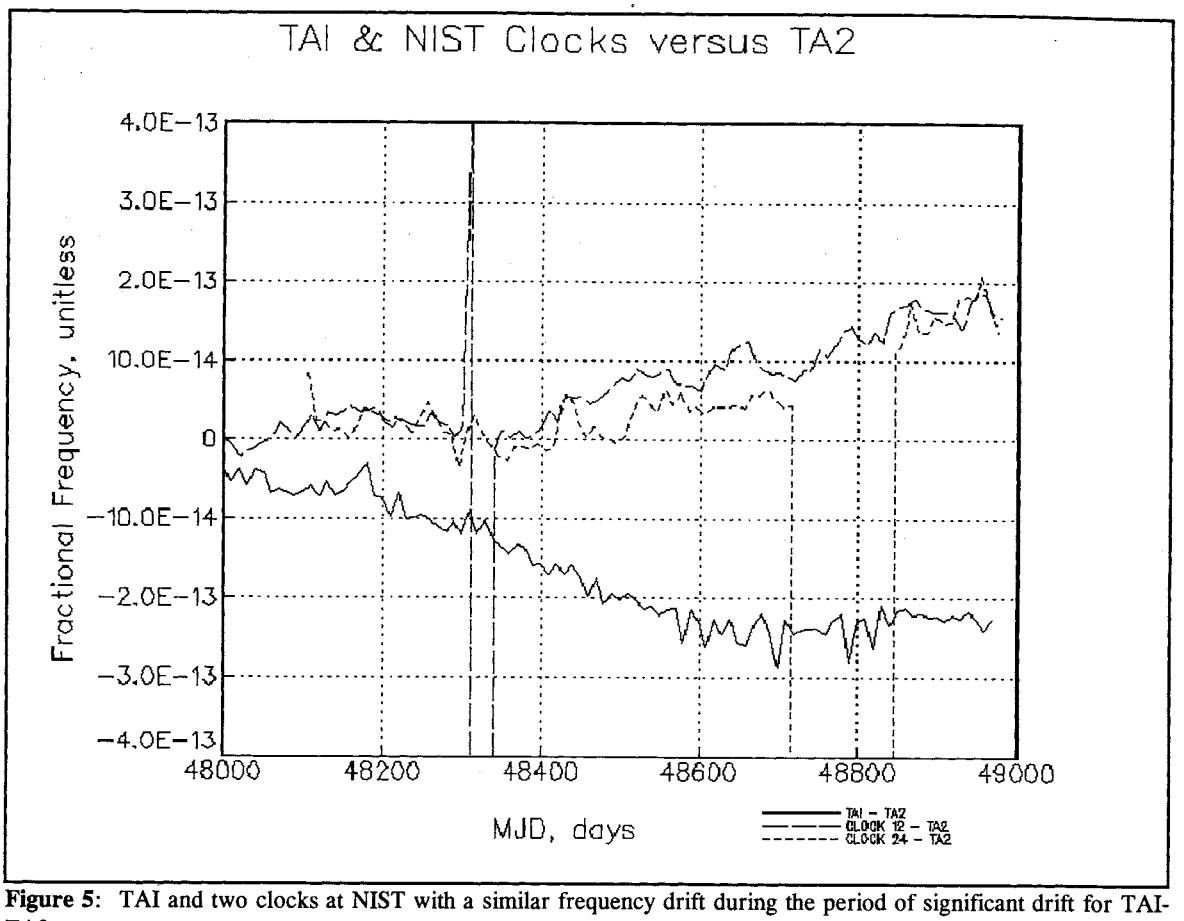


Figure 5: TAI and two clocks at NIST with a similar frequency drift during the period of significant drift for TAI-TA2.

takes a frequency step it has a smaller effect than in AT1, since TA2 is run in post-processing and AT1 is a real-time scale. Still, the two scales use very similar algorithms on the same major clocks. Again, these differences between the two reveal non-linearity of the time scales. Small differences and changes can produce large effects. Other significant features include a time step between TA2 and AT1 at MJD 46760 when the NIST (then NBS) measurement system failed temporarily, and a steering in AT1 of $3.44 \cdot 10^{-14}$ applied over a four day interval centered at MJD 48780, 00:00 hours UT.

Figure 2 shows the fractional frequency offset of TA2 from International Atomic Time, TAI [13]. We see annual deviations over the first five years which seem to attenuate after MJD 47527, the beginning of 1989. The period approximately corresponding to mid 1990 to mid 1992, from MJD 48140 to 48730, however, shows a steady drift of about $-3 \cdot 10^{-16}$/d.

Figure 3 shows TA2 frequency versus the primary frequency standards of the Physikalisch-Technische Bundesanstalt (PTB), Braunschweig, Germany, PTB CS1, PTB CS2 [14] and unsteered master clock of the U.S. Naval Observatory (USNO) [15]. Although each of these clocks is correlated with TAI, there is no reason to expect correlation among them. We might suspect some correlation between the two PTB clocks, because they are in the same laboratory. But both PTB CS1 and the unsteered USNO Master Clock show a similar frequency drift against TA2 over the interval from MJD 48140 to 48730.

We show the five dominant clocks contributing to TA2 during this period in figures 4 and 5 along with TAI-TA2. For these figures intial frequencies have been set to 0, and frequency steps which have been characterized and inserted into the scale have been removed from the plots. Thus, these are the frequency changes which the algorithm sees, except for modelled frequemcy drift. Figure 4 shows NIST clocks 1, 7, and 11, which seem to have some similarity in overall frequency drift. Clock 1 seems to have no consistent drift against TA2 over the interval MJD 48140 to 48730, though it significant random walk FM is visible even though it is off line for 176 d. Clock 7 does not seem to drift against TA2 over most of the interval, though it does seem to drop in frequency from about 48550 to the end of the TAI-TA2 drift interval at about 48730. Clock 11 seems to drop slowly, though it is off line for 148d. Figure 5 shows clocks 12 and 24 which seem to be drifting against TA2 opposite to TAI. In these cases we attemped to model drift based on internal comparisons between clocks. We concluded by modelling clock 12 (figure 5) with a drift of $5 \cdot 10^{-16}$/d from MJD 48340 to 48622, and by modelling clock 7 with a drift of $2.3 \cdot 10^{-16}$/d from MJD 48080 to 48400. It is possible we might have benefited from the maximum likelihood estimation of Jones and Tryon [19], though we would have had to assume a constant drift on all clocks over some interval. We conclude that the drift between TAI and TA2 from MJD 48140 to 48730 is a mean frequency drift among the clocks that dominate TA2 over

this period.

4. CHANGES IN TA2

There have been a number of changes in the TA2 algorithm since previous publications [1,2,3]. TA2 is the name given to the three-pass smoothing process of running the AT2 algorithm forwards and then backwards in time, and then combining the clock frequencies generated in those passes into a final third pass forward to generate clock time offsets from the ensemble time. In the next section of this paper, we present the complete equations of AT2. Here we discuss the motivations for some of the changes.

Several equation changes follow from our recognition of a need to differentiate between two different values of τ which we call τ_x and τ_y, for each clock in the ensemble. Because data for an individual clock may be missing for various reasons, τ_x is necessary to track the reference time since the last time update. Similarly, we need τ_y to track the reference time since the last frequency update, because we cannot estimate frequency on a cycle in which there is a time step. Hence we may have a time update without a frequency update.

The method we use to estimate the bias of a clock's prediction error due to the clock's correlation with the ensemble has changed to incorporate the recent work of Tavella, Azoubib and Thomas [16]. Through simulation we have verified that the current formula indeed gives a better value for the bias than the previous one.

We have moved the drift term into the frequency prediction, instead of in the time prediction. This better allows us to separate the computations of clocks' time offset, such as time step detection and correction, from those with frequency.

We have added the wct (weight control) parameter to enable limiting of a clock's contribution to the ensemble, while the algorithm continues to estimate the clock's prediction error. This is useful in two important situations. First, on the detection of a time step, whereas previously we abruptly removed the clock from the scale for that measurement cycle, now we gradually deweight a clock as the step size increases from three to four standard deviations from its predicted value. This move from a single-cycle discontinuous gap in a clock's presence in the ensemble, to a ramped-deweighting procedure increases the robust character of the algorithm, in that a small change in a measurement will not produce as large a change in the scale [17]. Second, this parameter allows us to enter a clock in the scale with zero weight while the scale still computes that clock's prediction error variance.

Besides these changes we have corrected some errors in the formulas as published previously, notably in the prediction error update ϵ^2 and in the statistic used for defining frequency steps σ_L. We have also introduced the parameters R and Q to simplify the equations for the Kalman estimate of frequency y, and to clarify the formulas used for converting stochastic parameters among various representations.

5. THE EQUATIONS DEFINING AT2

The AT2 algorithm estimates the deterministic parameters of the clock model, time and frequency offsets of each clock from ensemble time (and frequency). An estimate of frequency drift for each clock can be entered and used for frequency prediction, but it is not updated in the calculations. A time offset is a weighted average, the weight of each clock being determined adaptively by its prediction confidence, which is the normalized reciprocal of the prediction error. AT2 is based on the AT1 time scale algorithm with the addition of a Kalman filtering technique for estimating frequency.

In addition to the deterministic parameters, there are two stochastic parameters in the clock model, white frequency modulation (FM) σ_α and random walk FM σ_β [18-20]. These two parameters are consistent with atomic clock data, though there is evidence that flicker FM commonly exists in cesium clocks also [21-29]. AT2 reduces to AT1 in steady-state if the weights of clocks are chosen properly; hence AT2 inherits the ability in AT1 to model flicker FM also.

Like AT1, AT2 estimates time offsets using predictions from the previous measurement cycle and the current measurements. These predictions allow us to detect time steps in a clock by comparing the difference between its predicted and updated time, the time innovation, with that clock's prediction error. The Kalman technique provides a variance of frequency prediction, which also allows us to detect frequency steps.

In our formalism, we assume there is a constant reference time interval τ_0 between measurements. At each measurement cycle, all clocks are measured against a common clock with negligible measurement noise. Thus, we obtain a measurement between any pair of clocks by differencing the measurements of these two clocks against the common clock. If there are no data for a particular clock, the time interval between measurements can grow to some multiple of τ_0. We denote this interval as $\tau_{x,i}$, because we use x for a clock's time offset, and this interval is the duration of reference time between time updates for clock i. When the algorithm detects a time step, it does not update the frequency for the stepped clock, even though time is still updated. Hence the reference time interval between frequency updates can be different from $\tau_{x,i}$. We denote this second reference time interval as $\tau_{y,i}$, because y denotes the frequency offset of a clock from the ensemble.

5.1 Time prediction

The prediction of the time offset $\hat{x}_i$ of clock i from ensemble time for the current reference time, t, is calculated from the previously updated time offset x_i at reference time $t-\tau_{x,i}$ and the filtered frequency y_i,

$$\hat{x}_i(t) = x_i(t-\tau_{x,i}) + y_i(t-\tau_{x,i})\cdot\tau_{x,i} \quad (1)$$

5.2 Time update

Given measurements of time differences $x_{ji}(t)$ for clock j minus clock i, the time offset of clock i is updated in a weighted average, using the prediction, weight and time-difference measurement of every clock j, that is,

$$x_i(t) = \sum_{j=1}^{n} w_j \left[\hat{x}_j(t) - x_{ji}(t) \right]. \quad (2)$$

The weights w_j determined from the prediction confidence, are normalized to sum to 1. We assume in this equation that the measurement noise is negligible, so

$$x_{ij} - x_{kj} = x_{ik} \quad (3)$$

for all combinations of i, j, and k. If the validity of this assumption fails, then a pre-filter must be used to obtain estimates of x_{ij} that satisfy (3). We may view equation (2) as the definition of ensemble time, offset from one clock. The ensemble time offset from other clocks is simply

$$x_j = x_i + x_{ji}. \quad (4)$$

5.3 Adaptive Clock Weights

The clock weights used to update the time offsets are calculated from the variances of the time residuals ϵ_i^2 by

$$w_i = \frac{\epsilon_x^2}{\epsilon_i^2} wct_i, \quad (5)$$

where ϵ_x^2 is the ensemble prediction error variance, ϵ_i^2 is the prediction variance since the last time update of clock i, and wct_i is the weight control parameter for this clock. The weights adaptively match the relative stabilities of the clocks in the ensemble because ϵ_i^2 and ϵ_x^2 are estimated from the data. As mentioned above, the weight control parameter affords the algorithm a way of dealing with time steps robustly, by deweighting a clock smoothly as a function of the size of the step [17]. This is discussed further in the section on time steps. Setting wct to 0 allows a clock to be estimated against the scale without it actually contributing to the scale, which thus keeps that clock uncorrelated.

5.4 Ensemble Prediction Error

The ensemble prediction error is a numerical estimate of the stability of the ensemble over an integration time τ_x. It is calculated as

$$\epsilon_x^2 = \left(\sum_{i=1}^{n} \frac{wct_i}{\epsilon_i^2} \right)^{-1} \quad (6)$$

The dependence on t is suppressed in this equation. We see here that a poorly performing clock will not destabilize the ensemble; clocks can only improve ensemble stability.

5.5 Prediction Error Estimate

The prediction error of clock i over the interval $\tau_{x,i}$ is estimated from the difference between the time prediction and subsequent update, often called the innovation. This difference is adjusted for the average bias of this value due to the clock's correlation with the ensemble. The equation is

$$\hat{\epsilon}_i^2(t) = \left(x_i(t) - \hat{x}_i(t)\right)^2 \cdot K_i \quad (7)$$

where K_i is the expected proportion of the bias of clock i. Thus $\hat{\epsilon}_i^2(t)$ is a sample estimate of the prediction error variance of the clock, based on the current measurement cycle.

5.6 Bias of the error estimate

Because ensemble time is a weighted average of each of the individual clock times, the prediction error of a clock measured against the scale will be biased low, on the average. To correct for this biasing, we include the estimate of the multiplicative bias for each clock [16]

$$K_i = \frac{1}{(1-w_i)} \quad (8)$$

Thus the factor K_i is the proportion of the unbiased-to-biased average prediction error, relative to the scale.

5.7 Filtered Prediction Error

We assume a model for the prediction error of white PM and random walk PM. That is, the prediction error estimates vary around some true mean prediction error at the two hour measurements due to the white FM level of the clocks, but this white FM level varies slowly at about 20 d. Therefore, we exponentially filter the squared prediction error of each clock from the current measurement cycle to obtain a more statistically reliable estimate of an individual clock's variance. The filter is

$$\epsilon_i^2(t) = \frac{\hat{\epsilon}_i^2(t) + N_{\tau_{x,i}} \epsilon_i^2(t-\tau_{x,i})}{1+N_{\tau_{x,i}}} \quad (9)$$

Because we assume the noise characteristics of a clock are not stationary, past updates are deweighted in the filtering process. The time constant for the filter is typically chosen to be $N_\tau = 20$ d for cesium clocks, representing the time we expect the white FM level to remain constant. The initial value of ϵ_i^2 is estimated as $\tau_0^2 \cdot \sigma_y^2(\tau_0)$.

We identify the estimate of prediction error $\epsilon(t)$ with the white FM level σ_α (in units of time)--the expected amount of time dispersion after an interval τ_0. This is valid if the measurement interval τ is well within the range of integration times for which the white FM is dominant, as is the case with the system at NIST, which makes measurements every 2 h.

5.8 Frequency Estimate

Derived from a simple Kalman filter formalism [31,32], the method we use to calculate the frequency offset of each clock from the ensemble frequency involves an estimate of the nonstationary frequency changes y in the presence of white-noise frequency modulation. We estimate the flicker-plus-random-walk FM components of frequency based on the first difference of time. In this formalism the first difference of time offset is a measurement of y the white FM is the measurement noise R and the expected variance in frequency is the process noise Q. The Kalman formalism results in an exponential filter of the first difference of time, much as in the AT1 algorithm. In addition, the Kalman formalism provides a variance P of our estimate of the frequency, which we use to define a confidence on the current frequency. This method allows us to estimate frequency steps in AT2, as well as to combine forward and backward passes of AT2 to obtain the smoothed scale TA2.

The first difference of time, which is the average frequency of each clock at time t over the interval τ_x, based on the latest two time updates (when there are no time steps), is used as an estimate of that clock's current frequency y_{dx}. The calculation is

$$y_{dx_i}(t)=\frac{x_i(t) - x_i(t-\tau_{x,i})}{\tau_{x,i}}. \tag{10}$$

We predict y from previous time taking the last updated value of y and adding any deterministic frequency drift to obtain

$$\hat{y}_i(t) = y_i(t-\tau_{y,i}) + D_i\tau_{y,i}. \tag{11}$$

When a time step is detected, we cannot use the estimate y_{dx}, because the step in time inserts a bias. Hence there can be no frequency update until the next measurement cycle. At that time, the time since the last frequency estimate $\tau_{y,i}$ is larger than τ_x. We estimate y_{dx} over $\tau_{x,i}$, but we predict $\hat{y}$ over $\tau_{y,i}$.

Because we are estimating the non-stationary part of frequency that may be derived from averaging the white FM, the Kalman filter equations prescribe the formulas for combining our measurement y_{dx} with our prediction $\hat{y}$, based on previous estimates. These terms are weighted with the white FM level R as measurement noise and the predicted variance $\hat{P}$ as our confidence on y from the past:

$$y(t) = \frac{\hat{P}(t)\,y_{dx}(t) + R\,\hat{y}(t)}{\hat{P}(t)+R}. \tag{12}$$

Here, the Kalman formalism provides us with an exponential filter on y, as in AT1. In steady state, AT2 reduces to AT1 if the weights are chosen properly; AT2 thus inherits from AT1 the ability to model flicker FM.

The white FM level R is adjusted from σ_α^2, so that we may interpret it as the expected variation in frequency over $\tau_{x,i}$ due to the white FM process. This interpretation equates R to the average frequency variation over $\tau_{x,i}$ from the mean frequency offset of this clock from ensemble frequency.

This mean frequency is defined approximately because white FM is the dominant noise process. We assume that the random walk in frequency is slow compared to the white fluctuations in frequency. Thus, averaging the white FM defines an underlying mean frequency consistent with a pure random walk FM model. This definition of R equates it with the two-point, or Allan, variance and gives it units of frequency according to [18]

$$\begin{aligned} R &= \left\langle \left(\bar{y}(t+\tau)-y_{mean}\right)^2 \right\rangle \\ &= \frac{1}{2}\left\langle \left(\bar{y}(t+\tau)-\bar{y}(t)\right)^2 \right\rangle \\ &= \sigma_y^2(\tau) \\ &= \frac{\sigma_\alpha^2}{\tau_0|\tau_x|}. \end{aligned} \tag{13}$$

The brackets $< >$ in these equations mean "expected value." Because we estimate σ_α^2 as the "measurement noise," our system constitutes an adaptive Kalman filter, which allows P, the variance of the residuals of y, to evolve with changes in σ_α^2, from an initial value. Thus the integration time of the exponential filter on y, as expressed in equation (12), changes with time. This occurs when initializing a new clock and when a clock's white FM level changes.

5.9 Frequency Variance Prediction

We predict the frequency variance according to

$$\hat{P}(t) = P(t-\tau_y) + Q(\tau_y), \tag{14}$$

where $\hat{P}$ is the prediction of the variance of the residuals of y. The process noise in frequency Q is the variance in y due to random walk FM over the interval τ_y [20],

$$\begin{aligned} Q &= \left\langle \left(\bar{y}(t+\tau)-\bar{y}(t)\right)^2 \right\rangle \\ &= 2\sigma_y^2(\tau) \\ &= \sigma_\beta^2\frac{2n^2+1}{3n}, \end{aligned} \tag{15}$$

where n is the number of τ_0 intervals in τ_y,

$$n = \frac{\tau_y}{\tau_0}. \tag{16}$$

Q grows with τ_y as a function of σ_β^2, the expected variance in y over the interval τ_y due to flicker or random-walk FM.

5.10 Frequency Variance Update

As we obtain a measurement of frequency, so the confidence improves. Hence the variance P decreases as a function of the measurement noise. The Kalman formalism provides for our update of the frequency variance as

$$P(t) = \left[\frac{R\hat{P}}{R+\hat{P}}\right]_t. \tag{17}$$

6. TIME-STEP DETECTION

A time step is defined as an innovation greater than three times the prediction error, that is, when

$$|x_i(t) - \hat{x}_i(t)| > 3\epsilon_i(t). \tag{18}$$

The algorithm responds to a time step by deweighting that clock using the weight-control parameter, wct_i, for the current measurement cycle [17]. If we define the ratio of the innovation to the prediction error as the proportion

$$prop_i(t) = \frac{|x_i(t) - \hat{x}_i(t)|}{\epsilon_i(t)}, \tag{19}$$

then we set wct_i to

$$wct_i = \begin{cases} 1, & if\ prop_i \le 3, \\ 1 - (3 - prop_i)^2, & for\ 3 < prop_i < 4, \\ 0, & if\ prop_i \ge 4\ . \end{cases} \tag{20}$$

This allows a clock to lose weight smoothly as its time innovation increases from 3 to 4 times the prediction error. A quadratic dependence was chosen as the simplest way to decrease the weight with a zero initial rate of change.

7. FREQUENCY-STEP DETECTION

Frequency steps can be detected only some time after they occur [1,4]. Because our measurements are of phase differences between clocks with inherent white FM, frequency steps appear only after the accumulation of measurements. Flicker or random-walk FM can be understood as a series of stochastic changes in frequency, hence there is no mean frequency. Small frequency steps are indistinguishable from the stochastic noise. We define a frequency step as a shift in frequency greater than four times the standard deviation of the stochastic impulses. Empirically we have found that this detects frequency steps which are clearly significant, while leaving undetected smaller steps which can be interpreted as large stochastic impulses driving the random walk or flicker FM. This implies that the distribution of such impulses is not Gaussian. By contrast, a time step is a shift in phase of a clock that exceeds three times the current white FM characterization.

The detection of frequency steps may be obscured by the presence of time steps and the stochastic noise from white and random-walk FM. If a clock generates a time step, an average frequency spanning the time step will be biased by the size of the time step, divided by the interval between measurements. This bias can be large. Only after two measurements can we conclude that the new frequency after the time step is approximately equal to that before it and so distinguish between time and frequency steps.

It may also be useful to look backward beyond two measurements for the determination of small frequency steps. This procedure allows the white noise to be averaged down to the point where only random-walk or flicker FM remains. Thus the maximum useful extent of a search backwards for frequency steps is where the flicker or random-walk FM begins to dominate over the white FM.

We use as a measure of this time interval the time constant weighting the previous frequency estimates in the exponential filter for y (12). If we call this time interval L_{max}, the number of τ_0 intervals is then, from (12),

$$L_{max} = \frac{R}{\hat{P}}. \tag{21}$$

For $\hat{P}$ in (21), we use the steady-state value P_{ss} of P then add Q to it as indicated in (14). To obtain P_{ss} we set $P(t+\tau) = P(t)$ and solve, using equations (13) -(17). Also, substituting in the expressions for P and Q, equations (13) and (15), using $\tau=\tau_0$ we find

$$\begin{aligned} P_{ss} &= Q\left(\sqrt{\frac{1}{4} + \frac{R}{Q}} - \frac{1}{2}\right) \\ &= \sigma_\beta^2\left(\sqrt{\frac{1}{4} + 2\left(\frac{\sigma_\alpha}{\tau_0\sigma_\beta}\right)^2} - \frac{1}{2}\right). \end{aligned} \tag{22}$$

Adding Q to P_{ss} to obtain $\hat{P}_{ss}$, (14), substituting this into (21) and simplifying, we find

$$\begin{aligned} L_{max} &= \frac{1}{2}\left(\sqrt{1+4\frac{R}{Q}} - 1\right) \\ &= \frac{1}{2}\left(\sqrt{1+4\left(\frac{\sigma_\alpha}{\tau_0\sigma_\beta}\right)^2} - 1\right). \end{aligned} \tag{23}$$

We assume, for evaluating R and Q in (23), that $\tau=\tau_0$, because (23) has been derived from the steady state P.

Because we must measure clock differences, our best estimate of an individual clock frequency will be for a clock against the ensemble. But the ensemble itself must have stochastic processes, and these also need to be considered when testing for frequency steps.

The algorithm tests for frequency steps by iterating backward for each clock a range of measurements from two before the current measurement time up to L_{max} back. On each measurement cycle, for each clock, we maintain a buffer of the previous L_{max} values of all the necessary parameters for frequency-step detection. At each measurement, we search for frequency steps for each clock. We define the average frequency of a clock from the first difference over the time interval from the measurement L intervals back ($L < L_{max}$), x_{-L} at time t_{-L}, to the most recent measurement before the current time x_{-1} at time t_{-1}; that is,

$$y_{avg} = \frac{(x_{-1} - x_{-L})}{(t_{-1} - t_{-L})}. \tag{24}$$

We compare this frequency with the updated estimate of frequency $y(t_{-L})$ saved from time t_{-L}. That is, we compare a filtered estimate of frequency computed at the beginning of a fixed-time interval based on the clock's history with an estimate of the average frequency over that interval. If the frequency difference exceeds a test value, we conclude that the clock generated a frequency step.

The variance for the test is the sum of the expected variance for y_{avg} plus the variance for $y(t_{-L})$. The variance of y_{avg} is the sum of the expected variance of y against the ensemble over the interval from t_{-L} to t_{-1}, plus the variance of the ensemble itself. The expected variation of the ensemble is a function of both $\sigma^2_{\alpha,x}$ and $\sigma^2_{\beta,x}$ where

$$\sigma^2_{\beta_x} = \frac{1}{\sum_i \frac{1}{\sigma^2_{\beta_i}}}. \qquad (25)$$

For the variance of y against the ensemble over the interval, we use the maximum $P_{max}(t,L)$ of P either at the beginning or the end of the interval multiplied by M/L, because the variance decreases by the averaging time L.

The variance of the ensemble frequency due to white FM decreases as L, while the variation in ensemble frequency due to random walk FM increases with L. Finally, the variation in $y(t_{-L})$ is $\hat{P}(t_{-L})$. We use $\hat{P}$ instead of P, since we are testing for a frequency step. Hence we cannot use the measurement of frequency at time t_{-L} to decrease the uncertainty of frequency. For L large, typically the dominant term in σ_L is $\hat{P}$.

The test variance is

$$\sigma_L(t) = \sqrt{\frac{L_{max}}{L}\left[P_{max}(t,L)+\sigma^2_{\alpha_x}(t)\right]+L\sigma^2_{\beta_x}(t)+\hat{P}(t_{-L})}\ . \qquad (26)$$

We use four standard deviations of σ_L, simply because we empirically have found it appropriate. Smaller frequency steps have the same effect on the ensemble as random walk FM, which is already modelled.

Hence we define that a frequency step has occurred when

$$|y_{avg} - y(t_{-L})| > 4sig_L. \qquad (27)$$

If a clock generates a frequency step, usually that same step will appear over a contiguous range of test times. We look for frequency steps over all times in the allowable range and treat that having the largest inequality in (27) as containing the actual frequency step. If we find only one occurrence of a frequency step over a range of L values, we ignore it.

When a frequency step is detected, we re-run the scale from the time t_{-L} of that step with that clock deweighted. First we set the frequency and variance estimate of frequency at t_{-L} to

$$y(t_{-L}) = y_{avg},$$
$$P(t_{-L}) = \frac{\sigma^2_\alpha}{L} + \sigma^2_\beta \cdot L. \qquad (28)$$

We do not search for new frequency steps during this secondary run; thus there is no iterative search for frequency steps. The stepped clock remains deweighted until a time of L_{max} measurement cycles after the occurrence of the step. In other words, when we find a frequency step, that clock's frequency is re-initialized with an average frequency and the clock is removed from the scale until the algorithm adaptively learns the clock's new frequency.

8. TA2 FROM AT2

The smoothing algorithm TA2 essentially averages the filtered frequencies of clocks, generated by running AT2 forward and then backwards in time over the same data to obtain a "best" estimate of frequency for each clock. We run AT2 first forwards and then backwards in time over a fixed interval of data, saving the frequency values $y_i(t)$ of clocks, and their variances $P_i(t)$. In the third and final, forward pass through the data we average frequencies from the previous two passes for the predictions of x_i, for each clock as in (1). The frequency y_f from the previous forward pass is averaged with the frequency y_b from the backward pass using the reciprocals of their variances P_f and P_b respectively, for weights.

$$y_s(t) = \frac{\frac{y_f(t)}{P_f(t)} + \frac{\hat{y}_b(t)}{\hat{P}_b(t)}}{\frac{1}{P_f(t)} + \frac{1}{\hat{P}_b(t)}}. \qquad (29)$$

We use the prediction in the backward pass to avoid using the data at time t twice in estimating the smoothed frequency y_s. Then in the third pass, we generate ensemble time using these smoothed frequencies for prediction.

9. INITIALIZATION

When starting the algorithm with no ensemble history or else entering a new clock into an existing ensemble, we need an estimate of the deterministic parameters, offsets of time, frequency, and frequency drift from the ensemble, and also the stochastic parameters σ_α and σ_β of white and random walk FM.

When starting a new ensemble, the deterministic estimates must be consistent. If x_i and x_j are the initial estimates of time offset from ensemble time for clocks i and j respectively, then their difference must equal the measured difference; that is,

$$x_i - x_j = x_{ij}. \qquad (30)$$

Whereas this is the only requirement, any constant value may be added to all x_i, to predetermine the initial time value of the ensemble. Similarly, any constant may be added to all the clocks' frequency or drift estimates to predetermine the initial frequency or drift of the ensemble. Indeed, at any time, constants may be added (to all clocks) to steer the ensemble to any desired time, frequency or drift. The internal equations of the algorithm are not affected by such steering, which follows from our observation that we never measure the time of any clock, but only time differences between clocks. The algorithm optimizes stability among clocks, but provides no measure of accuracy.

The stochastic parameters can be estimated from an Allan or two-point variance of the clock, which is a variance that estimates fractional frequency stability [21-29]. In our formalism, $\sigma_\alpha(\tau)$ and $\sigma_\beta(\tau)$ are the expected dispersion of a clock after an interval τ, where the dispersion is one of time in ns due to white FM and of frequency in ns/d due to random walk FM, respectively. We enter the values for τ=1d, and internally convert them to $\tau=\tau_0$, two hours in our case. The conversion formula is

$$\sigma^2(\tau') = \sigma^2(\tau)\frac{\tau'}{\tau}, \qquad (31)$$

for both $\sigma_\alpha(\tau)$ and $\sigma_\beta(\tau)$. This can be derived using the formula relating them to the Allan variance from [18],

$$\sigma_y^2(\tau) = \frac{\sigma_\alpha^2}{n\tau_0^2} + \frac{\sigma_\beta^2(2n^2+1)}{6n}. \qquad (32)$$

If we then using the τ dependence of the Allan variance [21-27] for each of the two noise types separately in (32) we obtain (31).

11. CONCLUSIONS

We have shown the performance of the NIST post-processed time scale TA2 over a period of nine years and given the equations defining the algorithm generating these data. The algorithm itself shows properties similar to the real-time scale AT1 which has been running at NIST since the late 1960's. AT2 incorporates some changes from AT1 which should provide some advantages over AT1 for post-processing or for automatic error detection. We have also seen the performance of the NIST ensemble of atomic clocks over this period.

12. REFERENCES

[1] M. A. Weiss and T. Weissert, "TA2, A time scale algorithm for Post-Processing: AT1 plus frequency variance," IEEE Trans. Instrum. Meas., vol. 40, no. 2, April '91, pp. 496-501.

[2] M. A. Weiss and T. Weissert, "AT2, A new time scale algorithm: AT1 plus frequency variance," *Metrologia*, 28, pp 65-74, 1991.

[3] Weiss, M. A., Weissert, T., "Promise into Practice: Implementing TA2 on Real Clocks at NIST," Proc. 5th European Frequency and Time Forum, Besancon, March 12-14, 1991, pp. 442-448.

[4] Weiss, M. A., Allan, D. W., Peppler, T. K., "A Study of the NBS Time Scale Algorithm," *IEEE I&M*, vol. 38, no. 2, April 89, pp 631-635.

[5] D. W. Allan, D. J. Glaze, J. E. Gray, R. H. Jones, J. Levine, and S. R. Stein, "Recent Improvements in the Atomic Time Scales of the National Bureau of Standards," Proc. of the 15th Annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting, pp. 29-40, Naval Research Laboratory, Washington D. C., Dec. 6-8, 1983.

[6] NBS Monograph 140, 1974.

[7] M. A. Weiss, C. Hackman, "Confidence on the Three Point Estimator of Frequency Drift," *Proceedings of the 21st Annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting*, Nov. 1993, McLean, VA.

[8] D. A. Howe, D. W. Allan, J. A. Barnes, "Properties of Signal Sources and Measurement Methods," Proc. 35th Symposium on Frequency Control, pp.1-47, 1981.

[9] Barnes et. al. "Characterization of Frequency Stability," IEEE Trans. on Instrum and Meas., vol. IM-20, no. 2, pp. 105-120, 1971.

[10] R. H. Jones and P. V. Tryon, "Estimating Time From Atomic Clocks," Journal of Research of the National Bureau of Standards, vol. 88, no. 1, pp. 17-24, Jan - Feb 1983.

[11] Stein, S.R., Gifford, A., Breakiron, L.A., "Report on the Timescale Algorithm Test Bed at USNO," *Proceedings of the 21st Annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting*, Nov. 1989, Redondo Beach, CA, pp. 269-288.

[12] Jones, R. H., and Tryon, P. V., "Continuous Time Series Models for Unequally Spaced Data Applied to Modeling Atomic Clocks," *SIAM J. Sci. Stat. Comput.*, vol 8, no. 1, Jan 87, pp 71-81.

[13] Guinot, B., Thomas, C., "Establishment of International Atomic Time," *BIPM Annual Report*, 1988, Part D, pp. D1-D22.

[14] A. Bausch, et. al., "The New PTB Primary Cesium Clocks," special issue of IEEE Trans. on Instrum and Meas. for CPEM '92, in press.

[15] D. B. Percival, "The U.S. Naval Obervatory Time Scales," IEEE Trans. on Instrum and Meas., vol. IM-27, no. 4, December, 1978, pp.376-385.

[16] P. Tavella, J. Azoubib, C. Thomas, "Study of the Clock-Ensemble Correlation in ALGOS Using Real Data," Proc. 5th European Frequency and Time Forum, Besancon, March 12-14, 1991, pp.435-441.

[17] D. B. Percival, "Use of Robust Stastical Techniques in Time Scale Formation," 2nd International Symposium on Time Scale Algorithms, Boulder, Colorado, June, 1982, chapter 14.

[18] J. A. Barnes, R. H. Jones, P. V. Tryon, D. W. Allan, "Stochastic Models for Atomic Clocks," Proc. 14th Precise TIme and Time Interval Applications and Planning Meeting, pp 295-306, 1982.

[19] P. V. Tryon and R. H. Jones, "Estimation of Parameters in Models for Cesium Beam Atomic Clocks,"

Journal of Research of the National Bureau of Standards, vol. 88, no. 1, pp. 3-16, Jan.-Feb. 1983.

[20] J. A. Barnes, R. H. Jones, P. V. Tryon, and D. W. Allan, "Stochastic Models for Atomic Clocks," NASA Conference Publication 2265-- Proc. of the 14th Annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting, pp. 295-306, 1982.

[21] D. W. Allan, "Statistics of Atomic Frequency Standards," Proc. IEEE, vol. 54, no. 2, pp. 221-230, 1966.

[22] J.A. Barnes, A.R. Chi, L.S. Cutler, D.J. Healey, D.B. Leeson, T.E. McGunigal, J.A. Mullen, Jr., W.L. Smith, R.L. Sydnor, R.F.C. Vessot, and G.M.R. Winkler, Characterization of Frequency Stability, IEEE Transactions on Instrumentation and Measurement, IM-20, No. 2, 105-120, 1971.

[23] "Characterization of Clocks and Oscillators," Eds. D. B. Sullivan, D. W. Allan, D. A. Howe, and F. L. Walls, NIST Tech Note 1337, 1990.

[24] J. Rutman, "Characterization of Frequency Stability: A Transfer Function Approach and Its Application to Measurements via Filtering of Phase Noise," IEEE Trans. on I & M, Vol. IM-23, No. 1, March 1974, pp. 40-48.

[25] J. Rutman, "Characterization of Phase and Frequency instabilities in Precision Frequency Sources: Fifteen Years of Progress," Proc. of the IEEE, Vol. 66, No. 9, pp. 1048-1075, September 1978.

[26] J. Rutman and F. L. Walls, "Characterization of Frequency Stability in Precision Frequency Sources," submitted to Proc. of the IEEE, Vol. 79, No. 7, pp. 952-960, July, 1991.

[27] Allan, D.W., "Time and Frequency (Time-Domain) Characterization Estimation, and Prediction of Precision Clocks and Oscillators," *IEEE Trans. on UFFC*, vol. UFFC-34, no. 6, November, 1987, pp. 647-654.

[28] D. B. Percival, "Reappraisal of Frequency Domain Techniques for Assessing Stability Measures," Proc. of the PTTI, 1987, pp. 69-80.

[29] W. C. Lindsey and C. M. Chie, "Theory of Oscillator Instability Based Upon Structure Functions," IEEE 64, 1652-1666 (1976).

[31] A. Gelb, ed., *Applied Optimal Estimation*, M.I.T. Press, Cambridge, MA, 1974.

[32] J. Barnes,"Time Scale Algorithms Using Kalman Filters -- Insights from Simulation," Proc. 2nd Symposium on Atomic Time Scale Algorithms, NBS, Boulder, CO, June 23-25, 1982.

RUSSIAN NATIONAL TIME SCALE LONG TERM STABILITY ESTIMATION USING GLONASS TIME LINK

A.Al'shina, N.Koshelyaevsky, S.Pushkin and P. Daly*

Institute of Metrology for Time and Space, IMVP, NPO "VNIIFTRI"
Mendeleevo, Moscow region, 141570, Russia
*Department of Electronic & Electrical Engineering
University of Leeds, Leeds LS2 9JT, UK

ABSTRACT

The Russian national timing center at IMVP, NPO "VNIIFTRI" is the only timing laboratory in the world equipped purely with the ensemble of continuously operating super high stable H-masers. For the first time we have pleasure to report national timescale long term stability estimations based on H-maser intercom - parisons at the level better than $3x10^{-15}$ for averaging time up to 100 days. The time transfer link between IMVP, Mendeleevo and the University of Leeds, UK through signals of the Russian navigation system GLONASS reveals the superiority of national timescale UTC(SU) longterm characteristics relative to the wellknown time scale of US Naval Observatory - UTC(USNO) and confirms high stability relative to UTC.

KEY WORDS: time scales, H-maser, time transfer link, longterm stability.

INTRODUCTION

Long term stability characteristics are of the greatest importance for national timing centers. The limited number of laboratories use continuously operating laboratory Cs primary standards, lion's share of the others uses portable Cs beam clocks and a limited number of H-masers. The Russian national time laboratory located at Mendeleevo, near Moscow is as we know the only laboratory which uses during several decades for timekeeping needs the ensemble of H-masers.

Starting from second half of 1989 a new commercial H-masers were introduced into the ensemble. H-maser intercomparison data for the period 1990 - 1992 shows extremely high stability characteristics; the estimated UTC(SU) modified Allan variance for time interval from 1 to 100 day gives values better $3x10^{-15}$. The estimated frequency drift less than $3x10^{-17}$/day. These data may be to some extend overstated due to possible coherent phenomena inside the ensemble, that's why these need to be verified relative to independent high stable time scale.

Fortunately since 1990 we have precise enough time transfer link between UTC(SU) and UTC, and in that way with leading international timing centers. These time link is based on Russian navigation system GLONASS signal reception during overlapping sessions in the laboratories of IMVP and the University of Leeds and provides one UTC(SU) - UTC(USNO) time difference data at one day basis.

This time link have revealed the superiority UTC(SU) long term stability relative to the wellknown time scale of US Naval Observatory. The better longterm stability was achieved relative to UTC, at the level better $1x10^{-14}$ for averaging time 100 days. These data also didn't reveal detectable frequency drift at the level $1x10^{-16}$/day.

Perhaps the most striking result was obtained when comparing behavior of UTC(SU) relative to UTC(PTB) during 1992: UTC(SU) longterm stability was significantly better relative to UTC(PTB) than even to UTC. We hope in future we shall have opportunity to investigate this problem deeply.

1. RUSSIAN NATIONAL TIME SCALE UTC(SU)

About 30 years one have started to use H-masers for time keeping needs in Russian national time center at VNIIFTRI and more than 20 years national time scale is based purely at the ensemble of H-masers (Ref.1). During that time the ensemble was twice removed from one building to another. Starting from 1989 laboratory occupies new alone building which is located at a calm forest about 1 km from the main territory of VNIIFTRI. In the second half of 1989 an ensemble of commercially available H-masers was put into operation in this laboratory. Starting from 1990 the national time scale of Russia - UTC(SU) is being reproduced basing on these ensemble of H-masers.

The H-masers and the time and frequency comparison instruments are located at separate termostabilized better than 1 ^{o}K rooms. A humidity and a pressure are monitored at 1 day basis.

Each H-maser has two main output signals - sine wave 5 MHz and 1 pps. The frequencies and 1 pps signals of H-masers in the ensemble are deliberately shifted in order to prevent possible pulling effects. A relative frequency difference exceed some parts 10^{-13}

There are two ways of H-masers intercomparisons. One may is to compare 5 MHz signal using set of phase comparators, which multiplies phase difference between H-maser under investigation and reference 100 times and then produce 1 pps output signal with the increased phase difference. These signals are fed through multiplexer to a time interval meter with a resolution about 300 ps. The set of original 1 pps signal from H-masers are also fed through multiplexer to the same time interval meter.

In such a way one get time difference data between any pair of H-masers at 1 hour basis. At the one day basis the frequency resolution of these data is about $7\times10^{-15}$for original 1 pps signals and better 1×10^{-15} for phase comparator signals. The main drawback of the measuring system is the absence of a reliable and powerful enough computer for data collection. The only actually available data for long term analysis are ones related to original 1 pps signals at 1 day basis.

The national time scale of Russia - UTC(SU) is a "paper time scale" which originates from the above mentioned a free ensemble of continuously operating H-masers, (Ref.2). The definition of the free ensemble of the time-keeping instruments presupposes that the size of the National time unit has been fixed some years previously and then was kept stable as much as possible basing on data of H-masers intercomparisons only. This means that, despite the fact that UTC(SU) is referred to as a "coordinated time scale", it is actually an atomic time scale into which the "leap" second is introduced. Since the time when the National time unit was fixed, the membership of the original ensemble of clocks has of course changed considerably - nowadays there are no H-masers of original starting ensemble.

The time scale algorithm used for UTC(SU) generation is very simple. One estimate individual frequencies of H-masers relative to each other and to to averaged frequency of the ensemble. If one sees significant frequency change of the individual H-maser this standard is excluded from the ensemble. It may be included back to the ensemble if it doesn't manifest a frequency drift and it's new frequency is determined well enough. Needless to say that for UTC(SU) calculation one needs to take into account frequency corrections due to original frequency differences of the standards.

As a matter of fact times to times

one needs to redetermine frequency corrections. Actually it is done very seldom, for the best instruments rarely than ones per year.

2. H-MASERS INTERCOMPARISONS DATA AND UTC(SU) STABILITY ESTIMATION

At a moment there are four H-masers in the ensemble, nos.47, 49, 50, 51 (see Fig. 1). One new instrument no 53 belongs to a new generation of H-masers, it is under investigation and is located at a separate room. Fig. 1 presents linear detrended time differences original 1 pps signals of H-masers during second half of 1992. The readings are referred to 00 h UTC. These data were specially processed to remove all detected time steps up to he level of 0.7 ns (sometimes it happens due to a cycle skipping of phase locked loop VXCO across maser signal).

These data give the relative frequency stability of pairs of standards presented in Fig.2. The apparent deterioration of presented stability for one day averaging time may be explained by lack of sensitivity of the measuring system. The data show an individual relative frequency instability of H-masers at the level about 2×10^{-15} for averaging time up to 30 days and doesn't reveal detectable individual frequency drift at the level 3×10^{-17}/day.

Next portion of the data for the period of 1991-1992 is presented in Fig.3. These are 10 days data which were sent to the BIPM. The readings are referred to a standard date - 00 h UTC MJD xx9. The most significant time steps are removed from presented data, but not each step, especially less than 3-5 ns.

The corresponding relative frequency stability estimations are presented at Fig.4. The slightly worse stability level may originate from the missed time steps in the original 10 days data. Solid line stability estimation of UTC(SU) demonstrates flicker floor at the level about 2×10^{-15} and shows a tendency of frequency drift at the rate less than 3×10^{-17}/day.

It is worth to note that the above mentioned estimates don't reflect possible correlation in frequency changes between H-masers. To reveal such a correlated behavior one needs to refer these secluded ensemble time scale to stable enough independent time scale.

3. GLONASS TIME TRANSFER LINK

At the beginning of 1990 on the initiative and support of the BIPM, the laboratories at the University of Leeds and IMVP started joint work on GLONASS signal application for time transfer between UTC(SU) and UTC(USNO) and UTC itself.

The timing laboratory at Leeds is equipped with three receivers. The first receiver, a TRIMBLE 4000A produces at its output socket a 1 PPS signal locked to UTC(USNO) based on broadcast GPS navigational message. This signal is then used as reference for the second receiver (designed and built in Leeds), a GPS/GLONASS instrument. A third receiver (TRIMBLE 5000A until the end of 1991, Allen Osborne TTR6 after the start 1992) is used for time comparison of local clocks - UTC(LDS) - under the international GPS common-view schedule produced by the Time Section of BIPM. All day and every day the joint GPS/GLONASS instrument performs automated measurements of UTC(USNO) - satellite time for all available GLONASS and GPS satellites. The duration of the track time each day is about 0.5 day for GLONASS and 0.3 day for GPS. The raw data taken at the rate of 1 reading/s is averaged over 180 s and stored. Daily averages of UTC(USNO) - GPS/GLONASS system time are also calculated. All contributing receivers at Leeds employ the same omnidirectional antenna, the location of which has been determined with an uncertainty of 3 m in the WGS 84 reference frame (Refs.3,4).

At the laboratory of IMVP sporadic GLONASS signal reception was started in the second half of 1989; from the

beginning 1990 it began operating regularly according to its own internal schedule which was defined primarily for comparison of time scale of the laboratories located inside Russia at Novosibirsk, Irkutsk, Khabarovsk and Petropavlovsk - Kamchatski. A GLONASS navigation receiver of A-724M-01 type designed for aircraft was used at IMVP. It was updated to allow access to the difference "local time" and GLONASS System Time (Ref.5). During each day, there are 4 - 5 half hour manual sessions, producing about 25 readings from GLONASS satellites passing to the East of Mendeleevo. This raw data is averaged to produce daily averaged points of UTC(SU) - GLONASS system time. The location of antenna has been determined from about 50 navigation solution with uncertainty about 1.5 m in SGS 85 reference frame. Starting from the middle of 1991 GPS timing receiver of the type Allen Osborne TTR6A was installed at the IMVP, providing new opportunities for time comparison of UTC(SU) with international community (Ref.6).

4. UTC(SU) LONG TERM STABILITY ESTIMATION BASED ON INTERNATIONAL COMPARISONS DURING 1990-1992.

Subtracting two data streams from Leeds and Mendeleevo one obtains - see Fig.5 - direct readings of UTC(SU) - UTC(USNO). The depicted dependence confirms the high stability of two time scales under examination, showing relative time units difference about 1×10^{-13} Removal of the constant frequency difference between two time scales reveals the random part of the time scale difference shown in Fig.6. The natural question which Fig.6 raises is the following - which time scale is more responsible for the obvious frequency variation? To find a solution one needs more stable reference than the two time scales under test. To analyze the behavior, we take into account additional data disseminated by BIPM Circular "T" publication. The overlayed boxes curve at Fig.6 presents detrended dependence of UTC - UTC(USNO). The obvious correlation between the dependencies at Fig.6 allows one to conclude that the source of frequency changes originates mainly from UTC(USNO).

Fig.7 shows the detrended dependencies of UTC - UTC(SU) for the period of 1990 - 1992 obtained in two ways: via the time link Leeds - Mendeleevo and via official information from the BIPM Circular "T". From the beginning of June 1990 (MJD~48100), the source of Circular "T" data was the data produced during a one-way GLONASS session from Leeds, supplemented by the τ_c value transmitted on line 5 of GLONASS navigation message. This τ_c quantity is the predicted value of UTC(SU) - GLONASS system time. From February 1991, (MJD~48310), the Circular "T" published data were based on the time link between Leeds and Mendeleevo giving rise to a dramatic increase in precision. Starting from the beginning of 1992, the Circular "T" data used data from the GPS TTR6A receiver installed at IMVP. One can observe that the predicted value of UTC(SU) using τ_c quantity is much worse than the information on UTC(SU) available via Leeds - Mendeleevo time link.

The similar dependencies one may see in Fig.8 which depicts detrended series of UTC(SU) - UTC(PTB).

The estimations of long term stability for averaging time from 1 day to 300 days are presented in Fig.9. Two main conclusions leads from Fig.9

i. First of all the reported GLONASS time link which don't follow "common view" restriction is precise enough to compare one of the most stable time scales.

ii. It clearly demonstrates the superiority of the UTC(SU) long term stability over UTC(USNO).

The anomalous points for 1 year averaging time may follow from limited time interval - only three years.

Perhaps the most striking result was obtained during separate data processing for the period of 1992 year. During this year our laboratory has two operational

time links: i. GPS "common view" sessions under international schedule and ii. The above mentioned GLONASS time link. This gives one opportunity to estimate GLONASS time link precision limitations, see Fig. 10. The data of 1992 through GLONASS evidently show that UTC(SU) long term stability estimations look much more better relative to TA(PTB) than even to UTC! Concerning UTC(USNO) the result agrees with previous one for the period 1990 - 1992. The same data through GPS enhance this impression.

This on one hand pleasant and to other hand alerting result have to be further investigated in future.

CONCLUSION

For the first time using national means of comparison of the remote time scales: GLONASS signals we were able to demonstrate extremely high longterm stability characteristics of the Russian's national time scale which originate nothing but four H-masers.

Acknowledgements. The authors are greatly indebted to Director of BIPM Dr T.J.Quinn and Heads of BIPM Time Section Dr B.Guinot and Dr C.Thomas, for helpful support in arranging both GLONASS and GPS time links between IMVP and international timing community. They also would like to express there gratitude to technicians for their patient and qualified manual servicing of the GLONASS receivers.

REFERENCES

1. N.B.Koshelyaevsky, Proceedings of the 5-th European Frequency and Time Forum, Besanson, March 1991, p. 415.
2. N.B.Koshelyaevsky, S.B.Pushkin, Proceedings of the 22-nd Annual Precise Time and Time Interval (PTTI) Application and Planning Meeting, December 4-6, 1990, Tyson Corner, Va., USA.
3. P.Daly, G.T.Cherenkov, N.B.Koshelyaev - sky and S.B.Pushkin, Proceedings of the 4th Institute of Navigation Technical Meeting, 199-206, September 91.
4. P.Daly, N.B.Koshelyaevsky, W.Lewandow - sky, G.Petit, C.Thomas, Proceedings of the 23-rd Annual Precise Time and Time Interval (PTTI) Application and Planning Meeting, December 3-5, 1991, Pasadena, CA, USA.
5. Н.Б.Кошеляевский, С.Б.Пушкин и Г.Т.Черенков, Труды Советско - Западно германского совещания "Научное и промышленное сотрудничество при использовании спутниковой навигационной системы "ГЛОНАСС", 24-30 сентября, 1990 г., Ленинград, СССР.
6. G.T.Cherenkov, N.B.Koshelyaevsky, S.B.Pushkin, G.Petit, C.Thomas, Proceedings of the 24-th Annual Precise Time and Time Interval (PTTI) Application and Planning Meeting, December , 1992, USA, to be published.

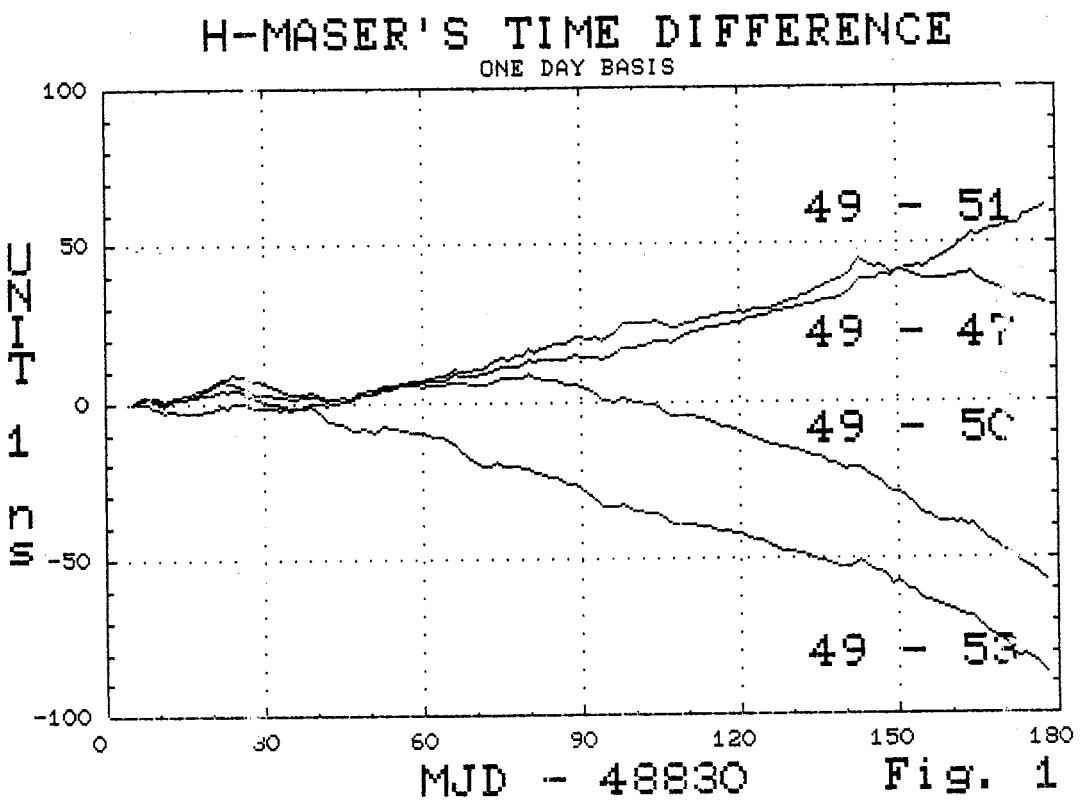

H-MASER'S TIME DIFFERENCE
ONE DAY BASIS
100
50
0
-50
-100
UNIT 1 ns
49 - 51
49 - 47
49 - 50
49 - 53
0
30
60
90
120
150
180
MJD - 48830

Fig. 1

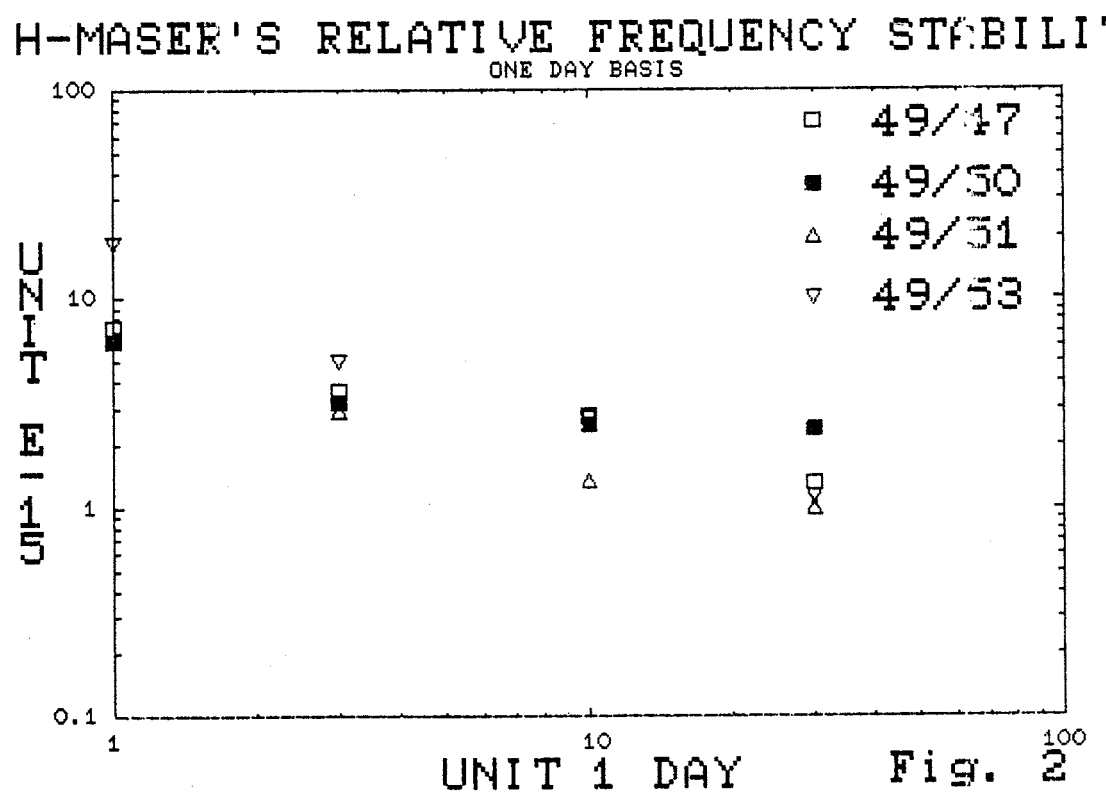

H-MASER'S RELATIVE FREQUENCY STABILITY
ONE DAY BASIS
100
10
1
0.1
UNIT E-15
49/47
49/50
49/51
49/53
1
10
100
UNIT 1 DAY

Fig. 2

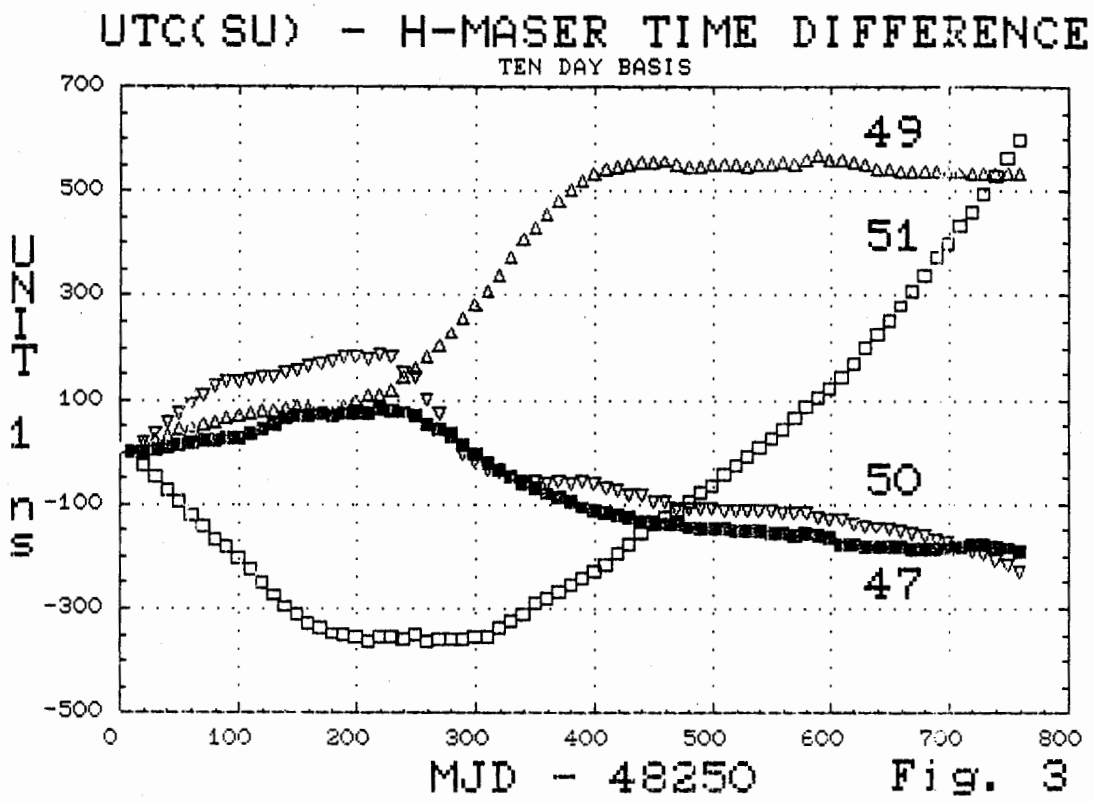

UTC(SU) - H-MASER TIME DIFFERENCE
TEN DAY BASIS
UNIT 1 ns
700
500
300
100
-100
-300
-500
0
100
200
300
400
500
600
700
800
49
51
50
47
MJD - 48250

Fig. 3

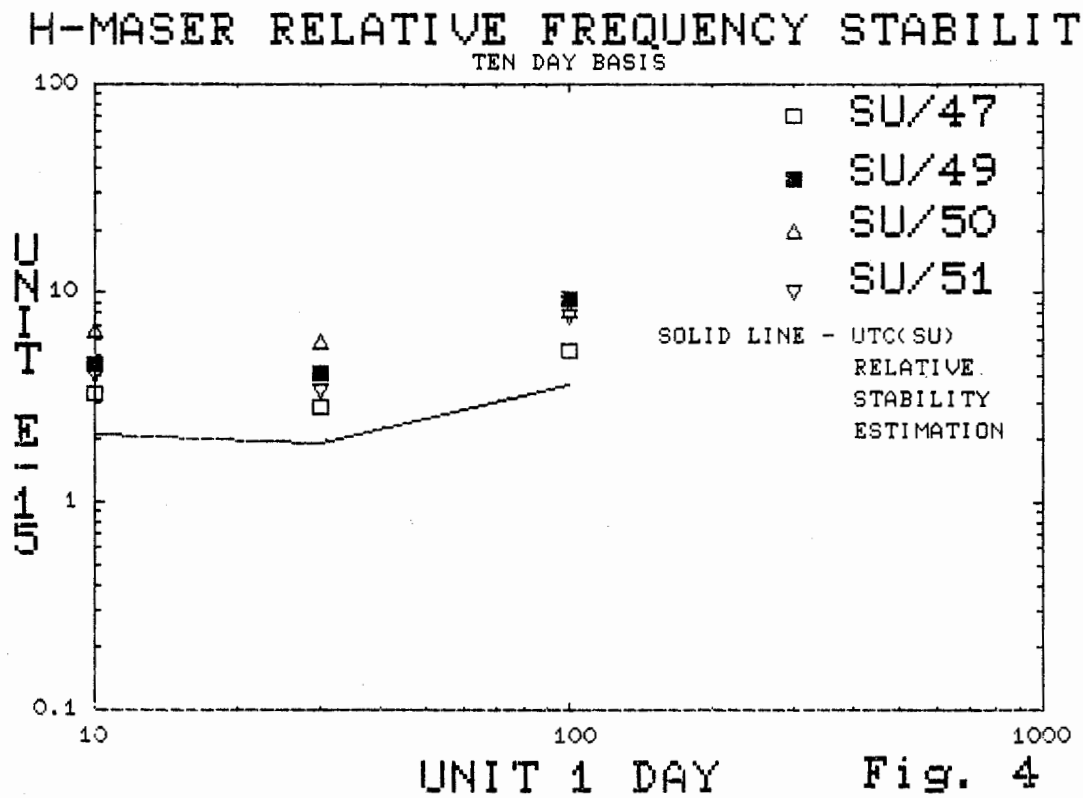

H-MASER RELATIVE FREQUENCY STABILITY
TEN DAY BASIS
UNIT 1 ns
100
10
1
0.1
10
100
1000
SU/47
SU/49
SU/50
SU/51
SOLID LINE - UTC(SU)
RELATIVE
STABILITY
ESTIMATION
UNIT 1 DAY

Fig. 4

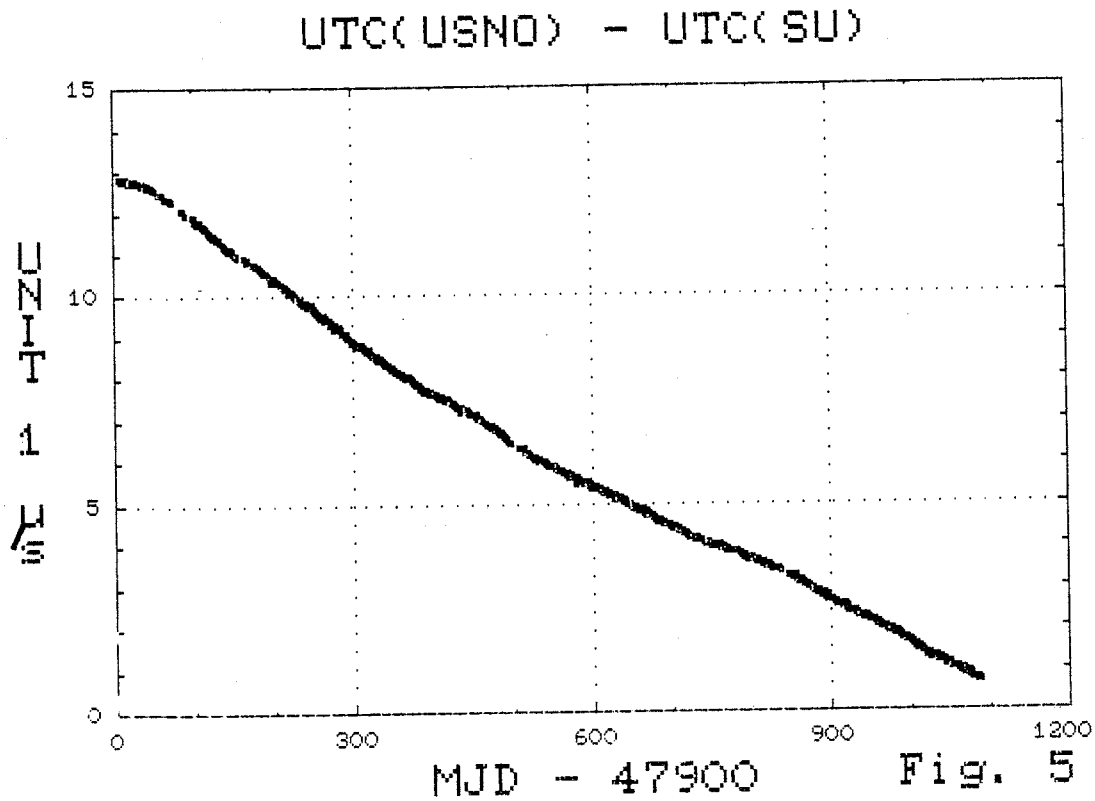

UTC(USNO) - UTC(SU)
15
10
5
0
UNIT 1 μs
0
300
600
900
1200
MJD - 47900

Fig. 5

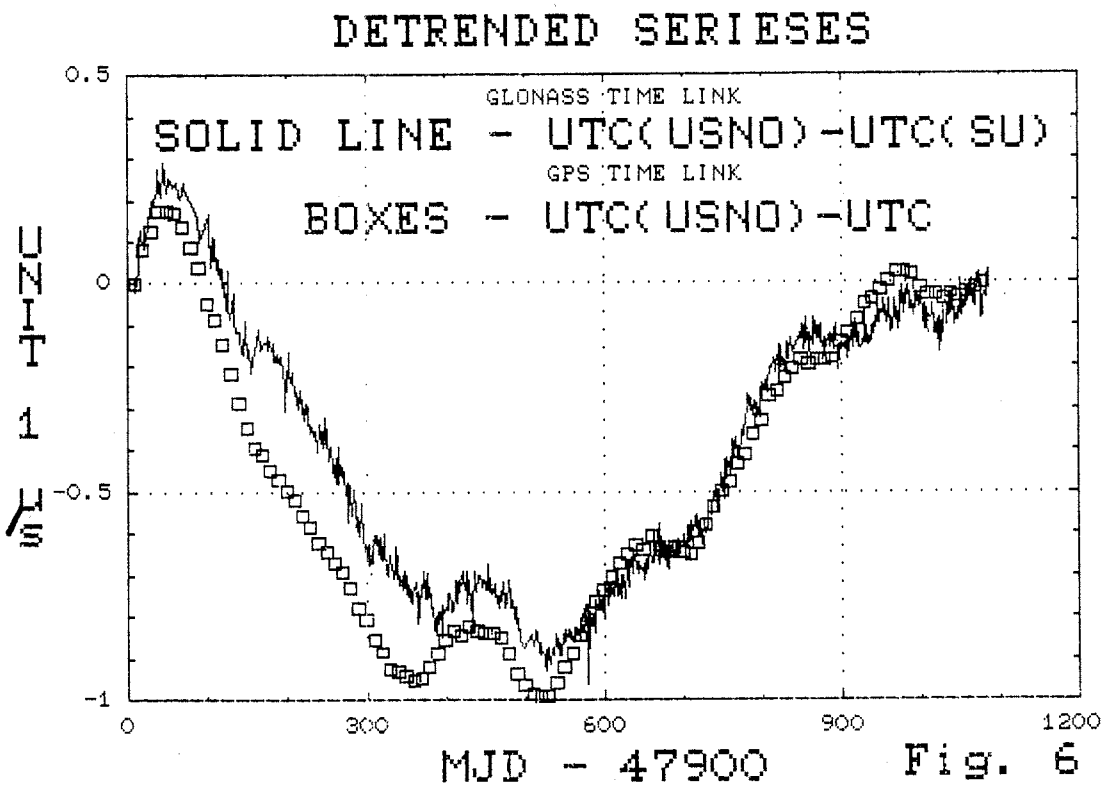

DETRENDED SERIESES
GLONASS TIME LINK
SOLID LINE - UTC(USNO)-UTC(SU)
GPS TIME LINK
BOXES - UTC(USNO)-UTC
0.5
0
-0.5
-1
UNIT 1 μs
0
300
600
900
1200
MJD - 47900

Fig. 6

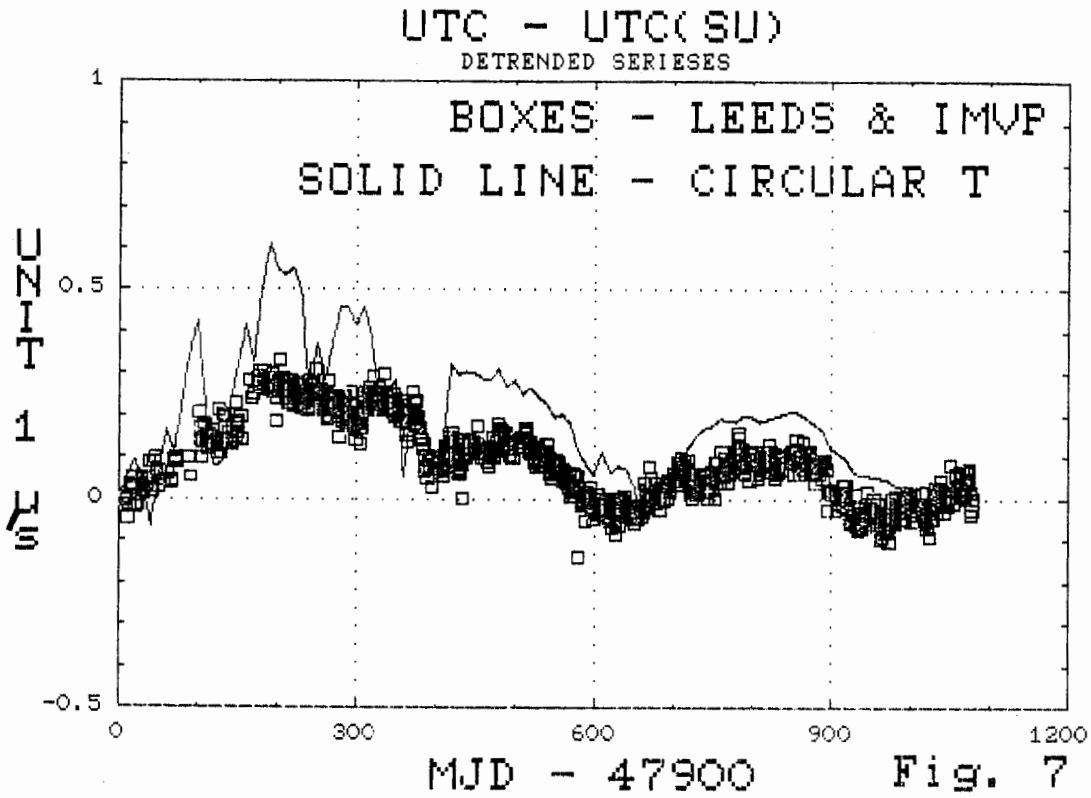

UTC - UTC(SU)
DETRENDED SERIESES
BOXES - LEEDS & IMVP
SOLID LINE - CIRCULAR T
UNIT 1 μs
1
0.5
0
-0.5
0
300
600
900
1200
MJD - 47900

Fig. 7

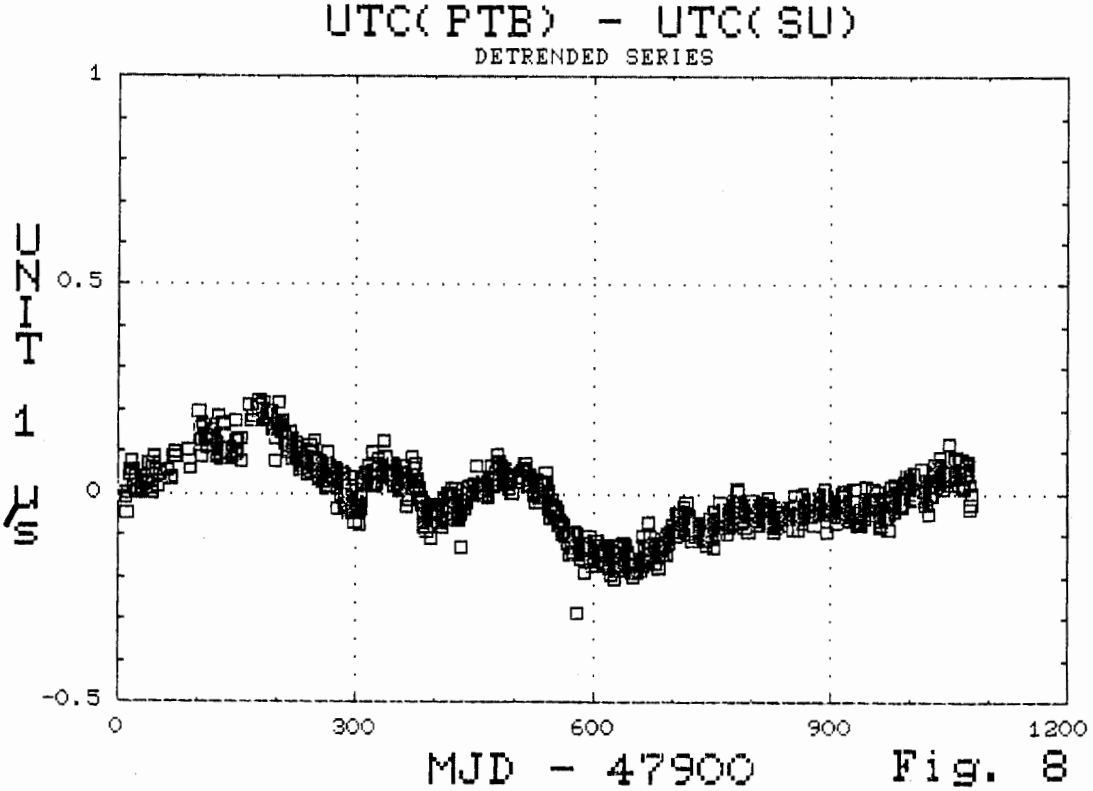

UTC(PTB) - UTC(SU)
DETRENDED SERIES
UNIT 1 μs
1
0.5
0
-0.5
0
300
600
900
1200
MJD - 47900

Fig. 8

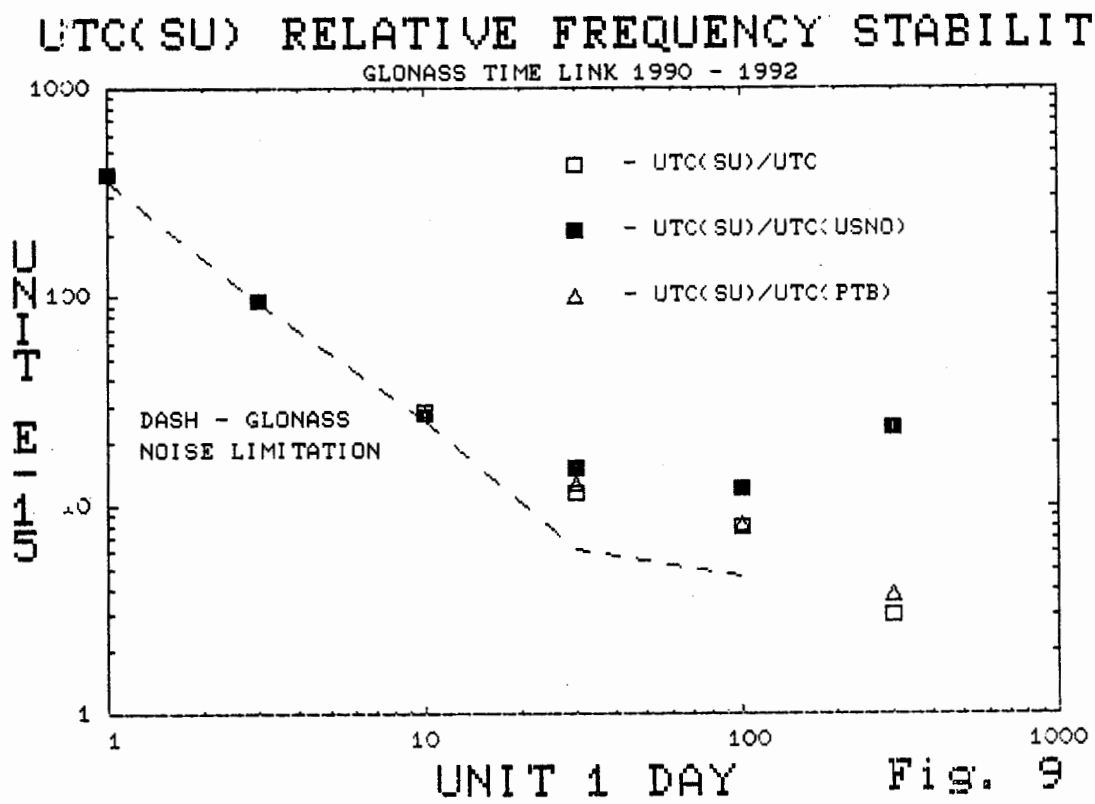

UTC(SU) RELATIVE FREQUENCY STABILITY
GLONASS TIME LINK 1990 - 1992
1000
100
10
1
UNIT E-15
□ - UTC(SU)/UTC
■ - UTC(SU)/UTC(USNO)
△ - UTC(SU)/UTC(PTB)
DASH - GLONASS
NOISE LIMITATION
1
10
100
1000
UNIT 1 DAY

Fig. 9

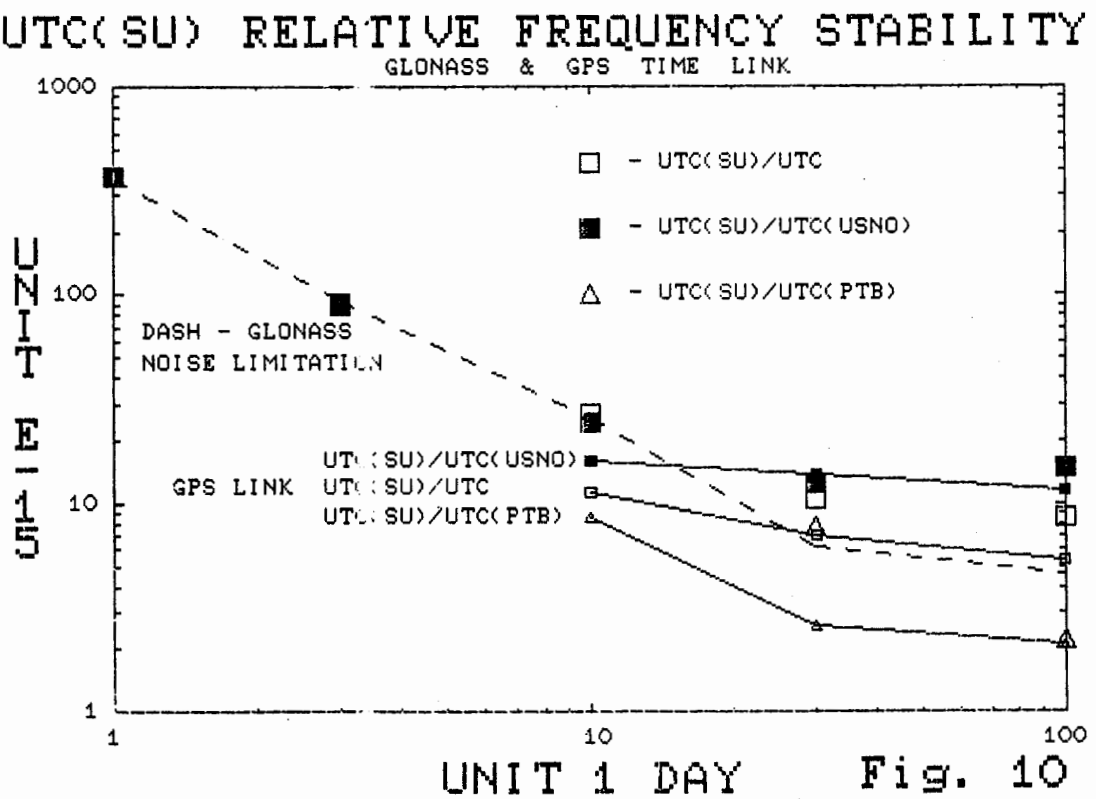

UTC(SU) RELATIVE FREQUENCY STABILITY '92
GLONASS & GPS TIME LINK
1000
100
10
1
UNIT E-15
□ - UTC(SU)/UTC
■ - UTC(SU)/UTC(USNO)
△ - UTC(SU)/UTC(PTB)
DASH - GLONASS
NOISE LIMITATION
GPS LINK
UTC(SU)/UTC(USNO)
UTC(SU)/UTC
UTC(SU)/UTC(PTB)
1
10
100
UNIT 1 DAY

Fig. 10

TRENDS IN INTERNATIONAL TIMING

by

David W. Allan
Allan's TIME
Fountain Green, UT 84632-0066

and

Alex Lepek
Consultant, Tech Projects
Jerusalem, Israel 96920

ABSTRACT

A long-term view of the trends in international timing reveals some astounding changes. These changes have come in response to several inventions and technological opportunities. It seems we are now at the threshold of another important change responding to current needs and new technologies. We now have, as it were, "Grand-father clocks" in the sky for all to look at. GPS and GLONASS (and in the future INMARSAT) provide such a service. In contrast to the sun as used anciently for timing, these "common suns," available to us all, are now shining 24 hours a day.

The current UTC(k)s as generated by the various timing centers around the world now differ by several microseconds. But there is a strong move to bring them into much closer synchronism in response to the growing need within the telecommunication community. Recently Dr. Ghani Abbas (System Design/International Standards Manager of a major telecommunications company (GPT) in the UK) said, "New technologies... are being introduced into the telecom transport networks. ...In order to meet the synchronization needs of these new emerging technologies, a convenient tie to UTC to better than 100 ns is required. In the long term this requirement will be a global one."

Several significant advancements have recently occurred, which make this 100 ns goal a near term possibility. The Bureau International des Poids et Mesures (BIPM), who is the custodian of UTC, is very actively working appropriate issues to improve this official international reference time scale. Along with improving UTC, one of the goals for the future is to make it more readily available. These advancements and some future possibilities will be discussed in the paper.

INTRODUCTION

Telecommunications capacities and techniques are currently advancing rapidly. Similarly, time and frequency continues to advance it's accuracies, stabilities and comparison techniques. These new techniques now available give the opportunity for a different perspective than could be had a few years ago. The impact of microprocessors and computers on clock systems is enormous. Rapid advances in telecommunications are anticipated to yield very high data rate transport. The advent of the Synchronous Digital Hierarchy (SDH) and the Synchronous Optical Network (SONET) are requiring the best of clock technology and transmission systems.

It has become much more important to take a system approach. In this regard a much closer relationship between the telecommunications community and the time and frequency community has developed with very useful synergistic outputs. Inasmuch as good time and frequency techniques are indispensable to improving telecommunications, seeing these techniques incorporated within telecommunications is a happy trend and state of affairs, but much more can and needs to be done.

In response to the current and anticipated needs within the telecommunications industry, the 1993 CCDS wrote a recommendation encouraging the study of the technical problems associated with the goal of 100 ns worldwide synchronization. As these problems are studied, it is anticipated that many of them will have solutions within currently-available resources. The resources within time and frequency have improved dramatically over the last decade, and there is good reason to believe they will continue to improve. The availability of GPS, alone, has been very significant to the telecommunications industries (telecom) as well as to time and frequency community.

If you ask the question of the telecom leaders, "What should be your primary reference for time?" They will invariably say, "UTC." If you then ask, "How many of you are using UTC?" The answer will almost always be no one! Everyone knows UTC is an outstanding time scale, but the problem is that it tells you what time it was two months ago, and, in addition, it is not readily accessible in any direct way. Indirectly, it is accessible via GPS, which also broadcasts UTC(USNO). In turn UTC(USNO) is now steered to within 100 ns of UTC. Some countries desire direct access to UTC.

In visiting recently with Dr. Ghani Abbas, System Design/International Standards Manager of GPT in the UK, he said,

> "New technologies (such as SDH/SONET/ATM) are being introduced into the telecom transport networks. These technologies require good quality synchronization as well as better short term clock stability. In order to meet the synchronization needs of these new emerging technologies, a convenient tie to UTC to better than 100 ns is required. In the long term this requirement will be a global one."

TIME AND FREQUENCY METROLOGY CONCEPTS AND DEFINITIONS

The following concepts and definitions will be useful for better understanding this paper. First, let us define time and frequency accuracy and stability. (See CCIR Glossary of definitions)

By **frequency accuracy** we mean how well the interval between the 1 pps (pulse per second) output of a clock conforms to the SI second. In other words, it is how well the clock's frequency conforms with the definition as given, at present time, by the cesium atom.

By **frequency stability** we mean how stable the frequency is as a function of averaging or integration time, τ. Intuitively, it is a measure of how much the frequency changes as averaged over one interval τ to the next interval τ. In the time-domain, by CCIR and IEEE recommendation, this measure is $\sigma_y(\tau)$.

Time accuracy has no meaning by itself, in an absolute sense, and takes on meaning with respect to a reference time scale or clock. The official-international reference time scale is UTC. The UTC(k)s, as generated by different timing centers in their respective countries, are, by international agreement, to steer their time scales to UTC. Many of the telecom servers have agreed to use UTC as their official reference. Time accuracy, then, is the degree of conformity to UTC as a relativistic coordinate time scale defined on the rotating geoid.

By **time stability** we mean how stable the time is as a function of the averaging or integration time, τ. Intuitively, it is a measure of how stable is the change in the time-difference between two clocks, averaged over an interval τ, with respect to a similar average taken just before and just after this interval. In the time-domain, both for telecom and for the time and frequency community, the recommended measure is $\sigma_x(\tau)$.

If one examines the fundamental physics of the noise of time and frequency metrology some important insights are gained. Whereas, $\sigma_y(\tau)$ is an optimum estimate of frequency change, in a second-moment sense, for the frequency of a clock limited by white-noise frequency modulation (FM); $\sigma_x(\tau)$ is an optimum estimate of the change, in a second-moment sense, in the average time readings in a network, or time distribution or measurement system where the limiting noise is white phase modulation (PM). In the case of most frequency standards, white-noise FM is the theoretical limiting noise for intermediate and long-term stability. In the case of networks, and time measurement systems, white-noise PM is the theoretical limiting noise. In practice, the spectrum will often also have lower Fourier-frequency components. In either case, these second moment measures can be transformed into the frequency domain for spectral density estimates.

The **rms time error of prediction** is very different from time stability. It will be a function of the prediction algorithm as well as of both the systematic variations and of the random variations. Different types of random variations exist in the output of clocks and in measurement systems. For the random variations, and under the assumption of optimum prediction, an estimate of the rms time prediction error is given by $k \times \tau \times \sigma_y(\tau)$, where k is equal to $1/\sqrt{3}$ for white-noise PM, 1 for white-noise FM and random-walk FM and $\sqrt{(2*\ln(2))} = 1.2$ for flicker noise FM. In practice, we usually have a composite of several different spectral-density types as well as different systematic perturbations which influence this error. An optimum prediction algorithm is complicated by this, and must take all these things into account, or the time prediction will be degraded.

Currently UTC differs from TAI by an integer number of "leap-seconds." The leap-seconds occur as announced by the International Earth Rotation Service (IERS) in conformity with the international agreed upon goal to keep UTC and UT1 (earth time) in agreement to within 0.9 seconds.

Within a microsecond, GPS time is TAI - 19 seconds. In finer detail it is steered to be in agreement with UTC(USNO) modula one second. The steering time-accuracy requirement is 1 μs, but in practice it is typically kept within about 100 ns. Additionally, the GPS broadcasts the correction information necessary to estimate UTC(USNO). The regulation states that this conformity shall be within 100 ns. In practice, the conformity is typically better than 20 ns.

FROM AMMONIA RESONANCE TO SYSTEM TIME

Dramatic have been the improvements since I.I. Rabii's dream was first physically realized with the ammonia maser. As an absorption cell, its stability was not much different than that of the spinning earth. Since then we have seen improvements in primary cesium-beam frequency standards at a rate of about a factor of ten per seven years, and we now have atomic clock stabilities a million times better than the spinning earth. R&D evidence would indicate this rate of improvement will continue with 1 X 10^{-18} in view for an optical transition in a cooled Hg ion standard. [1]

Similar improvements have occurred in commercial clocks. The comparison of the first commercial cesium beam frequency standard, the Atomichron , produced in the early sixties, with the Hewlett Packard model 5071A, illustrates similar dramatic change. Though the model 5071A was only announced in December of 1991, and is too new to judge life time and long-term reliability, the design criteria are in harmony with such goals. The model 5071A has demonstrated to date the best accuracy and stability for a commercial cesium standard. [2] The demonstrated accuracy appears comparable to many of the laboratory standards at about 1 X 10^{-13}, and the long-term stability is better than 1 X 10^{-14} [3]. The long-term noise spectrum is characterized by white-noise FM, hence the time error of prediction is proportional to $\tau^{1/2}$, where τ is the prediction time. It has a two-month rms optimum prediction error of about 20 ns, and day-to-day of less than 3 ns. Already about 50 of the model 5071As have been introduced into the computation of TAI and UTC and the effects are dramatic.

There are three significant innovative advantages of the model 5071As. First, its environmental insensitivity gives it a long-term stability better than 1 X 10^{-14}. The second advantage is that it's accuracy is about an order of magnitude better than any other model of cesium beam frequency standard. And third, its output can be steered. In other words, it can be used, at the same time, both as an independent reference clock for optimum time prediction, but also as a controlled clock. We will discuss later a future opportunity for this third advantage.

Hydrogen masers have also shown some dramatic improvements in recent time. Along with their outstanding short-term stability, those having cavity-tuning servos have demonstrated frequency drifts of the order of 1 X 10^{-16} per day. [3,4] Several have demonstrated long-term frequency stabilities approaching 1 X 10^{-15}, where the averaging time has been several weeks. These stabilities could provide an optimum time error of prediction of about 10 to 20 ns for two months. The long-term random-instabilities of H-masers are modeled by either 1/f FM or f^{-2} FM. This results in a time error of prediction proportional to τ or $\tau^{3/2}$, respectively.

The theory of time scale algorithms has also made significant advancements over the decades of atomic time keeping. It has been shown, in theory, that a computed time can have better stability than the best of the clocks in the contributing ensemble -- both in the short-term and the long-term. Further, it has been shown that even the least stable clock in the ensemble need not degrade the performance of the computed time and can enhance it -- even though it may be by a small amount. This theory has been verified experimentally. Robust features have also been built into algorithms as well as tests for abnormal behavior. The net of the R&D efforts is that algorithms are now available which can give near optimum performance while still maintaining robustness, reliability and excellent error and abnormal behavior detection and protection. [8]

The discovery of the first millisecond pulsar, PSR 1937+21, by Backer and Kulkarni has opened up another exciting dimension in time and frequency metrology. [5] After removing a linear frequency drift and removing the effects of solar system dynamics with a set of fitting parameters along with the effects of relativity, the stability of the residuals of this first pulsar are comparable to that of TAI in the long-term. [6,7] The possibility exists of using this area of metrology to detect gravitational radiation as well as long-term systematics in international timing.

Many have wondered about the value of millisecond pulsars for timing. Here, an understanding of the differences among fundamental measurement noise, time stability, time predictability, frequency stability and frequency accuracy are important. Considering these factors, in light of trends in atomic clock performance, leads to the logical conclusion that pulsars may never be a practical source for terrestrial timing.

In figure 1 the stabilities of various clocks, current and anticipated, are plotted along with that of PSR 1937+21 -- the most stable of the millisecond pulsars to date. At present, the Arecibo Observatory's 300 m spherical antenna is being upgraded. The improvements should yield a measurement noise of 100 ns white PM. The value of 2 X mod.$\sigma_y(\tau)$ gives the optimum estimate of how well the frequency can be measured for the integration time τ. Hence, one can see from figure 1 that it would take about 100 years of integration time to reach 1 X 10^{-18}. The data length would need to be at least three times this long. Technological extrapolation would indicate that 1 X 10^{-18} accuracy might be achieved in frequency standards within a couple of decades.

Outside of a major breakthrough in reducing the measurement noise and the fluctuations in path delay across the interstellar medium, atomic clocks should always be some ahead of millisecond pulsars in time predictability -- even in the long term. A measurement noise breakthrough is not inconceivable [A. Lepek, et. al.], and certainly, millisecond pulsar timing work should be encouraged. The fruits from scientific investigation in this area will undoubtedly continue to be impressive. For the timing community, we may be able to sort out some of the long-term instabilities currently observed between the international timing centers using millisecond pulsar timing as a comparison tool.

TRENDS IN TIME AND FREQUENCY TRANSFER TECHNIQUES

When atomic clocks were first invented, the most convenient means of time transfer was the HF broadcast stations with frequency stabilities of 1 X 10^{-7} to 1 X 10^{-8} and with about 1 ms time accuracy. Later came the LF and VLF techniques -- providing significant advancements. Until the GPS common-view technique was developed, the frequency transfer noise was often greater than the instabilities in the best atomic clocks being compared -- an undesirable attribute of a measurement or comparison system.

Unless we continue to make significant advancements in this area, we will find ourselves again in this undesirable state.

Figure 2 shows a plot of some of the currently most useful means of transferring time and frequency between sites remote to each other. Figure 3 shows the frequency stability of a variety of means of transferring same. Again, we see dramatic improvements in the techniques when viewed as a time sequence according to when the various techniques became available. Again, for white noise PM, 2 X mod.$\sigma_y(\tau)$ gives the optimum estimate of how well the frequency can be measured via that particular technique and for τ equal the data length.

In this regard the Two-Way Satellite Time and Frequency Transfer (TWSTFT) technique holds significant promise. Some recent equipment-test data reported by NIST at the CCIR Study Group 7A meetings showed a standard deviation of the time residuals over seven days of 60 ps, and the short term stability plot was consistent with that shown in figure 3 [ref.]. If white noise PM characterizes the instabilities ($\tau^{-1/2}$ on a $\sigma_x(\tau)$ plot), then the confidence on the frequency transfer accuracy of this measurement or comparison system is given by the following:

$$S_d = \frac{\sqrt{12}\ s}{\tau_o\ n^{3/2}} \qquad (1)$$

where s is the standard deviation, τ_o is the data sampling period, and n is the number of measurements. For example, if the white PM characterization persisted for a few hours, then using the TWSTFT technique would yield uncertainties of less than 1 X 10^{-15}. For the NIST seven-day equipment test run, if the residuals were white PM, then equation 1 would yield an uncertainty of 4 X 10^{-18}. A visual inspection of the residuals showed slight diurnal fluctuations, but basically the data were extremely well behaved, and appeared in character to be close to white PM.

As an interesting aside, equation 1 gives exactly the same result as 2 X mod.$\sigma_y(\tau)$ for white-noise PM. The advantage of a mod.$\sigma_y(\tau)$ diagram over equation 1 is that it gives a measure of frequency instability for most all of the other (both white and non-white noise) spectral density processes encountered in time and frequency metrology.

The GPS common-view technique is the main means of communicating time and frequency from timing centers in the world to the BIPM for the generation of TAI and UTC. For the GPS common-view technique, comparison stabilities of about 10^{-15} have been achieved for integration times of about a week (see figure 3). GPS is becoming more commonly used in the telecommunications industry (telecom) as an independent source of time and frequency. The power industries are also using GPS more and more for timing their systems.

NEW IDEAS FOR USING GPS FOR TIME AND FREQUENCY

Three new ideas toward better utilizing GPS are suggested here. Two of them could be very useful for telecom and the other for more accurate frequency transfer than the common-view technique.

Idea one, for more accurate frequency transfer, utilizes some techniques employed in geodesy for differential positioning. Locking on the carrier phase cuts down the multipath by about a factor of 1,500. With multichannel receivers being available and using common-view double differencing techniques, typical in geodesy, two sites may well be able to maintain a weighted common-carrier phase between them. If measured ionospheric delays were used along with nominally compensated tropospheric corrections then 1 X 10^{-15} may be reachable with integration times of the order of a day. Continuous, not once a day, measurements would be necessary to achieve this performance.

Idea two revolves around optimum filtering of the GPS SA spectrum. Range data were taken for the middle two weeks of Dec.'92 and a $\sigma_x(\tau)$ analysis of it was calculated. For those familiar with the time variance, TVAR, it is equal to $(\sigma_x(\tau))^2$. [9] Four different satellites were studied, and the time-domain spectrum from each was essentially identical. Figure 4 is a plot of the spectrum for three of these satellites' SA signals. Range data were taken with τ_o = 100 s. A nominal $\tau^{1/2}$ behavior or f^{-2} spectrum was a reasonable model up to τ values of a few hundred seconds. For τ > 400 s the behavior was $\tau^{-1/2}$ or f^0. For the longest τ values the behavior was like τ^0 or f^{-1}. Whether this 1/f noise is from the SA modulation or from the receiver is not known. If it is SA, that can be integrated out across several satellites. If it is the receiver, that can be fixed.

The noise and frequency offset of a HP model 5071A were simulated and an optimum filter was designed to process the GPS SA data with the spectrum shown in figure 4. Figure 5 shows the results. After a turn-on transient, the steady state mean error for the last week of the simulation was 0.6 ns and the standard deviation was 2 ns. In practice, the errors due to multipath + ionosphere + troposphere and other GPS systematics will be the limiting error factors. There are some other important underlying assumptions in this simulated result. These need to be verified -- such as: is the spectrum in figure 4 the same for all the satellites and does it change with time? What correlations exist across the different satellites' SA signals? And what is the source of the about 20 ns 1/f behavior in the long-term? Experimental testing and additional data are needed to answer these questions.

If the spectrum in figure 4 is valid, then we can also design an optimum filter for a rubidium gas-cell clock. The theoretical estimate for the standard deviation of the steady state residuals is less than 10 ns and the frequency stability and accuracy should be better than 1×10^{-12}. This needs to be simulated and tested experimentally. This particular implementation could be very useful for both the power industry as well as the telecommunications community.

Idea three utilizes the excellent short-term stability of quartz-crystal oscillators. This idea may be particularly appealing to telecom and the power industry because of the need for high reliability. In this idea, each node in a network would have a very inexpensive GPS receiver, quartz-crystal oscillator and data communication equipment. This network ensemble, along with reasonable data communication, an ensemble algorithm and a unique modified common-view approach can be shown in theory to yield network synchronization consistency to 10 ns and about 1×10^{-12} relative stability. The network could also be synchronized to UTC to better than 100 ns.

By 1995 a new set of INMARSAT geo-stationary satellites will be configured around the earth which could provide both high accuracy time and frequency transfer as well as be an excellent source for UTC. [10] The receivers for the same would use the GPS gold-code and should be quite inexpensive. How this system evolves will be very important for a large number of potential users.

APPLICATIONS, DISCUSSION AND CONCLUSIONS

As outlined above, we expect time and frequency sources to continue to improve in their quality, availability and reliability. Such improvements could benefit a large segment of society.

The BIPM by its charter is the ultimate reference for the SI units. Because of this great responsibility, they are, wisely, very conservative and cautious when they perform calibrations and comparisons. Such calibrations and comparisons often take a significant amount of time in order to obtain the minimum uncertainty attainable, and great care is taken in experimental procedure. Such is the case with the SI second, and the history of their performance is impeccable as they have delivered the SI second via TAI and UTC. The SI second delivered reflects the best accuracies available from the primary cesium-beam frequency standards in the world.

UTC, different from TAI, is both a source for the SI second and for time-of-day. For all of the SI units -- including the second -- it is not important that they be available at the current epoch. Ex-post-facto calibration values are the *modus*-operandi, which is perfectly acceptable. It is the opinion of the authors that the second dimension of UTC -- to be the ultimate source for time-of-day -- will best be served if it is available at the current epoch.

As we look to the future, we may anticipate a current-epoch UTC. The current UTC is not used for active dating of events because of the two month delay in its availability. A precedent has been set by several of the timing centers where both the SI second and time-of-day are generated at the current epoch with very good time and frequency accuracy. Each of the UTC(k)s which steer to UTC, do this by predicting forward, approximately two months, to the current epoch. Currently, for example, the USNO, with about twenty model 5071As and ten H-masers, is able to predict UTC to the current epoch with errors of the order of 20 ns. These errors, of course, result from instabilities in both the USNO clocks as well as in TAI and UTC.

In principle, the BIPM has access to more and better data than any individual timing center in the world. In practice, the limitations are the time and frequency transfer instabilities, the labor, expense and logistics of having current access to all of the world clocks. Many of these problems may be solved in the future, and the BIPM may be able to do better than any individual timing center in predicting TAI and UTC to the current epoch. Setting aside the logical problems and the measurement instabilities, the optimum-combined time prediction capability of all the clocks in the world is now about 3 ns for two months.

The practical problems to solve in order to realize an optimum UTC at the current epoch are: first, a reduction in the time required to generate TAI and UTC; second, a reduction in the instabilities in time and frequency transfer systems; third, an automatic, reliable, robust and accurate way of obtaining world data at the current epoch (perhaps once a day would be sufficient, if the expense to do so were reasonable); fourth, the algorithms necessary so that the BIPM staff could readily see that the generated time and the predicted estimate were consistent with the highest level of performance; and fifth, a method of making this BIPM optimum predicted value of UTC readily and conveniently accessible to all who need it and desire to use it.

Progress is being made by the BIPM and the supporting timing centers in all five of these practical-problem areas. The BIPM is actively studying how they can reduce the time necessary to compute TAI and UTC. The BIPM and several other timing centers are actively pursuing the reduction of the instabilities in time and frequency transfer. These efforts rest principally with GPS used in the common-view mode with a goal of 1 ns time transfer accuracy. As reported above, the TWSTFT technique may achieve better than 1 ns accuracy and better than 1×10^{-15} frequency transfer accuracy.

Regarding the third practical problem area, there is a strong trend to use INTERNET so that automatic data access can occur in either direction. Such a system as INTERNET has the advantages that the number of communication channels needs only increase as the number of timing centers, and data exchange can be fully automated. Whereas, if every timing center communicates with every other timing center to compare their data, there are $n(n-1)/2$ such combinations, which would be very inefficient, as it is now between several timing centers. This INTERNET concept has been suggested by Dr. J. Levine of NIST. A big part of the delay for the BIPM in calculating TAI and UTC is in getting the data from the timing centers. Automating the data transfer would help reduce the heavy work load of the BIPM. It would also have the significant advantage that TAI and UTC could be calculated much more responsively. TAI and UTC are accessible via the Circular-T bulletins. Fortunately, these can now be obtained electronically. Improvement in this area will clearly require cooperation on the part of the timing centers. In fact, the BIPM staff now spends a large amount of time dealing with problems in the data being sent by the timing centers.

As discussed earlier, significant progress has been made in the fourth practical-problem area of taking advantage of new algorithm developments. Further R&D effort is still needed -- especially in-lieu-of current new clocks and time and frequency transfer systems. We want to use the best algorithms which take advantage of the current best clocks and time and frequency transfer systems. A thorough and careful study -- including experimental results -- will be required before the BIPM staff will consider adopting a new approach.

As discussed in the new ideas section above, there appear to be some practical, robust and very efficient ways of using GPS as a time and frequency source as well as a transfer tool. If the BIPM had a steered physical clock output locked to their best estimate of predicted UTC, then this clock could be used as a measurement tool. This predicted UTC would behave with the long-term stability and accuracy characteristics of TAI. In the future, for example, it may be immediately available via INTERNET with an uncertainty of a few nanoseconds and using, for example, GPS, GLONASS and in the future INMARSAT as transfer standards.

As we look at how well we might do, given current time and frequency technology and circumstances, the optimum rms time prediction error for UTC to current epoch is about 3 ns. The measurement noise using an HP model 5071A optimally locked to GPS as outlined in the new idea section above is of the same order. The one day prediction error of the model 5071A is also of the same order. If the BIPM had a model 5071A phase-locked to the optimum estimate of UTC for the current epoch, and used its output to report an optimum filtered estimate of GPS time, then anyone similarly locked to GPS in another part of the world could know UTC to better than about 10 ns -- taking the square root of the sum of the squares of the contributing elements. This assumes ready access to the appropriate measurement data uploaded daily by the BIPM. This level of automaticity and data flow can certainly be anticipated in the future.

Instead of locking to GPS time, one could choose to lock to the GPS broadcast value of UTC(USNO). The latter time scale is not only more stable than GPS time, but is more tightly locked to UTC. Given current performance, and without having to call the BIPM, a remote receiver and model 5071A locked as suggested would be locked to UTC to about 20 ns. Using GPS time and/or the GPS broadcast value of UTC(USNO) as transfer standards yields more nearly direct access to UTC with potentially higher accuracy.

This future UTC, as it is now, would be optimized from a world data base and would continue to maintain high reliability and performance. In this scenario, however, it would be readily available at the current epoch. TAI, in this future scenario, continues to provide the ultimate SI reference for the second. A block diagram of this future prospect is shown in figure 6. As we anticipate the future, the importance of the BIPM's work in generating UTC will increase, and its utility will cover a wider spectrum of users throughout the world. This utility will arise, in large measure, because of it being more and more readily and conveniently available at the current epoch.

In telecommunications, support from the time and frequency community could significantly enhance the accuracy and the rate of information flow. UTC could become their real-time ultra-accurate reference on a global basis. Now most of the telecom servers generate their own timing, yet different servers have to communicate with each other, which often creates information flow conflicts. In practice, to avoid some of the conflicts in data flow, the servers tend to use the biggest server as a reference. Having an externally-unbiased and readily-available reference -- always somewhat better than their needs-- would mitigate many of the current problems. This external support from UTC as a real-time timing reference could significantly improve data flow efficiency. Such support may also reduce costs, since the servers would not have to use as much of their resources to supply precise timing.

It is interesting to note that as telecom moves toward sub-100 ns requirements for SDH, or other systems, if they use the reciprocity of optical fibers, for example, to synchronize different nodes, the relativistic Sagnac effect becomes significant. In other words, if the assumption of simple reciprocity is used to synchronize site A to site B on the rotating earth, and then B to site C, site C will not be synchronous with A. Most GPS receivers have compensation for this effect built into their software. In the future, the same will need to be done for optical transmissions. Going around the globe the effect can be of the order of 200 ns.

If UTC were to become the active primary timing reference for the telecommunication industry, they have inherent in the new systems being developed the ability to be a supplier of very accurate time and frequency information. A recent report by the Japanese delegation at a CCIR Study Group 7A meeting indicated compensated round-trip optical-fiber delay stabilities of $\sigma_x(\tau)$ of the order of one nanosecond. The range of τ values went up to several hours, and the path length for the signal was 2,000 km. In some countries, the telecommunications industries are already secondary suppliers of that countries UTC(k), adjusted for time zones.

For the power industry there are two important uses of accurate synchronization for their power networks. First is fault location. They would like to have a reliable and readily available 1 μs across their networks to time tag events so that both fault locations can be accurately determined as well as detecting failure modes. In the case of the latter, accurate timing could provide failure warnings and thus avoid them in some instances. Second, efficient and safe power flow depends on line loading and the power-factor. Accurate timing across a power system would allow them to better utilize the power generation systems and allow the lines that carry the power to be more efficient and safe. Thus, new generation systems and power lines would not have to be constructed as soon upon increase in future power demands.

The benefits from GPS to navigation, surveying, geodesy, position determination, space exploration and earth quake monitoring are already very obvious. The heart of GPS is the excellent time and frequency sources associated therewith. Since GPS also plays a very important role in the international timing community, and will continue to be very important, it is good to know that a Memorandum of Agreement has been signed between the Department of Defense and the Department of Transportation (effective 8 Jan.'93 until the year 2005) for the CIVIL USE OF THE GLOBAL POSITIONING SYSTEM.

Technology has evolved in proportion to the precision and accuracy of measurements coming from metrology. Time and frequency have provided the most precise and accurate measurements known to man. Other areas of metrology have leveraged off of the ease of making precise time and frequency measurements. This is true in R&D as well as in specific technological applications. As we continue to see the rapid improvements in accurate time and frequency, we can anticipate many significant direct as well as spin-off benefits.

ACKNOWLEDGEMENTS

The authors are in debt to a large number of people for some very meaningful conversations as well as for the data they have given us to analyze. Conversations with colleagues at the BIPM, INPL, PTB, NIST, USNO and VNIIFTRI have been extremely valuable. We wish to express our appreciation for their excellent work and support. The data support from NIST, PTB, USNO and VNIIFTRI is also greatly appreciated. In addition, Mr. Wayne Dewey was most helpful in supplying the GPS SA data for spectrum analysis.

REFERENCES

1. D.J. Wineland, W.M. Itano, and J.C. Bergquist, Absorption Spectroscopy at the Limit: Detection of a Single Atom, Optics Letters, 12, 389-391, 1987.

2. L. Cutler, et. al. " " Proceedings of the 1992 Frequency Control Symposium

3. Data from NIST and USNO

4. N.B. Koshelyaevsky, et.al., "Russian National Time Scale Long Term Stability Estimation using GLONASS Time Link," Proceedings of 16-18 March 1993 European Frequency and Time Forum.

5. D.C. Backer, et. al., Nature, Vol. 300, pp. 615-618 (1982)

6. L. A. Rawley, et.al., Millisecond Pulsar PSR 1937+21: A Highly Stable Clock, Science, Vol. 238, pp. 761-765, 6 Nov.'87.

7. D.W. Allan, Millisecond Pulsar Rivals Best Atomic Clock Stability, Proceedings of the 41st Annual Symposium on Frequency Control, Philadelphia, PA, 2-11, 1987.

8. Third International Algorithm Symposium, Torino, Italy; proceedings available from Time and Frequency Div., NIST

9. D. W. Allan, et. al., A Frequency-domain View of Time-domain Characterization of Clocks and Time and Frequency Distribution Systems, 45th Annual Symposium on Frequency Control, pp. 667-678, 1991.

10. A. Brown, et. al., Precise Time Dissemination Using the INMARSAT Geostationary Overlay (IGO), Proceedings of 1993 International Frequency Control Symposium, 2-4 June 1993.

FIGURE CAPTIONS

1. The frequency stability, mod.$\sigma_y(\tau)$, of the current best frequency standards along with the anticipated new ones. The frequency stability of the residuals of the millisecond pulsar, PSR 1937+21, is also plotted along with the anticipated stability limit, which is anticipated with the current upgrade of the Arecibo Observatory's 300 m spherical antenna.

2. A square root of TVAR, $\sigma_x(\tau)$, time stability plot as a function of averaging time for some of the most currently useful time transfer techniques. The short-term for the TWSTFT technique is characterized by white-noise PM at a level of about 300 ps $\tau^{-1/2}$. The diurnal and annual variations in Loran-C are evident. The bottom of the GPS common-view plot is the best obtained between two sites remote to each other, and is approximately characterized by 1 ns $\tau^{-1/2}$ with τ_o = 1 sidereal day. The top of this plot is a result of analyzing the data going all around the world over several years. One can see evidence of an annual variation -- probably due to errors in modeling the ionosphere.

3. The fractional frequency stability of most of the commonly available frequency transfer techniques are plotted. Where needed to remove the ambiguity occurring when $\sigma_y(\tau)$ behaves as τ^{-1}, the stability was plotted using mod.$\sigma_y(\tau)$. Those plotted using mod.$\sigma_y(\tau)$ are indicated using an asterisk "*" at the end of the descriptive label.

4. A time stability plot, $\sigma_x(\tau)$, of the GPS SA spectrum for three different satellites. Four were studied, and all four demonstrated essentially the same spectral density. The short term seems reasonably well modeled by a $\tau^{1/2}$ or f^{-2} out to τ values of a few hundred seconds. For times longer than about 400 s the characteristic model seems to be about 84 ns $\tau^{-1/2}$ down to a limit of about 20 ns. Whether this long-term flicker PM is due to SA or the GPS receiver is not known.

5. On the assumption that the SA spectrum in figure 5 is characteristic of all the satellites, some data were simulated both for the SA spectrum and for one of the new HP model 5071A cesium beam clocks. The simulated model for the model 5071A included a frequency offset of 1 X 10^{-13} along with a measured level of its typical white noise FM, $\sigma_y(\tau)$ = 8 X 10^{-12} $\tau^{-1/2}$. An optimum filter was designed to generate a real-time output from the model 5071A locked to truth defined as either GPS time or UTC(USNO). The truth was found to better than 1 ns after the turn on transient, and the standard deviation over the last week of the simulation was 2 ns.

6. A block diagram of the proposed TAI and UTC which may have the potential to greatly improve timing support to the telecommunication industries.

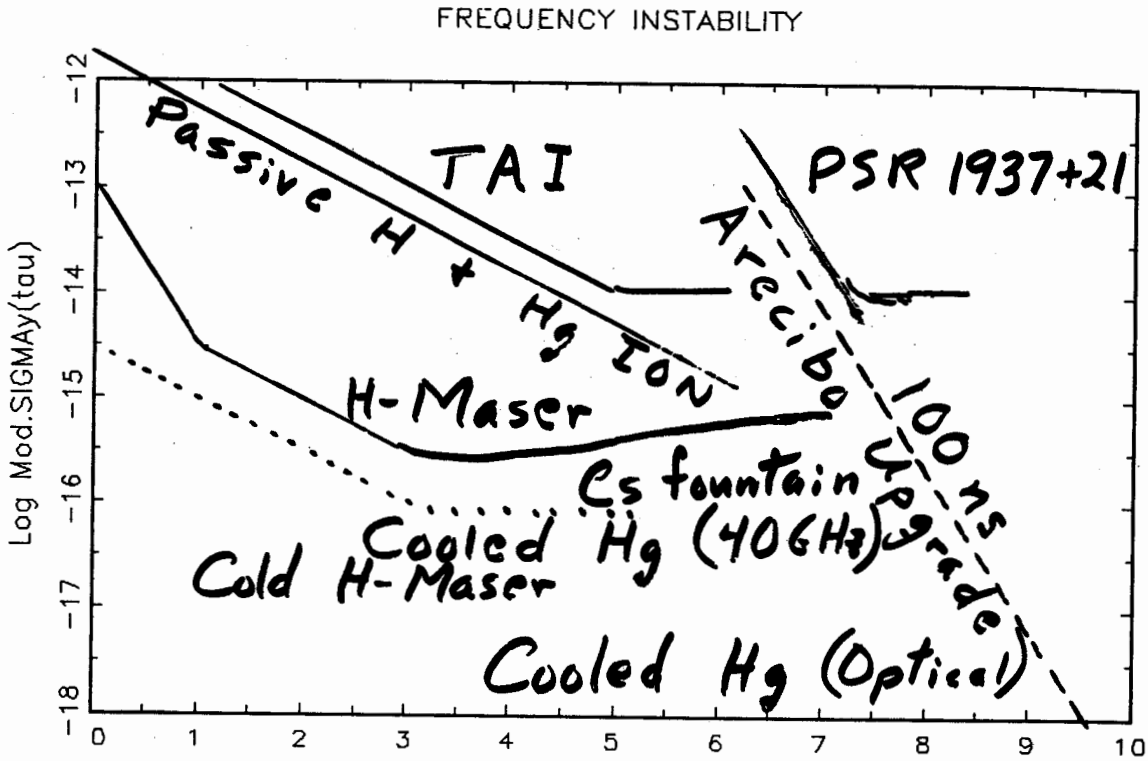


Figure 1

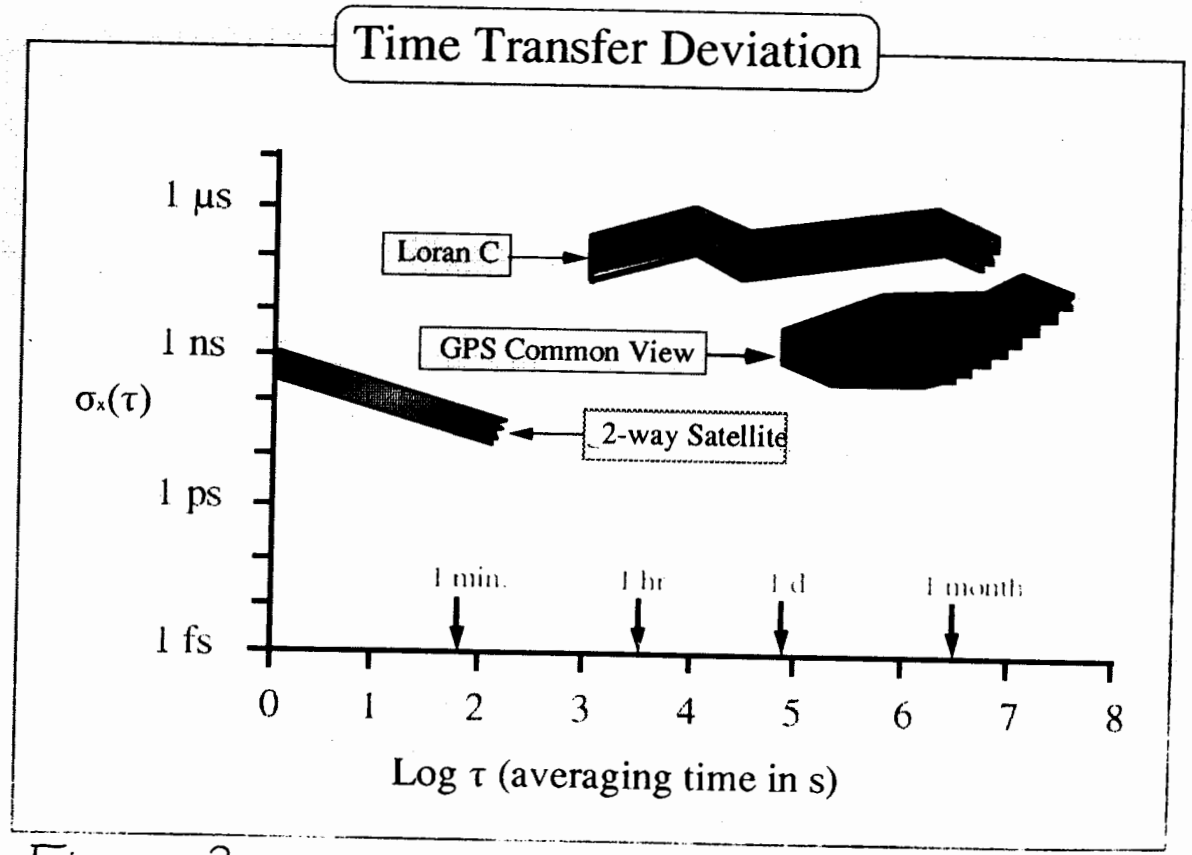


Figure 2

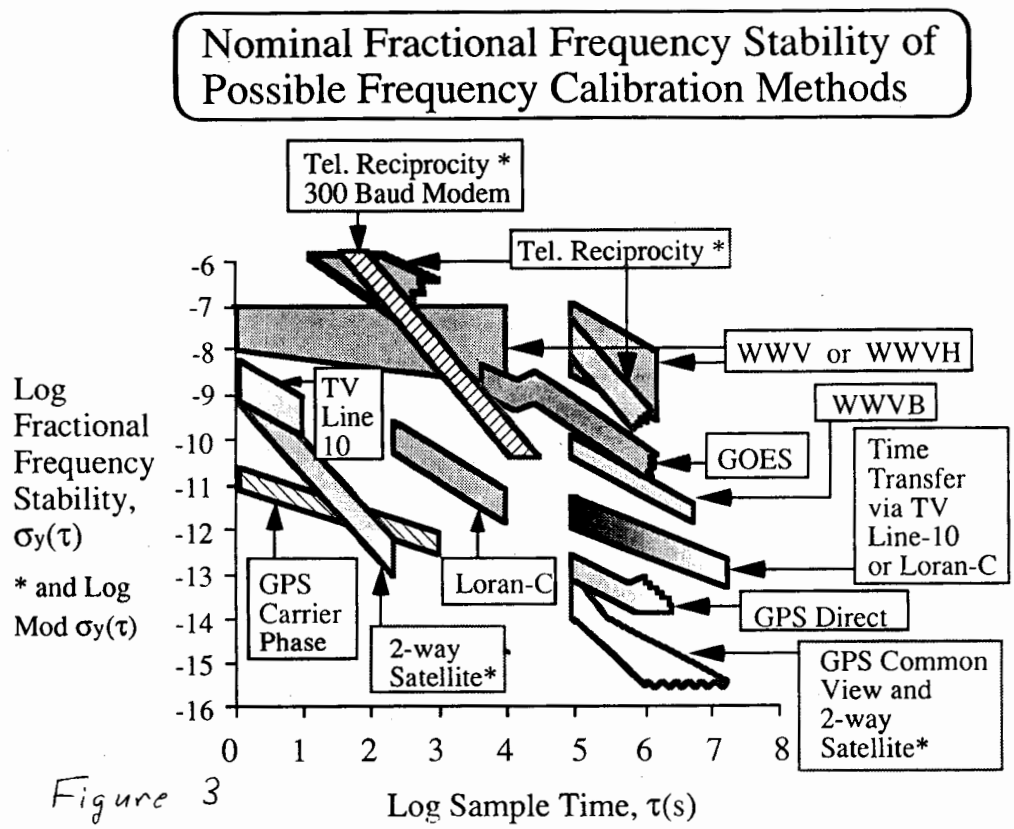


Figure 3

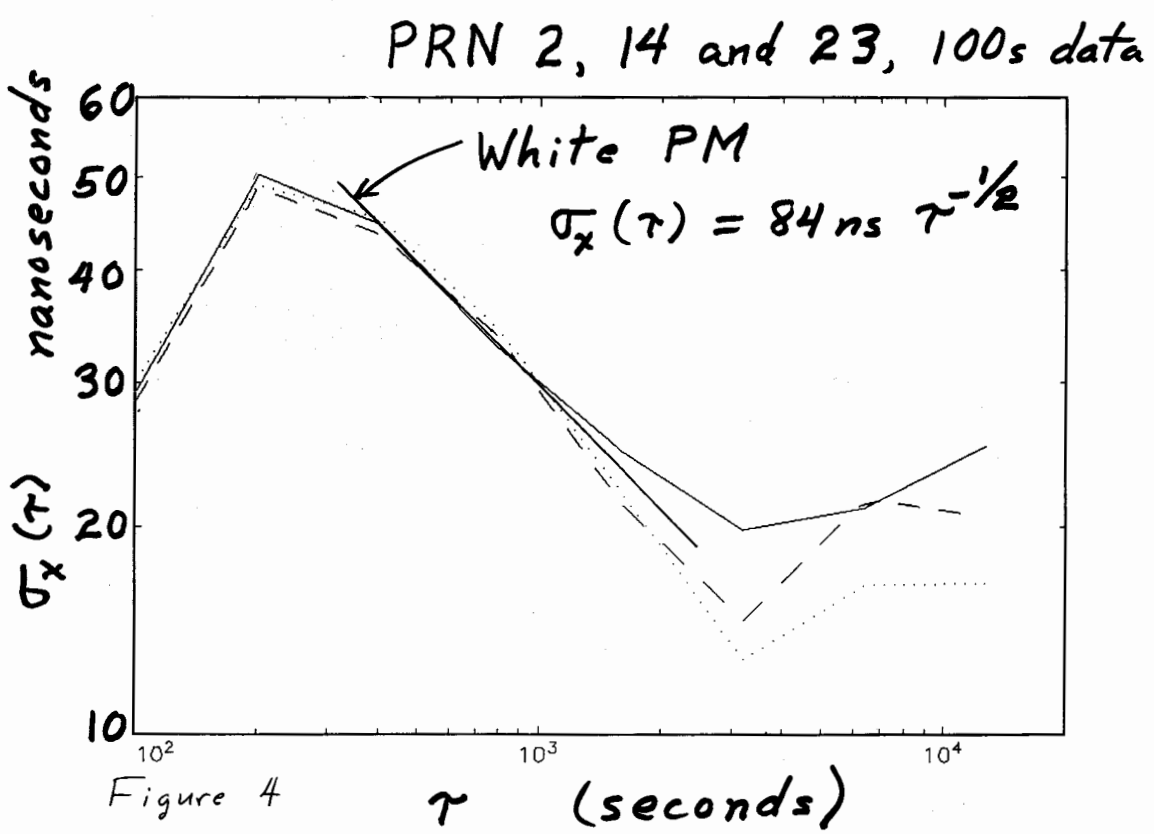


Figure 4

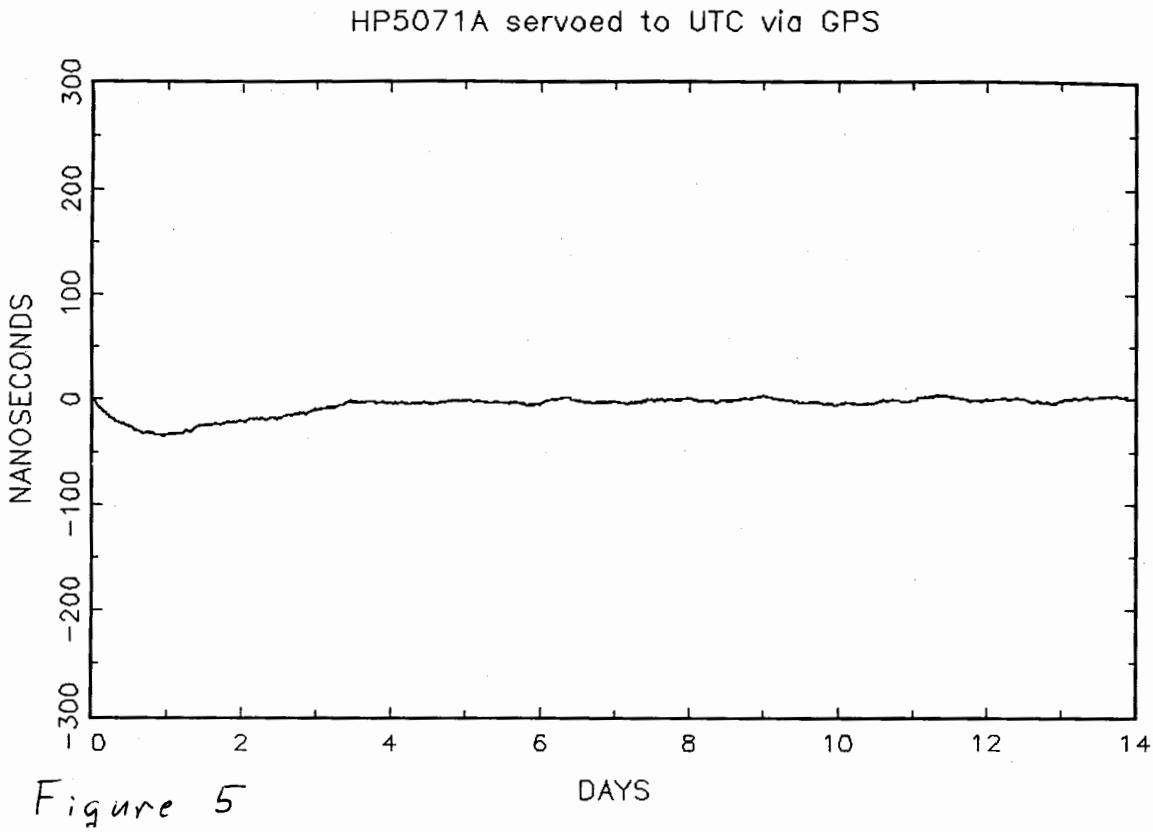


Figure 5

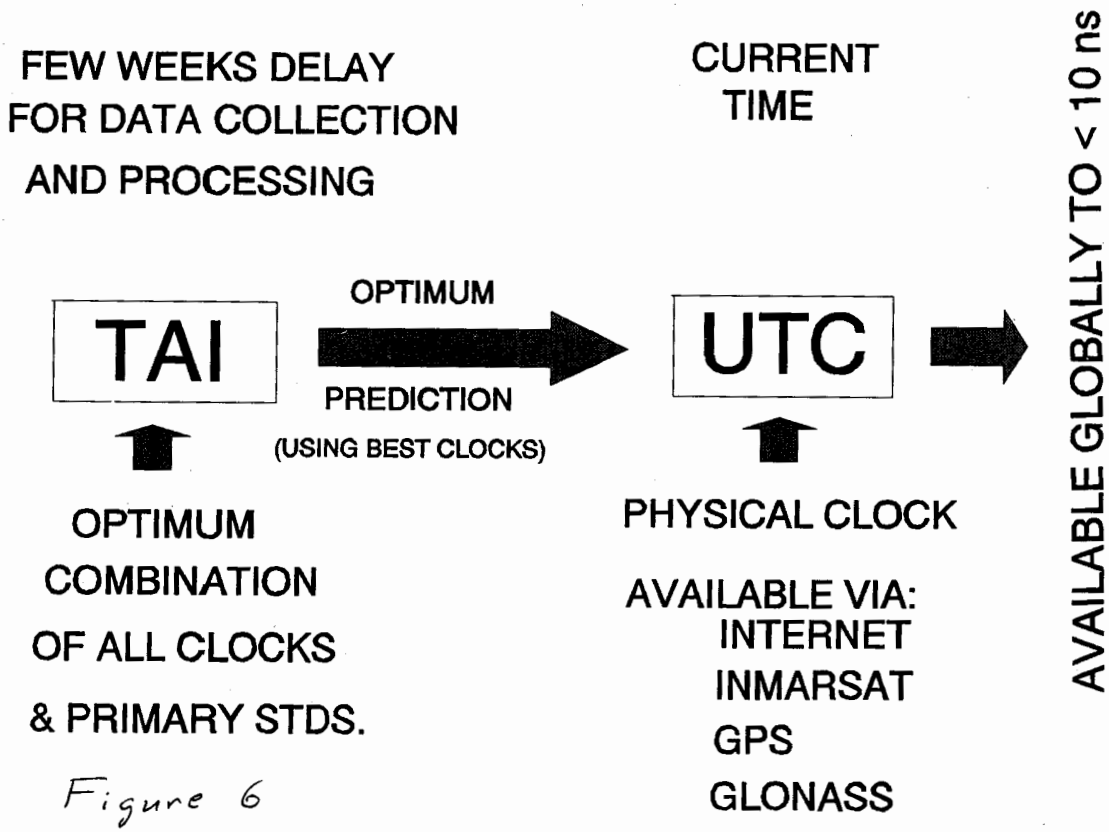


Figure 6

POSTER 1 - FREQUENCY CONTROL COMPONENTS AND MATERIALS

Chairman: W. Zingg

PIEZOELECTRIC MATERIAL CONSTANTS FOR SURFACE WAVE RESEARCH

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

U.S. Army Research Laboratory
AMSRL-EP-MD, Fort Monmouth, NJ 07703-5601

1. ABSTRACT

In this paper, we bring together and extract the various reported values of the material constants of interest for SAW device applications. We identify for the user community those sets of constants from which SAW design calculations can reliably be made, as well as those sets of constants requiring further refinement. The relevant material constants include: mass density ρ, elastic stiffness c_{ij}, piezoelectric stress e_{ij}, dielectric permittivity ϵ_{ii}, and the thermal expansion coefficients α_{ii}. The various data sets reported here can be used to predict theoretically the SAW velocity, piezoelectric coupling, temperature coefficients of time delay, and power flow angles for various SAW orientations. Data sets are given for crystals having symmetry classes mm2, 32, 3m, 4mm, and 6mm in both bulk and thin film formats. Compound families include: borates, titanates, niobates, tantalates, phosphates, silicates, PVDF-, KTP-, and wurtzite types. Only data published after 1978 are included, since the reference literature (Landolt-Börnstein) amply covers those years prior to 1978.

Keywords: Piezoelectric Material Contants, SAW Materials.

2. INTRODUCTION

The efficient generation of Rayleigh surface waves on piezoelectric substrates through interdigital transduction has allowed surface acoustic wave devices to impact a wide variety of frequency control systems and designs requiring high-frequency signal processing.

The physical properties of the SAW substrate and its crystal orientation determine the SAW efficiency, (piezoelectric coupling), SAW velocity, and temperature performance (temperature coefficient of delay). The limitations of a particular SAW device design will ultimately be determined by the values of the elastic, piezoelectric and dielectric tensor elements and their temperature coefficients.

These measurable properties affect the electrical (operating) characteristics of the SAW device. For example, there will be a coupling between the electric field applied through the metal fingers (interdigital transducer) and the mechanical strain of the substrate. This coupling is maximized at the frequency $f=v_s/p$, where p is the periodicity of the interdigital pattern. The velocity, v_s, is related to the material properties through the relation $v_s=(c/\rho)^{1/2}$, where ρ is the mass density of the material and c is the effective elastic constant of the SAW. The effective elastic constant c depends on the elastic and piezoelectric tensors and their temperature coefficients.

3. MEASURED PROPERTIES

The tables designated "Survey of SAW Materials and Their Measured Constants (published since 1978)" and "Measured Properties of SAW Materials," are submitted in a form similar to those found in the Landolt-Börnstein custom [1]. A more comprehensive survey of piezoelectric materials for SAW applications may be found in [2]. (The tables were placed at the end of the paper because of size and space limitations.)

4. CONCLUSION

This collection of piezoelectric material constants for surface wave research has attempted to identify those sets of constants from which SAW device design calculations can reliably be made. Two promising materials that have been recently examined are langasite ($La_3Ga_5SiO_{14}$) [7-10] and dilithium tetraborate ($Li_2B_4O_7$) [15-16] Large good-quality single crystals of these new materials are already available [10,21]. We expect SAW devices research on these materials to thrive over the next few years.

5. REFERENCES

[1] Landolt-Börnstein, Numerical Data and Functional Relationships in Science and Technology, Vol. III/1, 1966, Vol. III/2, 1969, and Vol. III/11, 1979, Springer-Verlag, Berlin, Heidelberg, New York.

[2] J. G. Gualtieri, J. A. Kosinski, and A. Ballato, Proc. 1992 Ultrasonics Conference, Tucson, AZ, in press.

[3] David K. Chu, John D. Bierlein, and Robert G. Hunsperger, Proc. 1992 IEEE Frequency Control Symposium, pp. 732-743.

[4] R. S. Wagers, Proc. 1980 Ultrasonics Symposium, pp. 464-469.

[5] D. S. Bailey, J. C. Andle, D. L. Lee, W. Soluch, J. F. Vetelino, and B. H. T. Chai, Proc. 1983 Ultrasonics Symposium, pp. 335-340.

[6] P. Krempl, J. Stadler, W. Wallnöfer, W. Ellmeyer, and R. Selic, Proc. 5th European Frequency and Time Forum, pp. 143-147, 1991.

[7] A. A. Kaminskii, I. M. Silvestrova, S. E. Sarkisov, and G. A. Denisenko, Phys. Stat. Sol., vol. 80, pp. 607-620, 1983.

[8] A. B. Ilyaev, B. S. Umarov, L. A. Shabanova, and M. F. Dubovik, Physica Status Solidi A, Vol. 98, No. 2, pp. K109-K114, Dec. 1986.

[9] I. M. Sil'vestrova, Yu. V. Pisarevskii, P. A. Senyushchenkov, and A. I. Krupnyi, Soviet Physics-Solid State, Vol. 28, No. 9, pp. 1613-1614, Sep. 1986.

[10] S. A. Ssakharov, I. M. Larionov, and A. V. Medvedev, Proc. 1992 IEEE Frequency Control Symposium, pp. 713-723.

[11] W. Soluch, R. Ksiezopolski, J. F. Vetelino, and A. Jhunjhunwala, Proc. 1979 Ultrasonics Symposium, pp. 602-605.

[12] R. V. Damle, J. Phys. D: Appl. Phys., Vol. 25, pp. 1091-1095.

[13] I. A. Dan'kov, E. F. Tokarev, G. S. Kudryashov, and K. G. Belobaev, Inorganic Materials (Plenum Publishing Corp.), Vol. 19, No. 7, pp. 1049-1054, July 1983.

[14] G. Kovacs, M. Anhorn, H. E. Engan, G. Visintini, and C. C. W. Ruppel, Proc. 1990 Ultrasonics Symposium, pp. 435-438.

[15] N. M. Shorrocks, R. W. Whatmore, F. W. Ainger, and I. M. Young, Proc. 1981 Ultrasonics Symposium, pp. 337-340.

[16] L. Bohaty', S. Haussühl, and J. Liebertz, Cryst. Res. Technol. Vol. 24, No. 11, pp. 1159-1163, Nov. 1989.

[17] N. M. Shorrocks, R. W. Whatmore, and F. W. Ainger, Proc. 1982 Ultrasonics Symposium, pp. 359-362.

[18] S. T. Liu and J. D. Zook, Ferroelectrics, Vol. 7, pp. 171-173, 1974.

[19] G. Carlotti, G. Socino, A. Petri, and E. Verona, Proc. 1987 Ultrasonics Symposium, pp. 295-299.

[20] K. Tsubouchi, K. Sugai, and N. Mikoshiba, Proc. 1981 Ultrasonics Symposium, pp. 375-380.

[21] J. G. Gualtieri, J. A. Kosinski, W. D. Wilber, Y. Lu, S. T. Lin, M. Murray, and W. Ruderman, Proc. 1992 IEEE Frequency Control Symposium, pp. 724-731.

SURVEY OF SAW MATERIALS AND THEIR MEASURED CONSTANTS (published since 1978)

Substance	Crystal Class	Chemical Formula	Mass Density (10^3kg/m^3)	c_{ij}	e_{ij}	ϵ_{ii}	α_{ii}	References	Remarks / Other Properties
Potassium Titanyl Phosphate	mm2	$KTiOPO_4$		+			+	[3]	piezocoupling coefficients
Rubidium Titanyl Phosphate	mm2	$RbTiOPO_4$		+			+	[3]	piezocoupling coefficients
Potassium Titanyl Arsenate	mm2	$KTiOAsO_4$		+			+	[3]	piezocoupling coefficients
Rubidium Titanyl Arsenate	mm2	$RbTiOAsO_4$		+			+	[3]	piezocoupling coefficients
Poly(vinyl-idene) Fluoride	mm2	$(C_2H_2F_2)_n$	1.807	+	+			[4]	PVF_2

+ Indicates complete constant set.

SURVEY OF SAW MATERIALS AND THEIR MEASURED CONSTANTS (published since 1978)

Substance	Crystal Class	Chemical Formula	Mass Density (10^3kg/m^3)	c_{ij}	e_{ij}	ϵ_{ii}	α_{ii}	References	Remarks / Other Properties
Berlinite	32	$AlPO_4$	2.57*	+	+	+		[5]	
Gallium Orthophosphate	32	$GaPO_4$	3.570	+	+	+	+	[6]	
Langasite	32	$La_3Ga_5SiO_{14}$	5.754	+	+	+	+	[7,8] [9,10]	
Sillenite	32	$Bi_{12}SiO_{20}$	9.2*	/	/	/		[11]	
Lithium Niobate	3m	$LiNbO_3$	4.628	+	+	+		[12,13] [14]	
Lithium Tantalate	3m	$LiTaO_3$	7.454	+	+	+		[14]	
Dilithium Tetraborate	4mm	$Li_2B_4O_7$	2.439	+	+	+	+	[15,16]	
Strontium Barium Niobate	4mm	$Sr_{.5}Ba_{.5}Nb_2O_6$	5.4	+	+	+	+	[17,18]	SAW coupling, TCD, and v_s
Aluminum Nitride (film)	6mm	AlN	3.512	+	+	+		[19]	
Zinc Oxide	6mm	ZnO	5.665	/	+			[20]	
Zinc Oxide (film)	6mm	ZnO	5.720	/	+			[20]	

* See Ref.[1], Vol. III/11, Ch. 3, p. 287.
\+ Indicates complete constant set. / Indicates partial constant set.

MEASURED PROPERTIES OF SAW MATERIALS

Substance	Formula	Symmetry	References
Potassium Titanyl Phosphate	$KTiOPO_4$	mm2	[3] KTP
Rubidium Titanyl Phosphate	$RbTiOPO_4$	mm2	[3] RTP
Potassium Titanyl Arsenate	$KTiOAsO_4$	mm2	[3] KTA
Rubidium Titanyl Arsenate	$RbTiOAsO_4$	mm2	[3] RTA
Poly (vinylidene) Fluoride	$(C_2H_2F_2)_n$	mm2	[4] PVF_2

Property (units)		KTP [3]	RTP [3]	KTA [3]	RTA [3]	PVF_2 [4]
Density (10^3kg/m^3)	ρ					1.807
Elastic Stiffness (GPa)	c_{11}	170				30
	c_{12}	33.33				7.8
	c_{13}	40				7.8
	c_{23}	35				7.8
	c_{33}	149.5 (-220)	140 (-118)	120	120 (-190)	14.5
	c_{44}	59.19 (-134)	56 (-166)	---	44 (-154)	3.7
	c_{55}	54.54 (-143)	75 (-122)	---	46 (-100)	3.7
	c_{66}	43.13 (-100)				3.7
Thermal Expansion*	α_{11}	(6.8)	(10.8)	(7.0)	(15.1)	
	α_{22}	(9.6)	(13.3)	(11.8)	(8.5)	
	α_{33}	(-1.3)	(-5.9)	(0.3)	(-7.0)	
Piezoelectric coupling (%)	k_t	34	31	34	38	
	k_{24}	16	7	---	5	
	k_{15}	14	8	---	9	

Note: Numbers in parentheses are temperature coefficients X 10^{-6}/K.
* For the temperature range (20° - 100°C).

MEASURED PROPERTIES OF SAW MATERIALS

Substance	Formula	Symmetry	References
Sillenite	$Bi_{12}SiO_{20}$	32	[11] BSO
Berlinite	$AlPO_4$	32	[5]
Gallium Orthophosphate	$GaPO_4$	32	[6]

Property (units)		BSO [11]	$AlPO_4$ [5]	$GaPO_4$ [6]
Density (10^3g/m^3)	ρ	9.2*	2.57*	3.570
Elastic Stiffness (Constant E) (GPa)	c_{11}	129	69.31(-150)	67
	c_{12}	29	10.51(-165)	21
	c_{13}		13.49(-722)	25
	c_{14}		-12.99(+90)	-4
	c_{33}		88.62(-253)	102
	c_{44}	24.6	43.02(-168)	37
	c_{66}		29.40(+119)	23
Piezoelectric Stress Constant (C/m^2)	e_{11}		0.14	0.20
	e_{14}	1.13	0.02	0.03
Dielectric Constant (ϵ_o)	ϵ^S_{11}		4.7 (0.71/K)	6.3
	ϵ^S_{33}	41	4.8 (1.10/K)	6.7
Thermal Expansion	α_{11}			(5.3)
	α_{33}			(1.2)

Note: Numbers in parentheses are temperature coefficients X 10^{-6}/K.
* See Ref. 1, Vol. III/11, Ch. 3, p. 287.

MEASURED PROPERTIES OF SAW MATERIALS

Substance	Formula	Symmetry	References
Langasite	$La_3Ga_5SiO_{14}$	**32**	**[7,8,9,10]**

Property (units)		**[7]**	**[8]**	**[9]**	**[10]**
Density (10^3kg/m^3)	ρ	5.754		5.743	5.751
Elastic Stiffness (Constant E) (GPa)	c_{11}	190.9	188.9(-58.7)	190.2(-47)	190.2(-47)
	c_{12}	106.3	104.6(-218)	106.3(-100)	106.3(-100)
	c_{13}	104.2	96.8(-101)	91.9(-130)	91.9(-130)
	c_{14}	15.2	14.3(-154)	14.7(-370)	14.7(-370)
	c_{33}	261.9	262.2(-135)	262.1(-94)	
	c_{44}	52.4	53.9(-79.7)	53.82(-30)	53.4(-30)
	c_{66}	43.2	42.2(139)	42.0(+36)	42.35(+36)
Piezo-Electric Stress Constant (C/m^2)	e_{11}	-0.45	-0.44(-2400)		-0.45
	e_{14}	0.077	0.07(-5000)		0.077
Di-electric Constant (ϵ_o)	ϵ^s_{11}	18.87			19.0(153)
	ϵ^s_{33}	49.32	49.10(-618)	-49(-760)$^+$	49.2(-760)
Thermal Expansion	α_{11}		(5.84)	(5.11)	(5.15)
	α_{33}		(4.03)	(3.61)	(3.65)

Note: Numbers in parentheses are temperature coefficients X 10^{-6}/K.
$^+$ See graph in [9].

MEASURED PROPERTIES OF SAW MATERIALS

Substance	Formula	Symmetry	References
Lithium Niobate	$LiNbO_3$	**3m**	**[12,13,14]LNO**
Lithium Tantalate	$LiTaO_3$	**3m**	**[14] LTO**

Property (units)		**LNO [12]**	**LNO [13]**	**LNO [14]**	**LTO [14]**
Density (10^3kg/m^3)	ρ	4.640	4.628	4.628	7.454
Elastic Stiffness (constant E) (GPa)	c_{11}	203.1	199.0	198.39	232.8
	c_{12}	53.0	53.8	54.72	46.5
	c_{13}	74.2	71.4	65.13	83.6
	c_{14}	8.5	7.85	7.88	-10.5
	c_{33}	241.3	237.2	227.90	275.9
	c_{44}	64.6	60.1	59.65	94.9
Piezo-electric Stress Constant (C/m^2)	e_{15}		3.61	3.69	2.64
	e_{22}		2.40	2.42	1.86
	e_{31}		0.28	0.30	-0.22
	e_{33}		1.59	1.77	1.71
Dielectric Constant (ϵ_o)	ϵ^s_{11}		44.5	45.6	40.9
	ϵ^s_{33}		26.1	26.3	42.3

MEASURED PROPERTIES OF SAW MATERIALS

Substance	Formula	Symmetry	References
Dilithium Tetraborate	$Li_2B_4O_7$	**4mm**	**[15,16]**
Strontium Barium Niobate	$Sr_xBa_{1-x}Nb_2O_6$	**4mm**	**[17]**

Property (units)		**[15]**	**[16]**	**[17]**
Density (10^3kg/m^3)	ρ	2.451	2.433	5.4[18]
Elastic Stiffness (Constant E) (GPa)	c_{11}	126.7(-125)	135.27(-80)	210.1(-120)
	c_{12}	0.5(14000)	1.09(20000)	65.7(-850)
	c_{13}	30 (350)	31.86(550)	35.5(680)
	c_{33}	53.9(354)	54.80(400)	116.6(-710)
	c_{44}	55.0(-23)	57.39(13)	66.3(74)
	c_{66}	46.0(-480)	47.38(-480)	68.9(19)
Piezoelectric Stress Constant (C/m^2)	e_{15}	0.36(-1300)	0.36	5.19(-150)
	e_{31}	0.19(-1300)	0.19	-1.89(3800)
	e_{33}	0.89(-1000)	0.87	9.81(4500)
Dielectric Permittivity (ϵ_o)	ϵ_{11}	8.97(-110)	7.81$^+$	370 (77)
	ϵ_{33}	8.15(-33)	10.25$^+$	282 (10200)
Thermal Expansion Coefficient	α_{11}		(5.1)	(11.3)
	α_{33}		(-2.3)	(-0.45)

Note: Numbers in parentheses are temperature coefficients X 10^{-6}/K.
$^+$ At constant T.

MEASURED PROPERTIES OF SAW MATERIALS

Substance	Formula	Symmetry	References
Zinc Oxide	**ZnO**	**6mm**	**[19]**
Aluminum Nitride	**AlN**	**6mm**	**[20]**

Property (units)		**ZnO [19]**	**ZnO (film) [19]**	**AlN (film) [20]**
Density (10^3kg/m^3)	ρ	5.665	5.720	3.512
Elastic Stiffness (GPa)	c_{11}	209	157	345
	c_{12}	----	----	125
	c_{13}	104.6	83	120
	c_{33}	210.6	208	395
	c_{44}	42.3	38	118
	c_{66}	44.25	34	110
Piezoelectric Stress Constant (C/m^2)	e_{15}	-0.48	-0.45	-0.48
	e_{31}	-0.57	-0.51	-0.58
	e_{33}	1.32	1.22	1.55
Dielectric Permittivity (ϵ_o)	ϵ_{11}			9.0
	ϵ_{33}			11.

CHARACTERIZATION AND TEMPERATURE SENSITIVITY OF SHEAR HORIZONTAL ACOUSTIC WAVES IN QUARTZ

M. Hoummady, E. Bigler, G. Martin

Laboratoire de Physique et Métrologie des Oscillateurs du CNRS
associé à l'Université de Franche-Comté-Besançon,
32 av. de l'Observatoire - 25044 Besançon Cedex - France

Abstract

Shear Horizontal Bulk Acoustic Waves (SH-BAW) and Shear Horizontal Acoustic Plate Modes (SH-APM) are generated and detected in quartz plates by Interdigital Transducers (IDT's) deposited on the same surface of the plate.

A model is presented to predict the resonant frequencies of SH-APM and their sensitivity to thermal effects is studied theoretically and experimentally. The thermal sensitivity depends, in addition to the quartz cut, on the mode order and on the ratio of thickness to wavelength. There is an additional degree of freedom to minimize the thermal sensitivity. New possibilities of temperature compensation have been found.

Introduction

Interdigital transducers (IDT's) deposited on piezoelectric materials are mainly used to generate Rayleigh Surface Acoustic Waves (SAW), but the same IDT are effective for launching various Bulk Acoustic Waves (BAW's).

The excitation of one particular elastic wave depends on the crystal symmetries and on the position of the IDT with respect to the crystallographic axis. The IDT position defines the electric field components and the crystal symmetries impose the non-zero piezoelectric constants. Both, lead to the stress-tensor which defines the polarization of the generated waves.

In a quartz plate, when an IDT is deposited with transducer fingers parallel to the X-axis, only the SH-BAW is generated. This wave is guided inside the plate by several reflections on top and on bottom sides of the plate (Fig. 1) [1]. If the plate thickness is weak, interference between incident and reflected waves, setup Shear Horizontal Acoustic Plate Modes (SH-APM). Those SH-APM can propagate only at discrete phase velocities and so at discrete resonant frequencies.

Devices using this propagation type have gained considerable importance in recent years, due to the fact that the use of the free surface (opposite side), where SH-waves are reflected, is interesting for measurements in liquid or gas phases [2-4]. Generation and detection of acoustic waves (unlike conventional Rayleigh wave sensors) are not perturbed by any kind of aggressive environment.

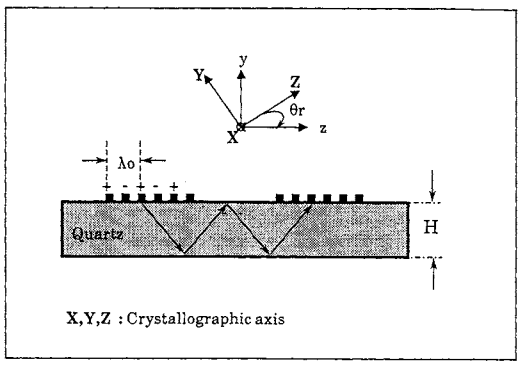


Fig. 1 : Quartz plate and IDT geometry to generate SH-BAW guided waves

In those applications, temperature behavior has not been examined. However, temperature effects appear as parasitic and can be an important limiting factor of the ultimate sensitivity of SH-APM sensors.

In this paper, a model to predict the resonant frequencies of SH-APM is presented. The temperature effects are studied theoretically and confirmed by experiments where an SH-APM device is used in an oscillator loop.

Modelling

In singly rotated quartz plates, SH-BAW are generated when IDT's are deposited perpendicularly to X-axis of quartz crystal (Fig. 2). Those waves are launched at an angle θ to the surface, when the phase matching condition of Eq. (1) is satisfied :

$$\lambda_\theta = \lambda_o \cos\theta \qquad (1)$$

where : λ_θ is the wavelength in θ direction
λ_0 is the period of IDT
θ is the emission angle.

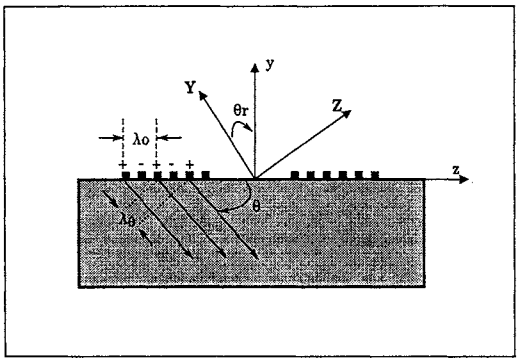


Fig. 2 : SH-BAW radiation satisfying the phase matching condition

The phase matching condition can be written as a function of the phase velocity v_p and of the operating frequency f

$$\lambda_o f \cos\theta = v_p(\theta) \tag{2}$$

Eq. (2), valid in the case of generated SH-BAW, shows the dependence of emission angle on the operating frequency.

SH-BAW can travel to the output IDT by several reflections. But, if the plate thickness is small, interference between incident and reflected waves will set up SH-APM only for a discrete number of frequencies.

In order to calculate the resonant frequencies of modes F_M, we use the boundary conditions on both sides (stress free)

$$T_{ij} = 0 \quad on \quad Y = \pm \frac{h}{2} \tag{3}$$

The values of frequencies for which the determinant of boundary conditions is equal to zero correspond to resonant frequencies F_M. Those frequencies are usually calculated numerically but in quartz crystal an analytical solution exists [5]. This calculation is similar to Auld's method [6] but taking into account the plate anisotropy. F_M can be written as a function depending both on crystal constants (stiffness C_{IJ}, mass density ρ) and on the ratio of thickness H to wavelength (period of IDT) λ_0

$$F_M = \frac{C_a}{\lambda_o}\sqrt{\frac{C_{66}}{\rho}}\sqrt{1+\left(\frac{M\lambda_o}{2\,C_a H}\right)^2} \tag{4}$$

where : M is the mode order (number of nodes)
C_a an anisotropic constant defined as

$$C_a^2 = \frac{C_{55}}{C_{66}} \cdot \left(\frac{C_{56}}{C_{66}}\right)^2 \tag{5}$$

The thermal sensitivity of SH-APM is derived from eq. (4) following an analysis similar to the classical model used for thermal sensitivity of Rayleigh waves SAW [7].

The variations of F_M versus the temperature T are obtained by changing as a function of T the value of all parameters in eq. (4), except the order of the mode M, supposed to be constant. This hypothesis (M constant) corresponds experimentally to the case of an oscillator using an APM device in the main oscillating loop. When the temperature changes, the frequency of the oscillator will slightly change but if temeprature variation are quasi-static, the oscillator will not jump from one mode to another.

The temperature dependence of all parameters in eq. (4) is modelled using :
- the temperature coefficients of the elastic constants of quartz C_{IJ}
- the coefficients of thermal expansion (variations of linear dimensions λ_0, H and mass density ρ).

Temperature coefficients of material constants for quartz are taken from Bechman [8] and defined as

$$C_{IJ}(T) = C_{IJ}(T_o) \times \left[1 + a^{(1)}_{C_{IJ}}(T\text{-}T_o) + a^{(2)}_{C_{IJ}}(T\text{-}T_o)^2 + a^{(3)}_{C_{IJ}}(T\text{-}T_o)^3\right] \tag{6}$$

where T_o is the reference temperature (25 deg. C) and $a_{C_{ij}}{}^{(1)}$, $a_{C_{ij}}{}^{(2)}$ and $a_{C_{ij}}{}^{(3)}$ are the temperature coefficients of the elastic constants, expressed in the material axes of the crystal.

The temperature dependence of the dimensions λo and H is computed for a singly rotated quartz plate as

$$\lambda_o(T) = \lambda_o(T_o) \times \left\{1 + \sin\theta_r\left[a_y^{(1)}(T\text{-}T_o) + a_y^{(2)}(T\text{-}T_o)^2 + a_y^{(3)}(T\text{-}T_o)^3\right] + \cos\theta_r\left[a_z^{(1)}(T\text{-}T_o) + a_z^{(2)}(T\text{-}T_o)^2 + a_z^{(3)}(T\text{-}T_o)^3\right]\right\} \tag{7}$$

$$H(T) = H(T_o) \times \left\{1 + \cos\theta_r\left[a_y^{(1)}(T\text{-}T_o) + a_y^{(2)}(T\text{-}T_o)^2 + a_y^{(3)}(T\text{-}T_o)^3\right] + \sin\theta_r\left[a_z^{(1)}(T\text{-}T_o) + a_z^{(2)}(T\text{-}T_o)^2 + a_z^{(3)}(T\text{-}T_o)^3\right]\right\} \tag{8}$$

where θ_r is the cut angle of the plate, and $a_y{}^{(i)}$ and $a_z{}^{(i)}$ are the coefficients of thermal expansion respectively along Y-axis and Z-axis.

The temperature dependence of the mass density is given by Jaillet [9].

Results

The model is used to predict the temperature dependence of the frequency for a given SH-APM. Examples of results relating to a BT-cut of quartz are shown on Fig. 3. the predicted temperature dependence is close to a parabolic shape.

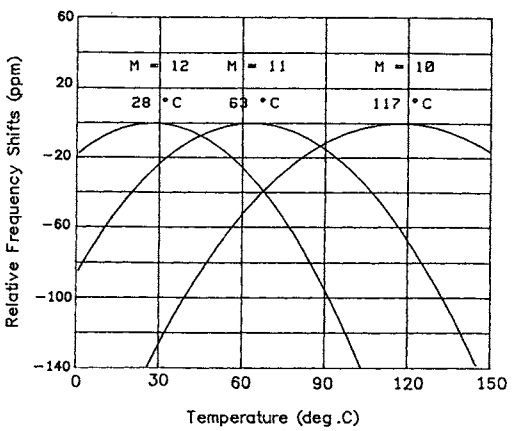


Fig. 3 : Predicted temperature sensitivity of SH-APM
-Simulation parameters : BT-Cut, H = 483µm, λo = 39.8µm

The inversion point of the parabola depends on the order of the mode and on the geometrical parameters of the device.

For a given quartz plate, a certain degree of freedom is allowed to choose the position of the inversion point by simply changing the order of operating mode. To choose a proper mode, the operating frequency has to be selected with a suitable passband filter inserted in the oscillator loop.

Experimental tests have confirmed the existence of parabolic temperature dependence of the resonnant frequency as shown on Fig. 4 and Fig. 5.

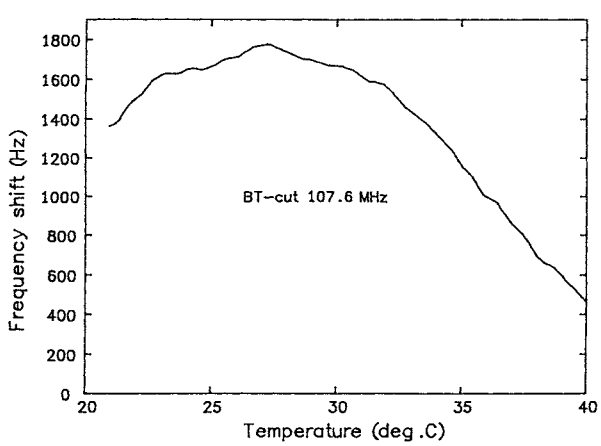


Fig. 4 : Experimental frequency-vs temperature curve for the 12th mode
-Experimental parameters : BT-Cut, H ≈ 500µm, λo ≈ 40µm

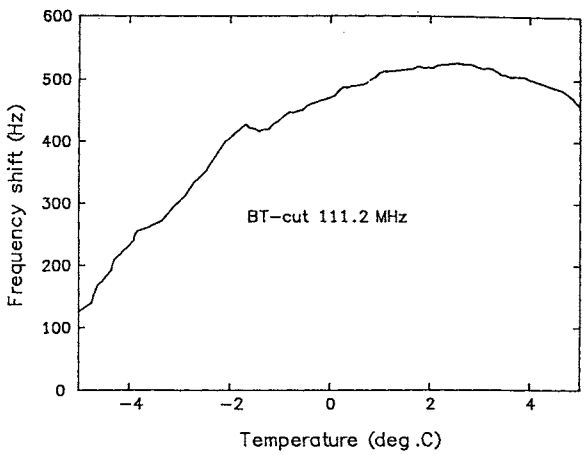


Fig. 5 : Experimental frequency-vs temperature curve for the 13th mode
-Experimental parameters : BT-Cut, H ≈ 500µm, λo ≈ 40µm

In the present state of the model, there is a difference between the measured values of the 2nd order temperature coefficient and the values predicted by the model of polynomial regression. It is known for Rayleigh surface waves that the polynomial regression does not give a good accuracy for 2nd order coefficients, but will give with a good approximation the proper cuts exhibiting zero first order temperature coefficient. The situation is probably the same for plate modes. the important point is in fact the existence of temperature compensated modes for plate waves.

It is interesting to investigate all the possibilities of 1st order temperature compensation of plate modes on singly rotated quartz cuts. The value of the 1st order coefficient can be extracted from the predicted frequency temperature curve by polynomial regression. Figure 6 shows the loci of zero first order temperature coefficient as a function of the cut angle of the plate and as a function of the order of the mode.

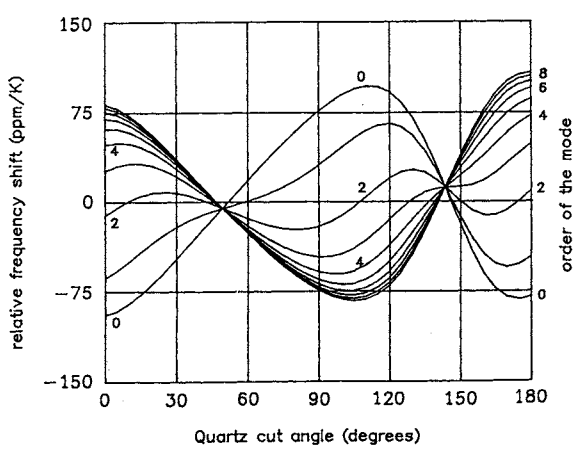


Fig. 6 : First-order temperature coefficient of frequency for various SH-APM in rotated Y-Cut quartz plate

It is shown that for many single rotated quartz cuts, possibilities of temperature compensation exist if a proper operating mode is selected. There is with such devices an additional degree of freedom (quartz cut/order of the mode) allowing to satisfy another criterion like : optimizing the electromechanical coupling constant or optimizing the sensitivity of the device for sensor applications.

Conclusion

It has been shown that SH-APM devices can exhibit a low sensitivity to 1st order temperature effects.

Intrinsic temperature compensation of SH-APM devices will greatly reduce thermal drifts and will increase the stability and sensitivity for sensor applications. Moreover, a proper choice of the operating mode by an external circuit allows to select a proper range of temperature without changing the crsytal cut, which is the case for conventional Rayleigh wave devices.

References

[1] M. Hoummady, B.A. Auld, D. Hauden, E. Bigler, "Excitation et détection, par des transducteurs interdigités, des ondes transversales horizontales guidées dans une plaque de quartz", 1er Congrès Français d'Acoustique, Lyon, avril 1990.

[2] F. Josse and Z. Shana, "Acousto-ionic interaction of SH surface with ionic solutions", IEEE Trans. on Ultrasonics, Ferroelectrics and Frequency Control, vol. UFFC-38, n° 3, pp. 297-304 (1991).

[3] M. Hoummady, F. Bastien, "Acoustic wave viscometer", Review of Scientific Instruments, vol. 62, n° 8 (1991).

[4] D. Rebière, J. Pistré, M. Hoummady, D. Hauden, P. Cunin, R. Planade, "Sensitivity comparison between gas sensors using SAW on shear horizontal plate mode oscillators", Sensors and Actuators B, vol. B6, n° 1-3 (1992).

[5] M. Hoummady, "Ondes élastiques transverses horizontales émises par des transducteurs interdigités déposés sur des plaques minces de quartz. Applications aux capteurs de viscosité et aux détecteurs de gaz", Thèse, Université de Franche-Comté-Besançon, 5 mars 1991.

[6] B.A. Auld, "Acoustic fields and waves in solids", vol. II, chapter 10H, pp. 136-138, Krieger publishing Co, Malabar, Florida, 1990.

[7] J.J. Campbell and W.R. Jones, "A method for estimating optimal crystal cuts and propagation direction for excitation of piezoelectric surface waves", IEEE Trans. Sonics Ultrason., vol. SU-15, pp. 209-217 (1968).

[8] R. Bechmann, A.D. Ballato, T.J. Lukaszek, "Higher order temperature coefficients of the elastic stiffnesses and compliances of alpha-quartz", Proc. IRE, p. 1812, (1962).

[9] G. Jaillet, "Ondes de surface sur substrats piézoélectriques. Etudes et réalisation d'un capteur de température", Thèse, Université Pierre et Marie Curie, Paris VI, 23 octobre 1981.

THERMAL TRANSIENT FREQUENCY BEHAVIOR OF SAW DEVICES WITH A LASER BEAM AS A THERMAL PROBE

Nobunori Oura, Naimu Kuramochi, Masahiro Takahashi,
Isao Kawasaki, and Tatsuya Nakase

Tokyo Insitute of Technology
4259 Nagatsuta, Midoriku, Yokohama 227, Japan

Abstract/Resume

From the viewpoint how the transient heat transfer from the surrounding to the SAW device has influence on the frequency temperature characteristics, the thermal transient frequency behavior of SAW devices is studied with a laser beam with 22.5 mW as a thermal probe.

A result shows that the frequency behavior to the beam is different from the delayline-type device to the resonator-type. The resonator-type device behaves in different ways, that is, the irradiation on the electrode causes the frequency excursion opposite to that of the non-electroded area. The equi-freqeucny maps of the frequency excursion of the devices like a contour map are useful for the thermal design of the devices and the application of the device to a laser powermeter.

Keywords:Thermal transient, SAW device, laser beam probe, frequency excursion, thermal transfer

1.INTRODUCTION

SAW oscillators are used widely for direct generation of frequencies higher than those in VHF band[Ref.1]. Their frequency stabilities have been approaching those of bulk crystal oscillators for common use.

Various kinds of angle of cut have been found for improving the frequency-temperature characteristics of SAW devices[Ref.2]. The frequency-temperature characteristics of the devices are divided into two categories, the static frequency-temperature characteristics and the dynamic ones[Ref.3-5], which take place when the devices are subject to an abrupt change in an environmental temperature. In this paper, from the viewpoint how the transient heat transfer from the environment to the device has influence on the device, the thermal transient frequency behavior of SAW devices is studied with a He-Ne laser[Ref.6,7] with 22.5 mW as a thermal probe for the improvement of the frequency stability with respect to transient temperature changes.

The devices used for the study are of the delayline-type devices at 157MHz, and the resonator-type ones at 145MHz made of quartz crystal and lithium niobate.

The measurements were made on the frequency changes with time in the oscillating frequency of the SAW device connected to an oscillator circuit with respect to the irradiation position of the beam on the device, beam power, and the environmental temperature as parameters, when the device was irradiated with the beam.

2.STATIC FREQUENCY-TEMPERATURE CHARACTERISTICS OF THE DEVICES

The frequency-temperature characteristics of the quartz crystal SAW devices and lithium niobate device are shown in Fig.1 and 2, respectively, which are plotted with data measured twice and drawn by least-square approximation.

Although the devices shown in Fig.1 have the same angle of cut, there are a 10°C discrepancy in ZTC(zero-temperature coefficient temperature) between the characteristics. This may result from the device type.

The characteristics of the lithium niobate device can be approximated with a straight line with a frequency-temperature coefficient of -73 ppm/°C. Note that the magnitude on the ordinate is 150 times larger than that in Fig.1.

3.FREQUENCY EXCURSION DUE TO THE BEAM IRRADIATION ON THE PRINCIPAL SURFACE OF THE DEVICE

A schematic measurement system of the frequency changes is shown in Fig.3. A beam from the He-Ne laser with 22.5 mW is converged to 0.08 mm in diameter on the plate surface of the device under test which is connected to an oscillator circuit. When the beam impinges on any position of the principal surface of the device glued on a base, the oscillating frequency changes abruptly. The transient frequency behavior is measured with a frequency counter provided with the clock frequency from a Rb frequency standard and recorded analogously on a recorder chart and stored in a computer at the same time. The irradiation position on the device was altered by moving the device fixed on an X-Y stage driven with a motor, while the beam is fixed. A transient frequency excursion due to the consecutive beam irradiation is shown in Fig.4, in which an amount of frequency change ΔF is defined as follows;

$$\Delta F = Fa - Fb \qquad (1)$$

where Fa is the maximum frequency attainable after 2s of the irradiation, Fb is the frequency in the absence of the beam on the device.

An equi-frequency change map like a contour map was drawn by connecting the points indicating the same magnitude of ΔF on the plan view of the device with line segments. The map for the quartz crystal delayline-type device is shown with a line separation of 5Hz in Fig.5. In Fig.6 is shown the

map with a separation of 10Hz for a quartz crystal resonator-type device. Figure 7 shows the map with a separation of 50Hz for a lithium niobate resonator-type device. The scrutiny of the maps gives information on the phenomena.

For quartz crystal devices the magnitude of ΔF for the non-electroded area is lager than that for the electroded area. The maximum of ΔF in the electroded area is 0.2 ppm for the delay-line-type device, while 0.5 ppm for the resonator-type device. The beam irradiation on any portion of the delayline-type device causes frequency increase. In contrast, the irradiation on the resonator-type device gives rise to increase or decrease in frequency in corresponding to the portion being irradiated with the beam.

For the lithium niobate resonator-type device, the irradiation of the beam on any portion of the device causes decrease in frequency. The maximum frequency excursion is 7.6 ppm, one order of magnitude greater than that of the quartz devices. Specifically the irradiation on the electroded area causes the frequency excursion inverse to that of the quartz crystal device.

The frequency variation in an oscillator with the lithium niobate device is ±10Hz measured with a 2s gate-time of the frequency counter, while the variation of the quartz device is ±1Hz.

The heat transfer to the propagation path of SAW yields thermal stress in the body of the device and has great influence on the frequency behavior of the device. The electroded area of the resonator-type device is smaller than that of the delayline-type device. Therefore the irradiation on the electroded area of the resonator-type device causes frequency excursion greater than that of the delayline-type device, since the thermal stress is concentrated in a small area which contributes to the traveling velocity of the SAW, provided that the thermal diffusion area a unit time is the same with respect to both the devices.

This fact suggests that the support, lead electrode, plate thickness of the device should be designed from the viewpoint of thermal transfer, too.

4.THE MEASUREMENT OF THE FREQUENCY EXCURSIONS DUE TO THE IRRADIATION REFERRED TO THE ENVIRONMENTAL TEMPERATURE AS A PARAMETER

The frequency excursion due to the irradiation by taking the environmental temperature as a parameter was measured to investigate the influence of the environmental temperature on the frequency excursions. The oscillator circuit and the SAW device were placed in an oven with a transparent glass window through which the beam irradiated the device surface. The frequency excursions were measured with respect to the crystal resonator-type device and delayline-type device at every time that the environmental temperature, that is, the oven temperature was changed. Figure 8 shows the frequency excursion vs. environmental temperature characteristics of two kinds of the devices. The characteristics are right descending linear lines , and independent of the increase in the environmental temperature, although the frequency in the static frequency vs. temperature characteristics increases with temperature between 10 and 20°C. These characteristics result in specific heat of quartz which is proportional to the increase in temperature between 5 and 10°C[Ref.8].

5.THE FREQUENCY EXCURSIONS DUE TO THE BEAM POWER WITH RESPECT TO THE ENVIRONMENTAL TEMPERATURE AS A PARAMETER

The measurements were made with the same system used for the measurements in the section 4. When the beam impinges the electrodes of the device through the glass window by keeping the oven temperature constant, the frequency excursion was measured every time with a step increase of the beam power with respect to the environmental temperature as a parameter. The frequency excursion vs. beam power characteristics of the crystal resonator-type and lithium niobate resonator-type device are shown in Figs. 9 and 10, respectively. The magnitudes of the frequency excursions are almost proportional to the beam power as for both the devices, while the slopes are inverse to each other. The slopes of the lines shown in Fig.9 become steeper with increase in the environmental temperature. The beam irradiation on the non-electroded area causes the frequency excursion with the beam power as well as the irradiation on the electrodes.

6.CONCLUSIONS

The thermal frequency sensitivity of the SAW devices to the laser beam irradiation on any portion of the principal surface of the devices was studied by measuring the transient frequency behavior of the device just after the irradiation. After the dependence of the frequency excursion on the beam power, irradiation position on the device, material of the device plate, environmental temperature was investigated, the equi-frequency excursion maps similar to a contour map and the frequency excursion vs. beam power characteristics were drawn. The maps are useful for designing the devices immune to an abrupt temperature change such as the configuration of the thin film lead on the device plate, and for applications of the thermal frequency behavior, for example, to a laser powermeter.

REFERENCES

[1]K.Shibayama ed.: "Surface Acoustic Wave Engieering, " Corona Press, 1983(in Japanese).

[2]Thomas E. Parker and Gary K.Montress: "Precision Surface Acoustic Wave(SAW) Oscillators," IEEE Trans. Ultrasonics, Ferroelectrics, and Frequency Control,Vol.35,No.3,pp.342-362,1988.

[3]D.Hauden, M.Michel,and J.-J. Gagnepain: "High Order Temperature Coefficients of Quartz SAW Oscillators, " Proc. 28th ASFC, pp.77-81,1974.

[4]D.Hauden, M. Michel, G.Bardeche, and J.-J. Gagnepain: "Temperature Effects on Quartz-Crystal Surface-Wave Oscillators, " Applied Physics Letters, Vol.31, No.5,pp.315-317,1977.

[5]G.Theobald and D.Hauden:"Theoretical Analysis of Dynamic Thermal Effects in SAW Devices," Proc. 38th ASFC, pp.294-299, 1984.

[6]Nobunori Oura,Naimu Kuramochi,Junichi Nakamura, and Takashi Ogawa: "Thermal Frequency Behavior in Contoured Quartz Crystal Plates Induced by Direct Irradiation of Laser Beam," Proc. 36th ASFC, pp.133-139,1982.

[7]R.J.Williamson: " Non-Destructive Means for Determinating Mode Shape in Thickness-Shear Resonators,"Proc. 44th ASFC, pp.424-436, 1990.

[8]R.B.Sosman: "The Properties of Silica,"Monograph Series, No.37,The Chemical Catalog Co., 1927.

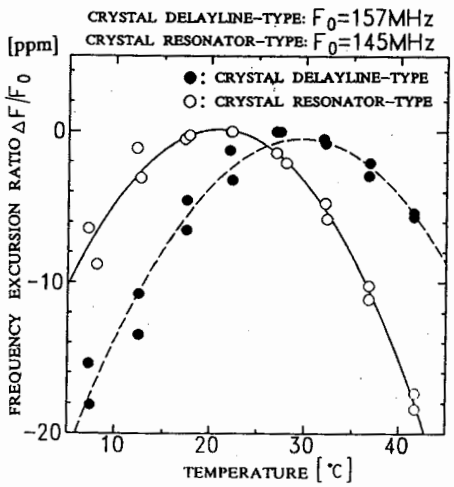


Fig.1 Static Frequency–Temperature Characteristics of Crystal Delayline–Type and Resonator–Type SAW Devices.

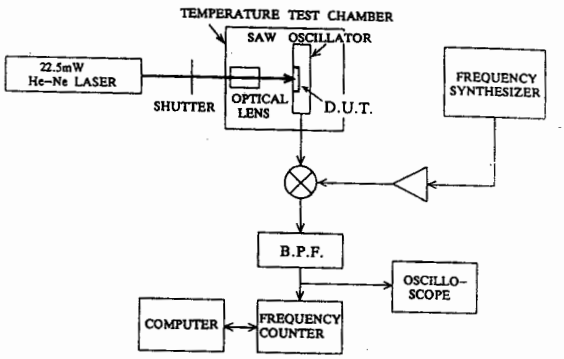


Fig.3 Laser Beam Irradiation and Frequency Excursion Measurement System.

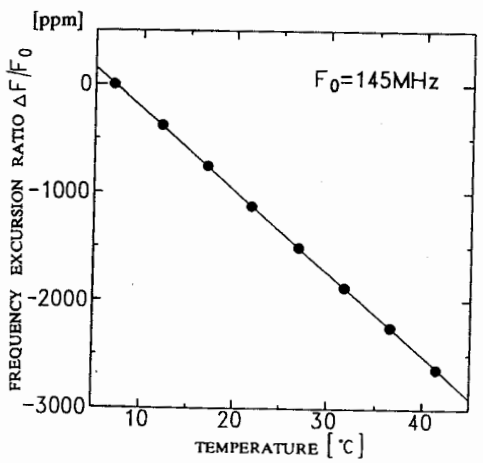


Fig.2 Static Frequency–Temperature Characteristics of a Lithium Niobate SAW Device.

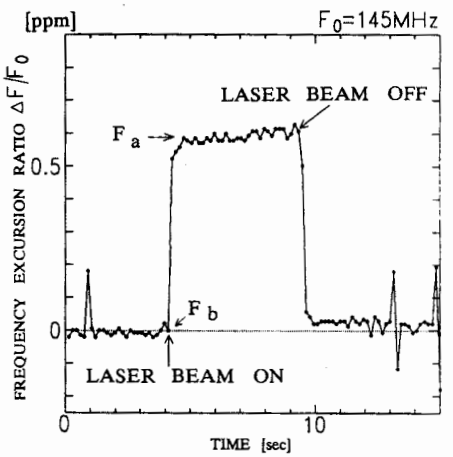


Fig.4 Thermal Frequency Behavior of the Oscillation Frequency due to Beam Irradiation.

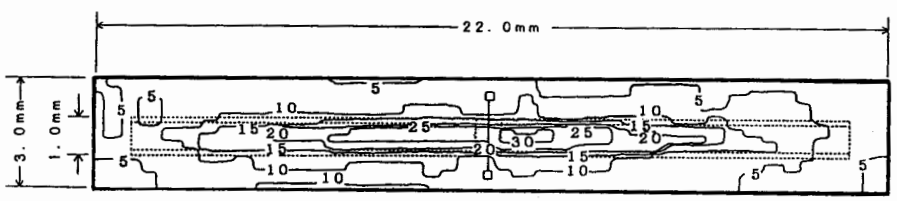


Fig.5 Equi–frequency Map of a Crystal Delayline–Type Device.

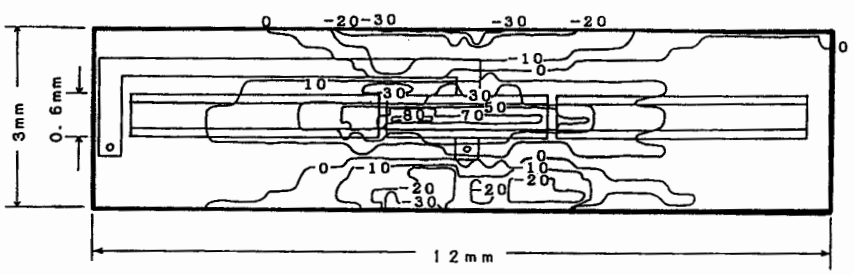


Fig.6 Equi–frequency Map of a Crystal Resonator–Type Device.

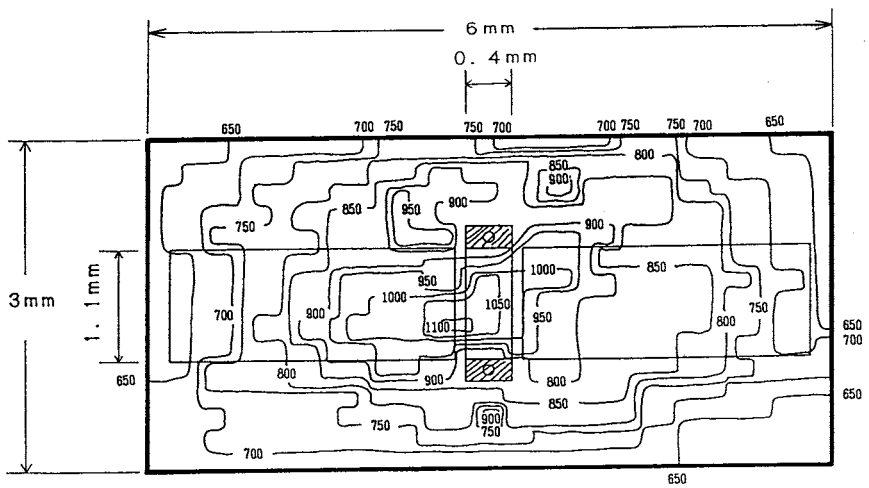


Fig.7 Equi–frequency Map of a Lithium Niobate Device.

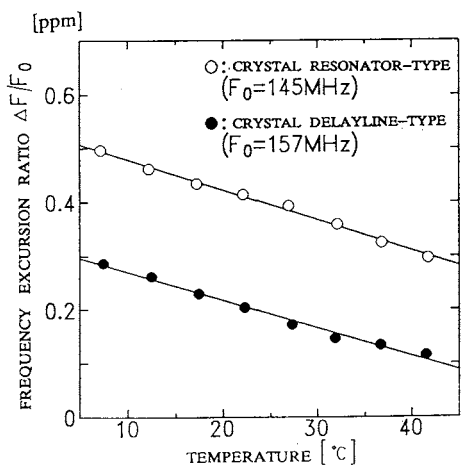


Fig.8 Maximum Frequency Excursion vs. Environmental Temperature.

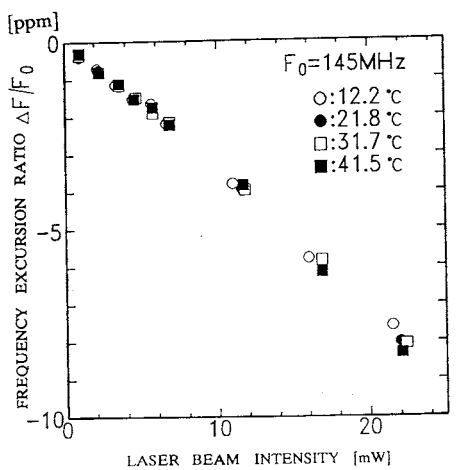


Fig.10 Beam Intensity vs. Frequency Excursion Characteristics of a Lithium Niobate Resonator–Type Device.

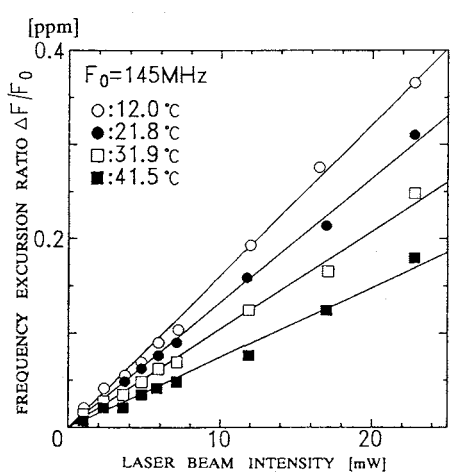


Fig.9 Beam Intensity vs. Frequency Excursion Characteristics of a Crystal Resonator Device.

DEVELOPMENT OF QUARTZ-RESONANT-SENSORS USING FEM

Christian Reichinger

TU München, Lehrstuhl für Elektrische Meßtechnik

ABSTRACT

The design of vibrating quartz sensors for force or pressure is a problem that can only be solved by computer simulation. One possibility is the use of finite element analysis.
The FEM program must be able to compute the piezoelectric coupling between the electric stimulation and the mechanical vibration of three-dimensional anisotropic bodies as well as the damping influence of the mechanical clamping. The determination of the force vs. frequency characteristic requires the implementation of the acoustoelastic effect which describes the change of the elastic wave velocity as a function of mechanical stress. Based on the standard FEM program ANSYS the necessary expansion was made. The simulation results of a typical quartz sensor are presentated.

Quartz resonant sensors, finite element analysis, piezoelectric simulation, acoustoelastic effect

1. INTRODUCTION

Quartz crystals as sensors for physical quantities that change the resonant frequency of the crystal directly have a high accuracy due to the high quality of the resonating quartz.

With the development of force or pressure sensors of this type the mounting must impact the load or pressure to the quartz resonator. A stable, loadable mounting will tend to stop any vibration in the quartz by damping. New quartz resonator geometries with a high quality in spite of the damping caused by the mounting are necessary to avoid this problem.

This design problem can be solved by using the finite element method (FEM). A new method of simulating the damping influence of the quartz mounting is introduced. Furthermore, the integration of higher order elastic coefficients in a finite element is described. The consideration of these higher order terms leads to a FE model that can calculate the change of the resonant frequency due to mechanical loads such as forces or pressures in a mounted quartz of arbitrary geometry in an arbitrary mounting.

For validation purposes the computed results with a quartz double-ended tuning fork are compared with measured values.

2. SIMULATION OF QUARTZ RESONATORS WITH ANSYS

The element library of ANSYS offers a three-dimensional multi-field solid which can calculate the piezoelectric effect. This effect describes the coupling between mechanical strain and electric fields in anisotropic materials. For the simulation the resonator geometry must be assembled with finite elements of this type. As the values of the elastic stiffness tensor, the piezoelectric tensor, the dielectric tensor and the mass density of quartz are known all undamped mechanical resonant frequencies possible in the quartz simulation model can be calculated. The resonant frequencies that can be stimulated piezoelectrically are determined by two calculations with different boundary conditions (i.e. open and shorted electrodes).

For the calculation of damped quartz resonators all possible parameters influencing the quartz must be considered. These parameters are the internal damping of the quartz material, the damping of the electrode material, the damping caused by the surrounding medium and the damping of the mechanical clamping. The first three damping influences can be considered when the material constants, especially the damping values, are known for the used materials. However, in most cases the damping influence of the mechanical clamping is much greater than the influence of the three others is. Because of that the implementation of the damping of the mechanical clamping is most important to get good results of the complete quartz damping.

With calculations including the above described damping effects the impedance of the quartz model as a function of the frequency is determined. From the impedance curve the four parameters C0, C1, L1 and R1 of the electric equivalent circuit diagram of a damped quartz resonator are found.

3. CLAMPING ELEMENT FOR QUARTZ RESONATORS

As the power losses of quartz resonators are mostly caused by mechanical clamping suitable models for the clamping must be developed.

One possibility is to model all parts of the clamping exactly. This leads to a very complex simulation model which is not necessarily more accurate with regard to the energy consumption. There is still the difficulty of defining all material parameters of the clamping, especially those which describe the damping.

Because of that a new clamping element is introduced which is as simple as possible. It must be attached at the quartz surfaces which are in contact with the mechanical clamping. The equation of the clamping element is a function of the contact area and only one other variable D. The basis of the equation is a mathematic consideration. This is the reason why the clamping element is absolutly independent of the clamping material. The damping of the clamping element as a function of D has a special characteristic, as can be seen in figure 1. The maximum damping of the clamping element at D = D* describes the worst case of damping. No mechanical clamping at the modelled surface can have a greater damping. This means that the simulated value of the resonant resistance R1 of the quartz electrical equivalent circuit diagramm as a measure of the complete damping is always greater than the measured value when the maximum damping of the clamping element is used in the simulation (D = D*).

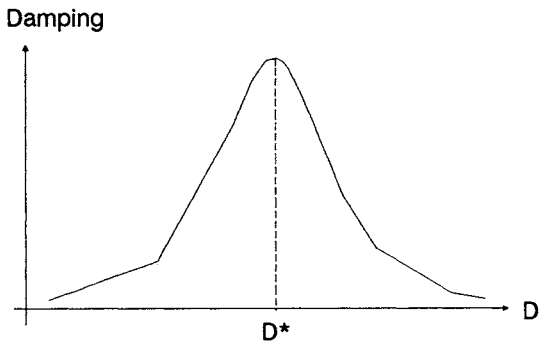


Fig. 1: Damping characteristic of the clamping element

On the other hand the simulation result depends on the exactness of the simulation model, too. To determine the exactness of a simulation model an error estimator for dynamic problems is used. This error estimator is similar to that given for static problems (Ref. 1). It weights the deviations of the simulated mechanical stress from a continuous stress profile. The deviation of the node stresses from the mean stress of a finite element in connection with the node displacements is used to compute the error energy. A common criterion for a good FE model is an error energy evenly distributed over all elements. With a very good simulation model the ratio E of the maximum error energy to the minimum error energy is nearly 1.

As it is normally impossible to create a simulation model with a value of E = 1 another way of determining the exact value of R1 must be found. The solution is to build two ore more simulation models with different values of

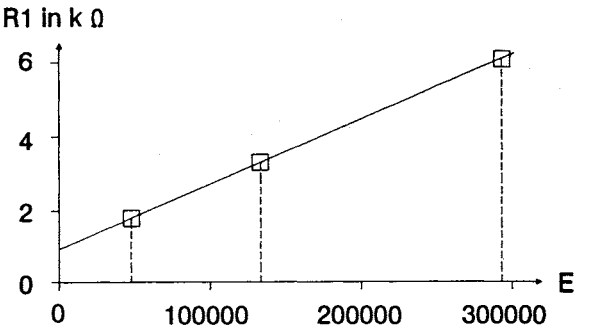


Fig. 2: Extrapolation of the resonant resistance R1

E. Then each model has another worst case value for R1. In figure 2 the E and the corresponding R1 values of three simulations are fitted by a function. The extrapolation of this fitted curve leads to the most exact value of R1 at E = 1. It must be remarked that the fitted curve is not always a straight line as in figure 2. There are curves of higher order depending on the complexity of the quartz resonator, too.

But by means of this new clamping element real quartz resonantors can be calculated with a high accuracy.

4. COMPUTATION OF THE ACOUSTOELASTIC EFFECT

Until now the equations of the simulation program were linear. They describe the stress-strain and the strain-displacement characteristic. In reality these equations are nonlinear.

The acoustoelastic effect is determined by the nonlinearities of these characteristics. Because of that two nonlinearities are added to the simulation program, first the material nonlinearity and second the geometric nonlinearity.

Material nonlinearity is called the fact that the elastic stiffness tensor which describes the linear characteristic between stress and strain is not independent from the stress but changes with the stress. The amount of the changing can be computed using the elastic constants of third order (Ref. 2).

Geometric nonlinearities must be considered when there are displacements which are not small compared to the resonator geometry. In this case the linear characteristic between strain and displacement is no more sufficient for good simulation results. The formulae that take into consideration the geometric nonlinearities can be found in Ref. 3.

The FE program ANSYS allows only the computation with geometric nonlinearities. For the consideration of the material nonlinearities another program was developed. Both nonlinearities change in a first simulation step the stiffness matrix of the simulation model. In a second simulation step the change of the resonant frequency due to the acting forces or pressures is determined by means of the updated stiffness matrix.

Beside this two nonlinearities there are two other quantities that can change the resonant frequency of a quartz resonator due to the acoustoelastic effect: the change of

the mass density and the change of the geometric dimensions of the resonator. With the quartz resonators of our interest these two additional effects are of minor importance as the simulation results prove in the following chapter.

5. SIMULATION RESULTS AND MEASURED VALUES

A typical example for a quartz resonant force sensor is a double-ended tuning fork. A resonator of this type is deliverable from Micro Crystal in Grenchen, Switzerland. Figure 3 shows such a sensor quartz. For the simulation of a quarter of the resonator a FE model with 240 nodes and 123 elements is used. The measured values are derived from a clamped resonator but the sensitivity value is taken from a data sheet.

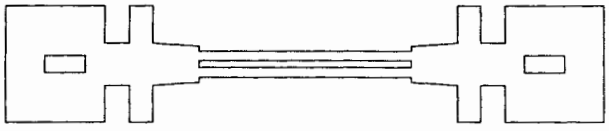


Fig. 3: Quartz double-ended tuning fork

The following table shows a comparison between measured and computed values.

Quantity	measured value	simulated value
fs	47194 Hz	49855 Hz
C0	5.4 pF	1.9 pF
C1	38 fF	8.1 fF
L1	300 H	1255 H
R1	17.5 kohms	24.983 kohms

The simulated resonant frequency fs is 5.6 % to high. This inaccuracy can be reduced by using finer FE meshes (more nodes and elements) but this would increase the calculating time drastically.
The measured value of C0 is higher than the calculated value because there are additional stray capacitances of the wires and the housing.
The deviation of the two quantities fs and C0 affects the values of C1 and L1. Improvements of the simulation with regard to fs and C0 automatically result in better values with C1 and L1.

The results commented above are nearly independent from the damping of the mechanical clamping and can be computed using the standard features of ANSYS. They do not vary by much when the clamping element is omitted. But the calculation of the resonant resistance R1 needs the new clamping element.

The simulation result of R1 is not the most exact value. The value of E is 604 with this simulation model and there are not enough other simulation models of the same resonator with other values of E. So the extrapolation to the value E = 1 is not possible yet. Another interesting thing is the composition of the R1 value. The portion of the quartz damping is 2 ohms, the portion of the air damping 408 ohms. The damping of the electrode material is not included. This means that 98 % of the resonant resistance R1 of the tuning fork are caused by the mechanical clamping when using this not optimal simulation model.

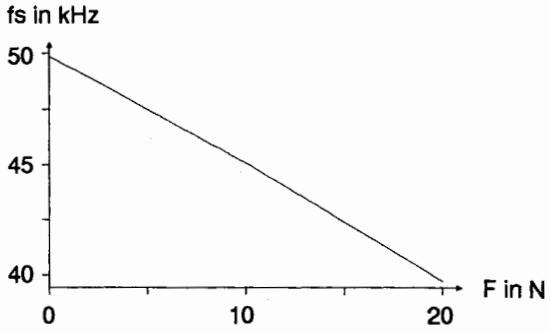


Fig. 4: Force vs. frequency characteristic

The most important quantity of a sensor is the sensitivity. The calculation of the force sensitivity of the tuning fork is made possible by the material and geometric nonlinearities. Figure 4 shows the force vs. frequency characteristic of the tuning fork loaded with a compressive force F. The relative sensitivity determined by simulation is -1.01 % / N. The data sheet of the quartz sensor has a value of -1 % / N. The good agreement of simulated an 'measured' value shows that the change of the mass density and the resonator geometry due to the compressive force F has no important influence on the change of the resonant frequency.

6. CONCLUSION

The expansion of the standard FEM program ANSYS with a new clamping element in connection with an error estimator for dynamic problems makes the simulation of real quartz resonators of arbitrary geometry in an arbitrary mounting possible. The additional consideration of material and geometric nonlinearities leads to a simulation program by means of that the design of quartz-resonant-sensors for force or pressure is possible. The development of the sensor mounting and housing is another task that can be solved with ANSYS, too.

7. REFERENCES

[1] Zienkiewicz, O. C.; Zhu, J. Z.: A simple error estimator and adaptive procedure for practical engineering analysis. - in: International Journal for Numerical Methods in Engineering, Vol. 24, pp 337-357, 1987

[2] Nalamwar, A. L.; Epstein, M.: Surface acoustic waves in strained media. - in: Journal of Applied Physics, Vol. 47, No. 1, pp 43-48, 1976

[3] Zienkiewicz, O. C.: Methode der finiten Elemente. - Hanser Verlag, München, 1975

ION EXCHANGE AND RADIATION RESPONSE OF HYDROGEN-RELATED POINT DEFECTS IN CRYSTALLINE NATURAL QUARTZ

Harish Bahadur
National Physical Laboratory
Hillside Road, New Delhi-110012, India.

ABSTRACT

Radiation effects in high-H and low-H natural quartz with similar aluminum concentration have been investigated upon irradiation at 77K and 300K. At low temperature irradiation the Al-OH bands equalize in their intensity. Among other major bands the alkali-related centers show a reduction in their strength at all stages of irradiation. Purely H-related centers show no steady state room temperature effects; only the low temperature irradiation reduces their band strength.

Keywords: Quartz Crystals, High-H Quartz, Low-H Quartz, Point Defects, Irradiation Effects, Infrared Absorption.

INTRODUCTION

Crystalline quartz is a strategic material for a variety of scientific and technological applications. The limitation of quartz crystal-controlled oscillators subjected to radiation in earth-orbitting satellites are well known. The crystal oscillators suffer from both transient and steady state frequency and Q offsets. It has been realized that these offsets occur due to radiation induced modification of impurity related point defects in crystalline quartz material. Therefore, considerable attention has been given so far to the study of radiation effects in crystalline quartz [1-3].

Among the impurities, aluminum is the most pervasive impurity in quartz and is present as Al^{3+} at substitutional sites of Si^{4+}. The charge compensation of aluminum is provided by one of the monovalent impurities such as Li^+, Na^+, K^+, proton interstitially placed in open c-axis channels or by an electron-hole on one non-bonding oxygen p-orbital. In addition, there are other types of point defects also present in quartz [4]. They include, e.g. oxygen vacancies, Ge-associated centers and numerous OH defects. Substitutional aluminum and and a group of unidentified hydrogen-containing defect centers, shown in Figure 1, are probably the most important point defects in quartz from the point of view of resonator performance. A variety of experimental techniques are used to characterize and monitor these defects.The OH related defects exhibit their OH stretching vibrational frequencies in the 3300-3700 cm^{-1} range and therefore can be studied using near infrared absorption. The absorption of quartz in this range observed at low temperature contains a large number of narrow bands superimposed on a broad background. These bands are as many as sixteen in natural quartz and four in cultured quartz. In a recent study, Bahadur [5] has shown that in a variety of natural quartz of different origins a set of major bands always present is at 3476, 3306 and 3367 cm^{-1}. The latter two bands represent respectively the short- and long-bond Al-OH centers. The 3476 cm^{-1} peak originally determined by Kats [6] to be due to Li-OH vibrations was studied by Bahadur [5]. While a specific model could not be ascertained, it was shown that the 3476 cm^{-1} band was not associated with Al-center. The unidentified near IR bands in natural quartz have been divided in four groups in which there is one major and many weak lines in each group [7]. Table (I) depicts the position, assignment and notations of IR bands in natural quartz in unswept, Na-swept and H-swept conditions.

In this paper we describe the IR spectral results on the major bands upon low and room temperature irradiation in unswept, Na-swept and H-swept conditions of high- and low-H natural quartz crystals.

EXPERIMENTAL PROCEDURE

Electrodiffusion (sweeping) was done at Oklahoma State University (OSU) using standard OSU recipe. Irradiation was done using the OSU van de Graaff electrostatic accelerator while the crystal tempearture was maintained in the Dewar at 77K. The crystals were irradiated at all stages in the Dewar itself without demounting. Spectral measurements were carried out on a Beckmann 4240 double beam spectrophotometer. The detailed description of electrodiffusion, irradiation and measurement techniques have been reported earlier [5,7].

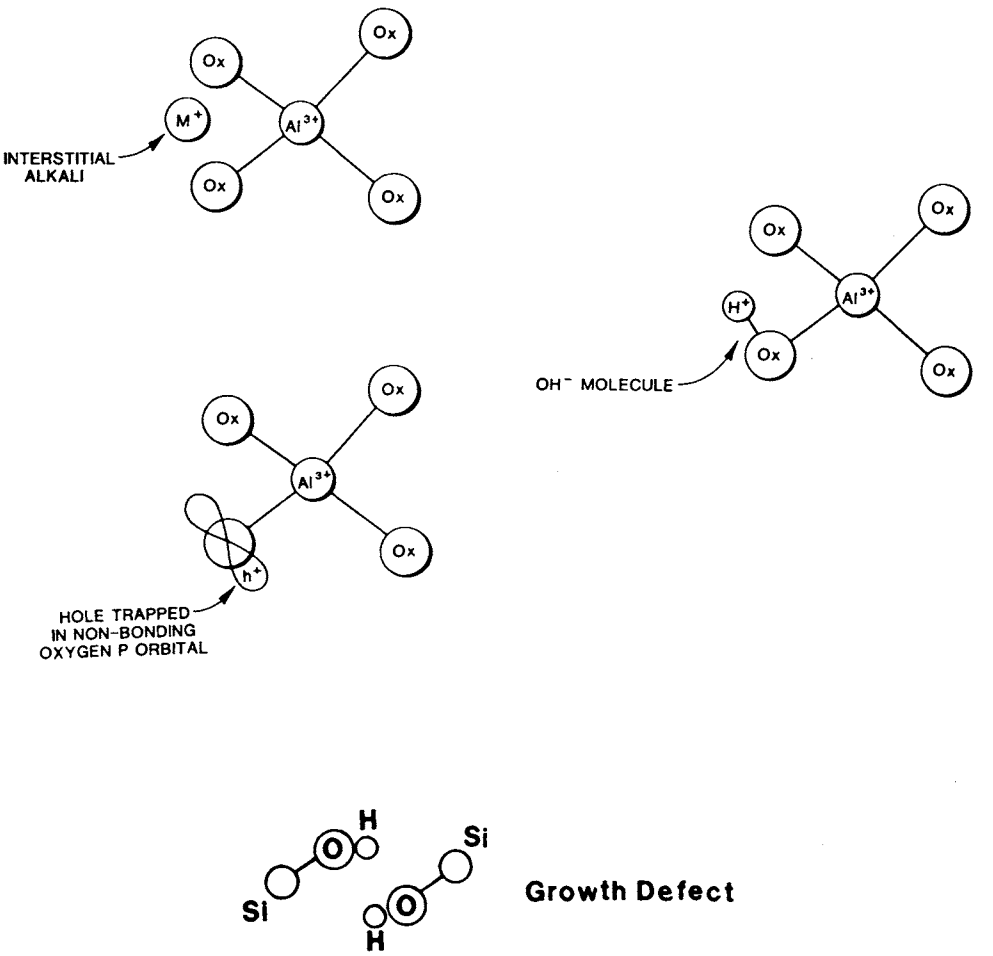


Figure 1. Schematic representation of aluminum and OH-realted pointed defects in crystalline quartz.

TABLE - (I)

Position, assignments and notations of IR bands in natural quartz in unswept, Na-swapt and H-swept conditions

UNSWEPT		Na-SWEPT		H-SWEPT	
Wavenumber (cm^{-1})	*Notations*	*Wavenumber (cm^{-1})*	*Notations*	*Wavenumber (cm^{-1})*	*Notations*
3240	-	3240	-	3240	-
3306	Al_s	3306	Al_s	3306	Al_s
3367	Al_i	3350	-	3367	Al_i
3383	-	3367	Al_i	3383	-
3395	*Li-n1a*	3375	*Na-n1a*	3395	*H-n1d*
3420	*Li-n2c*	3385	-	3415	*H-n1b*
3430	*Li-n2b*	3393	*Na-n1c*	3420	-
3437	*Li-n2a*	3395	*Li-n1a*	3425	*H-n1c*
3451	*Na-n2a*	3432	*Li-n2b*	3430	*H-n1a*
3460	*Li-n3c*	3435	-	3452	-
3468	*Li-n3b*	3437	*Li-n2a*	3463	*H-n2b*
3476	*Li-n3a*	3451	*Na-n2a*	3468	*H-n2a*
3480	*Li-n3d*	3468	*Li-n3b*	3476	-
3507	*Li-n4b*	3476	*Li-n3a*	3483	*H-n2c*
3520	*Li-n4a*	3507	*Na-n3a*	3535(broad)	-
3575	-	3520	*Li-n4a*	3535 (sharp)	-
3595	-	3555	*Na-n4a*	3570	-
		3585	-	3583	-
		3595	-	3595	-

RESULTS AND DISCUSSION:

Figure 2 shows the near IR spectra of the two kinds of crystals; Brazilian and Arkansas quartz. It may be seen that the Brazilian crystal has a large number of OH-related peaks and the Arkansas crystal has only three major peaks located at 3476 cm^{-1} and the Al-OH bands at 3306 and 3367 cm^{-1}. We shall consider only these peaks in our discusion. The peak has been found [5] to consist of a Li ion and a proton together. Electrodiffusion with H (also called H-sweeping) results in prominent bands at 3468 cm^{-1} and the usual Al-OH bands. Spectral changes on H-sweeping over that of the unswept condition of the Brazilian quartz are shown in Figure 3. It may be seen from the spectrum of H-swept crystal that both the Al-OH bands increased in their strength and the 3476 cm^{-1} peak is replaced with another 3468 cm^{-1}. Comparative results on unswept and Na-swept conditions are shown in Figure 4. Figure 5 depicts the near IR spectra for the Arkansas quartz in differently swept conditions. The Brazilian crystal was found to be hydrogen-rich and the Arkansas crystal as hydrogen-lean. We have in the following labelled the Brazilian crystal as 'high-H' and the Arkansas sample as 'low-H' quartz. Figure 6 shows the bar diagram for these samples indicating the relative strength of various impurity related bands.In the bar diagram the Al-OH band strengths have been plotted from zero up and other impurity related band groups down. According to the notations introduced before [7,8],the 3476 cm^{-1} peak has been labelled as Li-n3a meaning that it belongs to group no.3 and is the most intense line which has Li and H ions together. The 3468 cm^{-1} peak has been designated as H-n2a.

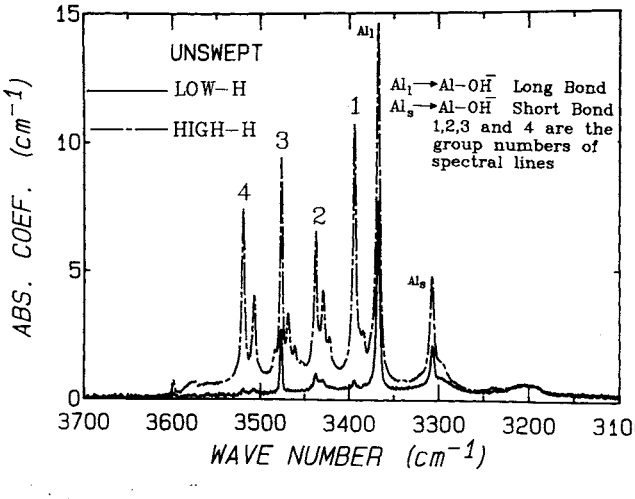


Figure 2. Near IR absorption spectra of Brazilian and Arkansas quartz crystals in their unswept conditions.

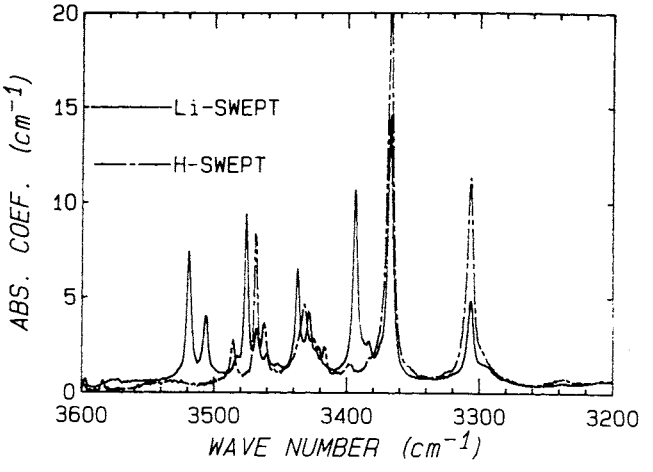


Figure 3. Near IR absorption spectra of high-H natural quartz in unswept and H-swept conditions.

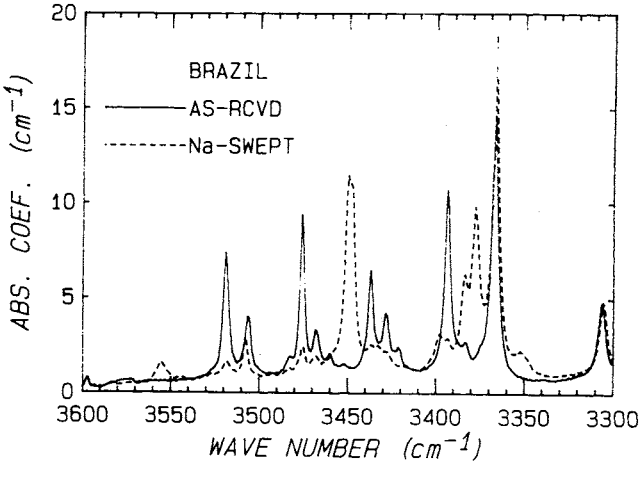


Figure 4. Near IR absorption spectra of high-H natural quartz in unswept and Na-swept conditions.

Figure 7 presents the spectra obtained after each step of irradiation sequence carried out for the high-H crystal in its unswept condition. Figures 8 and 9 represent the irradiation characteristics of the H-swept high- and low-H crystals. The complete results in the form of bar diagrams for both high-H and low-H samples are shown in Figures 10 and 11. From these Figures it may be seen

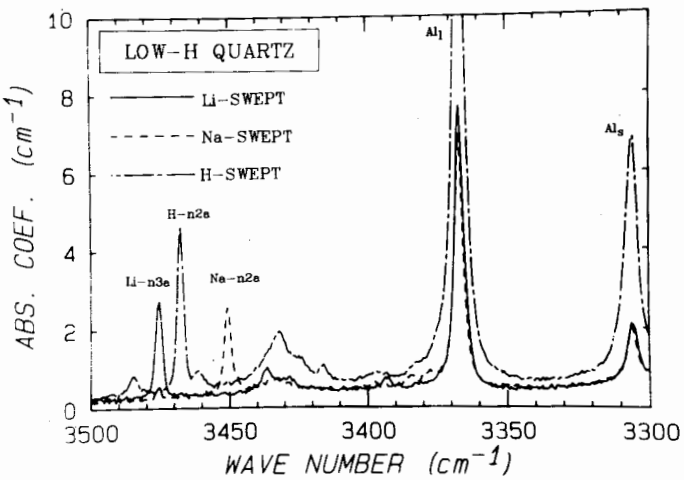


Figure 5. Near IR absorption spectra of low H quartz in its unswept, Na-swept and H-swept conditions.

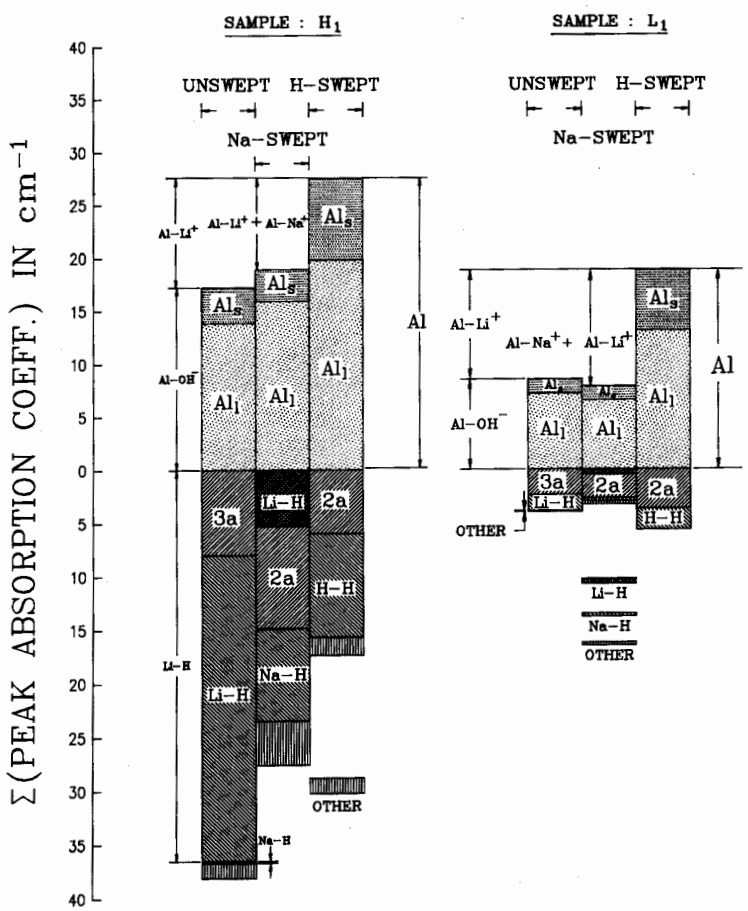


Figure 6. Bar diagrams showing the relative strength of various imperity - related IR bands in natural quartz crystals. H_1 is the high-H and L_1 is the low-H quartz.

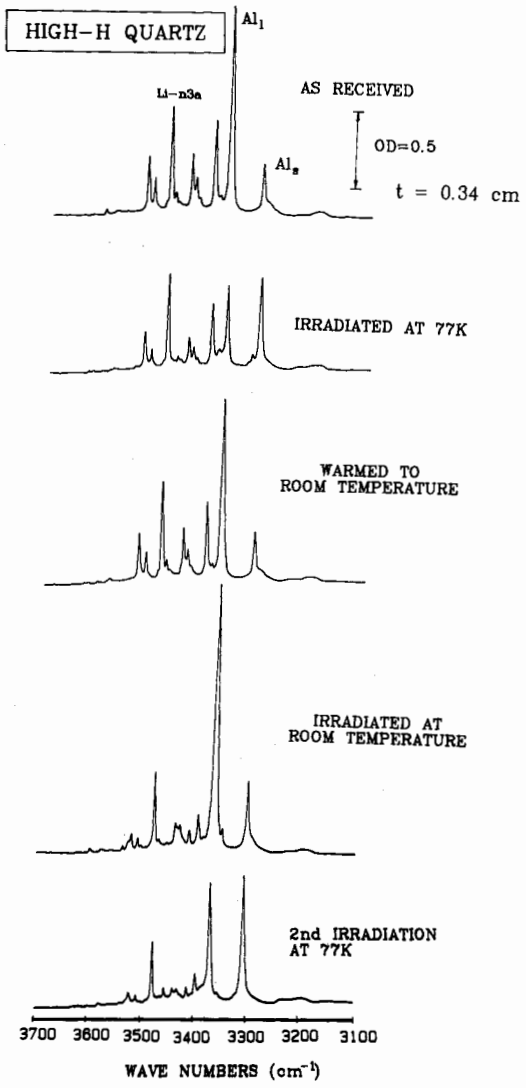
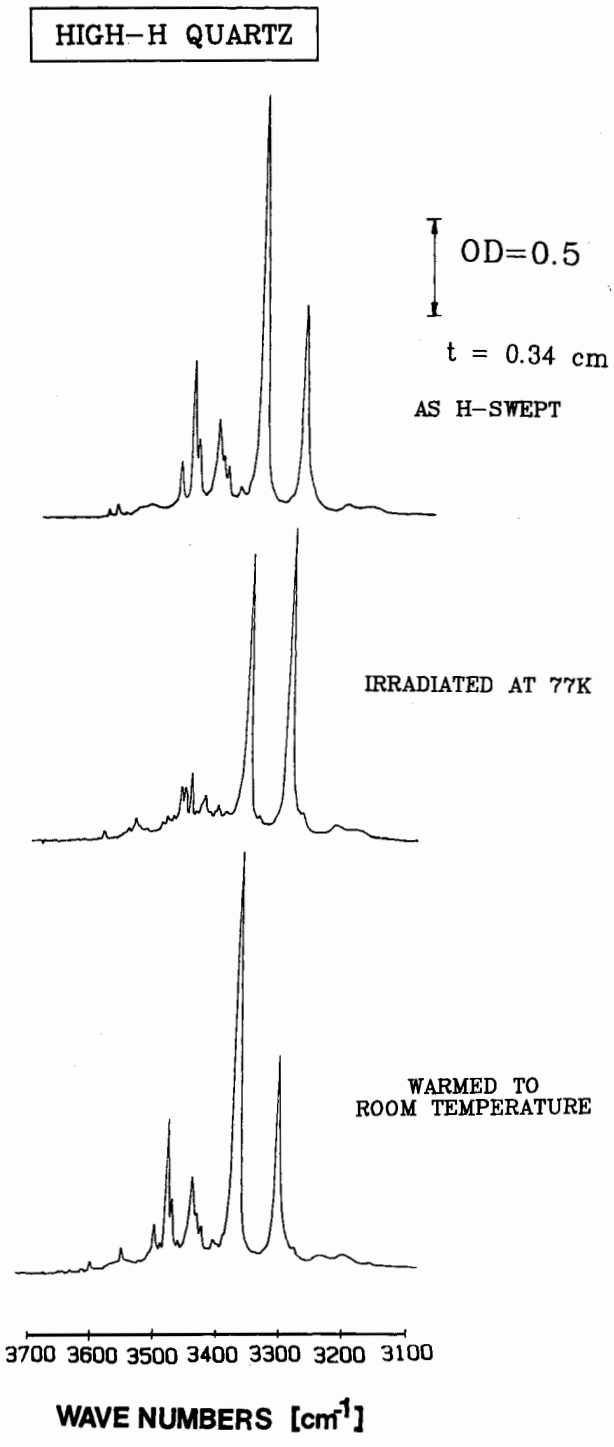


Figure 7. Near IR absorption spectra of high-H unswept quartz as a function of irradiation temperature.



This Figure is continued on next page.

that the low temperature irradiation effects result in equalization of Al-OH bands. This would be due to some protons moving away from Al_1 band and getting into Al_2 band. Warming to room temperature results in about 85-95% recovery in these and other bands. This shows that protons from these defects move to shallow traps. On the other hand, while the Li-n3a and Na-n2a bands showed a decrease in their strength at all stages of irradiation the H-n2a band exhibits significant reduction only at low temperature irradiation. This suggests that protons are held firmly at the H-n2a center and the room temperature irradiation does not allow the formation of $[H\text{-}n2a_{e+}]^0$-center from its precursors. Another important difference between the low temperature- and room temperature-irradiation is the fact that room temperature irradiation leads to trapping of protons and alkalis from their precursors by deep traps. Basically, similar irradiation effects have been found for both high-H and low-H natural quartz.

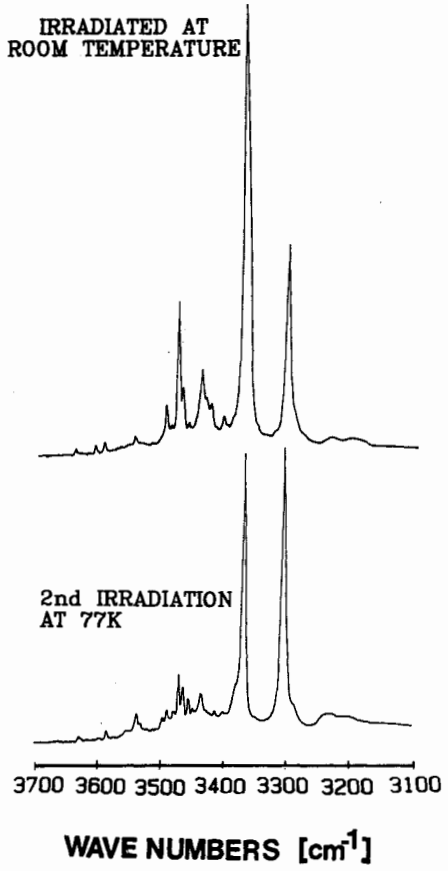


Figure 8.

Near IR absorption spectra of H-swept high-H quartz as a function of irradiation temperture.

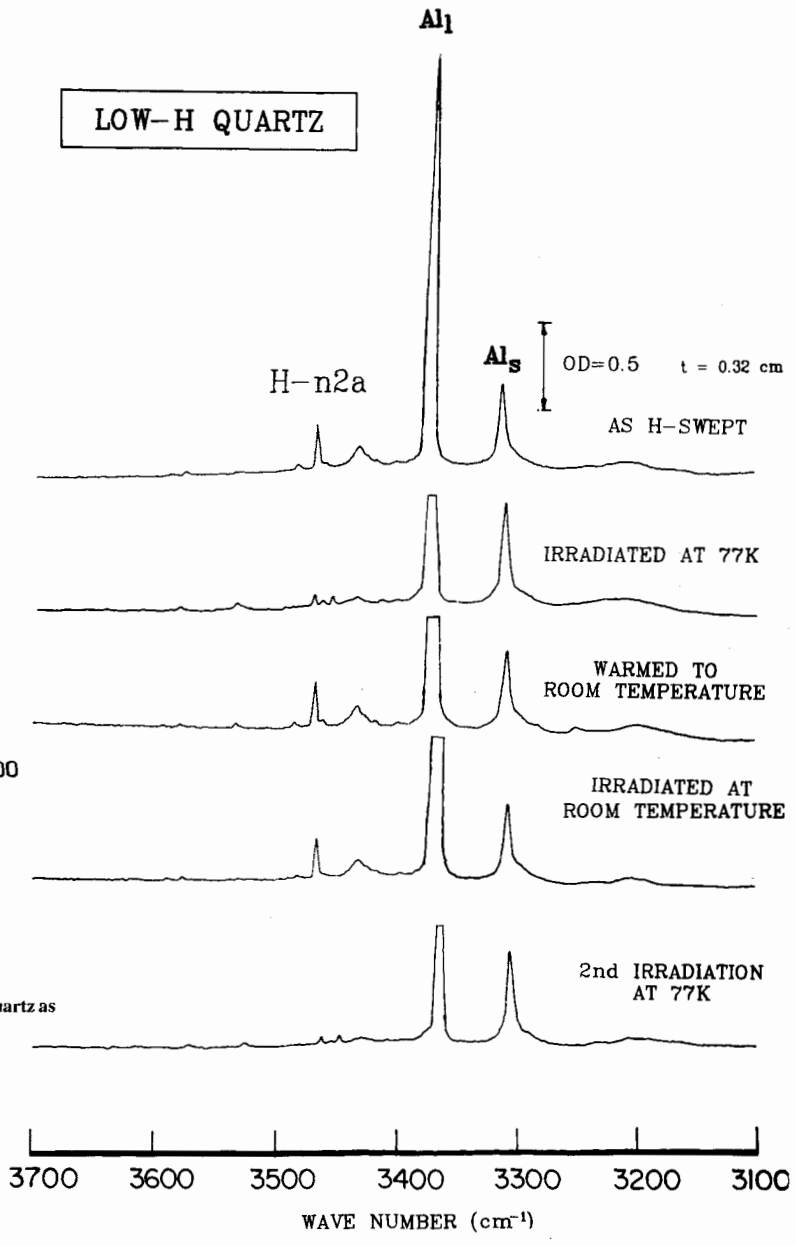


Figure 9. Near IR absorption spectra of H-swept low-H quartz as a function of irradiation temperture.

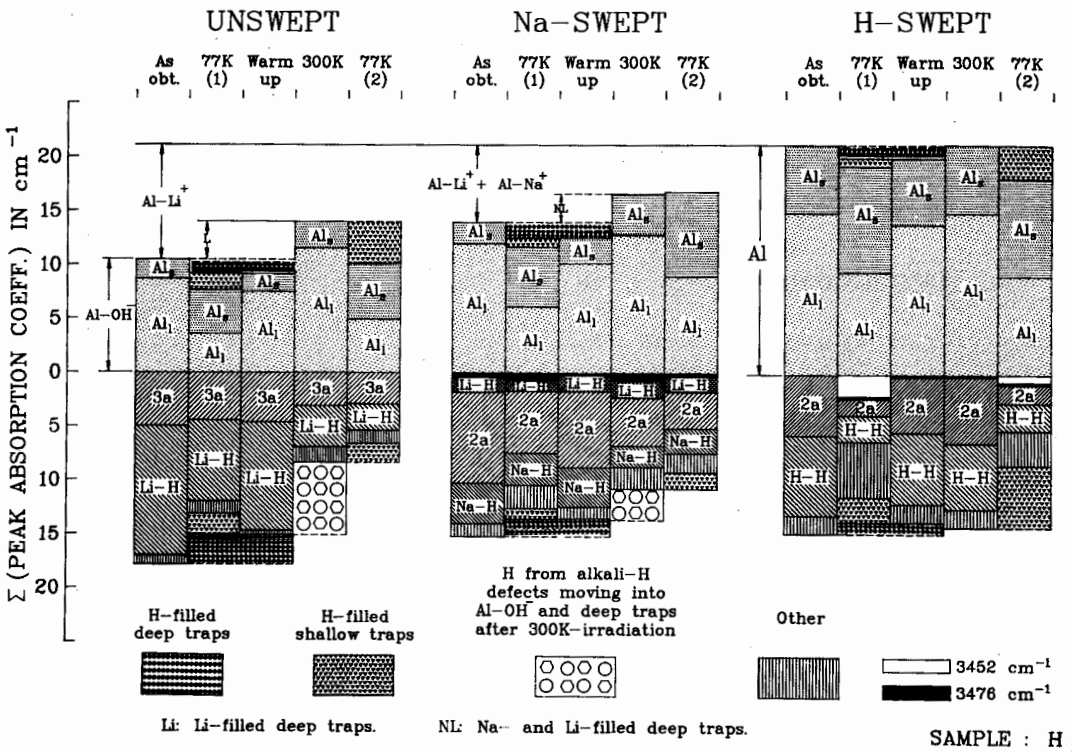


Figure 10: Bar diagrams (for high-H sample H_2) representing irradiation effects on different Al, Li, Na and H related OH defects in high-H natural quartz. Notations 'As_{obtd}' represents as-received unswept or obtained after Na- and H-sweeping, '77K$_{(1)}$' depicts results after first irradiation at 77K; 'Warm up' means warming to room temperature; '300K' shows results after irradiation at 300K and '77K$_{(2)}$' is the stage of second irradiation.

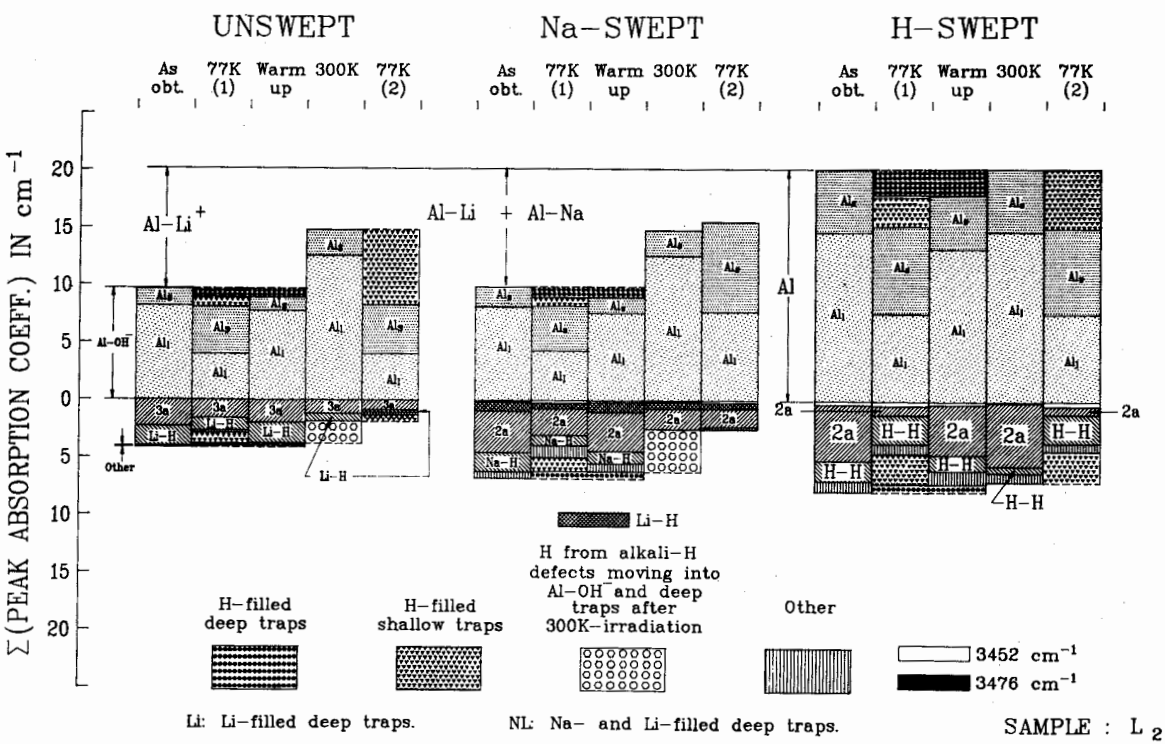


Figure 11: Bar diagrams (for low-H sample L_2) representing the irradiation effects on different Al, Li, Na and H related OH defects in low-H natural quartz crystal. Notations represent the same meaning as in Figure # 10.

CONCLUSION

Basically, similar irradiation characteristics were seen for both high- H and low-H natural quartz crystals. Alkali-H related infrared bands show reduction at all stages of irradiation while only H-related defects show no steady effects after room temperature irradiation; only the low temperature irradiation reduces their band strength.

ACKNOWLEDGMENTS

The author is grateful to Professor Joel J. Martin of Oklahoma State University , USA for supporting this work. This work was funded by Sandia National Laboratories and the Solid State Sciences Division of the Rome Air Development Command under the principal investigator Professor Martin. The author is thankful to Dr. J.C. King, Director, Materials Fabrication and Process Engineering of Sandia National Laboratories (now retired) for his encouragement to investigate the natural quartz crystals. The author is further thankful to Professor E.S.R. Gopal, Director, National Physical Laboratory, New Delhi and Dr. T.K. Saksena, Head, Ultrasonics Section, NPL for their encouragement to publish this work.

References

1. L. E. Halliburton, J.J. Martin and D.R. Koehler, "Properties of Piezoelectric Materials" in *Precision Frequency Control,* E.A.Gerber and A.D. Ballato, eds., Orlando, FL: Academic Press, Vol.1 pp. 1-46.
2. J.C. King and D.R. Koehler, "Radiation Effects on Resonators" in *Pression Frequency Control,* E.A. Gerber and A.D. Ballato, eds, Orlando, FL: Academic Press, Vol.1 pp. 147-159 (1985).
3. J.J. Martin, J. Appl. Phys. **68,** 5095 (1990).
4. J.A. Weil, Phys. Chem. Minerals **10,** 149 (1984).
5. Harish Bahadur, J. Appl. Phys. **66** 4973 (1989).
6. A. Kats, Philips Res. Rept. **17,** 133 (1962).
7. Harish Bahadur, Proc. 45th Annual Frequency Control Symposium, IEEE , **37,** (1991).
8. Harish Bahadur, IEEE Trans. Nucl. Sci. pp 2170-2177 (1992).

CONTROLLED CRYSTAL DISSOLUTION APPLIED TO QUARTZ

M.Deleuze, A.Goiffon, A.Ibanez, E.Philippot

LPMS, URA D0407 CNRS, case 003, UM II, F-34095 Montpellier cédex, FRANCE

ABSTRACT

Quartz chemical lapping of AT cuts was performed in NaOH.xH2O medium and this controlled dissolution resulted in reasonable quality of surface texture. Activation energy was calculated from etching rates measured against temperature. The initial surface texture of the samples influenced roughness parameter evolution but not the final Ra value. The electrical response of piezoelectric devices made by chemical lapping has been compared against others obtained by IBE process. The comparison showed that controlled dissolution process is at least as good as IBE for producing piezoelectric devices.

Keywords : Quartz, Chemical etching, Controlled dissolution, Roughness , Piezoelectric material, Basic médium.

I- INTRODUCTION

For quartz piezoelectric device manufacture, the crystalline quality of the material is of prime importance, and the atraumatic adjustement frequency of resonators a very important step. Indeed, the Q-factor is closely related to both crystalline quality and the final surface roughness of the sample. Thus lapping must produce a good surface state, and a definite thickness which corresponds to the required frequency.

Unfortunately, for sophisticated high frequency resonators, with thickness less than 30 µm, the mechanical lapping process cannot be used. Only the expensive ion beam etching technique (IBE) can be applied.

Much research has been devoted to another possibility, based on a chemical approach to these problems using fluoride media (Refs 1-3), but has not resulted in an industrial method of quartz chemical etching being devised. Indeed, the HF or NH4HF2 baths used are very traumatic for the material, even with surfactants (Refs 4-6).

In our department, we have first defined the crystal growth conditions of a quartz-like material, AlPO4, berlinite in acidic media , and its controlled dissolution in the same solvents (Refs 7-8). These two processes are closely related through the reverse thermodynamic relations (Refs 9-14), and these phenomena are even more controlled when experimental conditions are constitued by a solubility equilibrium, i.e. weak super- or under-saturation of solvents for crystal controlled growth or dissolution respectively. In this case, it appears an energy barrier, ΔG*, avoiding most etch pit formation in the dissolution process, figure 1.

Taking into account these considerations, we investigate the quartz problem, which includes a factor critical for piezoelectric device manufacturing. The quartz crystal growth being made in basic medium, our controlled dissolution has been carried out in a similar solvent, but inversely to the previous works (Refs 1-6). Drastic growth conditions (high pressure and high temperature) were not applied, but these basic media under standard conditions are rather atraumatic for the quartz. Indeed, it is a question of process reversibility: possible under certains conditions in basic medium, but impossible in an acidic one. We give here our first results for the AT-cut of quartz samples in NaOH solvent.

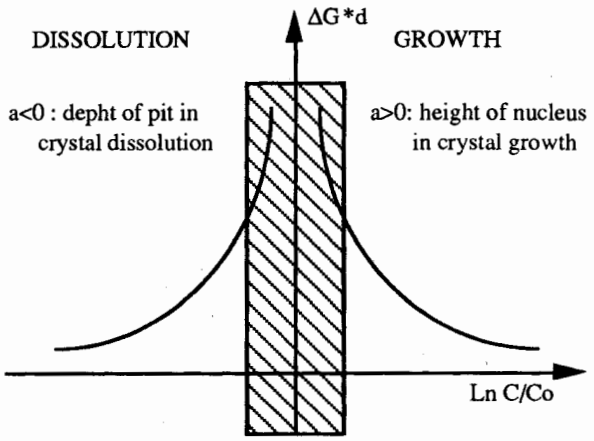


Fig 1 : ΔG*d energy barrier variation for growth and dissolution of a real crystal in terms of solute concentration, C (Co= solubility at chosen temperature).

II - EXPERIMENTS

II.1 - Samples

Only the AT-cut orientation has been investigated in this first study. All samples, provided by CEPE/Thomson, have been sawn from high-Q swept quartz bars. Discs are ground and polished, of 5 to 8 mm in diameter . Two sizes of final abrasive were used: 3 µm for ground wafers and 0.3 µm for polished ones. The used plates were chosen in the range 800 to 67 µm thick before chemical etching process. These thicknesses correspond to resonance frequencies ranging between 2 and 25 MHz.

II.2 - Experimental procedure

II.2.1 - Description of etching reactor.

In this investigation, glass containers have been replaced by PTFE ones to avoid chemical corrosion of (NaOH+H2O) mixtures. Fig 2 is an illustration of this system used for plates in basic medium.

II.2.2 - Etching baths.

Etching solutions were prepared from high purity sodium hydroxide pellets (MERCK pro analysis). Three different compositions were investigated:

	Molality (mol/kg solvent)
NaOH.H2O	55,6
NaOH.2H2O	27,8
NaOH.3H2O	18,5

II.2.3 - Measurement of plate thickness.

The decrease in thickness of the wafer was checked during etching by measurement of the resonance frequency given by :

$$F(\text{MHz}) = K(\text{MHz.}\mu\text{m}) / e(\mu\text{m})$$

where e is the thickness of the plate (given in μm) and K a constant related to the material and orientation studied. In AT cuts of quartz, K=1.67 MHz.mm . These measurements were made using a frequency meter which allows a broad range of frequencies from 0 to 110 MHz to be scanned.

II.2.4-Characterization of surface texture.

Surface texture was examined using a Perthometer W5B surface profilometer. This apparatus, fitted with a microprocessor, was used to calculate the roughness parameter Ra (or average roughness) defined mathematically by the following equation:

$$Ra = \frac{1}{L}\int_0^L |Z| dx$$

where L is the profile length and Z is measured from the center line, i.e. from the line so that the cross-section areas of asperities above and grooves below are equal. For reliability, all measurements were carried out with the same electric filter, λc=0.08 mm, and the given Ra value is the mean value of 4 different scanning directions (45°different from each other).

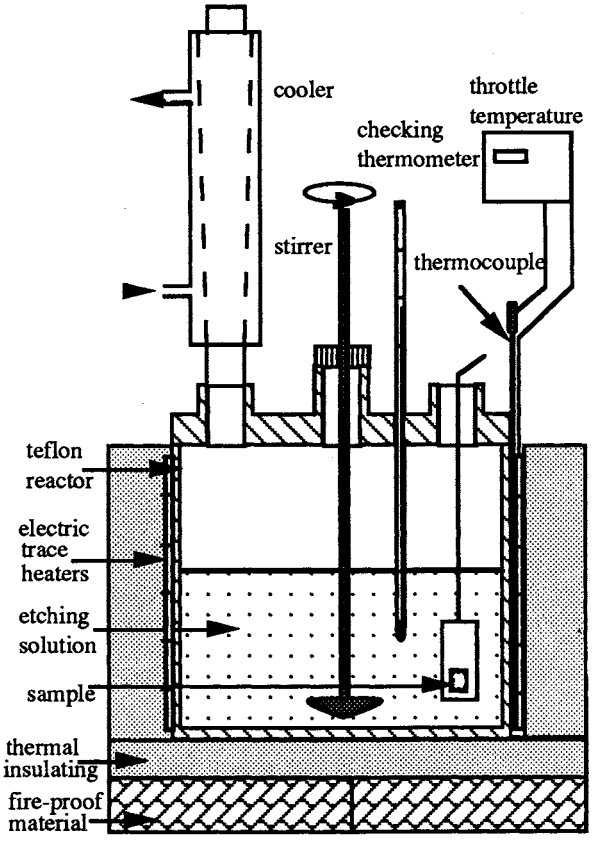


Fig 2 : System used to etch plates

III - RESULTS AND DISCUSSION

III.1 - Controlled dissolution of polished AT plates in NaOH,xH2O (x=1, 2, 3) mixtures.

The initial roughness value of the polished AT plates is Ra=0.004 μm, which corresponds to an optical polish surface state. These first experiments, at 170 °C, were undertaken to provide us with some knowledge on quartz behaviour in strongly basic conditions. The best etching solution seems to be the most concentrated one (fig 3). Indeed, for the more diluted medium, NaOH.3H2O, the Ra value increases quickly from 0.004 to 0.09 μm after a thickness decrease of only 23 μm. Light microscopic observations (x1100) show regular and oriented grooves.

On the other hand, for the most concentrated solvent, NaOH.H2O, the roughness of the plates and their surface texture (x1100) does not differ from the initial values. For this reason, this last concentration was used to carry out the investigation of ground AT-cut wafers in basic solutions.

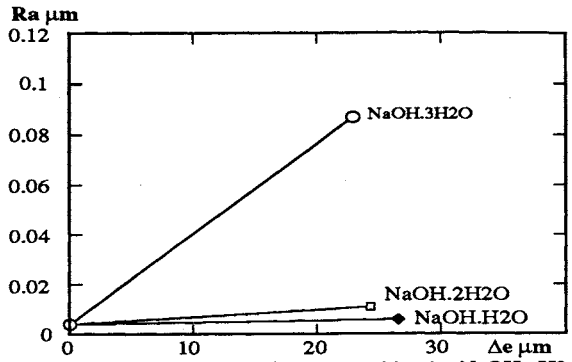


Fig 3: Polished AT plates etching by NaOH.xH2O

(Fig 3 is only an illustration of these tests, Ra is not a linear function of the decrease in thickness Δe).

III.2 - Controlled dissolution of ground AT plates in NaOH. H2O mixture.

This study was based on the influence of bath temperature and the initial roughness of the plates on dissolution kinetics and the final surface state of quartz AT plates.

III.2.1 - Temperature influence.

AT discs used were 6 mm in diameter, i.e. initial frequency ≈ 10MHz (e=167 μm), with a roughness value in the range 0.16<Ra<0.18 μm.

Some evolutions of the etching rates against the etching depth have been plotted in figure 4. As expected, these etching rates increase with temperature. All the three curves, shown in figure 4, can be divided in two parts:

- the first part relates to a rapid decrease of the etching rate. This step corresponds to the etching of the surface layer disturbed by cutting and polishing.
- In the second part, the etching rate becomes more stable. It might thus be the characteristic intrinsic rate of chemical etching of a given orientation of the material at the chosen temperature.

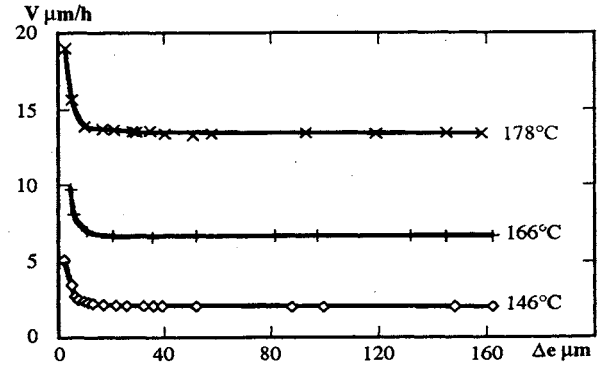


Fig 4 : Etching rate at different temperature of AT plates plotted against etching depth in NaOH.H2O

In figure 4, the etching rates become constant for a total etching depth of 7 μm, i.e. 3.5 μm per face. This last value appears to be the depth of the disturbed layer. The same result can be found for the roughness parameter, from the curves of Ra, plotted against the etching depth, Δe, figure 5. Previous studies on quartz (Ref 15) and berlinite (Refs 7-8) samples have already shown this behaviour.

Whatever the temperature, the Ra=f(Δe) plots show a similar behaviour: Ra reaches a maximum for an etching depth of about 7 μm (for the two faces) and then slowly falls to a low value close to 0.06 μm, which corresponds to "good chemical polishing" or "controlled dissolution" of quartz material.

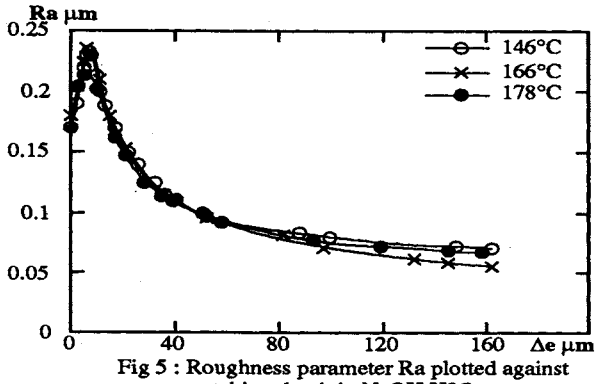


Fig 5 : Roughness parameter Ra plotted against etching depth in NaOH.H2O

The evolution of the surface texture was checked by light microscopy during the study, figure 6. Before etching, the surface was formed by numerous disordered hillocks, figure 6A. The irregularity maximum was obtained for the roughness maximum measured. Indeed, chemical etching tended to preferentially remove the weaker disturbed parts at the expense of the crystal bulk, leading to a maximum Ra value. In a second phase, we then observed the slow polishing effect of the solvent which improved the surface quality of the samples.

Since the roughness stabilization was very slow, we tried to define a final Ra value after a longer etching time. 630 µm were thus removed from a 800 µm thick wafer, and the Ra value decreased from 0.19 to 0.03 µm, and would certainly have continued to converge to a polish surface state. On the other hand the isotropic effect of this etching process should be noted, which is enhanced by no alteration in appearance of the circular shape of the quartz discs, even after removing several hundreds µm.

In conclusion, the NaOH.H2O medium is a very good one for a controlled dissolution of AT quartz ground samples. The temperature of the chemical etching has no effect upon the final roughness, but only upon the etching rate.

III.2.2 - Initial roughness (Rai) influence.

From previous experiments in NaOH.H2O medium, it appears that a slight change in the initial roughness of the wafers (0.16<Ra<0.18 µm) corresponds to a slightly different final surface state, figure 5. Thus, it seems that initial roughness influences the stabilization of the surface state in terms of etching depth. AT plates having different Rai values (table I) have been etched at 166°C.

	Ø mm	initial frequency MHz	Rai µm
plate 1	8	21	0,15
plate 2	6	10	0,18
plate 3	5	8	0,27
plate 4	5	25	0,004

Table I : Characteristics of plates etched at 166°C

The roughness evolution of these plates is plotted in figure 7. At first, we noted that the increase of Rai disappeared with the improvement of the surface state from 0.27 to 0.15 µm. We observed, however, that a low Ra value= 0.07 µm was reached after an etching depth of only 40 µm for sample 1 (Rai=0.15 µm). On the other hand, 160 µm must be removed before reaching the same value for sample 2 with worse initial roughness (Rai=0.18 µm). These two results confirm previous research on to quartz sample preparation (Refs 15-16) and emphasize its importance for the chemical etching process.

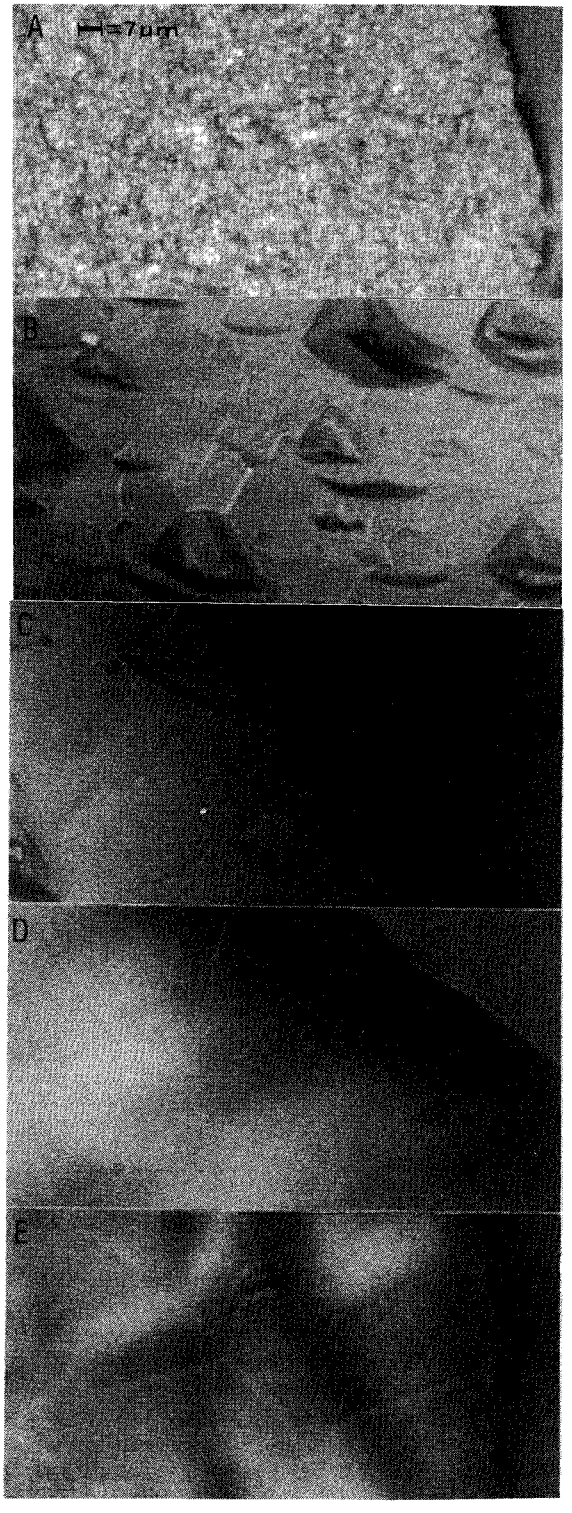


Fig 6 : Photographs of AT cuts during etching in NaOH.H2O at 166°C : A) prior to etching B) Δe=24 µm C) Δe=45 µm D) Δe=110 µm E) Δe=167 µm.

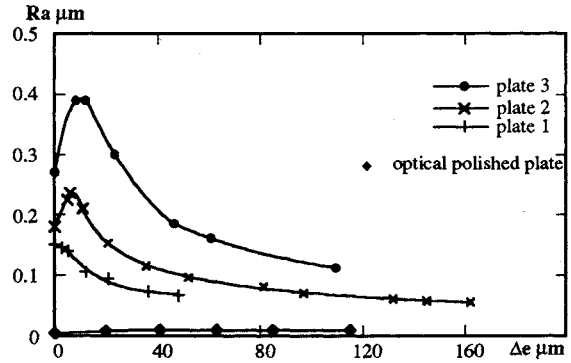


Fig 7 : Roughness parameter plotted against etching depth of differents AT plates in NaOH.H2O at 166°C

The effect of initial roughness is due to the slow polish effect of the solvent used. Indeed, a high initial roughness corresponds to a thicker disturbed layer and leads to a more irregular bulk surface after the removal of this disturbed layer as shown in figure 8.

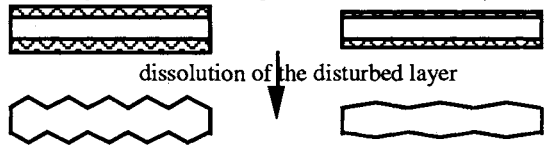


Fig 8 :crystal states under the disturbed layer

The etching depth, to reach the same surface state, must increase with initial roughness. In conclusion, the process of improvement of quartz AT plate roughness is significantly accelerated by a smoother surface state before controlled dissolution.

III.2.3 - Dissolution kinetics of AT quartz plates in NaOH. H2O.

We have plotted in figure 9 the evolution of dissolution rate against temperature.

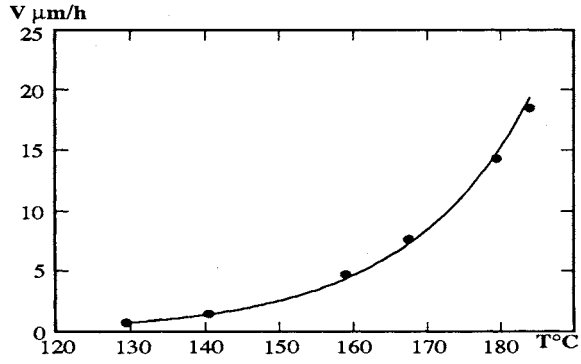


Fig 9 : Etching rate of AT plates in NaOH.H2O

The etching rate being strongly influenced by temperature, this parameter should be carefully checked during the adjustment of the resonator frequency. The exponential variation corresponds to a thermically activated mechanism, following Arrhenius's law $k_T = A. \exp(-Ea/RT)$, k_T being directly related to the etching rate. Ea can be calculated from $\ln(V) = f(1/T)$ with $V = A'. \exp(-Ea/RT)$ (V is the etching rate in µm/h, Ea: the activation energy in kJ, and A': the pre-exponential term including the frequency factor A and the solution concentration) . Indeed, the variation of Ln(V) against 1/T is a straight line (figure 10). This graph determinates the Ea value. Thus, the activation energy corresponding to this dissolution process is $Ea=91.3\ kJ.mole^{-1}$.

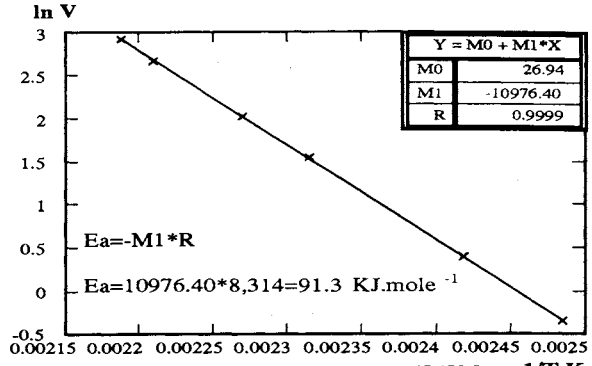


Fig 10 : Ln V=f(1/T) in NaOH.H2O

III.2.4 - Conclusion.

The NaOH. H2O medium is a very good one for a controlled dissolution of AT quartz plates. This dissolution corresponds to a thermically activated mechanism where temperature does not influence the final roughness but only the etching rate.

Therefore, in agreement with the diagonal symmetry around the crystallographic X-axis, the two sides of singly rotated cut (AT) show identical etch figures and relief, and initial surface texture has not influenced the final surface state, but only the rate of roughness parameter evolution.

In this last part, we present some results of AT quartz resonators whose frequency has been adjusted by controlled dissolution in NaOH.H2O.

III.3 - Piezoelectric characterization of resonators made from chemical lapping.

To check the efficiency of the chemical etching process, piezoelectric devices were produced after controlled dissolution of AT quartz plates. In a first step, some high frequency AT resonators were tested to check their Q-factor and the reproducibility of the process. Then, four resonators were chosen to test a four pole pass-band linear phase filter.

III-3-1- Sample preparation.

The objective was to realize 83 MHz resonators after metallization. i.e. 84.5 ± 0.2 MHz before (in fundamental mode). The final thickness of these samples, ~20 µm, required a stiffening of the external part of the plate as shown in figure 11. In practice, mechanical lapping first adjusted the resonator frequency to a value close to 25 MHZ, (~70 µm) and, after setting up a safety device on the external part, a 49 MHz central frequency was then reached by I.B.E. process.


Fig 11 : Well plate

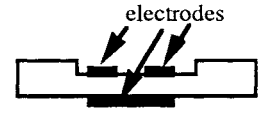


Fig 12 :2 pole resonator

At this step, controlled dissolution of these plates was performed in sodium hydroxide solutions at 170°C with a dissolution rate near 10 $\mu m.h^{-1}$ for both sides of the wafer.

Plates	Time mn	initial and final frequency MHz		etching depth µm
1	72+15+4	49,31	84,56	14,1
2	85	49,33	84,63	14,1
3	84+4,25	49,37	84,63	14,1
4	80+2,25	49,71	84,72	13,9
5	85	49,32	84,57	14,1

Table II:Well plates chemically etched

Table II shows that a reproducible etching rate was obtained and that final frequency could be achieved with one or several etching steps.

III.6.2 - Piezoelectric characterizations.

Two kinds of measurements were used:

- Air-gap method - The air gap method is a technique well adapted to the chemical etching process because sample measurements can be taken without adhesive metallic electrodes. Some results from this method are given in Table II where all the measurements are in the chosen range, 84.5 ± 0.2 MHz.

-Resonator measurements - For two pole resonators, metallic electrodes are deposited on the two faces of the plates by evaporating technique (as shown in figure 12) and connected to a network analyser.

Table III and figure 13 summarize comparative characteristics of resonators obtained from controlled dissolution and I.B.E. processes. For these first results, the less atraumatic chemical etching seems to give, specifications as least as good as that of the ionic beam etching.

-Linear phase filter - In a last step, four pole linear phase filters were made from two chemically lapped resonators and two I.B.E. ones. These filters are characterised by a frequency response which must be chalk up in a gauge (specification) defined by:

-a bandwidth for a maximum insertion loss (IL) allowed: Amax.

-an attenuated band for a minimum insertion loss allowed: Amin.

The characteristics of these filters are given by Fc, central frequency, BW, bandwidth at n dB and BA, attenuated band at n dB. Thus, these linear phase filters are also characterized by a constant propagation time τ (delay) in a part of bandwidth. Filters characteristics and specification are given in table IV, and frequency response in fig 14.

The characteristics here also confirm the attraction of an atraumatic controlled dissolution process (Ref 17). These piezoelectric results show that the surface state reached is sufficient for BAW devices to be made.

IV - CONCLUSION

Previous studies have suggested that controlled crystal growth and crystal dissolution must be governed by the same law (Refs 9-14). For this reason, we have investigated the chemical etching of quartz in the medium used for its crystal growth, i.e. a basic medium, although dissolution conditions are very different from growth ones. Dissolution is a reversible process under certain conditions, whereas a fluoride medium, for example, involves an irreversible chemical reaction.

Thus, in basic solvents, it is a question of "soft" or "controlled" dissolution, whereas in acidic solvents, "hard" etching must result in a rough surface state. It is, therefore, important to distinguish between"controlled dissolution" and "chemical etching".

Following from this, first experiments in sodium hydroxide solvent have demonstrated the merits of our hypothesis for adjusting the frequency of AT quartz resonators. It should be noted, however, that controlled dissolution is being used for other quartz plate orientations such as SC cuts and other mixtures.

Acknowledgements . The authors wish to thank Thomson / C.E.P.E for providing samples and piezoelectric device characterizations.

	resonators	
	IBE	chemically etched
Fr KHz	82873	82870
R Ω	28,59	27,02
L mH	2,894	3,3721
Q	52710	65000
Q.Fr	4,37E12	5,38E12

Table III : Electrical parameters of resonators.

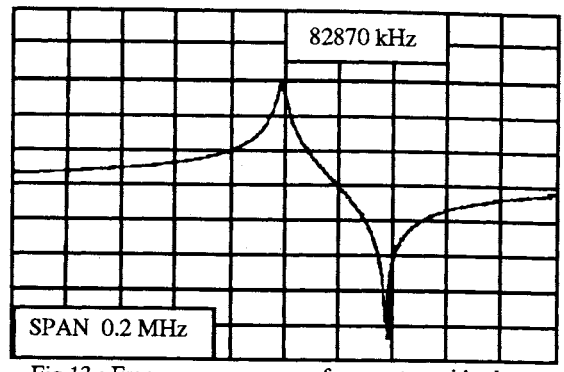


Fig 13 : Frequency response of resonator with plates chemically etched

			Filters	
		Specifications	Chemically etching	IBE
Fc KHz		83000	82986	83005
BW (KHz) à 3 dB		180	200	213
IL dB		5,5	3,9	3,0
BA (KHz)	25 dB	Fc+/-200,0	+175,0-183,8	≈+/-180
	40 dB	Fc+/-285,0	+252,9-252,9	≈+/-260
	50 dB	Fc+/-1000,0	+312,5-306,3	≈+/-330
Out of band attenuation dB		60 db at +/- 30,0 MHz	-87,3+89,0	> 80dB

Table IV : Electrical parameters and specifications of linear phase filters

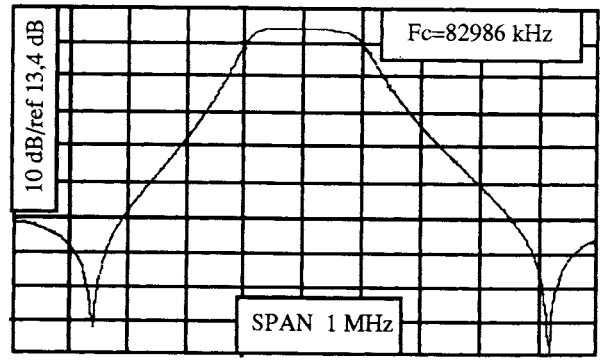


Fig 14 : Linear phase filter manufactured with chemically etched plates.

(1) : K.Bräuer, E.Müller, 1984, Crystal Res & Technol; 101.
(2) : K.H.Jones, 1987, 41st Annual Frequency Control Symposium; 199.
(3) : C.R.Tellier, 1984, Surface Technology; 21, 83.
(4) : A.J.Bernot, 1985, IEEE; 271.
(5) : R.J Brandmayr, J.R.Vig, 1985, 39th Annual Frequency Control Symposium; 273.
(6) : R.J Brandmayr, J.R.Vig, 1986, 40th Annual Frequency Control Symposium; 86.
(7) : O.Cambon, A.Goiffon, A.Ibanez, E.Philippot, 1993, J.solid.state.chem; 102. (to be published).
(8) : O.Cambon, A.Goiffon, A.Ibanez, E.Philippot, 1992, 6th EFTF; 377.
(9) : J Frenkel, 1945, J.Phys.USSR, 9; 392.
(10) : W.K.Burton, N.Cabrera, F.Frank, C 1949, Nature, 163; 398.
(11) : W.K.Burton, N.Cabrera, F.Frank, C 1951, Phil.Trans.Soc.London, 243; 299.
(12) : P.Bennena, G.H.Gilmer, 1972, J.Cryst.Growth, 13; 148.
(13) : W.G.Johnston, 1962, Prog.Ceram.Sci, 2;1.
(14) : N.Cabrera, N.Levine, 1955, Phil.Mag, 1; 450.
(15) : C.R.Tellier, J.L.Vaterkowski, 1985 ,Surface Technology, 15; 275.
(16) : Y.Sekiguchi, H.Funakubo, 1980, J.mater.sci; 15;3066.
(17) : Patent 07/19/1991 Ref 9109170.

A HIGH OVERTONE BAW RESONATOR AS A TOOL FOR ACOUSTIC CHARACTERISATION OF THIN FILMS AND LAYERS

G.D. Mansfeld, A.D. Freik, B.N.Krutov.

Institute of Radioengineering and Electronics,
Russian Academy of Sciences, Mokhovaya,11,Moscow,Russia.

Abstract/Resume

A method for measurement of BAW velocity and attenuation in solid matter with the help of a composite resonator using a layer or a thin film of investigated material was developed. The method is based on measurement of electromagnetic wave reflection of the resonator connected as a load to an electrical transmission line. The method provides an absolute attenuation measurement ranged from parts to hundreds decibel per centimeter with accuracy 10^{-2} and absolute velocity measurement with accuracy 10^{-4} in a wide frequency band.

Keywords: Sound velocity, attenuation, resonant method, composite resonator.

INTRODUCTION.

Wide-spread use of acoustic, acoustoelectron and acoustomagnetic devices, which employs new thin film materials and structures, in modern technologies requires simple and reliable methods for precise determination of velocity and attenuation of acoustic waves in these materials in a wide frequency range.

The present method is version of the general resonant method, adapted for a composite acoustic resonator and known as the method of the composite resonator (Refs. 1-3).

The measurement process is as follows. We build the composite resonator with a layer of the investigated material. As shown in Fig. 1, it consists of a thin (thickness d) parallel plate (l) of piezoelectric that excites the given mode. On one of the sides of the plate a thin film piezoelectric transducer with two metal electrodes (3,4) is deposited. The thicknesses of top and bottom electrodes are b_t and b_b respectively. The composite resonator has multiple resonant frequencies with the resonant peculiarities reveal whenever acoustic thickness of the resonator structure equals to integer number of halfwavelengths. As a result frequency dependence of absolute value of electric impedance of the resonator reveals a series of maxima ("parallel resonances") and minima ("serial resonances") as shown in Fig 2. This resonator is analogous with the Fabrie-Perot resonator in optics. Distance between the successive peculiarities on the module and phase-to-frequency dependencies of impedance of the structure and characteristic width of resonant peaks strongly depends on the acoustic wave velocities and attenuation in the materials that form the resonator and thickness of layers. Hence, it is possible to get information about attenuation and velocity of sound from impedance measurements.

As resonant peculiarities of a multifrequency resonator manifest itself in a wide frequency range, there is the unique opportunity to study sound attenuation and velocity in a wide frequency band. It is shown, that all the necessary information can be found from electromagnetic wave reflection coefficient measurement. The reflection coefficient measurement technique is very good developed. However in general case the relationship between the attenuation and velocity and measured frequency dependence of the reflection coefficient is very complicated and direct determination of these parameters from experimental data runs into serious computation difficulties (Refs.1,3). As it is shown below, it is possible to set some structure parameters and experimental conditions so that direct determination of attenuation and velocity from reflection measurement is possible with high accuracy. To ground the method we derive a simple relationship between halfwidth of the resonant peaks in the impedance vs frequency dependence and the acoustic parameters of the structure. Further we set some conditions providing simple and precise determination of halfwidth (and hence attenuation) from the frequency dependence of phase of the electrical signal reflected from the structure. Finally, some examples of practical utilisation of the method for study of attenuation in various multilayered resonant structures are given.

1. THE ANALYSIS OF ELECTRIC IMPEDANCE OF A COMPOSITE RESONANT STRUCTURE

According to (Ref. 4) there is the following expression for electric impedance of a piezoelectric transducer with two electrodes, a single piezoelectric film and an arbitrary acoustic loads:

$$Z_e = \frac{1}{i\omega C_0}[1 - Z_{sc}] \qquad (1)$$

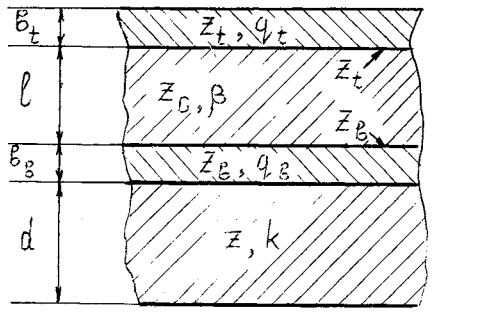


Fig 1.Schematic representation of the resonator structure: 1 - investigated layer, 2 - piezoelectric film, 3,4 - electrodes.

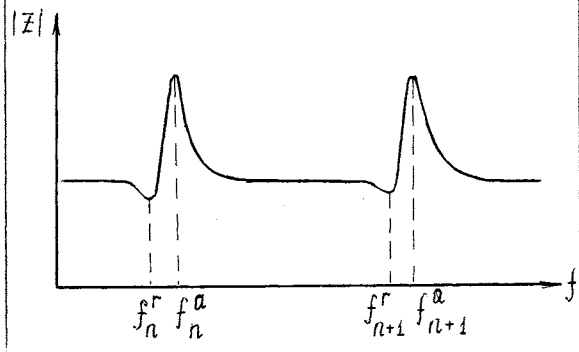


Fig 2. Absolute value of electric impedance of the resonator structure as a function of frequency.

Here C_0 is a transducer capacity at low frequency (or "stiffened" capacity) and Z_{ac} is so called "acoustic impedance":

$$Z_{ac} = \frac{K_t^2}{\beta l} \times \frac{i(z_b + z_t)\sin\beta l - 2(1-\cos\beta l)}{(1+z_b z_t)\sin\beta l - i(z_b + z_t)\cos\beta l}, \qquad (2)$$

$$z_b = \frac{Z_b}{z_{0c}},\ z_t = \frac{Z_t}{z_{0c}},$$

where K_t is coupling factor of the piezoelectric material used in the transducer, Z_b and Z_t - mechanical impedances of the top and bottom loads, z_{0c} is specific impedance of the piezoelectric film, β is complex wave number of the given mode in the transducer, l - its thickness.

If top and bottom loads are not matched loads for transducer, i.e. $z_b \neq 1$ and $z_t \neq 1$, we can represent Z_{ac} as composition of two factors, the first of which is a smooth function of frequency and the second determines resonant

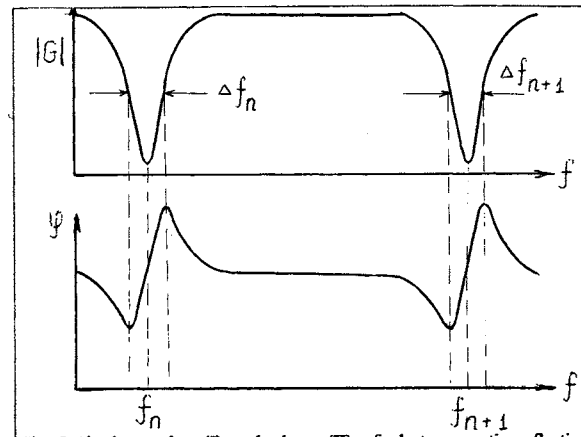


Fig 3.Absolute value (I) and phase (II) of electromagnetic reflection coefficient as a function of frequency.

properties of the structure:

$$Z_{ac} = SP,$$

$$S = 2i\frac{K_t^2}{\beta l}\sqrt{\frac{(z_b + z_t - 2)(z_b + z_t + 2)}{(1-z_b)(1+z_b)(1-z_t)(1+z_t)}} \times \sin\frac{\beta l}{2}\sin\left[\frac{\beta l}{2} - \arctan i\frac{z_b + z_t}{2}\right]$$

$$P = \csc[\beta l - \arctan iz_b - \arctan iz_t].$$

Further we are interested only in the resonant factor P. Argument of csc consists of the three terms, the first of which is phase difference of acoustic wave at the surfaces of the plate and the last two are the phase shifts due to reflections from the corresponding loads. It has clear physical meaning: resonant peculiarities occurs whenever real part of total phase increment is multiple of π. Note that contributions of transducer and loads in the total phase are additive.

The mechanical impedance of a nonpiezoelectric layer with an arbitrary load :

$$Z_{in} = z_0\frac{Z_l\cos kd + iz_0\sin kd}{z_0\cos kd + iZ_l\sin kd} = iz_0\tan\left[kd - \arctan i\frac{Z_l}{z_0}\right]$$

Here z_o is specific impedance of a layer material, Z_l is mechanical impedance of a load, k is complex wave number and d is thickness of a layer. Argument of tan also consists of phase difference of acoustic wave in a layer and reflection phase increment.

Applying the above formula to z_b and z_t, we get:

$$z_b = i\frac{z_{0b}}{z_{0c}}\tan\left[q_b b_b - \arctan\left[i\frac{z_0}{z_{0b}}\tan kd\right]\right] z_t = i\frac{z_{0t}}{z_{0c}}\tan q_t b_t$$

where z_{0c}, z_{0t}, z_{0b} and z_0 are specific impedances of the piezoelectric layer, the top and bottom electrodes and the

investigated layer respectively, b_t, b_b and d are thicknesses of electrodes and the layer. Wave numbers in the transducer β, top and bottom electrodes q_t and q_b and the investigated layer k contain imaginary components which correspond to the attenuation in these layers:

$$\beta = \beta' + i\beta'', \quad q_t = q_t' + iq_t'', \quad q_b = q_b' + iq_b'', \quad k = k' + i\alpha$$

$$\beta' = \frac{\omega}{v_c}, \quad q_t' = \frac{\omega}{v_t}, \quad q_b' = \frac{\omega}{v_b}, \quad k' = \frac{\omega}{v_s}$$

with total attenuation being assumed quite small:

$$\left|\beta'' l + k'' d + q_b'' b_b + q_t'' b_t\right| << 1.$$

Further we consider the electrodes being quite thin:

$$\left|q_t b_t\right| << 1, \quad \left|q_b b_b\right| << 1. \tag{3}$$

Then, confining to the linear terms of these small parameters (3), the P can be represented in a simple and convenient for further analysis form:

$$P = \frac{\csc\left[\beta l + \gamma_{tc} q_t b_t + \arctan\left[\gamma_c \tan kd\right] + \gamma_1 q_b b_b\right]}{\csc x} = , \tag{4}$$

$$\gamma_1 = \gamma_{bc} \frac{1 + (\gamma_b \tan kd)^2}{1 + (\gamma_c \tan kd)^2},$$

$$\gamma_{tc} = \frac{z_{0t}}{z_{0c}}, \quad \gamma_{bc} = \frac{z_{0b}}{z_{0c}}, \quad \gamma_c = \frac{z_0}{z_{0c}}, \quad \gamma_b = \frac{z_0}{z_{0b}}$$

As imaginary part of k is small, γ_1 can be treated as real positive coefficient.

The antiresonance frequencies are determined by an equation $\mathrm{Re}\, x_a^n = \pi n$, or

$$\mathrm{Re}\left[\arctan\left(\gamma_c \tan kd\right)\right] + \beta_n' l + \gamma_{tc} q_{tn}' b_t + \gamma_1 q_{bn}' b_b = \pi n$$

where n - an integer. For small α

$$\mathrm{Re}\left[\arctan\left(\gamma_c \tan kd\right)\right] \approx \arctan\left(\gamma_c \tan k'd\right)$$

The calculation shows that the differences between results concerning the location and shape of the antiresonant peaks obtained with the expressions (2) and (4) become considerable when relative width of the metal layers $\nu = (b_t + b_b)/(l + d)$ is greater than 0.1 (with error in frequency being about 2 %). In the practical realisations of the resonant structures ν may easily be made less than 10^{-3} and so the approximate expression (4) can be considered as strict.

2. ACOUSTIC CHARACTERISATION OF THIN LAYERS BY THE PECULIARITIES OF THE ELECTROMAGNETIC REFLECTION COEFFICIENT

Using the expression (4) one can easily derive the relationship for determination of sound velocity v_s in material investigated from the measured frequency difference between two successive antiresonances f_n, f_{n+1}. For the thin transducer ($l/d << 0.1$), we have

$$v_s = \frac{d}{\frac{1}{2(f_{n+1} - f_n)} \cdot K \cos^2 K\omega_n \left(\gamma_c + \frac{1}{\gamma_c} \tan^2 K\omega_n\right)} \tag{5}$$

$$K = \frac{l}{v_c} + \gamma_{tc} \frac{b_t}{v_t} + \gamma_1 \frac{b_b}{v_b}$$

To set the relationship between the form of the resonant peak and attenuation the expand to the series the expression (4) for Z_{ac} on the small term $\left(x - x_n^a\right)$ at the antiresonance points x_n^a. The first term is

$$Z_{ac} = S_n^a \frac{\mathrm{Re}\left(x - x_n^a\right) - i\,\mathrm{Im}\left(x_n^a\right)}{\mathrm{Re}^2\left(x - x_n^a\right) + \mathrm{Im}^2\left(x_n^a\right)}. \tag{6}$$

Similarly, admittance of the structure at the same approximation in the vicinity of the resonance point x_n^r resembles the expression (6) for Z_{ac} in the vicinity of antiresonance

$$Y_c = i\omega C_0 S_n^r \frac{\mathrm{Re}\left(x - x_n^r\right) - i\,\mathrm{Im}\left(x_n^r\right)}{\mathrm{Re}^2\left(x - x_n^r\right) + \mathrm{Im}^2\left(x_n^r\right)} \tag{7}$$

The analysis of (6) shows that there are two important relations for the bandwidth limits $x_{0.7}^a$ of amplitude of impedance Z_{ac} at antiresonance x_n^a and for the bandwidth limits $x_{0.7}^r$ of amplitude of admittance Y_c at resonance x_n^r:

$$\pm\,\mathrm{Im}\, x_n^a = \mathrm{Re}\left(x_{0.7}^a - x_n^a\right) = \mathrm{Re}\left(x_{0.7}^r - x_n^r\right) = \pm\,\mathrm{Im}\, x_n^r.$$

$\mathrm{Im}\, x$ corresponds to attenuation in all the layers of the structure:

$$\mathrm{Im}\, x = K_1 \alpha d + \beta'' l + \gamma_{tc} q_t'' b_t + \gamma_1 q_b'' b_b \tag{8}$$

$$K_1 = \frac{\gamma_c \left(1 + \tan^2 k'd\right)}{1 + \gamma_c^2 \tan^2 k'd}$$

Therefore, according to (6-8), the bandwidth of the

resonant impedance and admittance characters Δf_n at antiresonance characterise total attenuation in the composite structure

$$\pi\Delta f_n = \frac{K_1\alpha d + \beta'' l + \gamma_w\alpha''_t b_t + \gamma_1 q''_b b_b}{K_2\frac{d}{v_s} + \gamma_w\frac{b_t}{v_t} + \gamma_1\frac{b_b}{v_b} + \frac{l}{v_c}} \qquad (9)$$

$$K_2 = \frac{\gamma_c(\sec kd)^2}{1+\gamma_c^2(\tan kd)^2}$$

If attenuation in the layers and their thickness are such that all the terms except first in (8) can be neglected, from (9) we get the simple expression for the attenuation coefficient α in investigated material :

$$\alpha(f_n) = \frac{K_2}{K_1}\frac{\pi\Delta f_n}{v_s}. \qquad (10)$$

The distance Δf_n can be easily measured by a standard circuit analyser and it is this circumstance that makes this attenuation measurement method very convenient.

The direct method for determination of α and v_s is based on virtual determination of frequency dependence of electrical impedance Z_e that contains an information about these parameters. Determination of Z_e needs precise measurement of reflection coefficient G from investigated structure and following calculation with known formula:

$$Z_e = Z_0\frac{1+G}{1-G},$$

where Z_0 is specific impedance of electrical transmission line. Further procedure of determination α and v_s needs complicated calculations with formulas (1 - 2).

However under some special conditions we can determine α and v_s without determination of Z_e. All the necessary information we can get directly from form of amplitude and phase-to-frequency dependence of signal reflected from structure. Analysis of frequency dependence of Z_e and G carried by circle charts shows that under following relations between specific impedance of electrical transmission line Z_0 and impedance amplitude of the structure at antiresonance point $|Z_e|_n^a$:

$$Z_0\delta << |Z_e|_n^a < \frac{Z_0}{2} \qquad (11)$$

$$2Z_0 < |Z_e|_n^a << \frac{Z_0}{\delta} \qquad (12)$$

where $\delta = |\Delta G/G|$ - device resolution of G, resonant peculiarities of G are close to antiresonance and resonance of Z_e respectively where approximations (6) or (7) can be used. In this case amplitude and phase-to-frequency dependence of reflection coefficient has special form, shown in Fig. 3. Phase has two extrema at each peculiarity. Frequency difference between these extrema equals exactly to halfwidth of impedance Z_e at antiresonance or admittance Y_e at resonance with relative error in halfwidth with respect to (10) being equal to

$$\sigma_\alpha = \frac{1}{3}\left|\frac{Z_e}{Z_0}\right|^4 \qquad (13)$$

in case of (11) and

$$\sigma_\alpha = \frac{1}{3}\left|\frac{Z_0}{Z_e}\right|^4 \qquad (14)$$

in case of (12). It is convenient for approximate calculations to consider Z_e at antiresonance to be of the order of $K_t^2/\omega C_0\alpha l$. By varying dimensions of structure elements (in particular, thickness and area of transducer) it is easy to reach an accuracy of determination of half width and hence attenuation coefficient α of the order of 10^{-2}. Relative error of sound velocity determination for both cases (11) and (12) is usually less than 10^{-4}.

Among advantages of the method presented we notify wide range of attenuation measurement in the same structure: from 10^{-2} to 200 dB/cm according to our estimations.

3. EXPERIMENTAL TEST OF THE METHOD.

Scheme of the experimental installation for acoustic wave attenuation and velocity measurement from electric response of composite resonator is shown in Fig. 4. It includes high stability and precision generator and vector analyser that measures ratio of signals in reference and measurement channels. The necessary measurements can also be carried out by standard circuit analyser. Microvawe transformer can be used to fit in experimentally with the conditions (11), (12).

First example illustrates high attenuation measurement

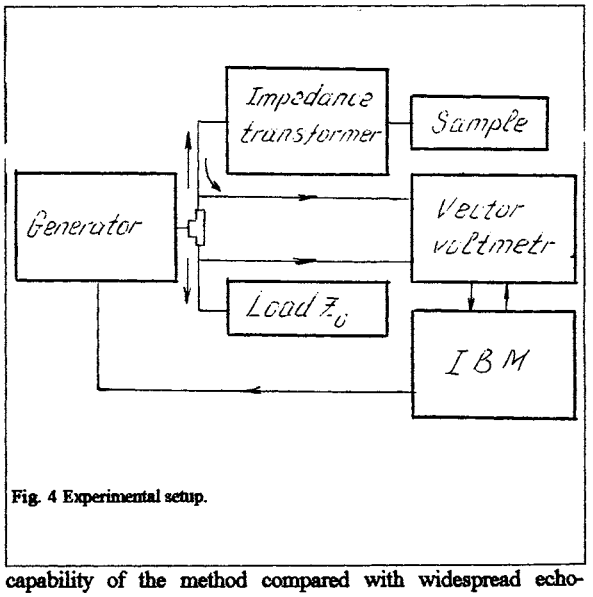


Fig. 4 Experimental setup.

capability of the method compared with widespread echo-

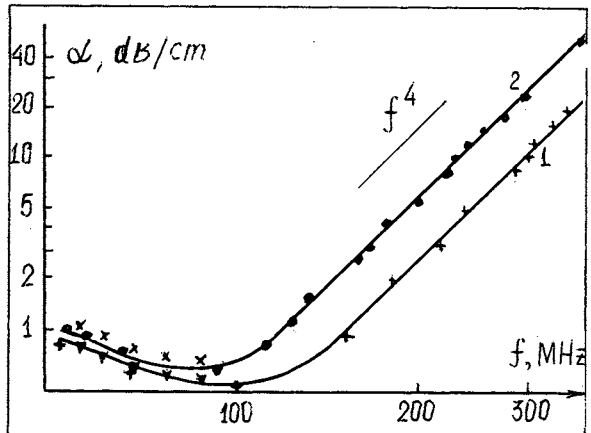


Fig. 5. An example of the experimental frequency dependences of the attenuation coefficient of longitudinal acoustic wave in Ni - Co polycrystalline ferrite. Below ~ 100 MHz the attenuation is governed by diffraction , at high frequencies - by Rayleigh scattering ($\alpha \sim f^4$). Attenuation depends on pores' concentration n and their mean diameter $\bar{d}$.

1. $n \cong 8\,10^{10}$ cm^{-3}, $\bar{d} \cong 1.2$ μm **(triangles for echo-method)**

2. $n \cong 5\,10^{10}$ cm^{-3}, $\bar{d} \cong 1.5$ μm **(crosses for echo-method)**

method (Ref. 5). Experimental conditions were set to satisfy relation (11). Investigated samples were nickel-cobalt spinel. Average grain and pores dimensions were 1.2 and 1.5 μm respectively and pore concentration 8 10^{10} cm^{-3} and 5 10^{10} cm^{-3} for samples 1 and 2 respectively. Samples were thin parallel plates (nonparallelizm of the order of 10″). On one of the sides a transducer was deposited consisting of 8 μm ZnO film and 1000 A aluminium electrodes that excites longitudinal acoustic waves.

It is clear *a priori* that attenuation contribution from ZnO film is less than contribution from ferrite (approximately 3 orders lower). Therefore data on attenuation in structure corresponds to acoustic losses in ferrite with high accuracy. Attenuation coefficient corresponding to nonmagnetic losses (Ref. 6) is shown in Fig. 5. It was measured in wide frequency range on the same sample. Forth power of attenuation frequency dependence is clearly seen and points on Rayleigh scattering mechanism (Ref. 5). Error of measurement is governed by (13) and less than 3%. We show echo-method experimental data (Ref. 5) on the same plot by crosses and triangles. It is seen good correspondence between data from both methods but it appears very difficult to overcome 100 MHz barrier with echo method. Using composite resonator method, the important conclusion on the real nature of the nonmagnetic losses in polycrystalline ferrite was made. It was rigorously proved that them attenuation is mainly due to Rayleigh scattering on pores.

The method can also be used for attenuation measurement in material with small acoustic losses i.e. when condition (12) is fulfilled. Frequency dependence of attenuation coefficient in composite resonant structure, consisting of thin layer of monocrystalline yttrium aluminium garnet (nonparallelizm of the order of 2″, [100] orientation), with small attenuation and ZnO transducer (0.7 μm ZnO film, 1000 A aluminium electrodes) is presented in Fig. 6. According to (9), bandwidth Δf_n determines attenuation in all layers of the structure. We can see that frequency dependence of total attenuation in structure is of quite complex character. In

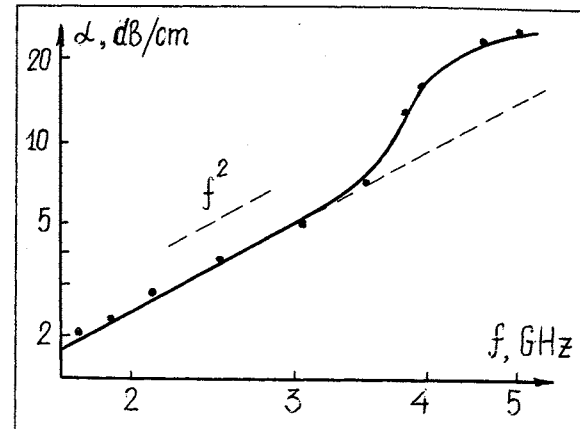


Fig. 6. An example of the experimental frequency dependence of the attenuation coefficient of longitudinal acoustic wave in resonator structure ZnO - YAG (0.7 μm ZnO film, [100] orientation YAG parallel plate with nonparallelizm 2″). At low frequency the attenuation is governed by Akhiezer mechanism.

low frequency range dependence is of square character, which testify to Akhiezer attenuation (Ref. 5). Dominant contribution in acoustic losses at these frequencies, as it can be seen from simple estimations, under selected layers' thickness, makes YAG and ZnO due to Akhiezer mechanism (Ref. 7). At higher frequencies increase of the total attenuation and then saturation are observed. Mechanism of attenuation is still not proved. Error in Δf_n according to (14) is 1 - 3 %.

Both examples illustrate possibility of wide frequency band measurements of one sample.

CONCLUSION.

We presented and tested simple method for acoustic wave attenuation and velocity measurement in layers of composite resonant structures. Method can be used for investigation of materials with both high and low losses.

Advantages of the method are simplicity, that comes from employing only relative measurement of frequencies at which there are peculiarities at phase characters, and opportunity to measure the same sample in wide frequency range. Unlike methods known before (Ref. 1-3) the present method does not require calibration procedure and complex calculations.

Authors thank I.M.Koteluansky for manufacturing of transducers for investigated samples and S.N.Ivanov for support this paper and valuable discussions.

References.

1.V.P.Demidov, A.A.Smirnov. Attenuation Of Longitudinal Waves In CdS Texture Films. Akusticheskii zhurnal,1973.v.19.#3.pp.339-344.

2.K.Kushida, H.Takeuchi. Piezoelectricity of c-axis oriented $PbTiO_3$ thin films. App.Phys.Lett. 50(25),pp.1800-1802.

3.E.Benes. Impruved quartz crystal microbalance technique.J.App.Phys.56(3),pp.608-626.

4.G.S.Kino. Acoustic waves: Devices, imaging, and analog signal processing. Prentice-Hall, Inc., Englewood Cliffs, NJ.

5.R.Truell, Ch.Elbaum, B.B. Chick. Ultrasonics Method In Solid State Physics. Academic Press,New York and London,1969.

6.A.F.Kabychenkov, G.D.Mansfeld, V.N.Nagirniak, A.A.Rubtsov. Peculiarities Of Attenuation Of Acoustic Waves In Polycrystalline FYG. Fizika tverdogo tela,1991,v.33,#3,pp.675-680.

7.A.I.Morozov, V.V.Proklov, B.A.Stankovskii. Piezoelectric Transducers For Radioelectron Devices. "Radio i sviaz'", Moscow, 1981.

Frequency control components and materials

THE TEMPERATURE VARIATION ON THE SURFACE AND ALONG THE CENTERLINE OF AUTOCLAVE DURING QUARTZ CRYSTALS GROWTH

Władysław Hofman

Institute of Electronic Materials Technology

01-919 Warszawa, POLAND

When quartz crystals are grown at the constant testing temperatures a gradual increase of solution pressure and a small decrease of heating power, required to maintain these temperatures,are observed. The above changes have to alter the temperature profiles of the system, especially near the vessel wall. This paper presents the experimental results of accurate temperature profiles measurements in a subsurface layer of the autoclave wall,i.e.along its height and around its periphe ry and the profiles along the center axis of the system at some stages of the hydrothermal growth of quartz crystals.

The hydrothermal synthesis was conducted in 8 litre vessel at the constant temperatures: 375° C under the bottom and 350° C under the cover of the unit and at a initial pressure of solution equal 1400 bar. The results show that small temperature fluctuations of short duration exist at a given point at the centerline and slow changes of the temperature, associated with heaters work, are present on the vessel wall. In spite of axisymmetrical heaters arrangement the local temperature deviations from axial symmetry were recorded and the asymmetry was greater for the dissolution zone. The small peaks of temperature, correlated with heaters positions, and the single deep minimum of temperature, between the lower and the upper chamber, on the vessel wall were also detected. As the crystals were grown the appreciable changes of axial temperature gradient were created in both the disolution and supersaturation chambers of the reactor.

A SEMI-QUANTITATIVE MODEL OF THE DYNAMIC BEHAVIOUR OF SPACE USOS UNDER FRAGMENTED AND CONTINUOUS IRRADIATIONS

J. Bourrieau (*)
M. Brunet (**)
JF. Dutrey(**)
B. Laporte (**)

(*) CERT - ONERA/DERTS
2 avenue E.BELIN, 31055 Toulouse, France.

(**) CNES - RA/TE/AE/IT/TF
18 avenue E. BELIN, 31055 Toulouse, France.

Abstract

On board satellites the Ultra Stable Oscillators (USO) are exposed to the ionizing radiations encountered in the Earth's radiation belts. Results obtained in lab show that the induced frequency drifts often exceed the limits specified for the mission.
In order to simulate the exposures on Low Earth Orbit (LEO) and at geostationary altitudes (GEO), CERT - ONERA/DERTS and CNES performed on height USOs a study of the frequency drifts induced during fragmented (as on LEO) and continuous irradiations (as on GEO) by gamma rays (Cobalt 60, dose rate 0.01 gray per hour). The radiation induced frequency drifts observed on all the USOs, during continuous and fragmented exposures, can be seen as the sum of two components. The one presents a saturation with a fast annealing (time constant between part of day and some days) at the end of the exposure ; on the other, with or without saturation occuring in the dose range studied (up to 10 grays), a long term (some months) annealing is probable but cannot be clearly evidenced until today.
The shape of the radiation response depends on the type of the resonator (BVA or QAS) and on the nature of the quartz (swept or unswept). A radiation damage model is proposed and physical processes involved are suggested.

Keywords: Oscillators/Frequency drifts/ Radiation/Modelling/Space

INTRODUCTION

This study was performed at CERT-ONERA/DERTS under CNES contract, with the technical support and the electronic facility furnished by CNES/RA/TE/AE/TT/TF.

The radiation sources in space

The primary source of energetic particles are the Earth's radiation belt (Ref.1), the Galactic Cosmic Rays (GCR) and the sporadic emissions of energetic electrons, protons and heavier ions during solar events. Considering the total dose effects on USOs, GCR and solar flux of heavy ions can be neglected.
Due to the structure of the radiation belt and to the effectiveness of the magnetospheric shielding against transiting solar protons, the hazard depends mainly on the orbit (Ref. 2).
On GEO sporadic exposure to solar protons (about ten events per year during solar maximum period) and continuous irradiation by trapped electrons and secondary bremsstrahlung can be expected. Taking into account the attenuation by the structures of the satellite the main part of the dose comes generally from bremsstrahlung and the exposure can be considered as continuous. Depending on the structural shielding, doses ranging between 10 to 100 grays (1 and 10

krads) can be expected inside the vehicule during a ten year mission. Less than 10% of this dose, due to energetic protons, can be induced in some days, during solar events.
On LEO, the offset and tilt of the geomagnetic field axis, referred to the rotation axis, bring the belt to low altitudes upon the South Atlantic. In this region (the South Atlantic Anomaly, SAA) high proton and electron flux are observed, as well as high electron population in regions corresponding to the polar horns of the radiation belt. The magnetospheric shielding effectiveness bringing down at high latitude, high flux of transiting protons can be expected on polar regions during solar events.
Inside the satellite, taking into account the protection by the structures, a periodic and discontinuous exposure during some tens of minutes (Ref. 3), can be expected in the SAA (dose rate about 1 to 10 grays per hour induced by protons) and in the polar horns (dose rate about 0.01 to 1 mgray per hour due to bremsstrahlung) ; sporadic large increases of the dose rate (up to about some tenths of grays per hour) can be observed at high latitude during large solar proton events.

Objective of the study

This work was performed in order to study the USOs frequency drifts induced by the radiations during LEO and GEO missions. As seen above, for a realistic simulation of the conditions encountered, continuous and fragmented irradiations must be carried out.
In order to evaluate the oscillator sensitivity variations due to the quartz quality, the resonator technology and conditionning by pre-irradiation, height USOs are ordered by CNES to CEPE ; their characteristics are summarized table 1.
Four USOs (M1/T1, M2/T2, M3/T3 and M4/T4) are specially manufactured with deported electronics, external lead collimators limit the irradiation to the resonator. In these oscillators, to measure transient temperature change during irradiations, a thermometric quartz (B mode, sensitivity about 10^{-4} K) was associated to the metrologic quartz (C mode).
The four other USOs, characterized during several years by CNES, have been manufactured by CEPE with BVA and QAS resonators (swept and unswept pre-irradiated quartz). Two of them (E67360 and 365) have been submitted to complementary irradiation up to 200 grays (20 krads) three months before the begining of this study.

Experimental conditions

The electronic facility assuming the sequential measurements of the USO frequencies, as well as of the temperature and humidity in the irradiation room, was developed by CNES around an HP9816 controler (figure 1) for the datation, the data handling and storage (Ref. 4).
Preliminary measurements of the frequency drifts have been performed during several weeks in the irradiation facility in order to evaluate the intrinsic drift (aging) of the various USOs before exposure.
The oscillators, with the flat surface of the resonator assembly normaly oriented in front of the gamma ray beam, have been irradiated at room temperature and atmospheric pressure in the DERTS facility (see figure 1) with a Cobalt 60 source.
Three series of tests have been performed (figure 2) at a dose rate equal to 0.01 gray (1 rad) per hour, a continuous exposure up to 10 grays, and fragmented irradiations (beam ON and OFF during respectively 40 and 72 minutes) before and after the continuous.

RADIATION INDUCED DRIFTS

Measurement results

The experimental results obtained are summarized on the figures 3 and 4 showing total drifts measured during fragmented and continuous irradiations.
Excepted for the USO M1, the intrinsinc drift seems to be relatively easely extracted in order to obtain the radiation induced response. The first features to be pointed out are the USO behaviour diversity and the fast dynamical response (saturation and annealing) evidenced during fragmented irradiations ; as transient temperature changes are not observed on the thermometric modes, this effect cannot be put forward.
The total radiation induced drifts should be interpreted as the sum of two components:

$$Y(t) = dF/f = y_1(t) + y_2(t)$$

The one, $y_1(t)$, is generally weak ; the USOs responses to fragmented exposure, as well as the initial drifts observed during the continuous irradiation, indicate a rapid saturation and annealing of this transient component.
The other, $y_2(t)$, dominates mainly during continuous exposures ; depending on the USO, saturation of the drift linked to this second

process can be evidenced. If annealing occures, its time constant is larger than some months and its effectiveness can be neglected on the period considered.
Assuming that the amplitude and the sign of the drifts induced by both processes depend on the USO, the competition between the components explain the various responses observed.
For M2, and probably also M1, only y_1 is evidenced and the response to the continuous irradiation can be obtained (see here after) by combining saturation and annealing.
For M3 saturation occurs on both components during the continuous exposure ; y_2 prevails (about a factor 10 at saturation) and induces a negative drift ; y_1 produces a positive frequency change. The response to the fragmented exposures, weakly negative initially, becomes largely positive after the saturation of y_2 during the continuous irradiation and annealing of y_1.
For M4 the sign of both drifts are changed and saturation of y_2 is not observed ; during the continuous exposure the main part of the drift, about 99%, is induced by this component (positive shift). The transient component (negative drift) prevails only at the begining of the continuous exposure (saturation in less than one day) and during the fragmented ; its annealing in some hours appears at the end of each type of irradiation.
The behaviours of the E67360 and 361 USOs are similar to the preceeding, with a reenforced y_2 component (factor 4 and 10 respectively) ; the saturation and annealing of y_1 is observed during the fragmented exposures.
For E67363 and 365 the response of both oscillators are equivalent ; the two processes induce positives drifts and a saturation tendancy of y_2 is observed, less than 10% of the total drift is linked with the transient process during the 1 krad exposure.

Radiation response modelling

Considering one degradation process linearly linked with a saturable primary effect, then the damage variation should be obtained from :

$$dy = [c-\alpha.y(D)].dD$$

with, during the exposure between t=0 and $t=T_m$:

$$D = \beta.t$$

Then the slope of the degradation is given by :

$$y'(t)=k.\exp(-a.t) \qquad (1)$$

Annealing produces an exponential decrease of the damage at the end of the exposure (see for instance the results obtained with M1, M2 and M3), or considering the instantaneous part annealed :

$$dy_{ann} = b.y(t).dt \qquad (2)$$

Taking into account both, saturation and annealing, and assuming that primary effect is not actually restaured, then from 1 and 2 :

$$y'(t) = k.\exp(-a.t)-b.y(t) \qquad (3)$$

Solution of 3 gives the variation of the degradation during the exposure ($0<t\leq T_m$) :

$$y(t)= K.[\exp(-b.t)-\exp(-a.t)] \quad (4)$$

After the end of the exposure the damage decrease is then given by :

$$A(t)=K.\exp(-b.t).\{1-\exp[(b-a).T_m]\} \quad (5)$$

Permanent process

Annealing of y_2 is not observed at the end of the exposures , the relation 4 becomes :

$$y(t)=K.[1-\exp(-a.t)] \quad (6)$$

Saturation is not evidenced up to 10 grays on the USOs E67360, 361 and M4 ; then, with $a.t<<1$ in the range, the relation 6 becomes :

$$y(t)=k'.t=k''.D$$

with :

- for USO E67360 : $k'=1.08.10^{-9}$ $(day)^{-1}$ and $k''=4.5.10^{-9}$ $(gray)^{-1}$
- for USO E67361 : $k'=2.6.10^{-9}$ $(day)^{-1}$ and $k''=1.08.10^{-8}$ $(gray)^{-1}$
- for USO M4 : $k'=2.4.10^{-10}$ $(day)^{-1}$ and $k''=10^{-9}$ $(gray)^{-1}$

Saturation occurs on the USOs M3, E67363 (figure 5) and E67365, the drift during the continuous exposure is then obtained from the relation 6 with :

- for USO E67363 : $K=6.65.10^{-9}$ $(day)^{-1}$ or $2.8.10^{-8}$ $(gray)^{-1}$ and $a=0.04$ $(day)^{-1}$
- for USO E67365 : $K=5.9.10^{-9}$ $(day)^{-1}$ or $2.5.10^{-8}$ $(gray)^{-1}$ and $a=0.045$ $(day)^{-1}$
- for USO M3: $K=-2.1.10^{-9}$ $(day)^{-1}$ or $-8.7.10^{-9}$ $(gray)^{-1}$ and $a=0.11$ $(day)^{-1}$

Transient component

Typical cases are the radiation induced damages observed on M2 and probably M1 ; on these oscillators the transient process is the

only one evidenced, saturation and annealing occur and the complete relations 4 and 5 must be used.
For M2 the time constant of the annealing is equal to 3.1 days (b=0,32) ; the best fit of the transient response (see figure 5) during continuous irradiation is given by:
$y_1(t)=-1.88.10^{-9}.[\exp(-0.32.t)-\exp(-0.03.t)]$

This relation defines an initial slope per gray equal to $-2.25.10^{-9}$, to be compared to the value obtained during fragmented exposures (about $-1.5.10^{-9}$ per gray).
For M3 after saturation of y2 during the 10 grays exposure, only the transient component $y_1(t)$ inducing positive frequency changes appears during fragmented irradiations. It represent about 10% of the total drift observed during the continuous exposure ; $y_1(t)$ is computed from the difference between the measurements, Y(t), and $y_2(t)$ defined above. Its shape is given figure 5 showing its exponential decrease (time constant about 1 day) and its saturation (level : 2.10^{-10}) in less than 10 days.
For M4 (figure 3), $y_1(t)$ induces negative drifts with an initial slope per gray equal to $-1.5.10^{-8}$; during continuous exposure its saturation at about -8.10^{-11} occures in less than 10 hours, and its annealing time constant is of the order of one hour (figure 4) .
Defined as the difference between the measurements and the model of the permanent component, the transient effect observed on the USO 67363 is shown in the figure 5. Positive, $y_1(t)$ saturates rapidly (in less than ten days at a level about 3.10^{-10}) during continuous exposure, its annealing time constant is about 0.7 day. Similar results could be obtained for the USO 67365.
On the USOs 67360 and 361 the transient component is very weak ; nevertheless $y_1(t)$ appears clearly during the fragmented exposures (figure 3). It induces negative frequency changes, with a saturation in some minutes at levels -1.5 and $-0.7.10^{-11}$ respectively. A fast and complet annealing (time constant <10 minutes) is observed between each irradiation step.

Limits of the modelling

Due to limitations of the storage unit the frequency measurements have been performed each 3 and 60 minutes during respectively the fragmented and continuous irradiations. Consequently the exact drifts induced by the transient component at the beginning of the 10 grays exposure is often difficult to know.
The total frequency changes measured are the sum of the aging and of two radiation induced effects ; the accuracy of the evalution and modelling of the last depend mainly of the characteristics of the first.
If aging drifts are weak and uniform they can be obtained by interpolation during the irradiation ; high dose rate, rather than the 10 mgy per hour used in this study, could reduce the incertainties observed on the evaluation of the permanent component. A possible annealing of $y_1(t)$ (time constant : several months), difficult to separate from the aging on long term periods, is not taken into account in the model.
Generally the saturation level and the annealing time constant of the transient component is preciselly obtained from the fast frequency changes at the end of the continuous exposure. In contrast, the uncertainty on the initial shape of $y_1(t)$, defined as the difference between the total drift and $y_2(t)$, is too large and its modelling doubtfull.

Technological impact on the USO sensitivity

The table 2 summarizes the results obtained on the various USOs studied.
Excepted M4 all the oscillators with swept quartz show complet or partial saturation of the component y_2 during the continuous irradiation up to 10 grays.
The permanent component is more than ten times larger on the pre-irradiated USOs (BVA) 67360 and 361 manufactured with unswept quartz, confirming the reduction of the radiation susceptibility by the quartz sweeping.
The resonators of the USOs 67363 and 365 have been manufactured with quartz irradiated in 1987 (648 and 848 grays respectively), in addition the second oscillator was exposed to 200 grays in December 1991. The partial saturation of the permanent component observed in less than 10 grays on both oscillators should indicate that the pre-irradiation conditionning of the quartz is uneffective and that a long term annealing should be possible.

CONCLUSION

In order to determine the behaviour of USOs exposed to the radiations of the space environment, irradiations have been carried out at low dose rate (10mGy per hour) with a Cobalt 60 source.
The simulations performed reproduce the transient and continuous exposures that can be suffered respectively during missions on LEO and GEO.
The saturation effects often observed show that for missions on GEO, during continuous exposure, the general tendancy is a long term reduction of the radiation response.
The rapid annealing of a part of the damage seen on various USOs indicates that only an uncomplete long term reduction, and in some cases an increase, of the sensitivity can be expected when discontinuous exposure occurs, on LEO for instance.
Dynamic models must be developed in order to forecast the USO behaviour in the various exposure conditions encountered in Space ; the results and the preliminary model furnished by this study show that the drifts observed have two origins, one inducing transient (rapidly saturated and annealed) changes of the frequency, the other producing a permanent (may be with a very long term annealing) damage.
The amplitude, sign and time constants of the phenomena associated to both components explain the various signatures observed here as well as those reported by J.R. Norton (Ref. 5).
In order to improve the modelling of the radiation induced drifts, a long term evaluation of the aging is in progress and fragmented irradiations will be performed with adapted doses rates, total doses and annealing periods.
In another way, the physical processes involved in the radiation induced frequency drifts are not well identified ; studies are in progress at DERTS on two possible origins, the one concerns charge deposition in the quartz, the other its dimensionnal change under irradiation.
As seen in this study from the measurements performed on the thermometric quartz, the transient component cannot be linked with a thermal effect ; the annealing time constants obtained, between some hours and some days, are compatible with charge relaxation period as observed during preliminary experiment on quartz samples.
A process, which could produce a permanent effect, is the radiation induced expansion of the quartz (Ref. 6). These dimensionnal changes have been observed on various materials (Refs. 7-8) as variing as the square root of the absorbed dose (ionisation effect) ; the mechanical stress induced and its relaxation, depending on the resonator processing, could explain the various behaviour observed.

REFERENCES

1. J.P. Philippon "Evaluation Théorique et Calcul des Flux de Particules dans la Zone Stable de la Magnétosphère", Thèse d'Université n° 258, Toulouse, March 1970.

2. J. Bourrieau"L'Environnement Spatial en Radiations Ionisantes", Conférences, Journées de Formation, RADECS 89, Publ. CEM, UST, Sept. 1989.

3. E.G. Stassinopoulos, J.M. Barth, R.L. Smith
"Metsat Charged Particle Environment Study, (Revised Edition) Method 2", Publ.NASA X600-87-11, August 1987.

4. J. Bourrieau, M. Brunet, JB. Laporte "Dérives de Fréquence Observées sur des OUS Soumis à des Irradiations Continues et Discontinues à Faible Débit de Dose", Pub. CNES, 92-050/RA/TE/AE/TT/TF, Oct. 1992.

5. J.R. Norton, J.M. Cloeren, J.J. Suter "Results from Gamma Ray and Proton Beam Radiation Testing of Quartz Resonators"
IEEE Trans.on Nuclear Science, Vol. NS-31, N° 6, p 1230, Dec. 1984.

6. W. Primak"Fast-neutron-Induced Changes in Quartz and Vitreous Silica", Physical Review, Vol. 110, N°6, p 1240, June 1958.

7. J. Bourrieau, M. Romero "Effect of Space Charged Particle Environment on Optical Components and Materials", ESA Symposium on Spacecraft Materials in Space Environment, Publ ESA SP 145, Dec.1979.

8. J. Bourrieau, G. Panabière "Etude des Variations Dimensionnelles du Niobate de Lithium sous Irradiations", Publ. CERT-ONERA/DERTS, CR/COMP/14, Oct. 1988.

USO Serie number/type	Frequency (MHz)	Manufactured in	Norme	Quartz type	Pre-irradiation Level/date
E67363/PMT-P	10	1988	Mil Std	QAS - SCP3 SAWYER (Swept)	64.8 krads/1987 (quartz)
E67365/PMT-P	10	1988	Mil Std	QAS - SCP3 SAWYER (Swept)	84.8 krads/1987 (quartz) 20 krads/1991 (USO)
E67360/PMT-P	10	1988		BVA - SCP5 SICN (Unswept)	84.8 krads/1987 (quartz) 20 krads/1991 (USO)
E67361/PMT-P	10	1988	Mil Std	BVA - SCP5 SICN (Unswept)	1.5 krads/1987 (quartz)
M1 T1 (thermometric)	10 11	1991		QAS - SCP3 SICN (Swept)	
M2 T2 (thermometric)	10 11	1991	Mil Std	QAS - SCP3 SICN (Swept)	
M3 T3 (thermometric)	10 11	1991	Mil Std	QAS - SCP3 SAWYER (Swept)	
M4 T4 (thermometric)	10 11	1991	Mil Std	QAS - SCP3 SAWYER (Swept)	

Table 1 : Characteristics of the USOs

USO	Component $y_1(t)$			Component $y_2(t)$		
	Saturation	Annealing	Initial Sensitivity (1E-11/rad)	Saturation	Annealing	Initial (final) Sensitivity (1E-11/rad)
M1	<1 krad	1 day	+1	***	***	***
M2	<1 krad	3 days	-1.5	***	***	***
M3	<1 krad	1 day	+1.5	<1krad	Not seen	-1 (0)
M4	<20 rads	<1 hour	-15	No	Not seen	1 (1)
67360	<1 rad	<<1 hour	<<0	No	Not seen	4.5 (4.5)
67361	<1 rad	<<1 hour	<<0	No	Not seen	11 (11)
67363	<150 rads	<1 day	+3	>1 krad	>3 months (?)	1.1 (0.02)
67365	<150 rads	<1 day	+3	>1 krad	>3 months (?)	1.1 (0.02)

Table 2 : Characterisitics of the radiation response of the studied oscillators

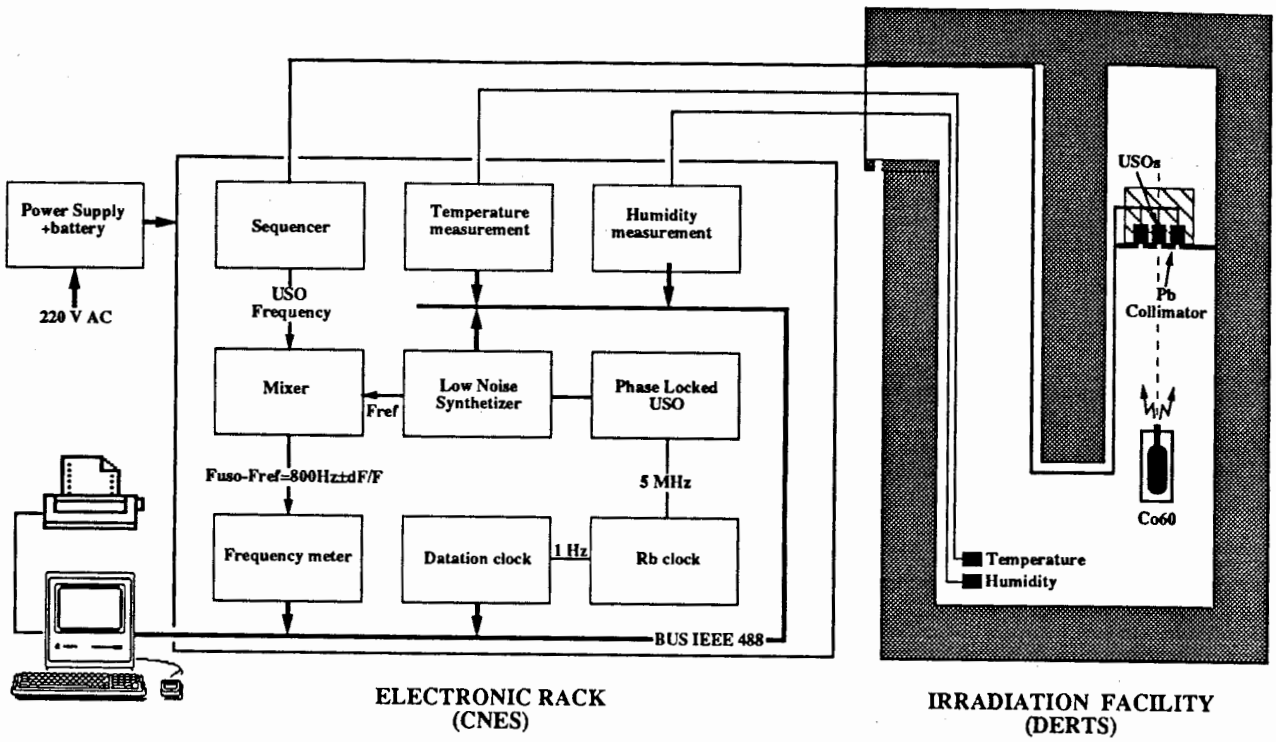


Figure 1 : Experimental facility

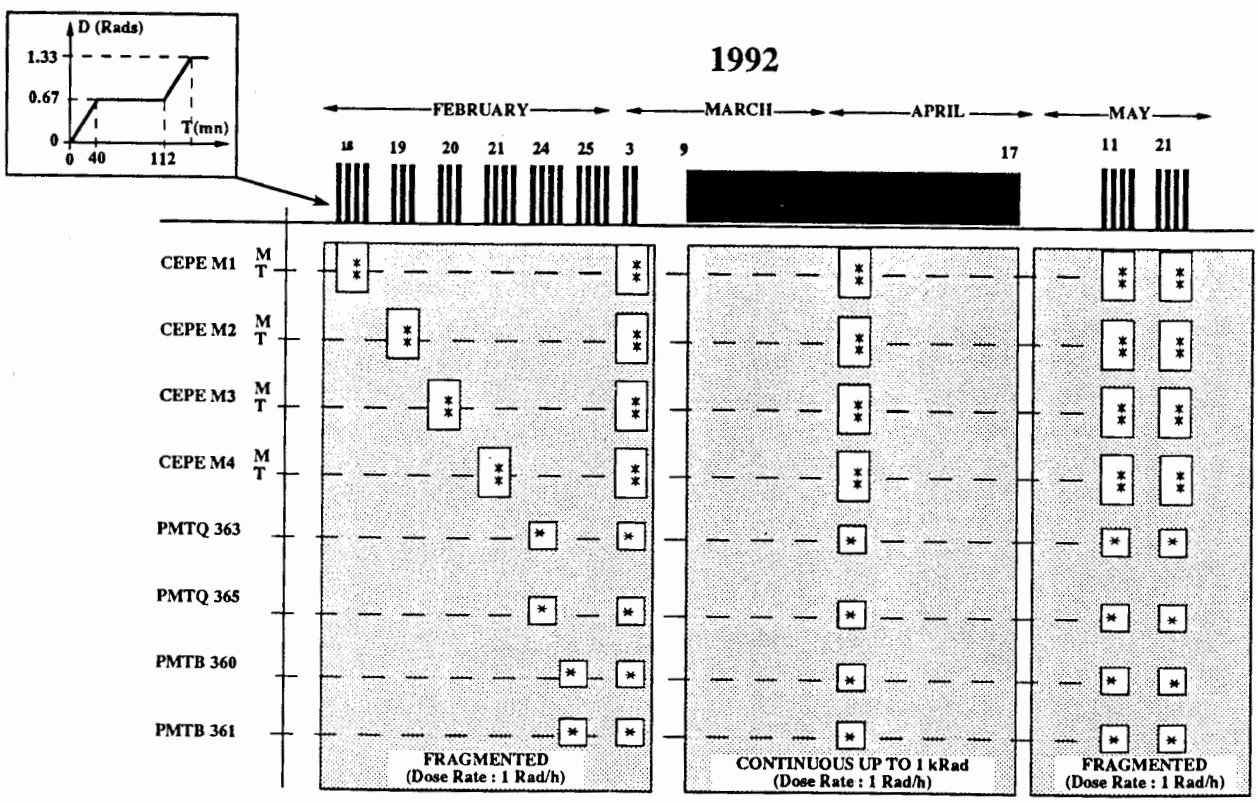


Figure 2 : Irradiation Plan

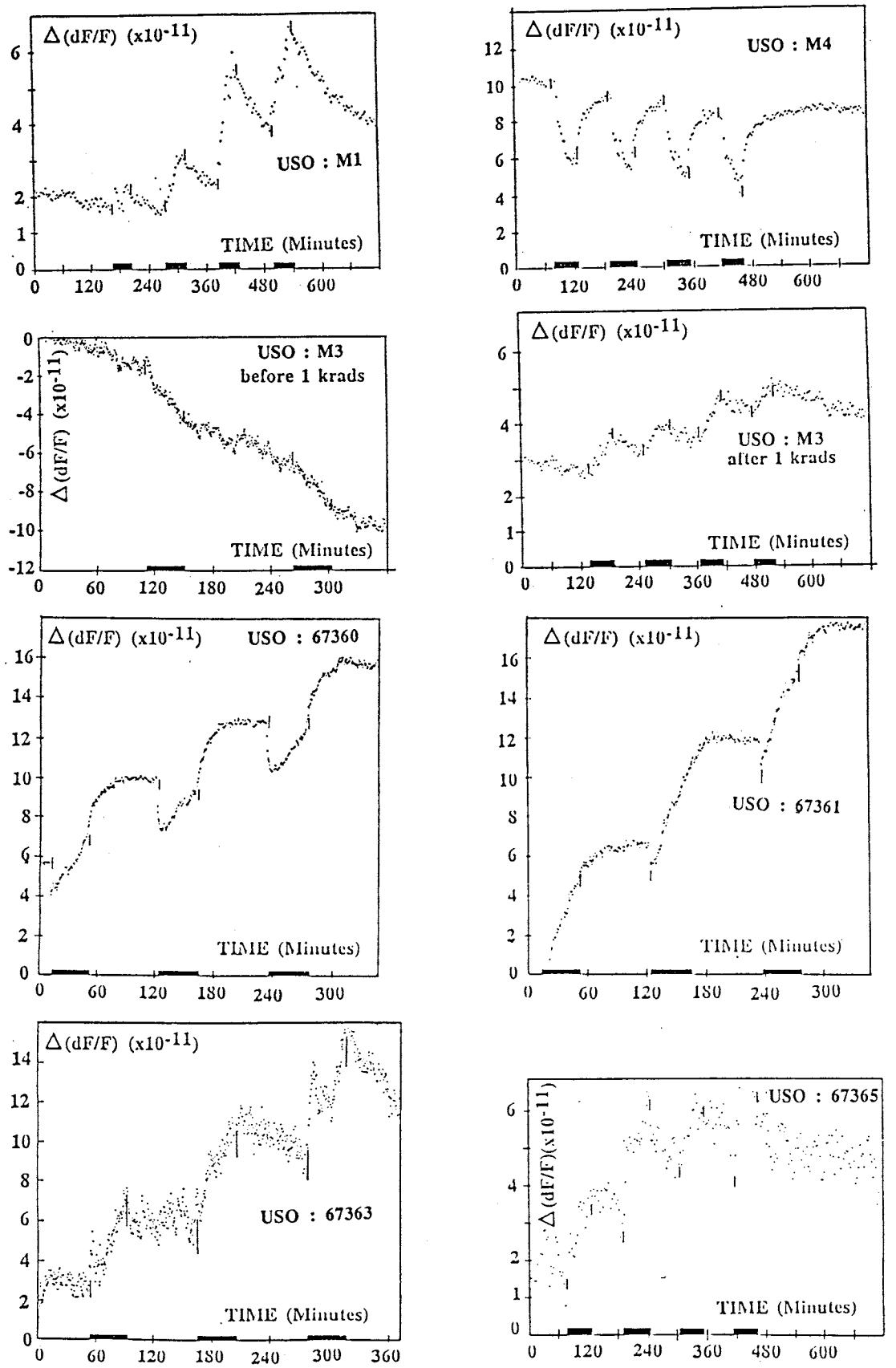


Figure 3 : Total drifts (aging+radiation induced) during discontinuous irradiations (Co60 / 10 mGy per hour)

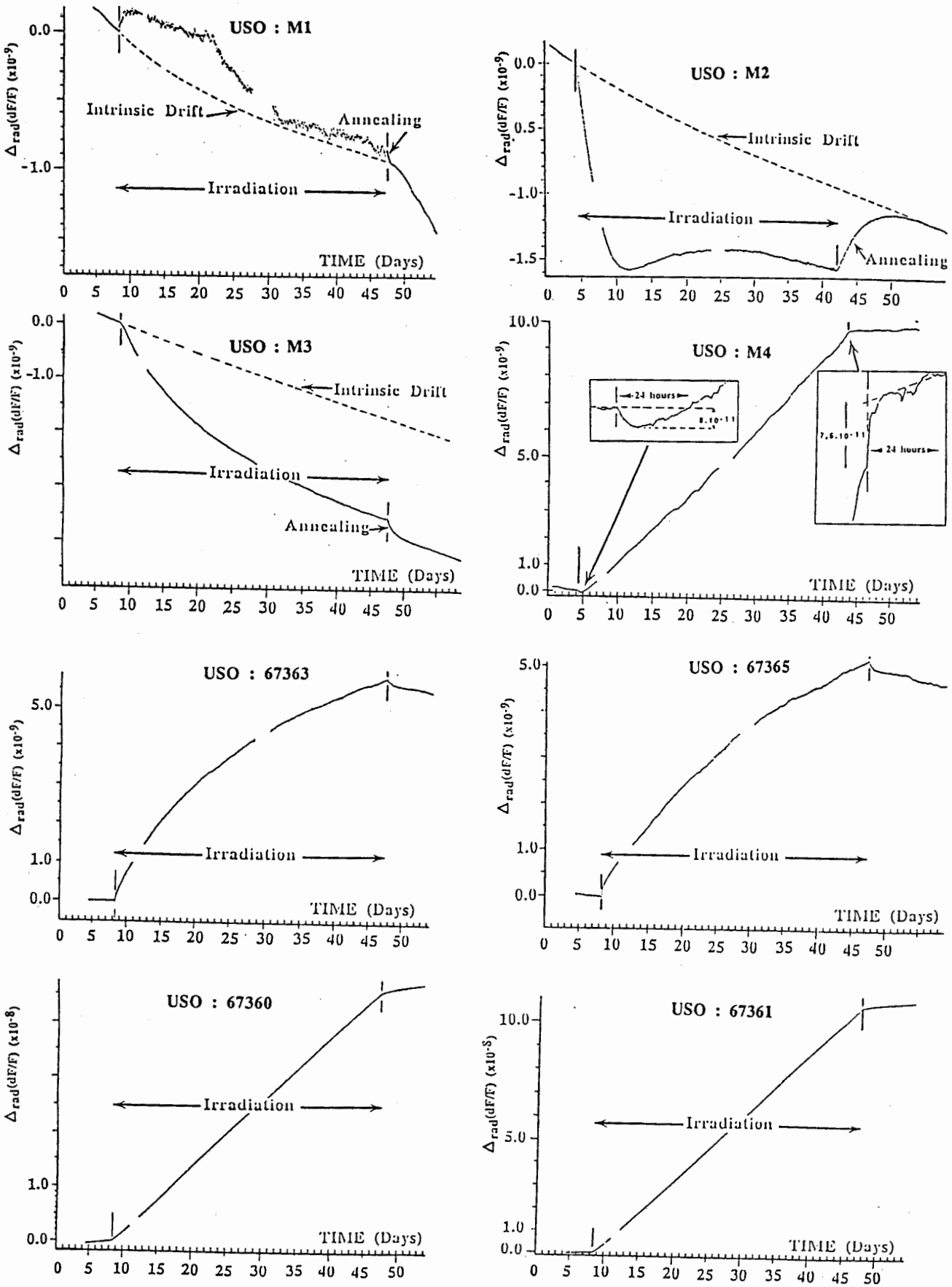


Figure 4 : Total drifts (aging+radiation induced) during continuous irradiation up to 10 Grays (Co60 / 10 mGy per hour)

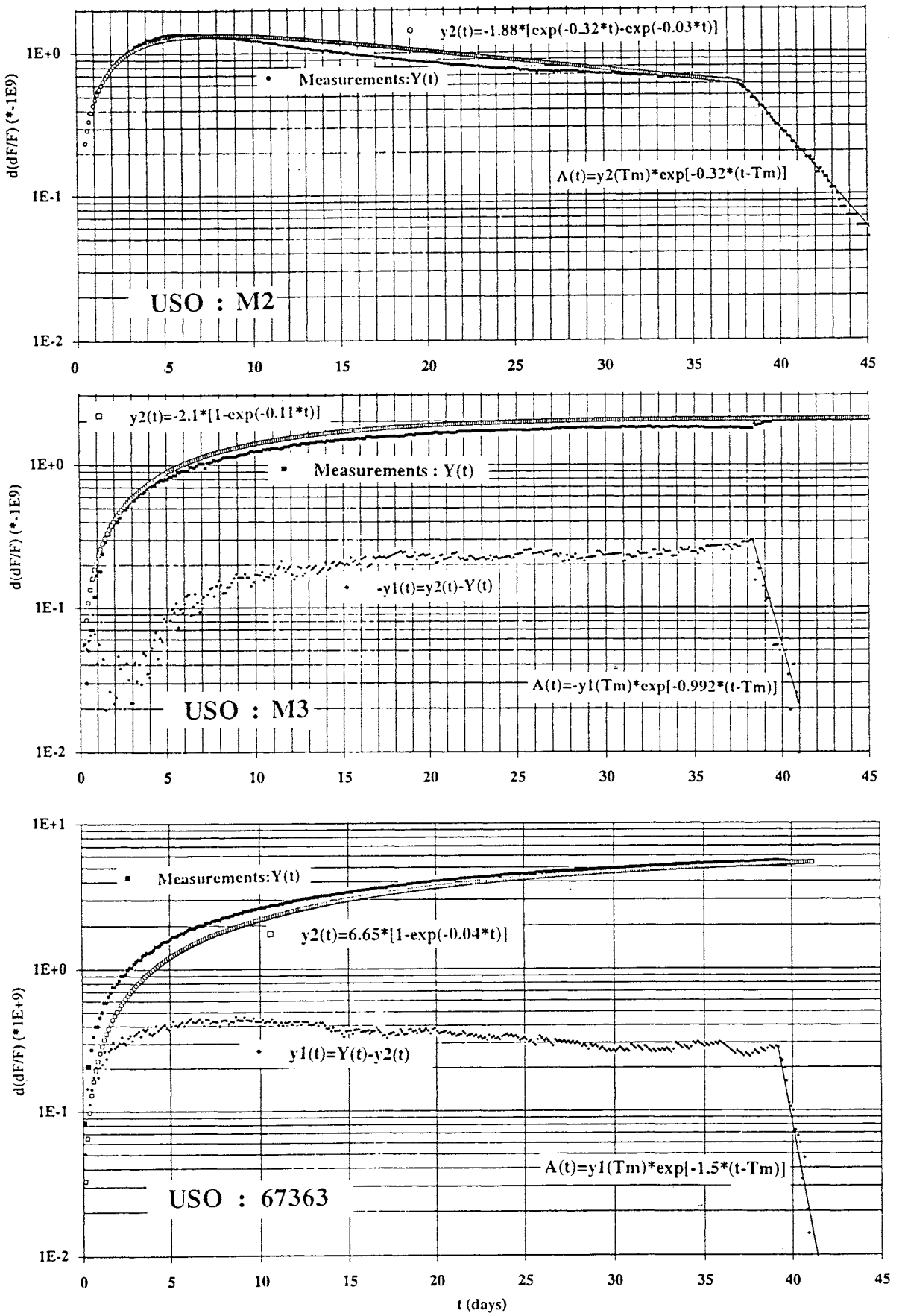


Figure 5 : Examples of radiation induced drift modelling

BEHAVIOUR AND MEASUREMENT OF DRIVELEVELDEPENDCY ON QUARTZ CRYSTALS

J.S. YERNA

YETECH APPLICATIONS

P.O.Box 674 7000 AR Doetinchem
The Netherlands

FAX + 31 8340 63601

A B S T R A C T

Since powerlevels with modern electronic circuits decrease steadily,the drivelevel for quartz crystals in an oscillatorcircuit is also decreasing.
Modern circuitry is working in regions of 1-10 µW nowadays.
The behaviour of a quartz crystal in a start-up situation will be discussed in conjunction with actual oscillator-circuitry.
Determination of driveleveldependcy and its measurement were items for long international discussion. Practical approach and investigation resolved in two realistic methods for measuring driveleveldependcy. Advice with respect to the match of the crystal to the circuitry is given and the two measuring-methods are prescribed.

Keywords: Drivelevel, Rr, -Rosc.,applicationmatch.

1. I N T R O D U C T I O N

1.1 Principle of an oscillator

An oscillatorcircuit is in principle given by fig.1
It consists of a passive and an active part.
The active part should be concidered as a -R part.
Condition for oscillating is:

-loopgain > 1
-fase-difference in/out=n x 360 °C

This condition must also be fullfilled during the start-up of the oscillator.
The -R of the active part can be concidered as a straight line and will be indicated as -Rosc.
Fig.2 indicates the start-up of a quartz crystal.

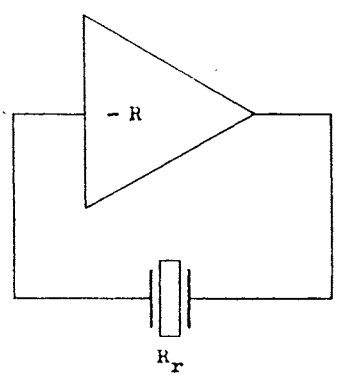


Figure 1

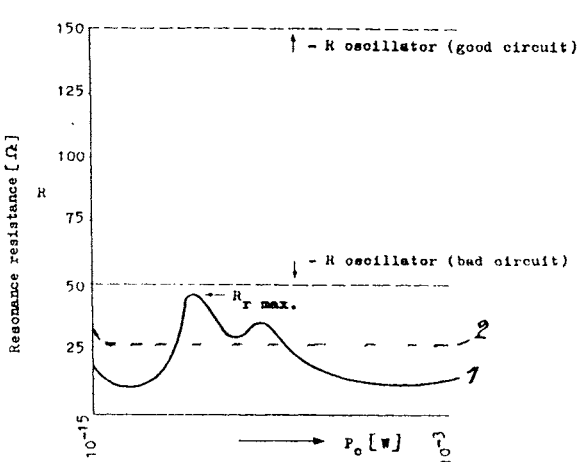


Figure 2

Theoretically an ideal quartz will show a straight line.(Curve 2) In practice we should consider start-up lines as indicated bij curve 1. Depending of the causes curve 1 might be reversibel but can also be irreversible. Moreover in time the irreversible effect can change again.
In the application should be considered that the -Rosc., is determinating the so called reject-level being the critical point where the oscillator will stop. The applicationmatch of the two parts is obvious.
Influences of temperature in combination with irreversible effects ask for sufficient distance between max. Rr of the crystal and the -Rosc. of the oscillator. International investigation and practice have resulted in an mandatory advice of IEC to design circuits with following ratio:

$$-Rosc. \geq 3 \times Rr$$

There is a tendancy in circuits for portable sets (pagers, telephones) to design circuits with extrem low loadcapacity in order to keep the currentdown value low.
International practice with circuitry based on many applications resulted in an IEC recommandation to use in circuits

$$C\text{-load} \geq 10\ pF$$

Figure 3 demonstrates the influence of the load-capacitance in relation to the minimum needed gain. From this graph again it is clear that designing circuits Cload must be ≥ 10 pF.
Moreover we should realise that the minimum needed gain-line is not a stable line as various IC-circuits show dependcy with temperature.
This effect only will become clear in practice in worst case conditions.
Again for this reason it will be clear that the IEC-recommandations should be executed in order to avoid practical problems.

1.2 Causes of Driveleveldependcy Effects (DLD)

Most reversible effects are due to excessive crystal loading,the irreversible effects are mostly due to imperfect production-techniques as:

- particles on the vibrator-surface
- mechanical damage of the vibrator,(e.g.fissures due to excessive coarse lapping abrasive)
- gas and oil inclusions in the electrodes(due to a poor vacuum or an inadequate coating rate)
- poor contacting of the electrodes at the mounting-place(due to inadequate baking procedures)
- mechanical stresses between mounting,electrodes and quartz element.

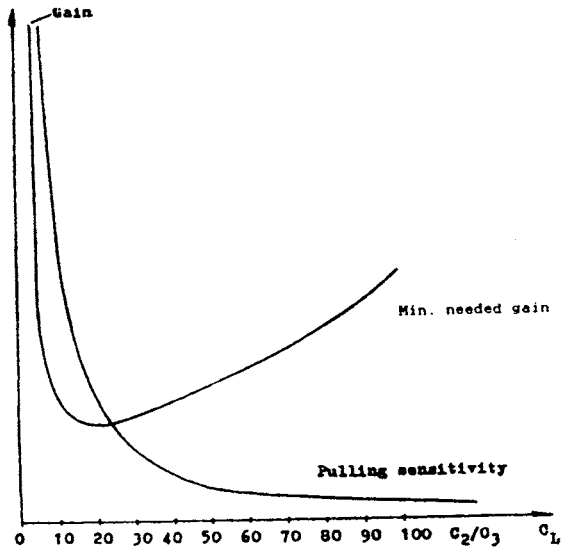


Figure 3

Pullability and min.needed gain as a function of the value of C-load.

2. MEASUREMENT

2.1 Introduction

There is an agreement on two ways of measurement within IEC-TC 49. Prior to measurement a storage of 24 hours on 105 °C is recommanded.
After the storage should be measured not earlier than 2 hours and not later than 1 week at roomtemperature.

2.2 Method A

This measurement is based on the pi-network according to IEC-444 while the crystal will be measured on three drivelevels being:

- on 50 μA and measuring Rr1
- on 1000 μA " " Rr2
- on 50 μA " " Rr3

Those drivelevels are derived for AT-cut crystals and based respectively on a velocity of Vmax=0.01 m/sec corresponding to 50 μA and a velocity of 0.2 m/sec corresponding with 1 mA.
The ratio Rr1/Rr2 = γ12 and should have a value which is smaller than the maximum γ given by the line in figure 4.
The allowed variation indicated by figure 4 is based on long term examinations of crystal manufacturers and proved to be a reliable indicator for crystal units showing start-up problems.
If necessary this method could also be extended by a large number of different drivelevels,which however proved not to be necessary in most cases in practice.

The ratio 13 should be smaller than (γ+1)/2, where the value of γ is taken from figure 4.
In the same way frequencydependence can be measured. The change (Ff3-Ff1) and (Ff1-Ff2) respectively measured on 50 µA.1000 µA and 50 µA should be specified in the detailspecification.

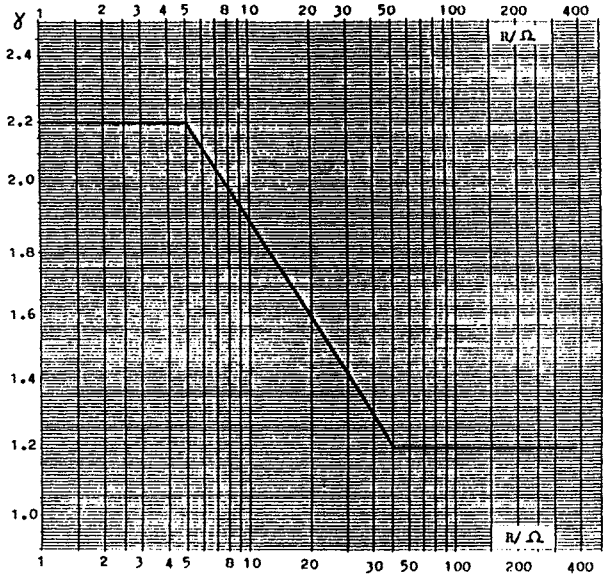


Figure 4

Maximum tolerable resistance-ratio γ for the drive leveldependence as a function of the resistances Rr2 respectively Rr3

2.3 Method B

This method is based on measurng the crystal in an active oscillator of special design.
The oscillator is designed in such a way that its -Rosc value is a true resistance of 200 ohms over a wide frequency-range from starting up point untill it reaches the nominal drivelevel of 1 mW. via a positive feedback network (see figure 5).
When the oscillator has reached the nominal drivellevel a Led indicates that situation.
The negative resistance and with it the DLD rejectlevel of the oscillator can be changed by connecting a positive resistor in series with the oscillator.(see figure 6)
In this way each reject-level can be installed between 0 and 200 ohms.
Connecting a quartz crystal unit with sufficient low Rr-value between the testclamps, the oscillator will build up starting from the initial noise level to its limitationpoint of 1 mW as schown in figure 7.
During the start up the Rr.max of the crystal is continously compared with the calibrated -Rosc. and the result is detected and transferred in a go/no go decision. If the crystal unit under test should show a certatain degree of DLD it is possible that the

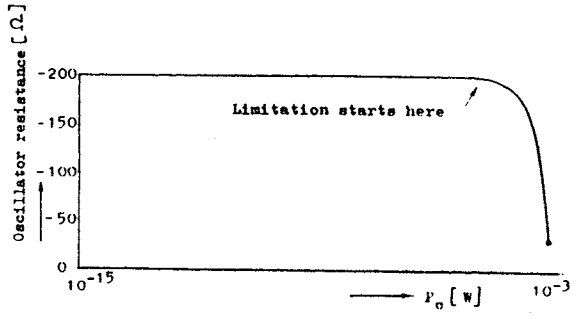


Figure 5

oscillation amplitude will not reach the 1 mW limiting point (point B in figure 7). In the example of figure 7 the building up of the oscillation is terminated at a much lower level of drive.(point A) When the Rr.max of the crystal unit will stay during start up under the installed reject-level, the oscillator arrives on point B and the Led indicates that the crystal under test did not exceed the DLD reject-level during its start up.

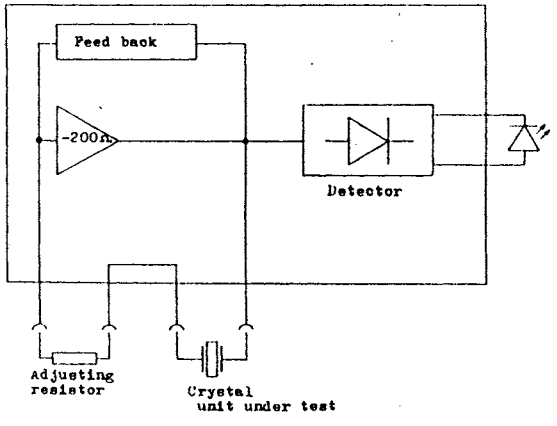


Figure 6

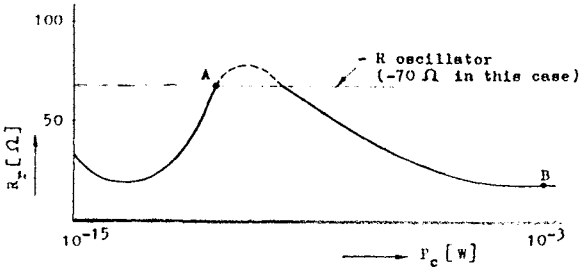


Figure 7

3. CONKLUSIONS

On designing oscillatorcircuits there is a IEC recommandation to design the oscillatorcircuit in such a way that -Rosc ≥ 3 x Rrmax.
Loadcapacitay should have a value ≥ 10 pF in accordance with IEC recommandation.
There is a standard way to measure DLD of a crystal unit available in accordance with IEC.
The latter IEC paper will be available end of 1993 under IEC 444-6 and will contain more detatiled background on the item.

Acknowledgements

The autor wishes to thank Mr.Neubig of Tele Quarz Group and Mr.Gericke of the Siemens Group for their investigations and their contributions in order to reach an agreement for a new international standard.

References

1. M.Bernstein: "Increased crystal unit resistance at oscillator noise levels" 21st Ann.freq.Symp.

2. Mil-C-3098 D through G and Mil-C-3098 H

3. IEC Publication 444

4. IEC Publication 122-2-1
   Part 2 Section 1 Quarz crystal units for
   µ-Processor clocksupply.

5. IEC 444-6 (In preparation)

6. J.S. Yerna: "Resistance measurement at very low drive levels" 38st Ann. Freq. Symp. Philadelphia.

A NEW APPROACH TO AGING INVESTIGATION OF QUARTZ RESONATORS

M. Wójcicki, A. Smolarski, A. Lisowiec

Tele and Radio Research Institute
Ratuszowa 11, 03-450, Warsaw, Poland

ABSTRACT

Long term frequency stability of quartz resonator is a basic parameter required by resonator users. Official normalizing documents recommend to measure the aging process for one month at temperature of 85°C that, it is thought, is equivalent to measuring the aging coefficient in one year at room temperature. However experiments that have been carried so far do not confirm this thesis. At Tele and Radio Research Institute there has been developed a measurement system enabling very thorough investigation of the aging process. The main goal of researches with the use of the system is evaluation in possible shortest time period specific quartz resonators aging parameters, which makes easier prediction of resonator frequency behaviour during its exploitation period. Results that have been already obtained seem to indicate that more rational approach to this problem than the one used till now, would be to measure the above mentioned parameters in a shorter than one month period at near room temperature by applying methods of possible high accuracy.
Keywords: quartz resonator, aging, measurement method.

1. INTRODUCTION

Aging of quartz resonators is a very complex internal process caused by physical and chemical phenomena and so far there is no sufficiently good theoretical model of aging mechanisms enabling accurate extrapolation of resonator behaviour in time. If one of the many factors causing aging is dominating, the description of aging phenomenon can be more easier and predictions more precise but it concerns only the selected types of resonators manufactured with specific technology. An additional difficulty in determining quartz resonator aging parameters, is their strong dependence on external factors like temperature, vibration, electrostatic field, ionising radiation, atmospheric pressure, etc. Moreover not only the actual level of external stresses but also their history is of great importance.

The basic method for aging measurement, still recommended by IEC documents, consists in measurement of the resonator frequency changes made over time period of 30 days at temperature of +85°C (for ovenized resonators at turnover temperature).

The above mentioned procedure is based on the assumption that the temperature increase accelerates the aging process to such an extent that changes in the resonator frequency observed over one month correspond to changes that would take place over one year if the resonator was placed at room temperature. However, many authors question the validity of this assumption, giving examples of the lack of such simple correlation between aging rate and temperature. It is pointed out, that some physical or chemical process that are responsible for aging are activated only above certain temperature (Ref. 1,2). Thus it is very probable that for many resonators operating at near room temperature the evaluation of aging characteristics carried out at temperature of 85°C will be misleading.

The other drawback of the above mentioned method is relatively long measurement time (30 days). It makes difficult the current evaluation of resonator aging parameters which is fundamental to the proper control and monitoring of the production process.

For that reasons new methods are being sought that will enable much faster and at the same time more credible evaluation of quartz resonator aging characteristics. For example in publication (Ref. 3) there are described some experimental technics for accelerated evaluation of aging. The key problem when applying such technics is taking the right mathematical model of aging that incorporates time as well as temperature dependence of aging parameters being measured.

In Tele and Radio Research Institute the investigation of quartz resonator aging phenomena has been carried out for many years (Ref. 4, 5), in that elaborating methods and systems for accelerated aging determination (Ref. 6). In the paper method for fast evaluation of exploitational aging parameters of medium stability quartz resonators is proposed.

2. METHOD OF AGING EVALUATION

It follows from the nature of the aging phenomenon that it is impossible, basing on short term post production measurements of short duration, to evaluate the resonance frequency after arbitrary period of time (e.g. 10 years) and there was no such an intention when elaborating this method.

The main goal of the proposed method was to create the possibility of relative, fast and objective evaluation of resonator aging properties immediately after production cycle. Such an evaluation is very essential for resonator users as well as for resonator producers because it generates the necessary feedback for eventual correction of the manufacturing process.

The following exploatational aging parameters of the quartz resonator are being proposed:

- momentary aging S, defined as the derivative of the characteristic $f(t)=\Delta f/f$ (t-time), after sufficiently long time period after production date, evaluated at stable exploitaional conditions not much different from stocking conditions
- influence of the defined changes of the exploitational conditions on the momentary aging $\Delta S/S$.

These parameters are evaluated independently according to the measurement procedures given below.

2.1 Determining the momentary aging S

Algorithm for determining the momentary aging S is presented in fig 1.

At the beginning, the measurement conditions are being settled down. It is essential to carry out measurements very carefully over given period of time with the highest possible accuracy, minimizing the influence of any other factors beside time. For that reason the temperature T of the investigation should not be very different from typical room temperature($20^{\circ}C \le T \le 40^{\circ}C$) and should be precisely

stabilised ($\Delta T \leq 0.001°C$). The measured resonators should not undegro vibrations, DC voltage polarisation and ionising radiation. The drive level should be kept at sufficiently low level. The proposed time period of investigation, t_m is from one to a few weeks (e. g. t_m=14 days), and measurements should be carried out regularly as often as possible.

After stabilisation of measurement conditions the measurements of quartz resonators resonance frequencies (and possibly other parameters) are being carried out over time period t_m and then the set of measurement points is approximated by selected mathematical functions describing probable mechanisms of aging changes.

The set of the mentioned functions contains: linear, exponential, multiexponential + linear, logarithmic, logarithmic + linear, multilogarithmic and other function of decreasing character.

From this set the best approximating function is chosen basing on the criterion of minimal root mean square deviation $\sigma=\sigma_{min}$. The minimum deviation σ_{min}, that has been obtained, carries information about degree of regularity of approximated dependance. The σ_{min} much larger than the measurement resolution indicates that the resonator has anomal aging characteristic and is not considered for further classification. From the optimum approximating function the momentary aging $\underline{S}$ is analytically determined for time t_1 sufficiently away from production date (e.g. t_1=1 year).

2.2. Determination of the influence of exploitational condition changes on momentary aging $\Delta S/S$.

Among many external factors determining aging characteristic the most important is temperature. The algorithm for determining the influence of temperature on aging is presented in fig. 2.

First, the basic conditions of measurement are defined. These are:

- initial temperature of aging investigation T_0, usually equal to the temperature at which momentary aging evaluation is made,
- time period t_m, determined as a minimal time period necessary for estimation the slope of $\Delta f/f(t)$ characteristic (in practice t_m=3÷7 days)
- temperature jump ΔT, corresponding to the possible temperature changes during exploitation (e.g. ΔT=10°C)

After stabilisation of initial temperature $T=T_0$ measurements of resonance frequency over time period t_m are taken analogoically as described in subclose 2.1. The aging S_{T_0} at T_0 is computed as the slope of the regression line approximating measurement points. Then a temperature jumpΔT is applied to the resonator and another set of measurement points is obtained, from which by method of linear regression the aging S_{T_1} at $T_1=T_0+\Delta T$ is obtained. The parameter $\Delta S/S$ is then computed as $\frac{\Delta S}{S} = \frac{S_{T_1} - S_{T_0}}{S_{T_0}}$.

3. A SYSTEM FOR INVESTIGATION OF THE QUARTZ RESONATORS AGING PROCESS

In Tele and Radio Research Institute there has been developed a special measurement system for aging investigation of medium stability quartz resonators, enabling the verification of the proposed fast method for evaluation of the exploitational aging parameters (Ref. 6). The system makes possible long series of measurements of resonance frequency (and other parameters) on large groups of resonators (up to 120) at a stable temperature from 40°C to 150°C range. The resonators can be in various enclosures and can have different nominal frequencies over the range of 1 to 200 MHz. Typical measurement accuracy is of the order of $1\cdot10^{-8}$ for resonators with $1\cdot10^5$ Q factor, which enables to determine momentary aging with accuracy ±(1 ppb/day+10% S) or ±(a few 0.1 ppm/year +10% S).

4. A SAMPLE OF MEASUREMENT RESULTS

Using the system described in close 3 a series of aging investigations of medium stability resonators have been carried out. During investigations, resonator types as well as temperatures and time periods of evaluation have been varied, Among others a comparative study of aging behaviour of the some resonators at temperatures +85°C and +40°C has been carried out. The result confirmed the thesis that there is no clear and simple dependence of aging on temperature which might justify the recommended method of investigation (for 30 days at temperature of +85°C). This is illustrated in fig. 3 which shows the distribution of the ratio of aging coefficient at +85°C to the aging coefficient at +40°C for a sample of 100 resonators of AT-cut and 35.5 MHz nominal frequency.

An example of 14 days aging frequency response of medium stability AT-cut 100 MHz resonator is shown in fig. 4. The best fitting of the experimental data is to a linear-exponential function $\delta f_{ap}(t)=A\cdot t+B+C\cdot\exp(-t/D)$ where A, B, C, D constants, with $\sigma_{min}=1.14\cdot10^{-8}$. Calculated momentary aging $\underline{S}$ for this example for time t_1=1 year is

$$S = \left(\delta f_{ap}(t)\right)'_{t_1=1\,year} = -13.2\,{}^{ppb}/_{day}.$$

5. CONCLUSION

The proposed method makes possible relative, fast and objective assessment of quartz resonator aging properties immediately after production process. It seems to be especially useful for quartz resonator producers because it generates possibility of process control.

Two types of aging parameters have been proposed:

- momentary aging $\underline{S}$, characterising resonator aging in stress less near stocking conditions and
- sensitivity of momentary aging rate $\Delta S/S$ due to any external stress, especially temperature.

Specialised commercially available aging measuring system has been elaborated for simultaneously quick evaluation of considerable large amount of medium stability quartz resonators.

REFERENCES

1. J. R. Vig, Resonator aging, Ultrasonic Symposium Proceedings, 1977.
2. R. L. Filler et al, Aging studies on quartz crystal resonators and oscillators, Proc. of 38th ASFC, 1984.
3. J. R. Vig, The aging of bulk acoustic wave resonators filters and oscillators, Proc. of 45th ASFC, 1991.
4. M. S. Wójcicki, System for investigation of quartz resonator aging, Prace ITR, vol 1(70), 1975 (in Polish).
5. M. S. Wójcicki, Technology and system for accelerated determination of quartz resonator aging, Final technical report, Contract ITR Z-582, 1987, (in Polish).
6. A. Smolarski, A, Lisowiec, A high accuracy automated system for investigation the aging process of quartz resonators, Proc, of TC 4 IMEKO Conference, Vienna 1992.

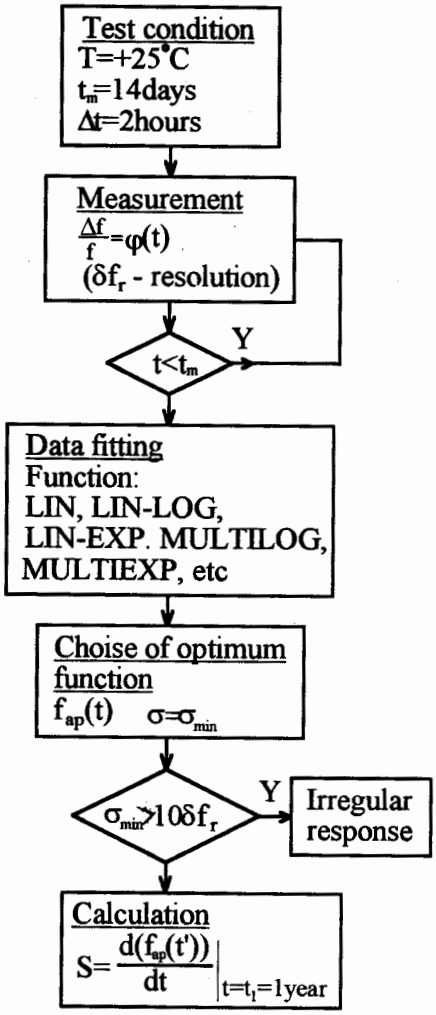


Fig. 1. Determination of the momentary aging S.

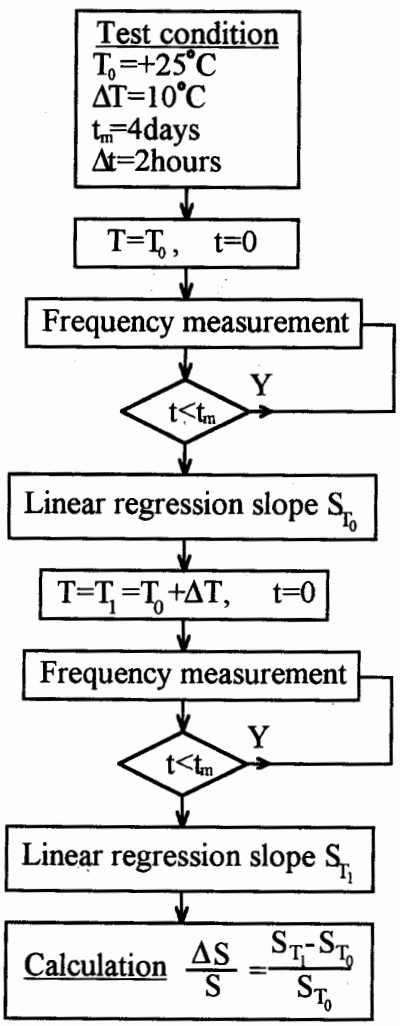


Fig. 2. Determination of momentary aging temperature stability ΔS/S.

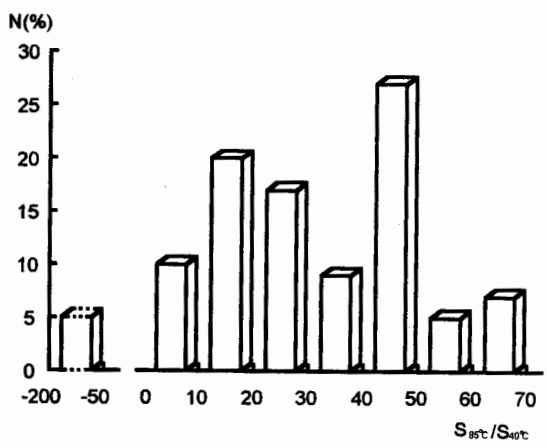


Fig. 3. Ilustration of temperature influence on resonators frequency aging rate (S) AT 35.5 MHz, 100pcs.

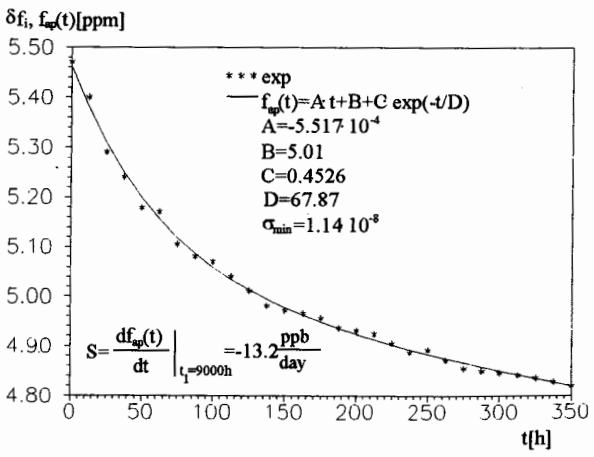


Fig. 4. 14-days post production aging characteristic of 100 MHz AT resonator.

CHARACTERISTICS OF TC-CUT CRYSTAL UNITS EXCITED BY LATERAL FIELD

I.V. Abramson

Omsk Research Instrumentation Institute, Omsk, 644063, Russia

1. ABSTRACT

Quartz resonators of TC-cut like RT-cut crystals have smooth f vs. T dependance and high Q-factor. However it's problem to use the crystals due to presence of strong B-mode near the useful resonance.

In this paper TC-cut crystals excited by lateral field are described. The optimal electrodes form providing reasonable suppression of the unwanted mode at the workable resistance of the useful mode has been defined.

Formulas for motional inductance of the crystals have been obtained.

The most important characteristics of the developed resonators have been found to be acceptable for modern XOs.

Key words: quartz crystal, TC-cut, B-mode, lateral field, C-mode.

2. INTRODUCTION

Quartz crystals of TC-cut (yxbl/13°54'/-33°) have been proposed in 1975 [1]. Like RT-cut (yxbl/15°/-34°) crystals they are very attractive for high quality XOs owing to high Q-factor (2-3 times greater than in AT-cut crystals) and excellent temperature-frequency stability.

However in spite of the evident advantages of TC-cut and RT-cut crystals they are not used widely at present. The reason is that the crystals have spurious high active B-mode near the main C-mode. As the B-mode is 2-3 times more active than the C-mode and frequency gap between them is less than 8% it needs special selecting circuit providing the oscillator operation on the C-resonance. That results in complexity in XO tuning as well as its stability degradation.

As it follows from numerous investigations one of promissing ways to improve the spectral characteristics of double-rotated crystals is use of lateral field for their excitation.

Application of appropriately directed lateral field in SC-cut plates provides sufficient suppression of B-mode and good electrical parameters on the main resonance [2]. However RT-cut and TC-cut crystals excited by lateral field are not discribed till now.

Subject of the present work is designing the TC-cut crystal units with the lateral excitation. A set of experimental and theoretical researches has been involved that provided good spectral and electrical characteristics of the crystals. Also their other important characteristics are found to be attractive for variety of modern XO applications.

2. DESIGN OF THE CRYSTALS

A design of TC-cut plate excited by lateral field is displayed in Fig. 1. It contains film electrodes deposited on one or both sides of the quartz blank separated by a gap. To select the useful C-mode the lateral field direction for TC and RT-cut crystals must be close to crystallographic X'-axis [3]. So the gap length is oriented almost along Z' axis. Owing to piezoconstant e'_{14} coupling the X'- directed field with the C-mode vibration is greater the piezoconstant e'_{16} coupling the X'- field with B-mode vibration the useful resonance is much more active than the spurious one.

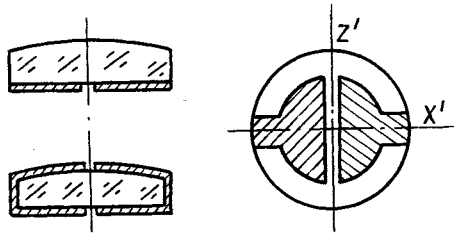


Fig. 1. The lateral field resonator designs.

To achieve the least motional resistance of the crystals the electrodes form appears must meet following requirements:

1.The electrodes sizes should be larger than the plate active area sizes. So the "energy trapping" in the crystal must be created not only by the film electrodes but also using the plate convex form.

2. The gap width must be far less than the plate thickness. For example, to get minimum resistance of 10 MHz crystals operating on fundamental mode the gap width must be less than 0.05 mm. One of the way to fabricate such gap is using the laser cutting the electrode film.

3. The two-sided electrodes are more effective than one-sided. However in some cases considered below one-sided electrodes may be preferabler as providing more simple fabrication and better stability of the units.

On the strength of these deductions TC-cut crystals have been developed. Their basic electrical characteristics are depicted in the table.

Frequency, MHz	Overtone	Electrodes form*	Parameters			
			C-mode			B-mode
			R, Ohms	L, H	Q, 10^6	R, Ohms
5	1	I	65	6	2.9	150
5	5	II	350	23	2.0	700
10	1	I	80	2.5	1.9	170
10	5	II	340	11	2.0	700

*I - one-sided; II - two-sided

As one can see from the data a suppression of B-mode is sufficient to excite the crystals in simple oscillators without circuit means to select the useful resonance.

It has been found that 3d overtone C-mode can't be excited in the crystals without regard to the plate thickness. That phenomenon may be ascribed to transformation of the vibration into 2d overtone A-mode vibration, which has almost identical frequency.As frequency of both the modes is in inverse proportion to the plate thickness the 3d overtone TC-cut crystals can't be used independently on their operational frequency.

It should be noted that the same problems occur with Y-cut crystals excited on the 3d overtone and BT-cut crystals operating on the 5th overtone. In both the cases the useful C-mode has almost the same frequency as 2 or 4th overtone A-mode.

To predict the lateral field crystals electrical parameters expressions for motional inductance vs. the plate geometry and the electrodes kind (two-sided or one-sided) have been obtained theoretically [4]:

$$L_2 = \frac{\rho\, t}{16\, e_{ij}^2} \cdot F\,, \qquad (1)$$

$$L_1 = \frac{L_2}{\left(1 - 0.24\cdot\frac{t}{\sigma_e} + 0.32\cdot\frac{t^2}{\sigma_e^2}\right)^2}\,, \qquad (2)$$

where ρ - density of quartz material;
t - thickness of the plate;
e_{ij} - piezoconstant;
F - coefficient depended on the crystal active area configuration; it equals to 0.8 for TC-cut;
σ_e - active area size in the lateral field direction where the vibrational amplitude degrades to 1/e of the maximum due to "energy trapping".

Obtained expressions provide the 20% accuracy of the calculations. Besides they lead to following important conclusions.

Two-sided electrodes result in less motional inductance and resistance of the crystals. However at small t/σ_e - ratio one-sided electrodes provide the resistance approaching the one at two-sided electrodes. So, for the crystals operating on fundamental mode with t/σ_e usually less 0.1 the one-sided electrodes are more acceptable as being more simple in producing and less affecting the crystal stability.

The lateral field crystals excited with two-sided electrodes have the inductance independing on active area sizes (if it remains less the electrodes size). That differs from the dependence for thickness-excited crystals and fits the experimental results [2].

3. THE CRYSTALS MAIN CHARACTERISTICS.

Besides the electrical parameters some other important characteristics of the TC-cut crystals have been measured.

The f vs. T dependence of the crystals like RT-cut crystals has parabolic form:

$$\frac{\Delta f}{f} = -8\cdot 10^{-9}(T-T_0)^2, \qquad (3)$$

where T - ambient temperature; T_0 - turn over temperature, which can be changed widely with the plate angle β (angle of rotation of plate around the X' - crystallographic axis).

From the above expression the temperature-frequency instability near T_0 is less than one for AT-cut crystals espessially at T_0 above 70°C. That makes TC-cut

crystals attractive for high stability TSXOs particularly if the crystals operate on 5th overtone. Moreover TC-cut resonators can be successfully used in S-TCXOs providing several ppm of frequency instability in wide temperature range (Fig. 2).

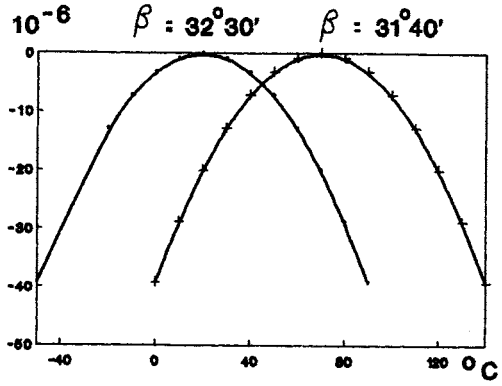


Fig. 2. Temperature-frequency characteristics of TC-cut crystals (YXBL/13 54'/-β).

The f vs. transient temperature stability of TC-cut resonators is 3-5 times better than one of AT-cut crystals. That provides improving the temperature-dynamic and warm up characteristics of the oscillators.

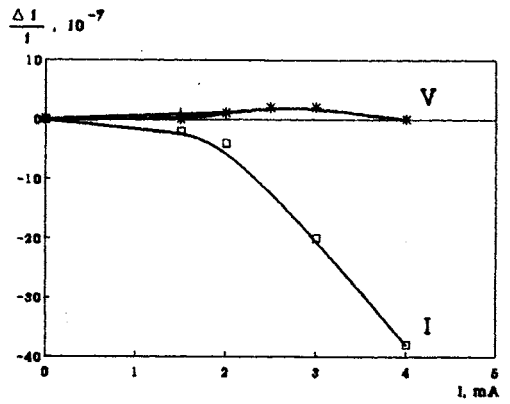


Fig. 3. The frequency vs. drive current for fundamental mode (I) and 5th overtone (V) crystals operating at 10 MHz.

The f vs. drive level are indicated in Fig. 3 to be different for the crystals operating on the fundamental mode and 5th overtone. While the first have f vs. drive current ratio $-0.6\ A^{-2}$ (in quadratic approach), the 5th overtone crystals are almost insensitive to the drive level. Such behavior of 5th overtone crystals has been discovered to be caused by nonlinear interaction of the useful vibration with 7th overtone A-mode. However more penetrating insight into that phenomenon demands further investigations.

To obtain high long-term stability of a quartz resonator a piezoplate must be clamped in places of minimum f vs. stress sensitivity. To define these places f vs. azimut of stress applied to the plate edges has been measured (Fig. 4).

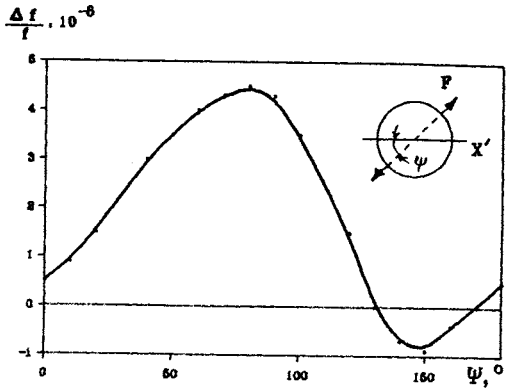


Fig. 4. The frequency vs. stress azimut dependance; the stress value is 150 gr.

As one can see from the figure zero sensitivity points exist and are located at 10° and 50° to the X' axis.

4. CONCLUSION.

1. Developed TC-cut crystals excited by lateral electrical field have good electrical and spectral characteristics and can operate in simple oscillator circuits.
2. The fundamental characteristics of the TC-cut crystals are shown to be accepted for variety of XO applications, for example, simple S-TCXO or high stability TSXO.
3. There are a few anomalies in the TC-cut crystal behavior that are not understanded perfectly and are subject for further investigations.

REFERENCE

1. Травкина С.Ф., Караульник А.Е., Шин В., Ярославский М.М. "Способ изготовления кварцевых резонаторов". А.с. N 474906, 1975, Россия.

2. Warner A.W., Goldfrank Ir. "Lateral Field Resonators", Proc. 39th Ann. Symp. on Freq. Control, 1985, pp. 473-474.

3. Ballato A., Hatch E.R., Mizon M., Lukaszek, Tilton R. "Simple Thickness Plate Modes Driven By Lateral Field". Proc. 37th Ann. Symp. on Frequency Control. 1985, pp.462-472.

4. Абрамзон И.В. "Высокостабильные кварцевые резонаторы SC и TC срезов с параллельным полем возбуждения ". Радиотехника, 1991, N 6. С. 85-86.

FINITE ELEMENTS METHOD MODELLING OF ACCELEROMETRIC AND BAROMETRIC SENSITIVITY OF QUARTZ RESONATORS

F. Deyzac

Office National d'Etudes et de Recherches Aérospatiales
BP 72, 92322 Châtillon Cedex, France

ABSTRACT

The resonant frequency of quartz crystal resonator is affected by stress modifications due to external disturbances (magnetic, pressure) or to accelerometric environment.
The BVA resonator was designed in particular to physically separate the vibrating crystal part from the metallic mounts.
The QAS takes advantage of this mechanical configuration.
The investigation by FEM clearly bring to light the interest of this concept for the reduction of the accelerometric sensitivity.
The use of a specific shape for mounts ("Muguet" mounts) proposed by ONERA a few years ago, makes the QAS quite insensitive to the acceleration.
This point is theoretically established through Finite Element Modelizations of different resonators taking into account manufacturing defects.
In conclusion, the QAS resonator associated with Muguet mounts appears to be a good and achievable solution for accelerometric sensitivity reduction.

Keywords: USO - Quartz - Resonator - Finite Elements Method - Pressure - Acceleration - BVA - QAS - "Muguet" mount.

1. INTRODUCTION

Applications for ultra-stable oscillators are increasingly numerous, and increasingly demanding in terms of the frequency stability characteristics required irrespective of environmental conditions.
Frequency stability is dependent on the quality of resonator behaviour under external loads. In a resonator, the mechanical resonance characteristics of a quartz crystal, which reflects environmental interference by the appearance of stresses and strains in the crystal, are used.
With the support of DRET, ONERA is engaged in studies designed to reduce the sensitivity of resonators to various factors such as acceleration, pressure, magnetism, etc.
This work has already led to improvements (Refs. 1, 2) which, despite their effectiveness, have not yet been exploited for reasons of fragility.
It is for this reason that we decided to develop a dedicated simulation tool, in order to achieve better comprehension of the sensitivity mechanisms of resonators, and assess the influence of manufacturing defects. This has also enabled us to define a non-degrading design concept for the improvements previously proposed.
This tool involves a finite elements method for modelling the resonator, and a theoretical stress/frequency model.

2. FINITE ELEMENTS METHOD MODELLING

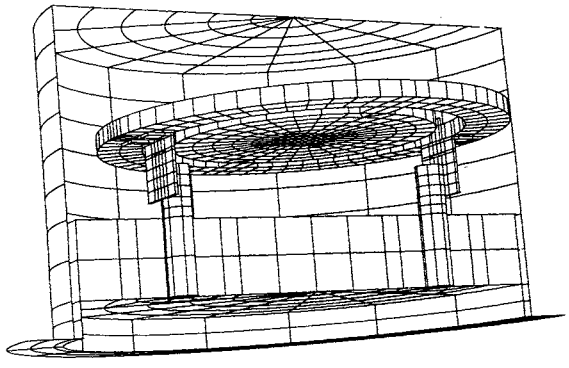
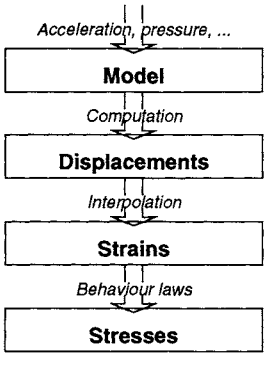


Fig. 1 - Cross-section of QAS quartz resonator and mesh.
Nickel mounts, kovar poles, brass case, glass base and quartz crystal.
The mesh has 3080 nodes and 1372 elements, and 22707 degrees of freedom.

FEM modelling involves discretization of the domain studied into a number of elements (Fig. 1), and parameterization, using an approximation technique, of the movement of distortable bodies (using form functions), to obtain a finite system of differential equations.
The system is then verified only at a number of points which are representative of the structure. These points are referred to as nodes. Displacements calculated at these nodes are then interpolated to obtain the global kinematics of the bodies, and strains and stresses via a behavioural model of the material.



The model can then be subjected to static loads, such as pressure, acceleration or temperature, and strain values at the nodes and stress states for each mesh (normally expressed at Gaussian integration points) can be obtained by computation.

3. THEORETICAL STRESS/FREQUENCY MODEL

This model is used to relate the stress state (computed by the FEM) to resonator frequency variation. The model takes account of the crystallographic cut (ϕ, θ), computing the wave polarization directions and rotated coefficients of elasticity. The model also takes account of crystal orientation (ψ) in the section plane.
The main hypothesis taken is to that the vibrating crystal of a resonator is designed so as to confine energy at its centre, which means that frequency variation depends on strain variations (T_i) in this central zone.

$$\frac{\Delta f}{f} = \sum_{i=1}^{6} \tilde{k}_i T_i$$

Stresses (T_i) and coefficients $\tilde{k}_i$ are expressed for the finite element model reference frame (Oxyz). The axes of this frame coincide with those obtained after triple rotation of the crystallographic reference frame $\left(Ox_1''x_2''x_3''\right)$ (Fig. 2).

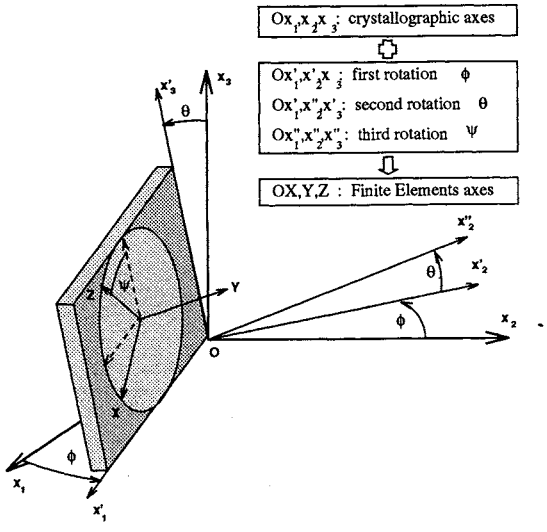


Fig. 2 - Crystallographic and finite elements reference frames

The procedure involves computing the elastic wave propagation directions in a crystal of any cut plane (appendix II), then applying a perturbation method (appendix III), assuming low stress-induced strain.

As the geometrical reference frame associated with the crystal (FEM reference frame) is used, coefficients of elasticity must be expressed for this frame (appendix I).

For example, in the case of SC-cut resonator (ϕ = 22.4°, θ = 34.3°), the coefficients used for a zero azimuth value (ψ = 0) are:

$\tilde{k}_1 = -2.38\times10^{-12}$ m²/N $\tilde{k}_4 = 39.85\times10^{-12}$ m²/N
$\tilde{k}_2 = -6.55\times10^{-12}$ m²/N $\tilde{k}_5 = 18.91\times10^{-12}$ m²/N
$\tilde{k}_3 = 1.35\times10^{-12}$ m²/N $\tilde{k}_6 = 43.86\times10^{-12}$ m²/N

4. BAROMETRIC SENSITIVITY OF A QAS RESONATOR

In this work, we examined a QAS resonator (ref. 7) with a self-suspended vibrating crystal, corresponding to the sticky electrode version of the BVA resonator (Ref. 8).
The form of the crystal used in shown in Figure I.1.
The barometric sensitivity of quartz resonators is a phenomenon which has been studied extensively at ONERA, and for which corrective modifications have already been proposed (Ref. 1).

The base of the resonator acts as a membrane distortable under the effect of pressure, and the poles rigid with this base transmit loads to the quartz crystal (Fig. 3).

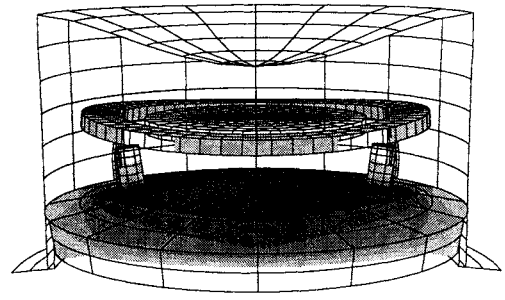
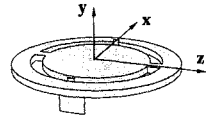


Fig. 3 - Barometric behaviour of a QAS quartz resonator.
Case strains are multiplied by 30, and quartz and base strains by 500. External pressure is 1 bar.

The principal stresses appearing in the vibrating crystal are of the extension-compression type.



Simulation of barometric behaviour of QAS under 1bar pressure		
Stresses are given at crystal centre Fractional frequency deviation is computed for SC-cut		
Tx (N/m²)	Tz (N/m²)	Δ f/f (/bar)
1500	4700	-1×10^{-8}

This type of resonator, used for the DORIS USO, has been fully qualified under experimental conditions (Ref. 3).
Mean barometric sensitivity measured is:

$$|\Delta f/f_{exp}| \approx 0.75\times10^{-8}/\text{bar}$$

indicating the good quality of the simulation.
Furthermore, the effect of pressure can be assimilated to diametral compression of the QAS crystal, and variation of barometric sensitivity according to the azimuth (ψ) [Fig. I.1] is a sinusoidal function, comparable with Ratajski curves.

5. ACCELEROMETRIC SENSITIVITY OF THE QAS RESONATOR

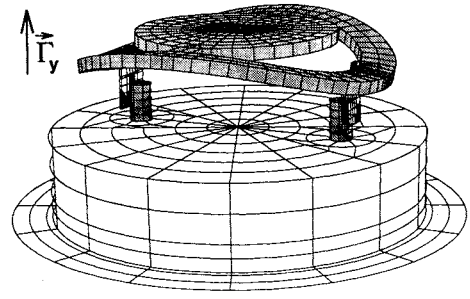


Fig. 4 - Accelerometric behaviour of a QAS quartz resonator.
Influence of 1 g axial acceleration on the QAS structure. A quartz ring is used as a mechanical filter for shear forces. Strains are multiplied by 10^5.

For accelerometric sensitivity, distinction is made between three cases corresponding to the three main directions in which acceleration is applied.
Figure 4 illustrates the effect of axial acceleration (applied on the Y axis).

Simulation of accelerometric behaviour of QAS 1g axial acceleration (Y axis)		
Stresses are given at crystal centre Fractional frequency deviation is computed for SC-cut		
Tx (N/m²)	Tz (N/m²)	ky (/g)
40	120	-2.7×10^{-10}

The advantage of a QAS type structure has been verified, in particular with respect to radial acceleration (Γ_x), by comparison with a conventional (non-QAS) resonator (Fig. 5).
Radial acceleration parallel to the mounts principally causes shear stress (Txy). The QAS structure reduces this stress by three orders of magnitude, and thus frequency sensitivity on this axis by the same amount.

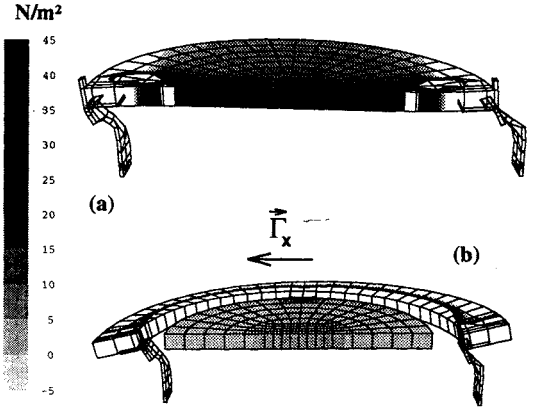


Figure 5 - Shear stress T_{xy} due to 1g radial acceleration Γ_x
Comparison of conventional resonator (a) with QAS resonator (b). Both have the same weight and thickness. Strains are multiplied by 3×10^5

In global terms, the QAS structure reduces sensitivity to acceleration in terms of the accelerometric sensitivity vector modulus, by a factor of 3.

Simulation of accelerometric behaviour of resonators 1g radial acceleration (X axis)		
Stresses are given at crystal centre Fractional frequency deviation is computed for SC-cut		
Resonator	Txy (N/m²)	kx (/g)
Conventional	26	10^{-9}
QAS	10^{-2}	5×10^{-13}

The mean accelerometric sensitivity of the QAS resonator was measured under experimental conditions (Ref. 3):

Accelerometric sensitivity influence of structure type	
Resonator	$k = \sqrt{kx^2+ky^2+kz^2}$
Conventional	10^{-9} /g
QAS	2.7×10^{-10}/g

$$|k_{exp}| = 2.4 \times 10^{-10}/g$$

Here again, simulation results are in full agreement with actual experience.

6. CONVENTIONAL MOUNTS OR "MUGUET" MOUNTS

The accelerometric sensitivity of quartz resonators has been studied at ONERA for many years (Refs. 4-6), and as in the case of sensitivity to pressure, this work has led to proposals for improvements (Ref. 1).
From this point of view, a new type of mount, referred to as the "Muguet" mount, was devised as an answer to the problem of D.C. accelerations (Fig. 6).

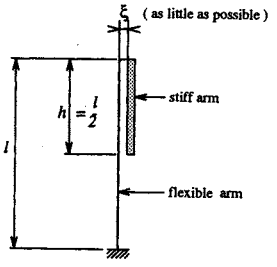


Figure 6 - Principle of "Muguet" mount

This mount has of two arms. The ascending arm is flexible with length l, and the descending arm is stiff and has length h = ½l

This type of mount offers two advantages:
- it avoids diametral compression in the case of Γ_y axial accelerations (Fig. 7),
- it avoids shear in the case of Γ_x radial accelerations (Fig. 8).

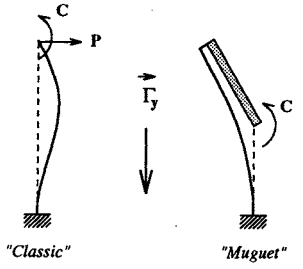


Figure 7 -
Mount behaviour under axial acceleration

The crystal imposes a rotation on the end of the mount, while remaining perpendicular to the attachment point With "Muguet" mount the flexible arm is subject to circular bending strain, and the only induced load is torque C, with no first order effect on frequency.

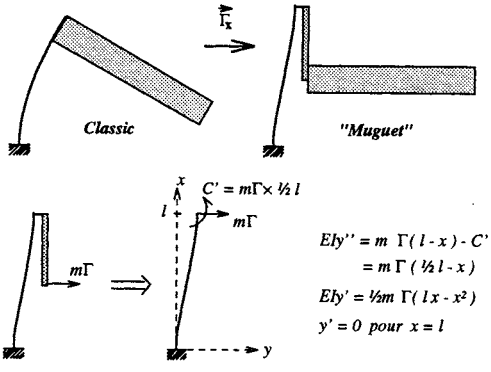


Figure 8 - Mounts behaviour under radial acceleration Γx
Assuming a fictitious resonator with a single mount, in the muguet configuration, the quartz crystal remains horizontal under the effect of Γx provided the relation h=½l is observed.

These results are checked by simulation (Figs. 9, 10) for industrial design "Muguet" mounts, defined to show eigenfrequencies at above 2 kHz.

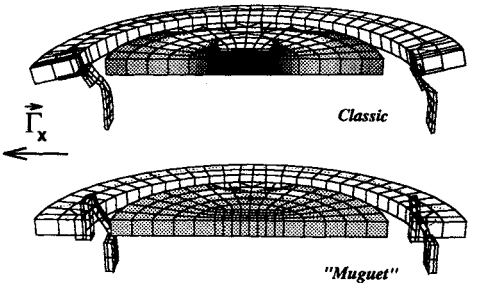


Figure 9 - "Muguet" mount efficiency under radial acceleration
QAS resonator subjected to 1g radial acceleration on the X axis. Shear stresses Txy are shown from 0 (white) up to .02 N/m² (black). Note the horizontal state of the quartz ring with the "Muguet" mounts.

Simulation of accelerometric behaviour of QAS 1g radial acceleration (X axis)		
Stresses are given at crystal centre Fractional frequency deviation is computed for SC-cut		
Mount	Txy (N/m²)	kx (/g)
Classic	10^{-2}	5×10^{-13}
"Muguet"	5×10^{-4}	2.5×10^{-14}

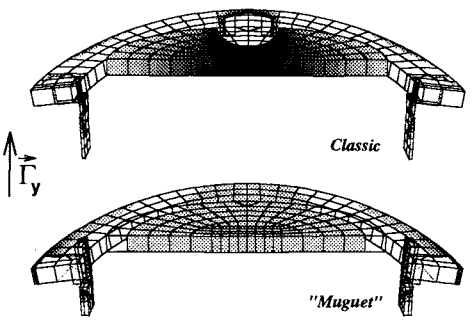


Figure 10 - "Muguet" mounts efficiency under axial acceleration
QAS resonator subjected to 1g axial acceleration on the Y axis. Stresses Tz are shown from 0 (white) up to 150 N/m² (black).

Simulation of accelerometric behaviour of QAS 1g axial acceleration (Y axis)			
Stresses are given at crystal centre Fractional frequency deviation is computed for SC-cut			
Mount	Tx (N/m²)	Tz (N/m²)	ky (/g)
Classic	40	120	2.7×10^{-10}
"Muguet"	.1	.3	5×10^{-13}

Thus we see that the "Muguet" mount can reduce the accelerometric sensitivity of a QAS resonator by two orders of magnitude.
Nevertheless, this demands excellent manufacturing quality for the mounts.

Accelerometric sensitivity Influence of mount shape	
Q A S	$k = \sqrt{kx^2+ky^2+kz^2}$
Classic	$2.7 \times 10^{-10}/g$
"Muguet"	$1.3 \times 10^{-12}/g$

7. CONCLUSION

It appears that the "Muguet" mount represents an effective solution to the problem of accelerometric sensitivity.
Furthermore, this mount offers a high degree of immunity to crystal orientation and assembly angular defects (Fig. 11).

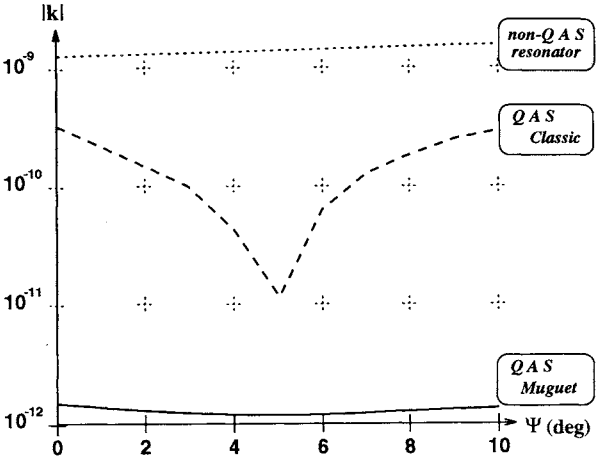


Figure 11 - Influence of azimuth angle on accelerometric sensitivity of quartz crystal resonator
Azimuth angle (psi) of crystal in cut plane (phi,theta). Psi is measured in the anti-clockwise direction. For a QAS, psi is the angle between the mounts and the x_3' axis. For a conventional resonator, psi is the angle between the mounts and the x_1' axis

REFERENCES

[1] J.P. Aubry, J.C. Craveur, F. Deyzac, M. Fragneau, D. Janiaud, *Analyse des mécanismes de sensibilité accélérométrique et barométrique des résonateurs à quartz de type QAS*, 1st European Time & Frequency Forum, Besançon, 1987, pp. 195-203.

[2] J.P. Aubry, B. Bonnotte, J.C. Craveur, F. Deyzac, D. Janiaud, *Optimisation de l'orientation des résonateurs autosuspendus*, 2nd European Time & Frequency Forum, Neuchatel, 1988, pp. 687-699.

[3] E. Gérard, R. Molle, *Design and qualification of an ultra stable oscillator for space applications*, 5th European Time & Frequency Forum, Besançon, 1991, pp. 347-357.

[4] M. Valdois, J.J. Gagnepain, J. Besson, *Influence of environment conditions on a quartz resonator*, 28th Annual Symp. on Frequency Control, 1974, pp. 19-32.

[5] D. Janiaud, L. Nissim, J.J. Gagnepain, *Analytical calculation of initial stress effects on anisotropic crystals*, 32th Annual Symposium on Frequency Control, 1978, pp. 169-179.

[6] R. Besson, J.J. Gagnepain, D. Janiaud, M. Valdois, *Design of a bulk wave resonator intensitive to acceleration*, 33th Annual Symposium on Frequency Control, 1979, pp. 337-345.

[7] J.P. Aubry, A. Debaisieux, *Further results on 5 MHz and 10 MHz resonators with BVA and QAS designs*, 38th Annual Symposium on Frequency Control, 1984, pp. 190-200.

[8] R. Besson, *A new piezoelectric resonator design*, 30th Annual Symposium on Frequency Control, 1976, pp. 78-83.

Appendix I - Elastic coefficients rotations

Let C_{mnpq} and C_{pqrstu} be two tensors, fourth and sixth order. One can compute C_{ijkl} and C_{ijklmn} after one rotation by:

$$C_{ijkl} = a_i^m a_j^n a_k^p a_l^q C_{mnpq} \quad \text{with } m,n,p,q = 1,2,3$$

$$\text{and } C_{ijklmn} = a_i^p a_j^q a_k^r a_l^s a_m^t a_n^u C_{pqrstu} \quad \text{with } p,q,r,s,t,u = 1,2,3$$

The first two rotations (ϕ,θ) define the cut plane (Fig. I.1):

$$\underline{1\text{st rotation}}\ \phi \text{ about } x_3 \begin{vmatrix} \cos\phi & \sin\phi & 0 \\ -\sin\phi & \cos\phi & 0 \\ 0 & 0 & 1 \end{vmatrix} \qquad \underline{2\text{nd rotation}}\ \theta \text{ about } x_1' \begin{vmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & \sin\theta \\ 0 & -\sin\theta & \cos\theta \end{vmatrix}$$

The third rotation (ψ) corresponds to the azimuth angle of the quartz plate (that is EF axes XYZ) in the cut plane (ϕ,θ):

$$\underline{\text{3rd rotation}} \quad \psi \text{ about } x_2'' \qquad \begin{vmatrix} \cos\psi & 0 & \sin\psi \\ 0 & 1 & 0 \\ -\sin\psi & 0 & \cos\psi \end{vmatrix}$$

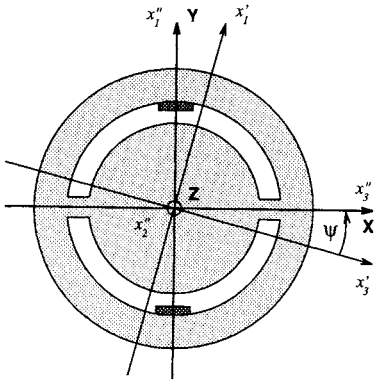


Figure I1 - Azimuth angle (psi) of QAS in cut plane

Appendix II - Propagation directions

Thickness shear vibration in a plate is well known:

$$\rho_0 \frac{\delta^2 u_i}{\delta t^2} = \frac{\delta T_{ij}}{\delta a_j} \qquad \begin{array}{l} \rho_0 = \text{density} \\ T_{ij} = \text{stresses in natural state} \\ u_i = \text{displacements} \end{array} \qquad (1)$$

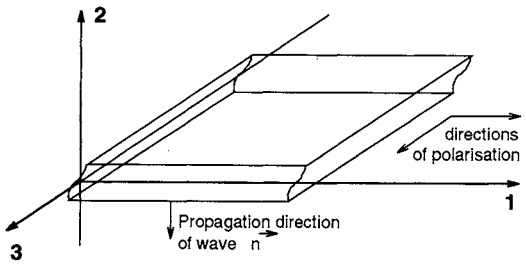


Figure II.1 - Thickness shear vibration in a plate

Assuming that damping and piezoelectricity are negligible, stresses can be obtained by $T_{ij} = C_{ijkl}\, S_{kl}$ and strains S_{kl} by the relation:

$$S_{kl} \frac{1}{2}\left(\frac{\delta u_k}{\delta a_l} + \frac{\delta u_l}{\delta a_k}\right)$$

The relation (1) becomes:

$$\rho_0 \frac{\delta^2 u_i}{\delta t^2} = C_{ijkl} \frac{\delta^2 u_k}{\delta a_i \delta a_k} \qquad (2)$$

We look for a solution in the form of a plane wave:

$$u_i = u_i^0 \exp\left[j\omega\left(t - \frac{\vec{n}\cdot\vec{a}}{v}\right)\right] \qquad \begin{array}{l} v = \text{speed of wave} \\ a = \text{position vector} \\ n = \text{propagation direction} \end{array}$$

and, putting u_i in relation (2), it comes $\rho_0 v^2 u_i = C_{ijkl} n_l n_j u_k^0$ where $C_{ijkl}\, n_l\, n_j$ is the Christoffel tensor Γ_{ij}:

$$\rho_0 v^2 u_i^0 = \Gamma_{ik} u_k^0 \qquad (3)$$

Then it's easy to show that polarization vectors (that are particle displacement directions) are eigen vectors of the Christoffel tensor. With $u_i^0 = u_k^0 \cdot \delta_{ik}$, relation (3) becomes:

$$u_k^0 \left(\Gamma_{ik} - \rho_0 v^2 \delta_{ik}\right) = 0 \qquad (4)$$

Calculation is then restricted to transverse excitation:

$$\vec{n} = \begin{vmatrix} 0 \\ 1 \\ 0 \end{vmatrix} \rightarrow n_s \cdot n_r = n_2^2 = 1 \rightarrow \Gamma_{ik} = C_{i2k2}$$

and by condensing sub-signs we have:

$$\Gamma = \begin{vmatrix} C_{66} & C_{62} & C_{64} \\ C_{62} & C_{22} & C_{24} \\ C_{64} & C_{24} & C_{44} \end{vmatrix}$$

Appendix III - Computation of $\tilde{k}_i$ coefficients

Natural state - Stressed state:

The plate defined in appendix II is now submitted to external forces; the equation of motion becomes:

$$\bar{\rho} \frac{\delta^2 u_i}{\delta t^2} = \frac{\delta}{\delta x_s}\left(\bar{B}_{iskr} \cdot \frac{\delta^2 u_k}{\delta x_r}\right) \qquad (5)$$

while in natural state the relation is:

$$\rho_0 \frac{\delta^2 u_i}{\delta t^2} = \frac{\delta}{\delta x_s}\left(C_{iskr} \cdot \frac{\delta u_k}{\delta x_r}\right)$$

This leads to write $\bar{\Gamma}_{ik}$, modified Christoffel tensor:

$$\bar{\Gamma}_{ik} = \bar{J} \cdot \bar{B}_{iskr} \cdot n_s \cdot n_r \qquad \bar{J} = \frac{\rho_0}{\bar{\rho}}$$

Perturbation

Assuming that resonant frequency can be written:

$$f(m) = \frac{M}{2e}\sqrt{\frac{C(m)}{\rho_0}} \qquad \begin{array}{l} M = \text{overtone} \\ e = \text{plate thickness} \\ C = \text{eigen value of m mode} \end{array}$$

The fractional deviation due to initial stress is expressed by:

$$\frac{df(m)}{f(m)} = -\frac{de}{e} + \frac{1}{2}\cdot\frac{dC(m)}{C(m)} \qquad (6)$$

where the variations of eigen values are determined with a perturbation method:

$$\bar{\Gamma}_{ik} = \Gamma_{ik} + \Delta_{ik}$$

and $\bar{C}(m) = C^0(m) + dC(m)$ ← eigen vectors

Since propagation directions are not really modified by initial stresses (eigen vectors $U^0 \simeq U$), then:

$$dC(m) = \Delta_{ik} . U_i^0(m) . U_k^0(m)$$

The fractional thickness variation de/e is equal to transversal strain expressed in EF axes $\bar{S}_2''$, and finally:

$$\frac{df(m)}{f(m)} = -\bar{S}_2'' + \frac{1}{2C^0(m)} . \Delta_{ik} . U_i^0(m) . U_k^0(m) \qquad (7)$$

$\tilde{k}_i$ coefficients computation:

In relation (7) we have expressed Δ_{ik} :

$$\Delta_{ik} = \left(J \cdot B_{iskr} - C_{iskr}\right) \cdot n_s \cdot n_r$$

or $\Delta_{ik} = J \cdot B_{i2k2} - C_{i2k2}$ for $n_s \cdot n_r = n_2^2 = 1$

General form of Δ_{ik} becomes:

$$\Delta_{ik} = \bar{J}\cdot\delta_{ik}\bar{\Gamma}_2 + C_{i2k2uv}\bar{S}_{uv} + C_{i2k2q}\frac{\delta\bar{u}_2}{\delta a_q} + C_{imk2}\frac{\delta\bar{u}_2}{\delta a_m} + C_{j2k2}\frac{\delta\bar{u}_i}{\delta a_j} + C_{i2p2}\frac{\delta\bar{u}_k}{\delta a_p} + C_{i2k2}\left(\bar{S}_1 + 2\bar{S}_2 + \bar{S}_k\right)$$

with the condition in EF axes: $\Delta_{ik} = \Delta_{ki}$

The relation (7) takes the form:

$$\frac{\Delta f}{f} = -\sum_{j=1}^{6} S_{2j} . T_j + \frac{1}{2C_{22}} \sum_{i=1}^{3}\sum_{j=1}^{3} \Delta_{ik} . U_i . U_k$$

leading to:

$$\frac{\Delta f}{f} = \sum_{i=1}^{6} \tilde{k}_i . T_i$$

LOW AGEING HC - 36/U ENCAPSULATED THIRD-OVERTONE AT-CUT 16.384 MHz CRYSTAL RESONATORS AND OSCILLATORS

K.K. Tuladhar and G. Jenni.
OSCILLOQUARTZ SA, NEUCHATEL, CH - 2002.

ABSTRACT

Although the ageing phenomena is one of the most researched topics in the history of quartz crystal resonators and oscillators, the ageing of resonators is still not well understood. To minimise ageing, a number of researchers in this fields have resorted to SC-Cut, ceramic package and associated mounts, QAS design etc. However the frequency was often limited to 10 MHz.

The aim of this paper is to demonstrate low-ageing resonators and oscillators (i.e better than 5×10^{-10} / day) up to 16.384 MHz using conventional HC-36/U encapsulated third-overtone AT-Cut crystals. This has been made possible by using an optimum resonator design and improved processing techniques.

And finally it will be shown (with statistical data) that the low ageing resonators and oscillators have not only been realised in the controlled laboratory environment but also in the routine production environment as well.

Keywords : Ageing, AT-Cut Resonators and Crystal Oscillators.

1. INTRODUCTION

Oscillator ageing has been defined as the systematic change in frequency with time due to internal changes in the oscillator, when factors external to the oscillator are kept constant (Refs. 1, 2). The ageing phenomena is still not well understood although it has been one of the most researched topics in the history of quartz crystal resonators and oscillators.

In this paper firstly some of the known ageing mechanism and factors affecting ageing will be given. Then a number of methods of minimising ageing during design phase will be described by taking into consideration of the constraints imposed by the practical resonator / oscillator.

A number of researchers in the field have resorted to SC - Cut (Ref. 3). Ceramic package and associated mounts (Ref. 4), QAS design (Ref. 5). However, often the frequency was limited to 10 MHz for the ageing (in the production environment) better than 5×10^{-10} / day. One of the main aims of this paper is to demonstrate low ageing resonators/oscillators up to 16.384 MHz using conventional HC - 36/U encapsulated third-overtone AT - Cut Crystals. This has been made possible by using an optimum resonator design at 16.384 MHz and improved processing techniques. And finally it will be shown (with statistical data based on some 400 units) that the low-ageing resonators/oscillators have not only been realised in the control environment but also in the routine production environment as well.

2. AGEING MECHANISMS AND FACTORS AFFECTING AGEING

Although the ageing mechanism and the factors affecting ageing behaviours of the crystal resonators / oscillators are described in detail recently by Vig and Meeker (Ref. 6), however the listing of these are done in an ad-hoc fashion without any clear sequences to resonator / oscillator design and processing. In this section, the above shortfall will be avoided. The listing in a proper sequence is as follows :

2.1 Quartz Material and Types of Cuts

a) Imperfections, Surface and Point defects, Dislocations, Impurities, Inclusions, Twins.
b) Outgasing of Quartz and Changes in Quartz Material.

2.2 Resonator Design

a) Frequency and Overtone Mode
b) Shape and Size of the Quartz Blank and Electrodes.
c) Contour and Dimensional Ratios.
d) Operating Temperature Range and $df/f \approx T$ Curve and Its Turn-over-Temperatures.
e) Electrode Material and Its Purity.
f) Plateback, ie Fractional Frequency Change.

Note that for the state-of-the-art resonator, the resonator designer must also have a clear understanding of the oscillator life-period, frequency pulling, etc. And similarly the oscillator designer must also be aware of the resonator limitations.

2.3 Resonator Processing and Packaging

2.3.1 Quality and cleanliness of the finished blanks. This depends upon Sawing, Lapping, Etching, Contouring processes and in between cleaning processes.

2.3.2 Metallisation

a) Deposition rate, Vacuum and Vacuum System, Gas Impurities present during deposition.
b) Adhesion between the quartz and electroded material.
c) Curing Temperature for stress relief and methods of curing.

2.3.3 Mounts and Mounting Techniques

a) Strain and stresses in the resonator due to mount type, bonding paste, location of mounts or clips (Stress free mounting is desired !).
b) Methods of outgassing the paste.
c) Methods of cleaning metallised and mounted crystals before final frequency adjusting (including UV cleaning if necessary).

2.3.4 Final Frequency Adjusting

Methods of depositing extra material on the electrode in the clean Ultra - High Vacuum environment.

2.3.5 Sealing Techniques

Methods of sealing final frequency adjusted crystal under vacuum after high-temperature baking (pressure changes in the resonator should be avoided due to leaks and outgassing).

2.3.6 Miscellaneous

Even after all the care taken to manufacture resonators, the following effects are still possible.

a) Contamination Transfer Effects, Adsorption, Desorption, Oxidation, Permeation.
b) Electrode Diffusion Effects.
c) Chemisorption due to Gases present inside the enclosed package.

2.4 Oscillation Maintaining Circuit Related

a) DC - Voltage on the quartz.
b) Drive level to quartz plus its fluctuation due to AGC instability.
c) Tuning capacitors (Electronic and / or Mechanical) and ageing of these capacitors.
d) Oven set-point temperature and its inherent instability and ageing.
e) Other electrical changes such as change in load reactance due to matching problem, power supply fluctuation.

2.5 Operating Environment related (for both resonators & oscillators)

a) Pre-ageing and acceleration of ageing.
b) Temperature cycling, rate of change of temperature and temperature range.
c) Low and high temperature soaking.
d) Interruption frequency and duration.

3. AGEING MINIMISATION

In general a reactive component is used in series with the crystal resonator, firstly to set the oscillator frequency to a desired frequency and then the same reactive component is used to re-adjust the oscillator frequency shift due to ageing. In the-state-of-the-art oscillator the same reactive component stability will also be a factor in achieving low ageing.

Now consider the well known equation,

$$\delta f = \frac{df}{f} = \frac{C_1}{2(C_0 + C_L)} \quad \cdots (1)$$

Where C_1 = motional capacitance of the crystal
C_0 = static capacitance of the crystal
and C_L = series capacitance

The same equation can be re-written,

$$\delta f = \frac{df}{f} = \frac{1}{2} \cdot \frac{1}{C_0/C_1} \cdot \frac{1}{(1 + C_L/C_0)} \quad \cdots (2)$$

This equation (2) clearly indicates to have low δf due to change in C_L, one must choose a resonator having a large C_0 / C_1 and must also use highest possible C_L value.

The ratio, static to motional capacitance, $r = C_0/C_1$ is in fact one of the fundamental properties of the crystal resonator and for the practical plano-plano fundamental AT-Cut resonator, the value is somewhere in the range 180 to 240. And for the contoured and overtone mode (m) AT-Cut crystal resonator,

$$r_m = k\, m^2\, r_1 \quad (3)$$

Where the value of k is dependent on the contour (for plano-plano case k = 1) and could be greater than 5 for the practical third and fifth overtone AT-Cut Plano / Bi - convex crystal resonators.

The effect of minute change (ie. instability) in C_L due to ageing of C_L or due to some other conditions has been apply demonstrated for the plano-plano fundamental crystal at about 20 MHz (Ref. 7).Now consider a plano-plano AT-Cut 20 MHz third-overtone crystal (Ref. 8), the pulling sensitivities of the same crystal at fundamental and fifth overtone modes are given in Table 1. This clearly demonstrates that the third and fifth overtone modes are 8 and 23 times less sensitive than the fundamental mode. Similarly consider a typical plano-convex fifth-overtone AT-Cut crystal at 7 MHz (Refs. 9,10,11), the pulling sensitivity at fifth and the pulling sensitivity of the same crystal at third-overtone are given in Table 1. This again

clearly demonstrates the plano-convex third and fifth overtone crystals are 12 and 54 times less sensitive than the plano-plano fundamental crystal

Now consider a typical AT-Cut crystal, pulling curve, Fig. 1. The chosen value for C_L will depend on three factors, A, B and C.

The pulling requirement for a given oscillator is governed by the ageing of the oscillator and its life time. However, the final frequency adjusting tolerances at the turn-over temperature is dependent on final frequency adjusting processes. And therefore one could optimise the value of "C_L" for minimum pulling sensitivities by adjusting the crystal to as precisely as possible. Finally, the design margin will be dependent on one's confident levels with respect to overall resonator / oscillator designs, measurements and operations.

Note also that a resonator (AT-or-SC-Cut or lateral-*field)* although might have a low ageing having a large C0 / C1 ratio, but without being able to adjust the oscillator frequency either to its desired frequency and / or to re-adjust to account for the frequency shift due to ageing for say 10 years, is of no use for the state-of-the-art oscillator applications.

4. A BRIEF DISCUSSION ON 16.384 MHz RESONATOR

Although it has been demonstrated that in general at a given frequency fifth overtone crystal will have a better ageing performance than the third overtone crystal. However, in this paper only the third-overtone crystals will be considered. Currently in the commercial and routine production environment the crystal oscillators having ageing better than $5x10^{-10}$ / day using third - overtone crystals are firstly limited in the frequency range (usually up to 10 MHz) and often utilised SC-Cut crystals, QAS design with associated mounts, fixtures and packaging. The main aim of this paper is to push the frequency to 16.384 MHz (one of the most commonly used frequencies) and to utilise existing but improved processing techniques and to maintain the ageing still better than 5×10^{-10} / day. Therefore in this section a brief discussion on 16.384 MHz third-overtone AT-Cut resonator and the associated processing will be described.

4.1 DESIGN

The design of third-overtone 16.384 MHz Plano-Convex AT-Cut crystal resonator has been carried out using the techniques described in the published papers (Refs. 9,10,11). The resonator has C0/C1 ratio of about 6500. And the resonators have been processed using Oscilloquartz optimum processing techniques and where-ever possible BVA resonator processing techniques were also utilised (Refs. 12,13). These processes include :

- Use of pure Z - Quartz material (with minimum inclusions) having Q greater than 2.5 M.
- Use of angle-correction techniques for quartz plaquettes produced by the sawing machine.
- Laped, etched contoured and polished with due care to keep the finished blanks as clean as possible.
- Metallisation of gold electrodes on the quartz substrate were carried out using clean vacuum deposition processes.
- HC - 36/U holders were used to mount metallised blanks using epoxy paste, and these were outgassed in an optimum way.
- Final frequency adjustments were carried out with a good precision (see results below) in clean high-vacuum environment.
- As pointed out by Vig and Meeker (Ref. 6), high temperature baking has also been carried out during resonator fabrication before sealing.
- And finally crystals were sealed using proper and clean HC - 36 /U covers in clean high-vacuum environment using cold weld press.
- Encapsulated crystals are usually pre-aged at an elevated temperature for weeks before evaluation.

In general, the demand made on crystal resonators for the state-of-the-art oscillator applications will be rather stringent and therefore crystal resonators must be free of Activity-Dips or Bandbreaks (Refs. 14,15) and must also not suffer from Drive-Level-Dependence, DLD or Second-Level-of-Drive, SLD (Ref. 16). Methods exist to screen resonators for Bandbreaks or Activity Dips (Ref. 17) and DLD (or SLD) effects (Ref. 18). Some preliminary theoretical modellings (Ref. 19) have also been carried out. Although screening processes have been widely used in the crystal industry, the authors of this paper strongly believe, these screening processes are rather combersome and costly and to be carried out only during resonator development phase. The problems related to DLD and Bandbreaks must be solved if necessary iteratively for the-state-of-the-art oscillator application. These problems can be solved for most resonators (have been solved for the third overtone 16.384 MHz AT-Cut Crystal Resonators) as the DLD effects are dependent on crystal manufacturing and cleaning processes. And similarly activity dips (bandreaks) are dependent on the resonator design parameters, size and shape of the quartz blank and electrodes, radius of curvature of the plano-convex crystal blank, plateback (i.e fractional change in frequency due to mass-loading), mounting locations and numbers of mounts & mount types. Finally resonators must also have low hysteresis effects (Refs. 20, 21).

4.2 MEASURED RESONATOR PARAMETERS

Currently third-overtone AT-Cut resonators at 16.384 MHz are produced in small volume (about 60 pieces) at a time. The resonator parameter measurements at the upper-turn-over temperature,

Pi are usually carried out using well known IEC-444 techniques. The mean and standard deviation of the resonator parameters for a typical batch are given below :

	mean ($\bar{x}$)	stdv (σ)
1. Pi (°C)	88.0	1.0
2. $\Delta f/f$ (ppm)	0.3	0.5
3. C_1 (fF)	0.67	0.007
4. Q	680 K	37 K

the above data clearly demonstrates :

1. angle tolerances and mounting positions were well controlled.

2. final frequency adjustments were carried out with good precision, as ($\bar{x} \pm 3\sigma$) value of $\Delta f/f$ is less than 2 ppm.

3. as the standard deviation, σ of the motional capacitance is only 0.007 fF ($\cong$ 1 % of the mean), the radius of the curvature of the quartz blank was well controlled and the electrode definition was also good.

4. one way of testing a goodness of the crystal resonator is to find the product of the frequency and the quality factor. For the state-of-the-art fifth over- tone 5 MHz conventional or BVA AT-Cut resonator, the value is about $(5 \times 10^6) \times (2.5 \times 10^{+6})$ [$= 12.5 \times 10^{+12}$]. Now similar product for our third-overtone 16.384 MHz resonator is $(680 \times 10^3) \times (16.384 \times 10^6)$ [$= 11.14 \times 10^{12}$]. This value is 89 % of the value for the fifth-overtone 5 MHz resonator. Therefore, we can consider this third-overtone resonator is of an optimum design.

Although we have measured ageing on the limited number of crystal resonators during development phase, currently we do not measure ageing on the resonators. The whole resonator selection procedure for the third-overtone 16.384 MHz crystal resonators is based on the resonator parameters, as we carry out quasi-vertical integration within Oscilloquartz as far as the crystal oscillator production is concerned.

5. DISCUSSION ON MEASURED AGEING RESULTS

Third overtone 16,384 MHz AT-Cut crystals have been installed in oven-controlled-crystal oscillators (OSA Model No 8660 series). The oscillator frequency is usually measured four-times a week to characterise the ageing behaviour of the oscillators. Once the slope of the ageing curve (linear regression between the last 4 sets of data) is less than 5×10^{-10} / day, the ageing measurement is stopped and the oscillators are put-in the stock for shipment.

First consider the distribution, Fig. 2., the time taken to meet the ageing of 5×10^{-10} / day for some 403 oscillators / resonators. Fig. 2 clearly demonstrates that some 88 % met the ageing of 5×10^{-10} / day within first 30 days. It can also be seen that around 32 % met the ageing of 5×10^{-10} / day during the periods, 13 to 18 and 25 to 30 days. Note also that some 12 % met the ageing during first twelve days and remaining 12 % took more than 30 days.

Now consider the detailed ageing data for the two batches meeting the ageing of 5×10^{-10} / day during 13 to 18 days and 25 to 30 days, Figs. 3a and 3b. Although majorities have positive ageing (in general oscillators having positive ageing are preferred to the ones having negative ageing), some 5 % have negative ageing. This could have been caused due to a variety of reasons and are currently under investigation and will be reported in due course. One point to note is that the mean (see Figs 3a and 3b) for both batches is nearly the same, however some 40 % have ageing in the range 4 to 5×10^{-10} / day for the case of Fig. 3a and only 25 % for the case of Fig. 3b.

Typical ageing curves for two resonators / oscillators during first 16 and 29 days having mean ageing of about 3×10^{-10} / day are given in Figs. 4a and 4b. Medium term ageing for the resonators / oscillators taking more than 30 days to meet 5×10^{-10} / day have been given in Figs. 5a and 5b. These two figs, 5a and 5b show the ageing slope decreased to 2.9×10^{-10} / day from about 4.6×10^{-10} / day after further ageing.

6. CONCLUSIONS

The crystal resonator material, design, manufacturing techniques & process related factors affecting the ageing and relevant ageing mechanisms are listed in a logical way. It has been demonstrated that the precision final frequency adjustment is required to minimise ageing and to choose optimum and higher value load capacitance. The load capacitance or any other load reactances must also be stable one for the state-of-the-art oscillator application.

All aspects of the crystal resonator such as material, design, and packaging, although could be conventional, but processing techniques must be improved and optimised where-ever possible to achieve better ageing. It has also been high-lighted that the troublesome activity-dips (or bandbreaks) and Drive-Level-Dependence (or Second-Level-of-Drive) effects associated within the crystal resonator must be solved during resonator development phase and screening for bandbreaks and DLD is not really satisfactory. A method of testing goodness of the crystal resonator by evaluating the product, frequency & Q- factor and then the way of comparing the result with a similar product of the state-of-the-art fifth-overtone 5 MHz has also been given in this paper.

Based on some 400 resonators / oscillators, it has been demonstrated that the ageing better than 5×10^{-10} / day has been achieved for HC - 36/U encapsulated third-overtone AT-Cut crystals. The ageing statistics also clearly demonstrated that some 88 % achieved ageing better than 5×10^{-10} / day

within first 30 days. And therefore it is safe to conclude that the resonators / oscillators below 16.384 MHz (at third-overtone) would have even better ageing characteristics. In the near future, we hope to publish statistically significant ageing data for other resonators / oscillators operating in the frequency range 4 to 16 MHz.

7. ACKNOWLEDGEMENT

The authors wish to thank Mr. Y. Schwab of Oscilloquartz for carrying out most of the ageing measurements.

8. REFERENCES

1. CCIR Recommendation No. 686, "Glossary", CCIR 17th Plenary Assembly, Vol VII, "Standard Frequency and Time Signals (Study Group)", 1990.

2. "IEEE Standard Definitions of Physical Quantities for Fundamental Frequency and Time Metrology", IEEE, Std. 1139-1988, IEEE NY, 10017,

3. Kuster, J.A, and Adams, C.A. "Production Statistics of SC (or TTC) Crystals", Proc. 34th AFCS, May 1980, pp. 167-174.

4. Wilcox, P.D, Snow. G.S, Hafner, E, and Vig, J.R, "A New Ceramic Flat Pack for Quartz Resonators", Proc. 29th AFCS, May 1975, pp. 202-210.

5. Aubry, J.P, and Debaisieux, A,"Further Results on 5 MHz and 10 MHz Resonators with BVA and QAS Designs", Proc. 38th AFCS, June 1984, pp. 190-200.

6. Vig, J.R, and Meeker, T.R, "The Ageing of Bulk Acoustic Wave Resonators, Filters and Oscillators", Proc. 45th AFCS, May 1991, pp. 77-101.

7. Vig, J.R, "Quartz Crystal Resonators and Oscillators for Frequency Control and Timing Applications - A Tutorial", May 1990, SLCET - TR-88-1 (Rev. 3.3)

8. Werner, J.F, and Dyer, A.J, "The Relationship between Plateback, Mass Loading and Electrode Dimensions for AT-Cut Quartz Crystals having Rectangular Resonators Operating at Fundamental and Overtone Modes", Proc. 30th AFCS, June 1976, pp. 42-53.

9. Vangheluwe, D, "The Frequency and Motional Capacitance of Partially Contoured Crystal Resonators", Proc. 34th AFCS, May 1980, pp. 412-418.

10. Tiersten, H.F, and Smythe, R.C, "An Analysis of overtone mode in Contoured Crystal Resonators", Proc. 31st AFCS, June 1977, pp. 44-47.

11. Sauerbrey, G, "Investigation of Resonant Modes of Plano-Convex AT-Plates", Proc. 21st AFCS, April 1967, pp. 63-71.

12. Peier, U.R, and Jenni, G, "Space Qualified 100 MHz BVA-VHF-Resonator", Proc 2nd EFTF, March 1988, pp. 701-713.

13. Graf, E.P, and Peier, U.R, "BVA Quartz Crystal Resonator and Oscillator Production - A Statistical Review", Proc. 37th AFCS, June 1983, pp. 492-500.

14. Frank,C, "On Activity Dips of AT Crystals at High Level of Drive", Proc. 21st AFCS, April 1967, pp. 436-454.

15. Wood, A.F. B, and Seed, A, "Activity Dips in AT-Cut Crystals", Proc 21st AFCS, April 1967, pp. 420-435.

16. Knowles, J.E, "On the origin of the Second-Level-of-Drive Effects in Quartz Oscillators", Proc. 29th AFCS, May 1975, pp. 230-236.

17. Ballato, A, and Tilton, R, "Ovenless Activity Dip Tester", Proc. 31st AFCS, June 1977, pp. 102-107.

18. Yerna, J.S, "Resistance-Measurements of Quartz Crystals at very low drive levels", Proc. 38th AFCS, June 1984, pp. 499-506.

19. Drorsky, L, and Kinsman, R, "A Simple Model for Quartz Resonator Low Level Drive Sensitivity", Proc. 45th AFCS, May 1991, pp. 148-155.

20. Tuladhar, K.K, Adams, B. J, and Fletcher, E.D, "Hysterisis Effects in Quartz Crystal Resonators", Proc. IERE, International Conference on Frequency Control and Synthesis, University of Surrey, Guildford, UK, April 1987, pp. 19 to 27.

21. Kuster, J.A, and Vig, J.R, "Thermal Hysterisis in Quartz Resonators - A Review", Proc. 44th AFCS, May 1990, pp. 165-175.

TABLE I. FREQUENCY CHANGE DUE TO CHANGE IN CL DUE TO AGEING OR SOME OTHER CAUSES

OVERTONE MODE	FREQUENCY (MHz)	MOTIONAL CAPACITANCE C1(fF)	df/f DUE TO CHANGE IN CL by 10 ppm
PLANO - PLANO (Co = 2.5pF, CL = 20 pF)			
FUNDAMENTAL	≃ 20/3	8.00	15.80 E-10
THIRD	20.000	1.00	1.98 E-10
FIFTH	≃ 100/3	0.35	0.69 E-10
PLANO - CONVEX (Co = 3.5pF, CL = 20pF)			
THIRD	≃ 21/5	0.75	1.36 E-10
FIFTH	7.000	0.16	0.29 E -10

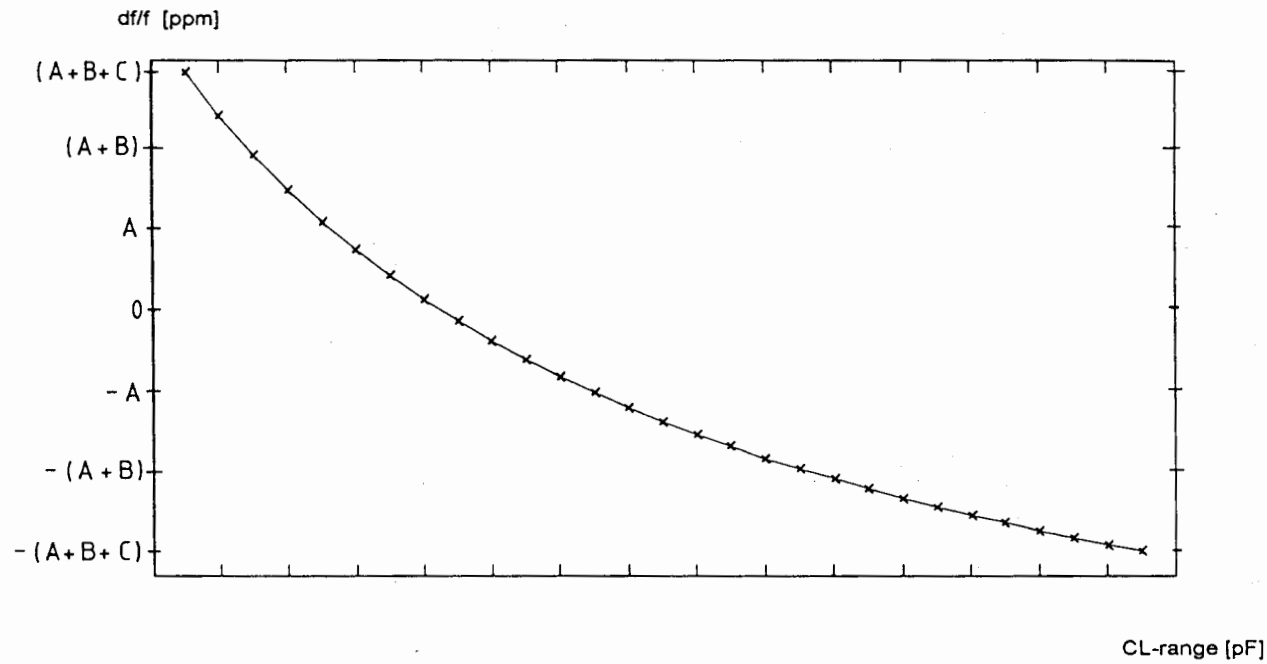


Fig. 1. TYPICAL FREQUENCY PULLING CURVE

A: FINAL FREQUENCY ADJUSTING TOLERANCE AT TURN-OVER-TEMPERATURE
B: FREQUENCY PULLING REQUIREMENT FOR THE GIVEN OSCILLATOR
C: DESIGN MARGIN

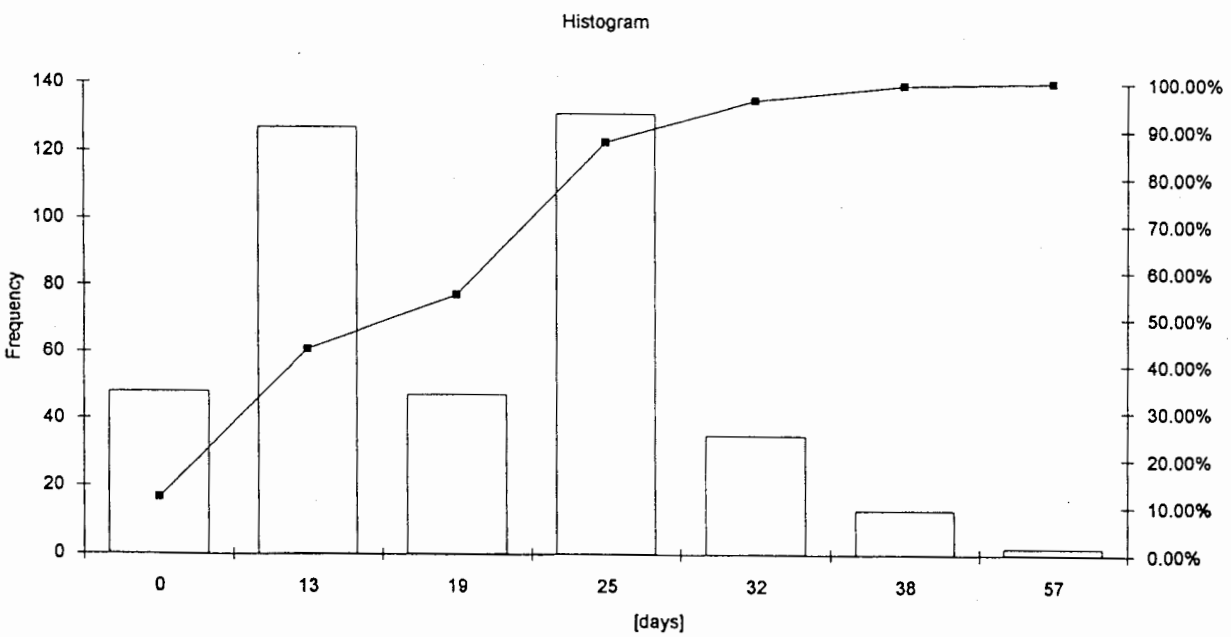


Fig. 2. Distribution for the resonators / oscillators having aging better than 5E-10 per day

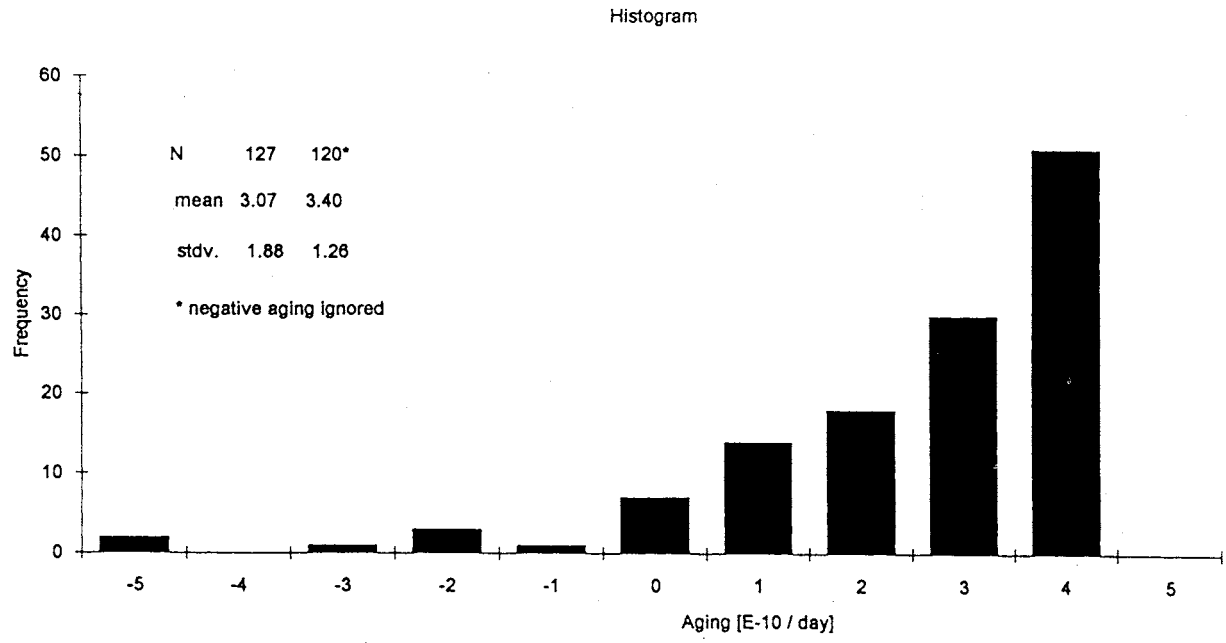


Fig. 3a. Distribution for the resonators / oscillators having aging less than 5E-10 / day during 13 to 18 days

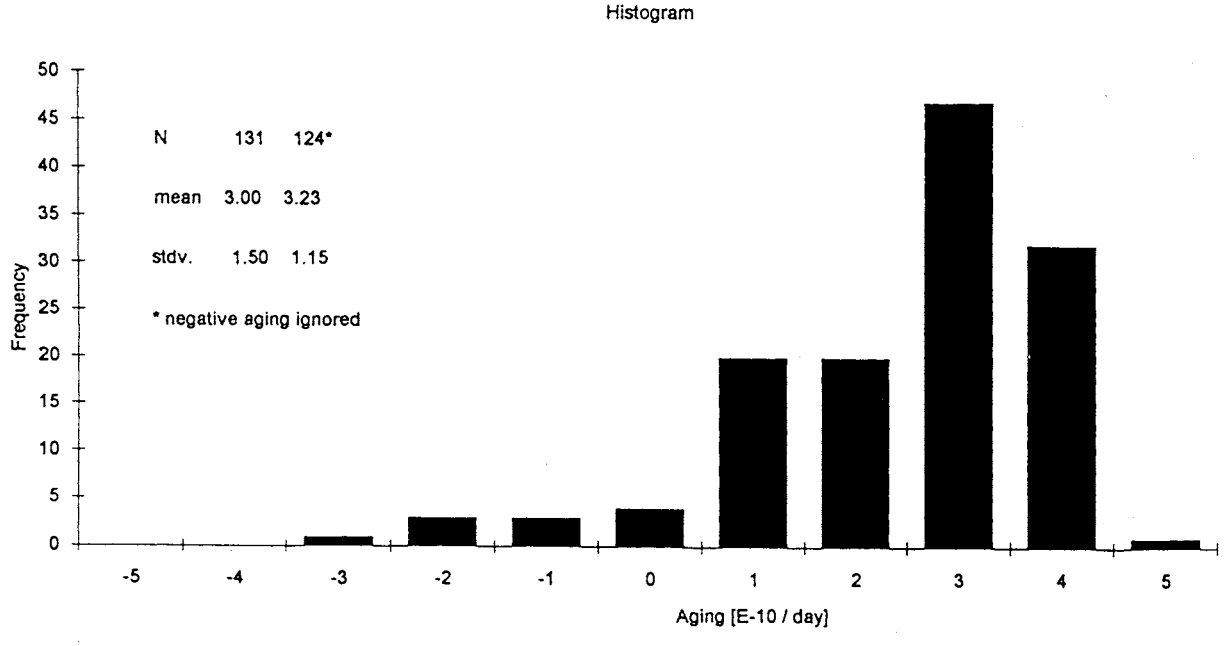


Fig. 3b. Distribution for the resonators / oscillators having aging less than 5E-10 / day during 25 to 30 days

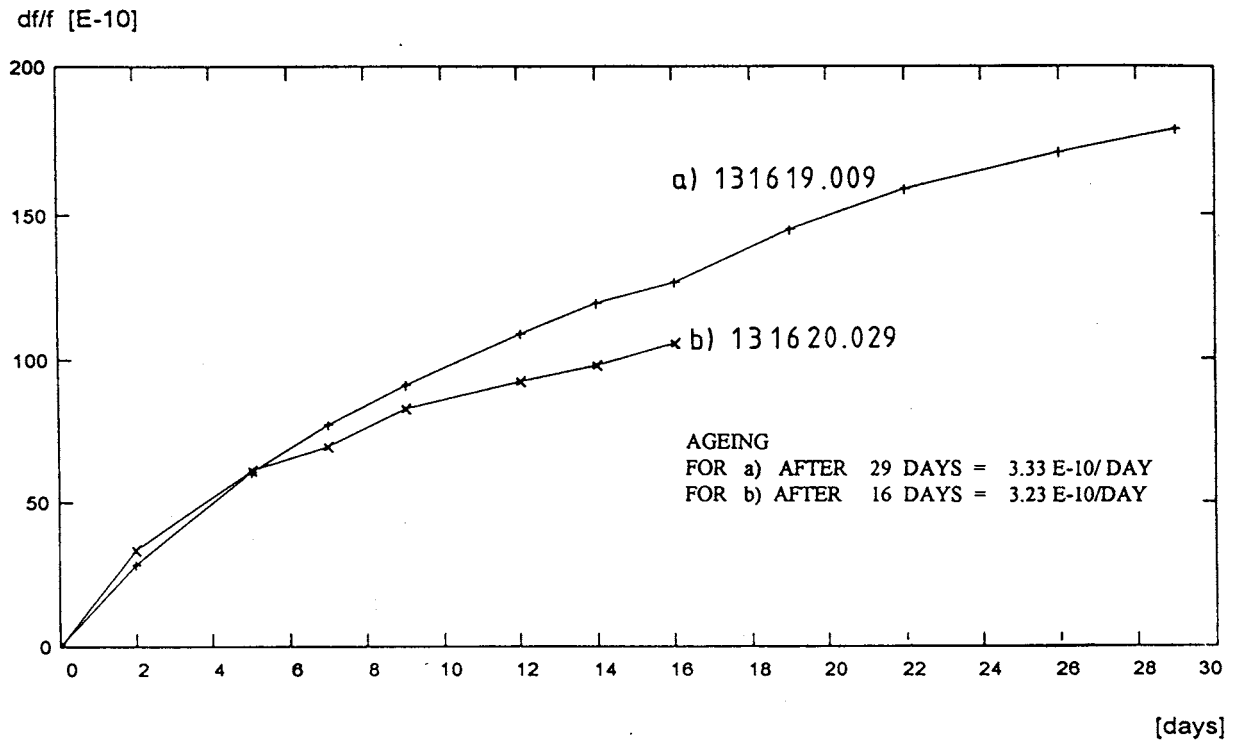


Fig. 4. Typical aging curves during first 16 and 29 days

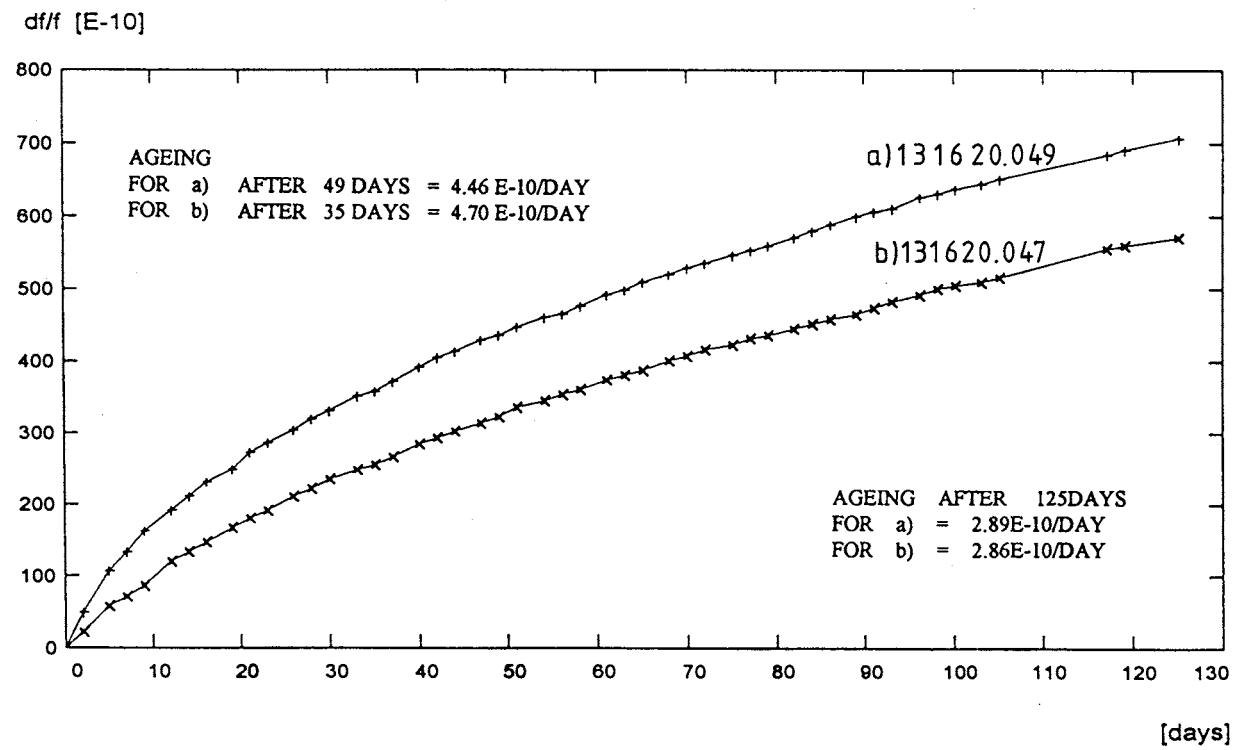


Fig. 5. Typical aging curves during first 125 days

STABILITY OF CR/AU LAYER ON QUARTZ SUBSTRATES

M. ZAFRANY AND F. MEYER*

THOMSON/CEPE
44, Av de la glaciere - BP 165
95105 Argenteuil - France

* INSTITUT D'ELECTRONIQUE FONDAMENTALE
UNIVERSITE PARIS-XI
91405 ORSAY

ABSTRACT

Stress relaxation and mass load effects are the major mechanisms involved in the ageing of quartz crystal resonators. The stability of Cr/Au electrode has been studied in terms of stress and resistivity measurements after different annealing temperatures. The diffusion of Cr atoms through the Gold layer is responsible of the variations of stress and resistivity; in particular a stress free interface has been evidenced.
The thermal stress calculation partly supports the results.

I/ INTRODUCTION

Long term stability of quartz crystal oscillators is attributed to a large extend to the performances of the resonator itself. The ageing of a quartz crystal is macroscopically defined as the change in frequency relative to the initial frequency over a given period. However on the microscopic scale a number of physical effects can take place in the resonator made of several materials, (i.e. quartz, metallic electrode, silver-filled adhesive polymer, nickel ribbon, copper cap and kovar base with gold, nickel-platinum, or copper clads), which have intrinsic characteristics that can be perturbed by the fabrication process and by the conditions of use in the final device.

A standard process for high stability resonators includes the following steps :

- Orientation and cutting
- Mechanical lapping and polishing
- Metallisation
- Chemical cleaning
- Mounting
- Annealing
- Frequency adjustment
- Sealing

Along this type of process, internal stress relief in the quartz introduced during lapping, inter diffusion at the metallic electrode, thermal stress during curing cycles of the conductive polymer, outgassing of the adhesive polymer, are some of the identified factors of instability.

Each of these effects causes a perturbation on the equilibrium state of the resonator which in turn modifies its resonance frequency.Miljkovic [1] and Dybwad [2] have proposed mathematical models to apply to the frequency shift, recorded during ageing, in order to go back to the time constant of the different physical processes involved.

The main effects suspected to have a significant influence on the ageing are : mass load related phenomenon and stress relaxation.

We limited our scope of interest at the quartz/metal interface stability and particularly to the Cr/Au metal layer used for high performances resonators.

Belser [3] indicated that resonators with electrodes of copper ,silver,aluminium or gold can yield a comparable ageing performance which is not improved by annealing. For these metals a penetration of 0.5 to 1 micron in the quartz substrate is evidenced, but the correlation with the frequency shift has not been clearly established.

Hoffman [4] gave a complete list of the properties of thin metal film on quartz crystals affected by the growth conditions. A special interest is given to stress at the interface consisting of two components : thermal stress due to the difference in thermal coefficients of expansion and intrinsic stress arising from the structure of the layer (grain

stress arising from the structure of the layer (grain size, grain boundary areas). The former is material choice dependent while the latter is expected to be controlled by the deposition parameters.

II/ EXPERIMENTAL

The samples were constituted by plano-plano AT cut blanks of .25 mm thick and 14 mm in diameter. The 100 A chromium layer and 600 A gold layer were deposited at room temperature in a classical vacuum system, supplied with a liquid nitrogen trap, under a pressure of 5 10^{-6} Torr.

Different annealing treatments were performed at temperatures ranging from 250 °C to 450 °C for 12 hours in a high vacuum system (10^{-6} Torr).

Resistivity measurements were performed using a four-probes system. To insure an ohmic behaviour of the contact, the following procedure is applied for each sample : the voltage between the two inner probes is recorded for a given direct current delivered by the two outer probes. Then the current is doubled and the voltage is checked to double.

Stress in the electrode was determined from the change in the deflection curvature of the quartz substrate after film deposition. The curvature of the substrate was measured using a stylus profilometer.

When the film thickness tf is small compared to the thickness of the substrate ts, the film average stress is given by the relation:

$$\sigma = Es / 6(1-vs) \ .ts^2/ tf \ .(1/R - 1/Ro) \qquad (1)$$

where Es and vs are Young's modulus and Poisson's ratio of the quartz substrate respectively.

Rc and Ro are the radius of curvature of the substrate after deposition and before deposition.

ts and tf are the thickness of the substrate and the film. Positive and negative values represent tensile and compressive stresses respectively.In our case, the metallic layer was removed by chemical etching in order to measure Ro.

III/ RESULTS AND DISCUSSION

3.1./ STRESS MEASUREMENTS

The as-deposited sample profile is shown on figure 1-a; the positive radius of curvature indicates a tensile stress which is usually observed for the deposition technique used here [3,5]. The sample thickness is directly measured around 700 A.

The same sample is measured after removing the layer, in order to estimate the initial radius of curvature (Figure 1-b). The surface of the substrate is almost flat within experimental precision;from this result we take, for all the samples, Ro = ∞.

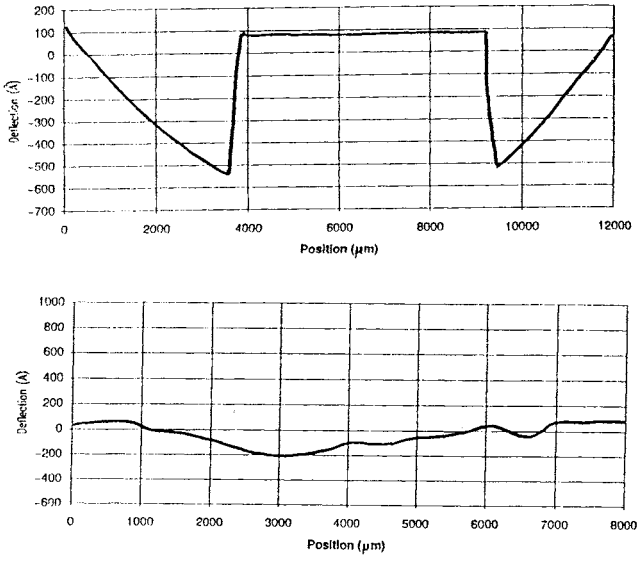


Figure 1a 1b

The 250 °C annealed sample exhibits a smaller radius of curvature compared to the as-deposited sample (Figure 2) but still of positive sign.

The 300 °C annealed sample is characterised by a relatively low Rc, but Figure 3 clearly shows a negative sign for Rc, i.e. the film passes from a tensile stress to a compressive stress. The Rc sign is still negative for the samples annealed above 300 °C (see figure 4 for Ta = 450°C) with a strong decrease of the radius of curvature with temperature.

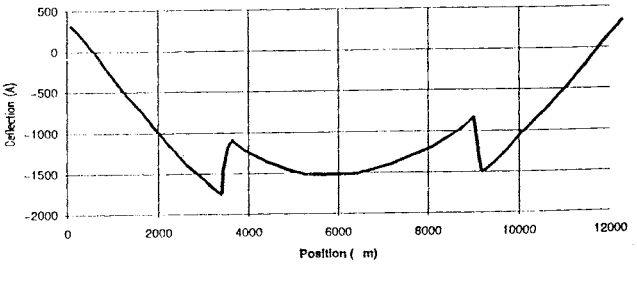


FIGURE 2

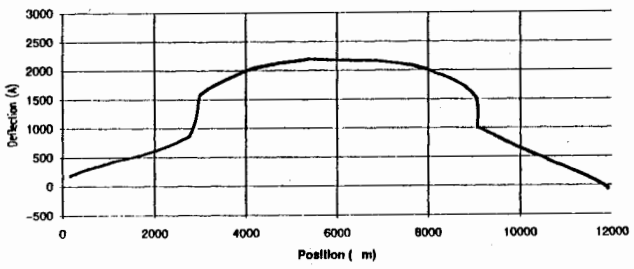


FIGURE 3

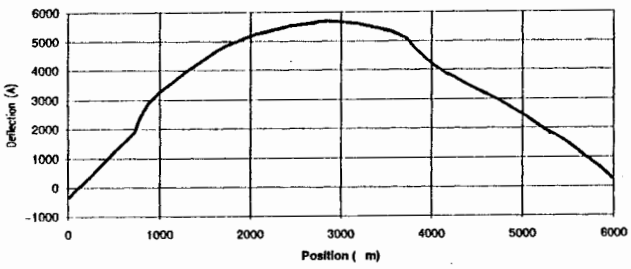


FIGURE 4

From the measurements presented figure 1 to 4, we investigate the shape changes of the electrode by substracting the deflection of the substrate. The results indicate that the electrode undergoes changes both on diameter and thickness - the diameter is taken at half maximum of thickness-(Figure 5). These changes have to be accounted for the computation of the stress and resistivity where tf, the film thickness is involved.

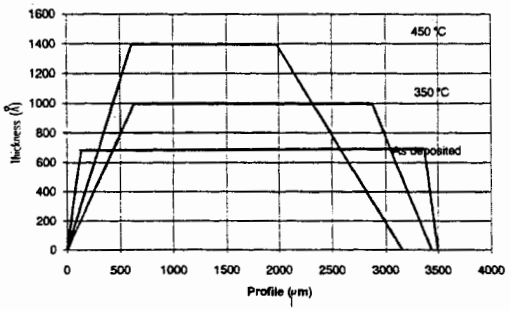


FIGURE 5

Stresses calculated from equation (1) are plotted versus temperature of annealing in figure 6. The film holds a tensile stress of 150 +/- 50 MPa up to 250°C. For higher temperatures the film has a compression stress of 100 +/- 30 MPa at 300 °C, estimated with a relatively good precision compared to the stresses at 350 °C and 450° (for which the thickness of the film is not precisely determined).

The transition from tensile to compression stress occurs at a temperature of 280 +/- 10 °C, meaning that for the samples studied here,the stress at the quartz/metal interface equals zero.

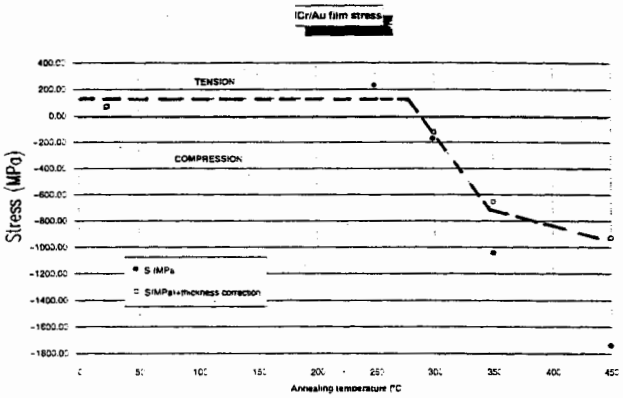


FIGURE 6

Vianco [5] located such a transition between room temperature and 380 °C since the compression stress measured after 380 °C and 450 °C treatments were found around 100 MPa and 200 MPa respectively, compared to the as-deposited tensile_ stress of 100 MPa.However the absolute values of stress are probably not comparable for the 450 A Cr / 1800 A Au deposited by electron gun evaporation at room temperature he studied and our samples.

3.2./ RESISTIVITY MEASUREMENTS

The resistivity dependence of the Chromium-gold layer with annealing temperature (Ta) is given in Figure 7. The resistivity of the gold film increases with Ta to a maximum value (80 µΩ cm) and then decreases to almost its original value (5 µΩ. cm).

The lower value corresponds to the 2.5 µΩ.cm of pure gold resistivity but the peak is largely exceeding the 14 µΩ.cm of pure chromium. Taking into account the thickness correction only affects the peak value at 83 +/- 15 µΩ cm.

The observation of the high temperature annealed samples evidences the formation of small aggregates on the gold layer surface .

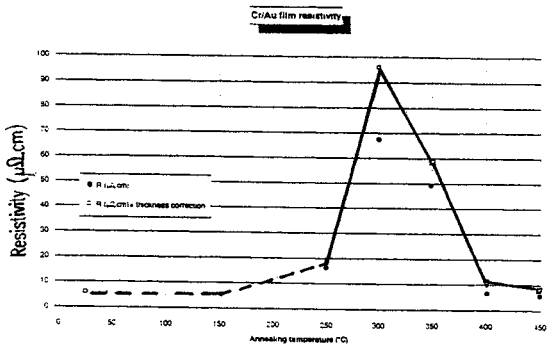


FIGURE 7

The variation of resistivity is attributed to the chromium diffusion into and trough the gold layer.

Vianco [5] identified this effect by Auger analysis for annealed samples and pointed out that the Cr is not oxidised in the gold layer although oxygen was also present. Hirvonen [6] observed, trough Rutherford Back-Scattering measurements ,a migration of the Cr atoms for temperature of annealing above 250 °C. The amount of Cr reaching the surface is almost independent of the annealing time at a given temperature.

The classification of the different parameters according to their importance in the resistivity changes is given by Gadenne [8] for $Cr_xAu_{(1-x)}$ co-evaporated alloys:

- The presence of oxygen at the grain boundary
- The formation of oxide interface
- The grain size of the film
- The chromium content

Munitz [7] proposed a complete comprehension (based on resistivity measurements, electron probe analysis and transmission electron microscopy) of the diffusion process summarised as follows : the chromium diffuses through the gold film and reacts with the oxygen at the top surface to produce Cr2O3 single crystals. This process continues until the chromium source vanishes. A peak value of 88 +/- 17 μΩ cm is thus reported to be in good agreement with the highest value achievable for this type of sample corresponding to the chromium solubility limit in Gold : 24 wt %.

3.3./ DISCUSSION

The transition temperature of annealing that yields a stress free interface corresponds to the peak resistivity position (figure 8).The inversion of the stress in the film is more likely due to the diffusion in the layer rather than an oxyde effect on the gold top layer since for the annealing time considered here (12 hours) the Cr_2O_3 oxyde is formed by discrete single crystals ([7]). In other words, a non continuous layer could not sustain any stress.

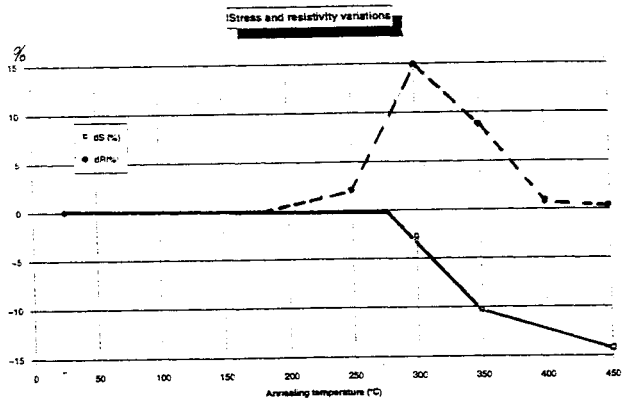


FIGURE 8

The thermal stress induced by annealing is investigated in order to have a better understanding of the behaviour of the Cr/Au film. Thermal stress, calculated from equation 2, is presented for the film studied here and for some other metals currently used in electronic thin films applications (Table 2)

$$\sigma th = \underbrace{Es / (1-vs) . (\alpha f - \alpha s)}_{\beta} . \Delta T \quad (2)$$

With Es : Young modulus of the film

vs : Poisson coefficient of the film

αf : linear expansion coefficient of the film

αs : linear expansion coefficient of the substrate

ΔT : difference between deposition temperature and Annealing temperature

	E (10^4 MPa)	vs	α (ppm/°C)	β (MPa/°C)
Au	4.2	0.45	14	0.137
Cr	32	0.16	4.5	-2.21
Mo	36	0.27	4.8	-3.67
W	39	0.28	4.5	-4.16
Ni	14	0.37	13.5	0.28
Ag	4.3	0.42	19	0.51
Al	6.3	0.37	23.2	1.1
SiO2	7.8	0.3	12.2	

TABLE 2 (FROM [9])

The annealing treatments provokes a tensile stress for Au film (β is positive) while a compressive stress is found for Cr film. Assuming a linear dependence of the coefficient of expansion during the diffusion of chromium atoms, the film is characterised by :

$$\alpha_{film} = (1-x).\alpha_{Au} + x.\alpha_{Cr} \qquad (3)$$

where x is the Cr content of the gold film.

$x=0.18$ offers a solution to have $\alpha_{film} = \alpha_{quartz}$ leading to a zero thermal stress on annealing.The existence of a stress free interface could be so demonstrated by the evolution an coincidence of the different linear coefficients of expansion.For $x < 0.18$ α_{film} is lower than α_{quartz}, leading to a tensile stress while for $x > 0.18$, the stress is compressive. The calculated stress values are comparable (Fig 9) with the measured ones for the two higher annealing temperatures (Ta > 300 °C); at lower temperatures (Ta < 250 °C), the stress value are of the same order of magnitude than that of pure gold layer.
This result correlated with the minimum diffusion temperature of Cr evaluated between 250 and 300 °C suggests that the diffusion of Chromium atoms controls not only the resisivity of the layer but also the stress at the interface.

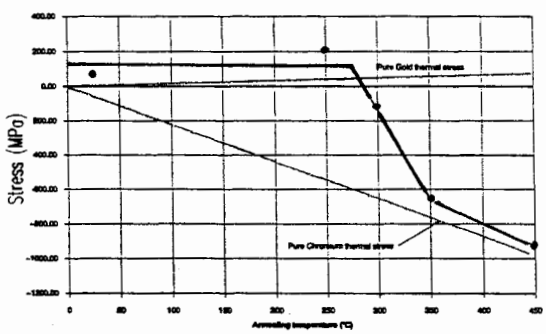


FIGURE 9

From a resonator design point of vue, it is to note that gold produces the minimum thermal stress among the metals currently used in resonator fabrication.

An evaluation of the quartz crystal orientation does not change the above description. The linear coefficient of expansion varies with θ from 8 to 14 ppm /°C [10], the amount of Cr atoms needed to have a stress free interface would extent from $x=0.63$ to $x=0$. The orientation for $\theta > 50°$ would need to exceed he solubility limit , that is to say for these orientations a stress free interface is not achievable.

IV/ CONCLUSION

The stability of the metallic electrode can be obtained in two opposite ways :

- Preventing diffusion and leave the electrode under constant stress. In this case ,using Mo as bonding layer or as diffusion barrier could be of interest, since no diffusion occurs through the gold layer under 400° C [5].
- Promoting the diffusion of Cr in order to reach a free stress interface; however the bi-layer is not in an equilibrium state, the Cr can still migrates at the surface to react with Oxygen.

REFERENCES

[1] M.R. Miljkovic, G.L Trifunovic and V.J Brajovic
Proc. 42 nd AFCS, 1988. ,pp 404
[2] G.L. Dybwad
Proc. 31nd AFCS, 1977. ,pp 144
[3] R. Belser and W.H. Hiccklin
Proc. 23 nd AFCS,1969. ,pp 132
[4] D.R. Hoffman
Proc. 28 nd AFCS,1974. ,pp 85
[5] P.T. Vianco, W.R. Conley, and J.K.G. Panitz
Proc. 44 nd AFCS,1990 . ,pp 132
[6] J.K. Hirvonen, W.H. Weisenber, J.E. Westmoreland, and R.A Meussner
Appl. Phys. Lett., Vol 21 ,No. 1, 1 July 1972
[7] A. Munitz and Y. Komem
Thin Solid Films, 37 (1976) 171-179
[8] P. Gadenne, C. Sella, M. Gasgnier, A. Benhamou
Thin solid Films, 165 (1988) 29-48
[9] A.Segmuller and M. Murakami
Treatise on materials science and technology vol 27- 1988 Academic Press
[10]V.E. Bottom "introduction to quartz crystal design"

VCTCXO for Synchronous digital hierarchy (SDH) applications

Wolfgang Balzer

KVG GmbH, Prodcutmanagement crystal components, D-6924 Neckarbischofsheim

Abstract

The network node interfaces of present digital hierarchy in Europe, North America and Japan are different from each other. Therefore direct interconnection of networks between these aeras are impossible.
Synchronous digital hierarchy (SDH) is a network node interfaces which allowes direct wordwide network interconnections, but also, at the same time sophisticated network operation ,as well as efficient transmission of future broadband ISDN(B-ISDN) signals. SDH was standardized by CCITT in 1989. Since then, efforts have been made to transform national trunk transmission systems into SDH-compatible systems.
The SDH application requires for the multiplexers stable oscillators with wide tunning range. This paper describes VCTCXO's with a special compensation network to fulfill the SHD application.

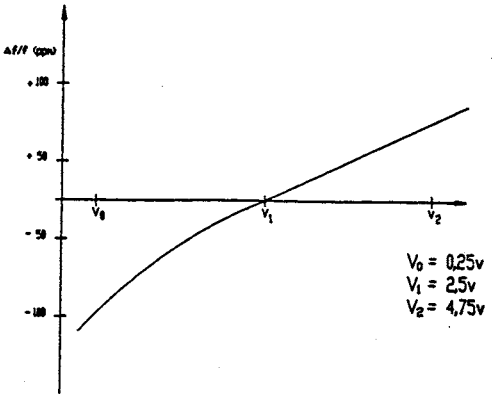
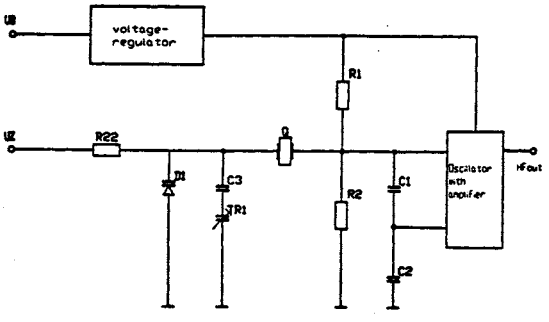


Fig.1 VCXO block diagram and tunning characteristic

Indroduction

VCXO are applied in communication and electronic equipment mostly as frequency modulators and as a part of phase locked loops(PLL). The output frequency of these oscillators is changed or modulated by an external voltage.The control range is dependent on the control reactances of the oscillator circuit and the equivalent data of the crystal.The load capacitance of the crystal will be changed by using a variable capacitance diode (Fig.1).



The load capacitance influence the temperature behaviour of an AT-cut crystal (Fig.2).

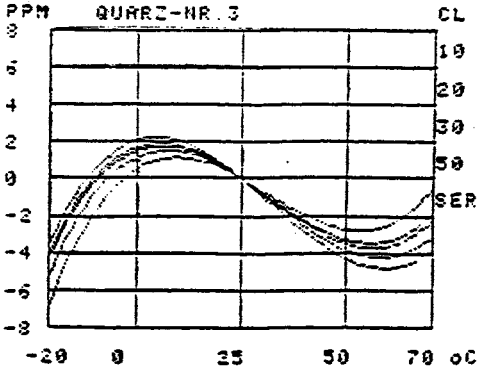


Fig.2 Temperature behaviour of an AT-cut crystal with different load capacitance (CL = 10 pF to 50 pF).

The temperature stability of a VCXO depends on the control voltage value, the crystals, variable capacitance diode and oscillator circuit temperature drift and the tuning range limits (Fig.3).

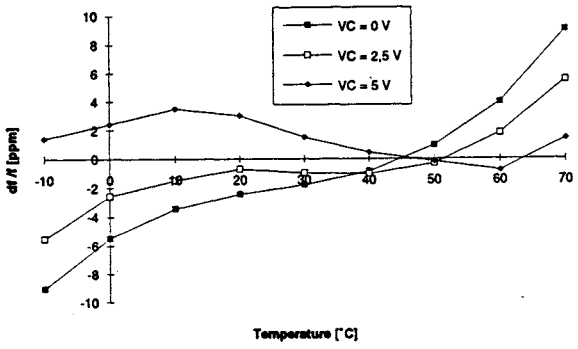


Fig.3 VCXO temperature stability at different control voltages

In a TCXO the frequency variation over the temperature range is stabilized by means of an analogue or digital compensation method (Fig.4).

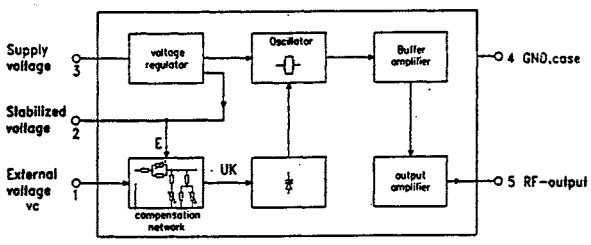


Fig.4 TCXO schematic diagram

A thermistor-resistor network (compensation network) generates a temperature-varying voltage that changes the variable capacitance diode in a predetermined manner to compensate the crystal's thermal drift .
Each oscillator will be measured at 5 temperatures. The tuning voltage UK, the stabilized voltage E and the resistance values for the thermistors for these 5 temperatures will be stored in a computer. The values for the thermistor-resitor network are calculated by an interpolation program. The resistors will be mounted into the oscillator and a second temperature test is made (Fig.5).

VCTCXO design

The TCXO application requires only a small control range, for example +- 5ppm to +- 12 ppm to compensate the crystal aging drift. The temperature stability will be obtained only at a fixed control voltage. The multiplexers for SDH needs oscillators with wide control range, f.ex. +- 30 ppm to +- 50 ppm or even more, and at each control voltage the oscillator must meet the required temperature stability.

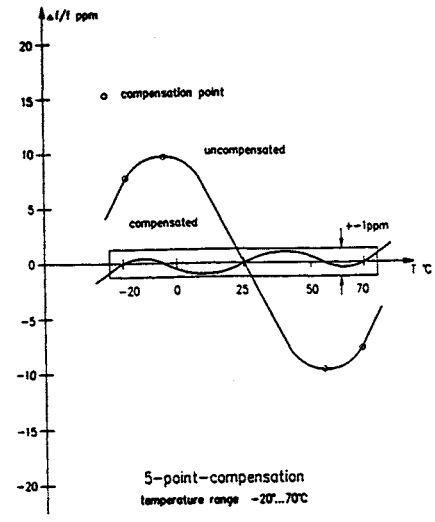


Fig.5 Frequency temperature characteristic of an oscillator with analogue compensation

The two functions, temperature stability and wide control range, are contrary. In the TCXO application the operating point of the variable capacitance diode will be changed only in a small control voltage range (f.ex. 2 V to 3 V) to minimize the compensation errors (Fig.6). To get the tuning range for a VCXO the operating point of the capacitance diode will be changed in a wide control voltage range (f. ex. 0,25 V to 4,75 V).

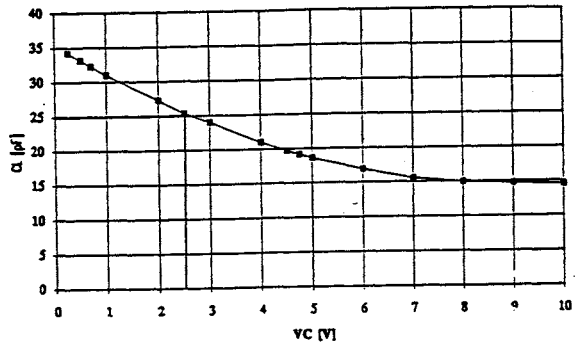


Fig.6 Tuning function of the variable capacitance diode BBY 42

If the temperature behaviour of an oscillator is compensated at Vc = 2,5V to meet a stability of +- 2 ppm in the temperature range -10 °C to + 70 °C and the control voltage is changed for example to 1 V or to 4 V, the oscillator will loose the temperature stability (Fig.7). This results from the nonlinear tuning function of the variable capacitance diode and the load changes influence of the crystal's thermal drift.

Fig.6 Tuning function of the variable capacitance diode BBY 42

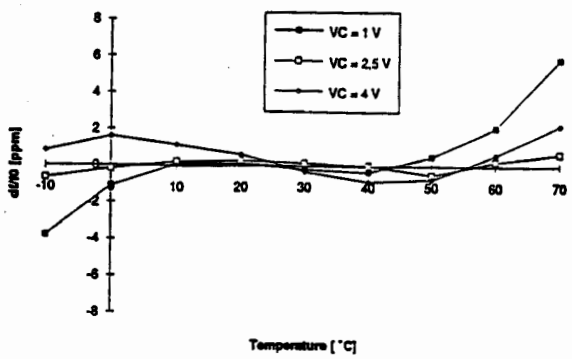


Fig.7 TCXO temperature stability at different varicap operation points

KVG has developed a circuit where the compensation network and the control network works nearly independent from each other (Fig.8).

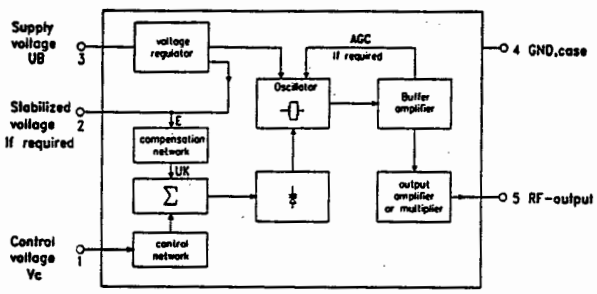


Fig. 8 VCTCXO schematic diagram

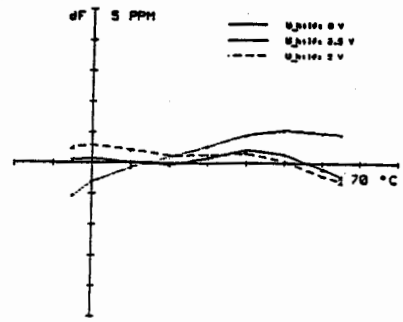
The control voltage influence the variable capacitance operating point is reduced and the nonlinear tuning characteristic is compensated. Table and temperature curves for two 19,44 MHz VCTCXO(specification in annex 1) with a temperature stability of +- 2 ppm and tuning range between +- 50 ppm and +- 70 ppm are shown below.

PLATZ NR. 30 TCXO NR._____

	1. Hilfsspannung		2. Hilfsspannung		3. Hilfsspannung	
T[°C]	F[kHz]	dF[PpM]	F[kHz]	dF[PpM]	F[kHz]	dF[ppm]
-5	19438.7556	-1.05	19440.3006	+.11	19441.1709	+.52
0	19438.7644	-.60	19440.3009	+.13	19441.1721	+.58
10	19438.7735	-.13	19439.9987	+.02	19441.1688	+.41
20	19438.7797	+.19	19439.9977	-.03	19441.1655	+.25
30	19438.7868	+.55	19440.0018	+.17	19441.1650	+.27
40	19438.7937	+.91	19440.0066	+.42	19441.1666	+.30
50	19438.7956	+1.05	19440.0038	+.28	19441.1620	+.07
60	19438.7952	+.98	19439.9950	-.17	19441.1529	-.40
65	19438.7944	+.94	19439.9902	-.42	19441.1494	-.58

ABLAGE (0.0V): -62.9 PpM OFFSET (20.0°C): -.33 59.790814249 PpmABLAGE (5.0V):

-1222.3 Hz 1162.3 Hz

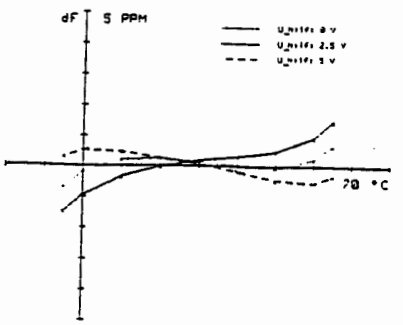


PLATZ NR. 31 TCXO NR._____

	1. Hilfsspannung		2. Hilfsspannung		3. Hilfsspannung	
T[°C]	F[kHz]	dF[PpM]	F[kHz]	dF[PpM]	F[kHz]	dF[ppm]
-5	19438.9831	-.67	19439.9558	-1.48	19441.2608	+.31
0	19438.8929	-.17	19439.9664	-.93	19441.2647	+.52
10	19438.9005	+.22	19439.9781	-.33	19441.2638	+.47
20	19438.9011	+.25	19439.9843	-.01	19441.2593	+.23
30	19438.8989	+.14	19439.9884	+.20	19441.2553	+.06
40	19438.8965	+.02	19439.9908	+.33	19441.2517	-.16
50	19438.8951	-.05	19439.9941	+.49	19441.2463	-.43
60	19438.9010	+.25	19440.0027	+.94	19441.2447	-.52
65	19438.9093	+.67	19440.0132	+1.48	19441.2491	-.29

ABLAGE (0.0V): -56.0 PpM OFFSET (20.0°C): -.01 55.3404156092 PpmABLAGE (5.0 V):

-1088.3 Hz 1270.2 Hz



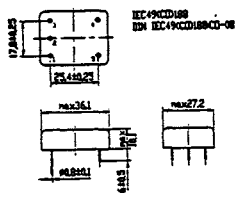
Standard frequencies for VCTCXO's in SHD applications are 16,384 MHz, 19,44 MHz, 38,88 MHz, 51,84 MHz and 155,52 MHz. The typical control voltage ranges are -3 Volt to + 3 Volt, -2,5 Volt to +2,5 Volt, 0 Volt to + 5 Volt or 0,25 Volt to 4,75 Volt. The typical control range is between +- 20 ppm to +- 50 ppm.

References

1. Frerkin, Marvin Crystal oscillator design and temperature compensation, Van Nostrand Reinhold Company 1978.
2. KVG Catalogue, Technical Indroduction.

19,44 MHz VCTCXO

BF 9



1. Pin configuration

1. Control voltage Vc = 0V ... +5V
2. N.C.
3. Supply voltage
4. RF-Output
5. Ground, case

2. Specification

2.1 Nominal frequency	19,44 MHz at Vc = 2,5 V +- 0,2 V
2.2 Frequency stability in the temperature range - 5°C to + 65°C	max. +- 2,0 ppm at VC = 0V to 5 V
2.3 Supply voltage U_B	5 V +- 0,25 V
2.4 Current consumption	max. 15 mA
2.5 Output voltage	HCMOS, low < 0,4 V, high > 4,0 V
Load	2 TTL, 10 HCMOS (1kOhm// 20 pF)
Duty cycle	40% ... 60%
2.6 Frequency stability	
vs. supply voltage changes +- 5%	max. +- 0,1 ppm
vs. load changes +- 5 %	max. +- 0,1 ppm

				KVG GmBH
				P.O.B 61
				6924 Neckarbischofsheim
1		12.06.91	W.Balzer	Tel.: 07263/6480 Fax.: 07263/6196
ED	Description	Date	Name	Sheet 1 of 2

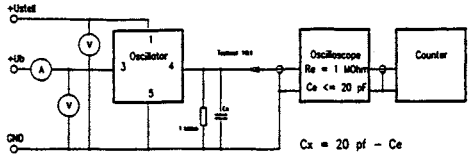
19,44 MHz VCTCXO

2.7 Aging	max. +- 1 ppm first year max. +- 15 ppm in 15 years
2.8 Control range	+- 50 ppm to +- 70 ppm
2.9 Control voltage Vc	0 V to + 5 V
2.10 Slope	positiv
2.11 Linearity	+- 20 %
2.12 Operable temperature range	- 20°C ... + 75°C
2.13 Storage temperature range	- 55°C ... + 105°C

3. Enviromental conditions

3.1 Shock	DIN IEC 68-2-27,Test Ea, 100 g, 6ms, Half-sine, 3 pumps in 3 main directions
3.2 Vibration	DIN IEC 68-2-6, Test Fc, 10-500Hz, 10g, 2 h in 3 main directions
3.3 Humidity	DIN IEC 68-2-3, 40°C/93%RH, 21 days
3.4 Solderability	DIN IEC 68-2-20 only for wire leads, Methode 3: Solder globule at + 235°C
4.Marking	Manufacturer's name, date code(week/year Specification Center frequency

5.Test circuit



				KVG GmBH
				P.O.B 61
				6924 Neckarbischofsheim
1		12.06.91	W.Balzer	Tel.: 07263/6480 Fax.: 07263/6196
ED	Description	Date	Name	Sheet 2 of 2

HIGH-STABILITY QUARTZ OSCILLATORS FOR SARSAT/COSPAS PROJECT AND OTHER SATELLITE-BASED GLOBAL POSITIONING SYSTEMS

Yakov L. Vorokhovsky, Inna G. Petrosian, Boris G. Drakhlis, Alexander A. Volkov

MORION Mfg. Co.
13a Kima Avenue, St.Petersburg 199155, Russia

ABSTRACT

Quartz oscillators based on Internally Heated Quartz Resonators (IHQRs) allow to combine high frequency stability with low power consumption, fast warm-up, small size, resistance to severe operational conditions. The developed and commercially manufactured quartz oscillators (ГК40 - ТС type) for SARSAT/COSPAS system distress beacons meet with sufficient margin all the requirements of the international specification. In the 5.0 MHz to 5.2 MHz frequency range quartz oscillators were also developed for use in non-distress satellite systems for continuous positioning of various mobile objects in which the desired coordinates are determined on the same basis as in SARSAT/COSPAS system.

Keywords: frequency, resonator, oscillator, ovenizing, positioning.

1. INTRODUCTION

In the SARSAT/COSPAS search-and-rescue system the reference quartz oscillator for 406 MHz Emergency Position Indicating Radio Beacon (EPIRB) must meet stringent performance requirements. On the one hand there must be high medium- and short-term frequency stability under sharp changes of ambient temperature in the wide temperature range. On the other hand - fast warm-up, low power consumption and small size. For use in EPIRB for a wide class of applications the oscillators must additionally be resistant to severe mechanical effects (shocks of 500 to 1000 g).

On the basis of approaches developed in SARSAT/COSPAS system other - non-distress - satellite systems for continuous positioning of various mobile objects in which desired coordinates are determined on the same basis as in SARSAT/COSPAS system are being developed. However, because of a large number of objects being positioned quartz oscillators for such systems must be fabricated at a number of frequencies according to the satellite frequency channels allocated to various objects (also in 406 MHz range). In this case requirements placed on oscillator's total frequency tolerance over the life time are much more stringent implying both higher long term frequency stability and better initial frequency tolerance.

Analysis and numerous experiments showed that temperature compensated quartz oscillators don't allow to meet the whole set of required performance parameters in given operating conditions or at least this approach is connected with serious difficulties. For the same reasons the high frequency quartz oscillator versions were rejected for the time being and the choice was made for an ovenized quartz oscillator in the 5 MHz frequency range followed by times 80 or 81 frequency multiplication in the beacon's circuitry.

Good opportunities for realization of this ovenized quartz oscillator open Internally Heated Quartz Resonators (IHQRs) (Ref. 1). On their basis the authors have developed mass production quartz oscillator for 406 MHz distress beacons of SARSAT/COSPAS system as well as much more accurate oscillators for the above mentioned non-distress satellite positioning systems.

2. INTERNALLY HEATED QUARTZ RESONATORS (IHQRs)

Functionally and technologically, IHQRs combine in one design a quartz resonator proper and an oven. Main features and basic design principles of IHQRs are presented in (Ref. 1) and briefly described in (Ref. 2).

The outline of the developed basic design of IHQR is showed in Fig.1. The heater, the temperature sensor, and heat-distributing elements of the inner oven are mounted on the holder together with the piezoelement, resonator vacuum being a natural and most efficient heat insulation of the inner oven, and the resonator enclosure being at the same time the enclosure of the entire device. Semiconductor positive TC thermoresistors are used as the heater and the temperature sensor. Accurate temperature stabilization necessary for ensuring frequency stability of the order of $1x10^{-7}$ and better is provided by means of the oven control circuit (Ref. 3) which is placed outside the IHQR case.

One of the main objectives was significant reduction of heat losses, i.e. power consumption of IHQR. For the residual pressure less than $1x10^{-4}$ mm Hg the heat transfer through the residual gas is negligible (Ref. 1), and the

thermoreflective coating of the IHQR enclosure greatly reduces the radiation heat losses. Heat transfer by conduction through the mounting supports is the most difficult to reduce because the ways to achieve this reduction often lessen the mechanical strength of the construction. Significant opportunities here are opened by the choice of materials, improvement of design (including shape) of the heat-removing parts of the holder, optimization of holder design in accordance with required level of mechanical effects. The obtained results can be seen in Fig. 2 by comparison of curves 1 and 3. It can be seen that the total heat losses were almost halved as compared with the initial version of the basic design bringing them to the level of approximately 155 mW at -60°C and 65 mW at +25°C. It is very important that in addition to that the resistance to mechanical shocks with acceleration of 1000 g is ensured.

3. QUARTZ OSCILLATORS ON IHQRs

A structural diagram of the quartz oscillator on IHQR for 406.025 MHz SARSAT/COSPAS system distress beacons is presented in Fig. 3. Depending on the frequency multiplication factor in the beacon circuitry operating frequency of these oscillators is 5.075312 or 5.012654 MHz.

Oscillators for other - non-distress - positioning systems are designed on the basis of the same structural diagram (Fig. 3). However, these oscillators are fabricated at a number of frequencies in the 5.0 MHz to 5.2 MHz frequency range. And, most important, "non-distress" oscillators have significantly - approximately by an order of magnitude - better total frequency tolerance over the life time than oscillators for SARSAT/COSPAS system.

This is ensured by considerably improved initial frequency tolerance and long term frequency stability of IHQR together with frequency adjustment of quartz oscillators during manufacturing.

Due to circuit design deliberations and economical considerations quartz oscillators are designed using standard printed circuit technology and widely available electronic components. The developed design of the quartz oscillator (Fig. 4) is a simple at sight sandwich construction, arranged along the IHQR axis with printed circuit boards placed ortoganally to the IHQR axis. Free inner space of the quartz oscillator is filled with plastic foam providing reliable fixing of elements of the construction (IHQR including), effective protection of the quartz oscillator from mechanical and climatic effects, high reliability.

4. MAIN PERFORMANCE CHARACTERISTICS OF DEVELOPED QUARTZ OSCILLATORS

Typical performance data for medium-term frequency stability (drift) for 15 min and for root mean square error of the operating frequency relative to the least square straight line fit to the measurement points (the residual noise) versus time after turn on in the most severe working conditions (at -40°C) are shown in Fig.5. As it should be expected the frequency drift rapidly decreases with time and after 10 min is less than 5×10^{-10}/min. The residual noise 10 min after turn on is practically not dependent on time, and is about 2×10^{-10}. Short-term frequency stability (Allan variance) for 100 ms 15 min after turn on does not exceed 5×10^{-10}.

The necessity of ensuring the required level of quartz oscillator performance in conditions of fast changes of the working temperature in a wide range imposes rather strict limitations on the oscillator's temperature frequency stability. Major improvements in the oscillator's circuitry allowed to realize frequency stability better than $\pm5\times10^{-8}$ in the temperature range from -40°C to +60°C with entirely unovenized electronic circuit (except for piezoelement in IHQR).

Due to effective system of oven control and thermal insulation of IHQR, the developed quartz oscillators practically maintain the output signal parameters in conditions of working temperature gradient of 9°C/hour and also 15 min after the direct thermal shock of 30°C made at the moment of oscillator's turn on. It should be noted that good "steady-state" temperature frequency stability - about 1×10^{-7} or better - is the necessary condition for ensuring the required small frequency drift under continuous (dynamic) changes of environment temperature. It could be easily shown that having an oscillator with the "steady state" frequency stability $\pm(1 \div 1.5)\times10^{-6}$ in the temperature range from -40°C to +60°C it is practically impossible to ensure frequency drift for 15 minutes within $\pm1\times10^{-9}$/min limits even at the lowest permitted rate of environment temperature change equal to 5°C/hour, not to speak of higher rates or reaction to thermal shock of 30°C.

As is clearly illustrated in Fig.5, even in the most severe conditions (-40°C) the developed oscillators meet all accuracy specifications of (Ref. 4) 10 minutes after turn on.

Taking into account the results obtained for IHQR (see Fig.2, curve 3) typical values of current consumed by the quarts oscillator as a whole are: 30 mA at - 40°C, 25 mA at 0°C and 20 mA at 25°C. Supply voltage may change from 9 to 18 V. The oscillator ensures the required parameters at 3 V intermittent drops ("dips") of the supply voltage without transients or frequency "drift".

The developed oscillators successfully withstand mechanical shocks with peak acceleration of 1000 g. Their design ensures resistance to long (56 days) effects of 100% humidity at +40°C, and to the effect of fungous mold and salt (sea) fog. Extreme storage temperatures are - 60°C and +85°C.

Overall dimensions of oscillators are 26.5 x 26.5 x 81.5 mm, weight about 90 g.

The problem of reliability of small-size ovenized quartz oscillators with vacuum heat insulation is of principal importance, especially in case of long operating life time. One of the most important prerequisites of reliability here is maintaining of vacuum in IHQR. Long-term tests and experience of use of a large number of such units in very important equipment prove that here this problem has been successfully solved.

5. CONCLUSIONS

Thus 2 types of commercially manufactured

quartz oscillators have been developed:

ГК40-ТС - for the SARSAT/COSPAS system meeting with sufficient margin all the requirements of international specification (Ref. 4) for 406 MHz Class 1 EPIRB;

ГК40-ТС-М - for radio beacons of non-distress satellite systems for continuous positioning of mobile objects in which the desired coordinates are determined on the same basis as in SARSAT/COSPAS system. These oscillators manufactured at any required frequency in 5.0 - 5.2 MHz frequency range have total frequency tolerance over the whole life time (10 years) better than $\pm 5 \times 10^{-7}$.

The developed technology is based on successful experience of production of precision quartz oscillators for important electronic equipment, and is a guarantee for mass production of highly reliable oscillators.

Model ГК40-ТС quartz oscillators are used in EPIRBs manufactured by plants in Russia and Ukraine and approved by COSPAS-SARSAT Secretariat in London (Ref. 5).

REFERENCES

(1) Я.Л.Вороховский. "Резонаторы-термостаты и пьезоэлектронные устройства на их основе". - В кн: Пьезоэлектрические резонаторы. Справочник. - Москва: Радио и связь, 1992.- Стр. 318 - 386.

(2) Ya.L. Vorokhovsky,I.G.Petrosian,B.G.Drakhlis, A.A.Volkov, "The Further Development of Internally Heated Quartz Resonators and the Oscillators on Their Basis for Use in Up-To-Date Search-and-Rescue and Communication Systems". in Proc.of the 1992 IEEE Frequency Control Symposium, USA, 1992. pp.505-512.

(3) Я.Л.Вороховский, Б.Г.Драхлис. "Выбор и построение терморегулятора для резонатора-термостата прецизионного кварцевого генератора", Электронная техника. Сер. Радиодетали и радиокомпоненты, 1987, вып.2 (67), с.67-70.

(4) Specification for COSPAS-SARSAT 406 MHz Distress Beacons. - C/S T.001, Issue 2 - Revision 3, September 1991.

(5) COSPAS-SARSAT System Data No.15, November 1992. Publ. by the COSPAS-SARSAT Secretariat, London, UK.

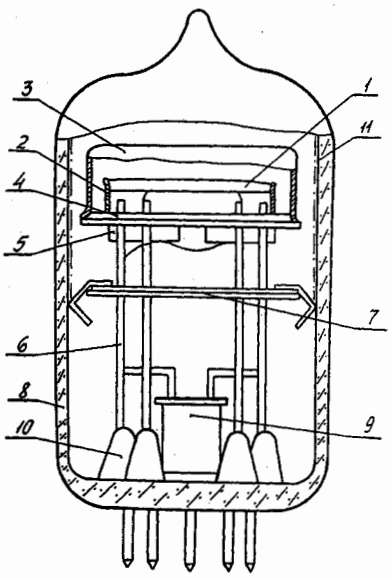


Fig.1. The basic construction of vacuum IHQR with positive TC thermoresistors as the heaters and the temperature sensor: 1 - the piezoelement; 2 - the crystal holder; 3 - the cover ~ the thermal screen; 4 - the base of the crystal holder; 5 - the heaters and the tempe - rature sensor on the positive TC termo- resistors; 6 - the holder; 7 - the in - termediate fixing base with the centring springs; 8 - the tube; 9 - the solid - - state getter; 10 - the base of the case with the terminals; 11 - the ther- moreflective coating.

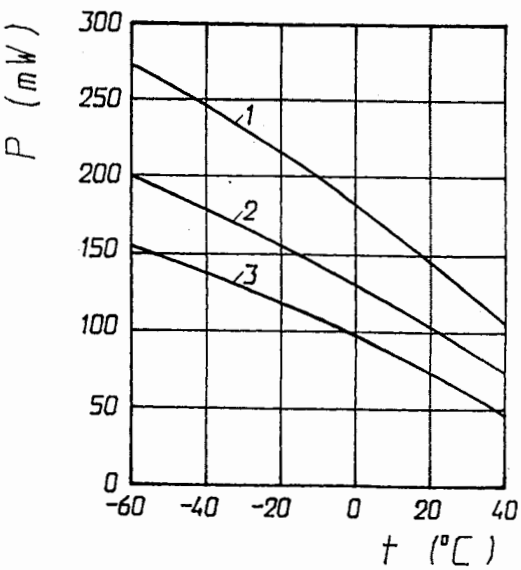


Fig.2. Typical IHQR power consumption /P/ in the range of ambient temperatures /t/: 1 - initial version of the basic design 2,3 - consequently improved version with reduced conduction through the holder /and leads/.

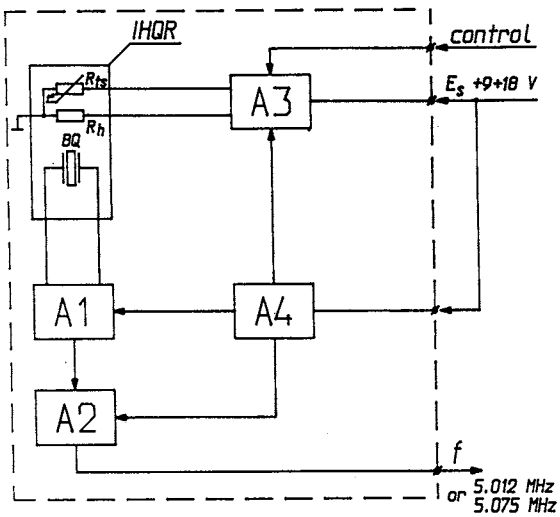


Fig.3. Structural diagram of quartz oscillator for radio beacons of SARSAT/COSPAS project:
A1 - active part of oscillator
A2 - former of TTL-signal
A3 - proportional oven control circuit
A4 - reference voltage source.

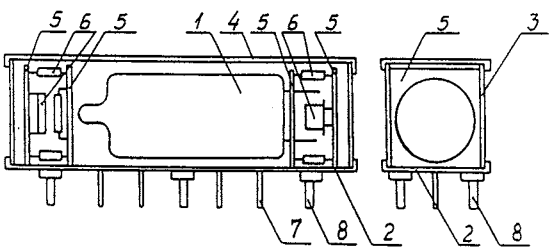


Fig.4. Miniature version of mechanical design of quartz oscillator on IHQR:
1 - IHQR
2 - base
3 - case
4 - cover
5 - printed circuit board
6 - radio elements
7 - lead
8 - stud; vacant inner space is filled with plastic foam.

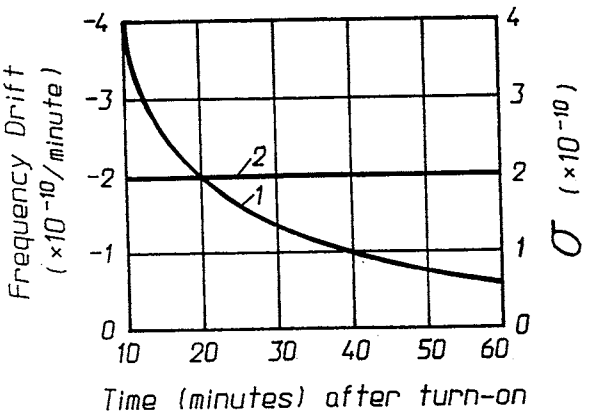


Fig.5. Typical characteristics of mid-term frequency instability /drift/ in 15 min and residual noise /root mean square error of the operating frequency relative to the least square straight line fit to the measurement points/:
1 - frequency drift
2 - residual noise /σ/.

DESIGN OF HIGH FREQUENCY OSCILLATORS USED IN TELECOMMUNICATIONS

Vincent CANDELIER, Philippe CATHELIN, Bassam NASREDDINE, Gérard GUITEL

THOMSON/CEPE
44, Av. de la Glacière - BP 165
95105 ARGENTEUIL - FRANCE

ABSTRACT

High speed communication systems require high frequency references. In order to design a high frequency oscillator, three major methods are available :

- The use of a high frequency crystal for example, with ion etched crystal.
- The use of a "medium" frequency crystal and analog frequency multiplication.
- The use of a low frequency crystal and a phase locked loop to multiply the frequency.

The method to be used will depend on the required performances. Sample devices will be presented at 139.264 MHz, 155.520 MHz and 622 MHz. The range of validity of each approch and the possible performances will be discussed. The main points will be : Thermal sensitivity, phase noise, long term stability, frequency shift amplitude and bandwidth of frequency control.

1./ INTRODUCTION

High speed communication systems like RNIS (Réseau Numérique à Intégration de Services), ATM (Asynchronous Transfer Mode) require high frequency references. We can use a quartz crystal oscillator with three major configurations (figure 1) :

- The use of a quartz crystal on the direct high frequency mode (design "direct" in figure 1).
- The use of a quartz crystal on a medium frequency mode with analog frequency multiplication (design "multiplier" in figure 1).
- The use of a quartz crystal at a low frequency with a phase locked loop to multiply the frequency (design "PLL" in figure 1).

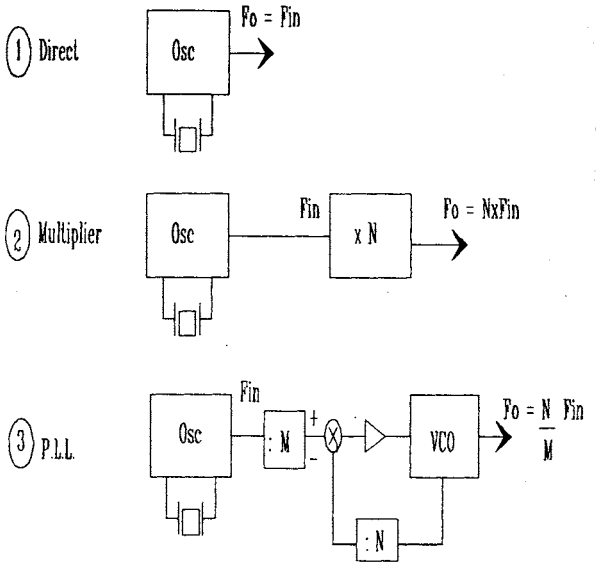


FIGURE 1 : DIFFERENT DESIGNS TO CREATE HIGH FREQUENCY REFERENCE, FO

2./ REQUIREMENTS AND DESIGN COMPARISON

This article intends to show which method can be chosen for required performances.
The usual oscillator requirements are detailed in table 1.

Nominal frequency range	130 MHz to 900 MHz
Frequency stability versus temperature	± 1 ppm to ± 15 ppm
Temperature range	0°C 70°C
Aging over 15 years	± 3,6 ppm to ± 30 ppm
$\mathcal{L}$ (100 Hz)	- 80 dBc/Hz
Spurious and harmonic ratio	- 40 dBc
Bandwidth of the frequency control	1 KHz to 50 KHz
Frequency control range	± 30 ppm to ± 100 ppm

TABLE 1 - OSCILLATOR REQUIREMENTS

The main difficulty is to associate a large bandwidth for frequency control (following the frequency correction of the telecommunication network) and good "stand alone" frequency stability (required when network synchronization is off).

We base our study on some sample devices at 139.264 MHz, 155.520 MHz and 622 MHz designed at CEPE. The table 2 presents a qualitative comparison of the three designs. The different main parameters required for high speed communication oscillator are : frequency stability versus aging and temperature, nominal frequency, phase noise (near the carrier and the floor), spurious and harmonics spectrum, frequency control (range and bandwidth).

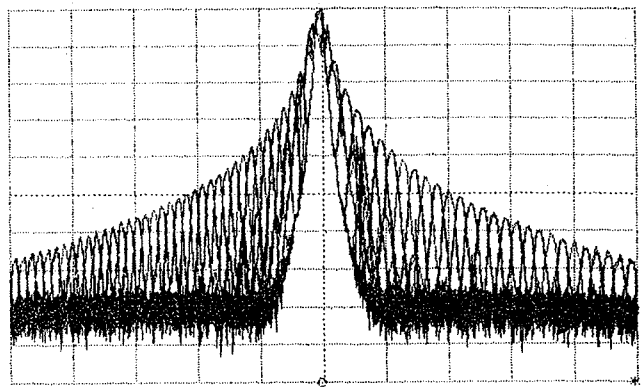
3./ RESONATOR

The first problem is the choice of the crystal.

For the design "direct", the resonator is an ion etched fundamental mode quartz crystal. For the other designs the resonators are mechanically lapped quartz crystals. We obtain large frequency shift (greater than ± 100 ppm) and large bandwidth (greater than 50 KHz) with fundamental mode. With overtone mode, the frequency control stays below 30 KHz. In fact the modulation envelope shows peaks at frequencies corresponding to the spurious mode of crystal resonator and for overtone mode, the spurious mode are nearer the carrier than for fundamental mode. Figure 2 shows the difference between the two modes.

On the other hand, the aging of overtone mode and low frequency fundamental mode (i.e. 77 MHz 3rd 0T and 16.384 MHz fund) crystals is better than high frequency fundamental mode (i.e. 139 MHz ion etched) - typical performance these designs can achieve are :

77 MHz 3rd 0T : 1 ppm/year
16,384 MHz Fund : 1-2 ppm/year
139,264 MHz Fund : 10 ppm/year



2a) FUNDAMENTAL MODE CRYSTAL

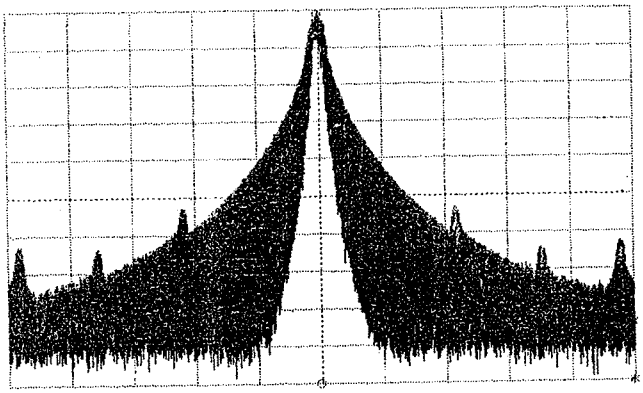


FIGURE 2 - SPECTRAL FREQUENCY MODULATION RESPONSE OF QUARTZ CRYSTALS OSCILLATOR

For all designs, the quartz crystal cut is AT. In that case, frequency versus temperature behavior of low frequency fundamental mode or overtone mode resonator .is better than the one of high frequency

DESIGN / PARAMETERS	DIRECT	MULTIPLIER		P.L.L.	
Fo	139.264 MHz	139.264 MHz	155.520 MHz	139.264 MHz	622.08 MHz
Fin	139.264 MHz	46.33 MHz	77.76 MHz	16.384 MHz	16.384 MHz
Fin overtone mode	Fund	Fund	3 rd 0T	Fund	Fund
Multiplication rate	1	3	2	8.5	37.96
Frequency stability versus temperature (0°C, 70°C)	± 15 ppm	± 8 ppm	± 1 ppm	± 1 ppm	± 5 ppm
Frequency stability versus aging (ppm/year)	15	4	1	1	2
$\mathcal{L}(f)$ near the carrier	XX	XX	XX	XXX	XXX
$\mathcal{L}(f)$ floor	XXX	XX	XX	XXX	XXX
Low spurious spectrum lines	XXX	XXX	XXX	XX	XX
Sub-harmonic ratio	XXX	X	XX	XXX	XXX
Bandwidth of the frequency control	XXX	XX	X	XXX	XXX
Frequency control range (ppm)	± 40	± 40	± 30	± 30	± 130

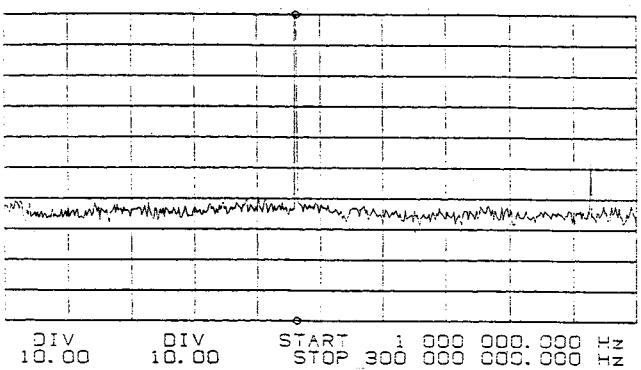
X Low performance XX Average performance XXX High performance

TABLE 2 - QUALITATIVE COMPARISON OF THE THREE DESIGNS

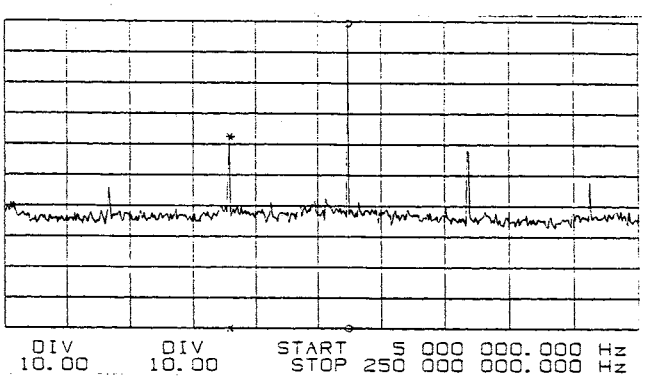
fundamental mode. With a crystal plate thickness the crystal, the frequency temperature relation is not very repetitive. For example, at 139,264 MHz, fundamental mode, it is difficult to achieve a temperature stability within (O°C, 70°C) better than ± 15 ppm. Crystals, working on third overtone mode or on fundamental mode frequency (lower than 40 MHz), exhibit a temperature stability better than ± 5 ppm.

Furthermore temperature compensated crystals oscillators require, for a target of 1 ppm, regular and reliable third order frequency temperature crystals behaviors.

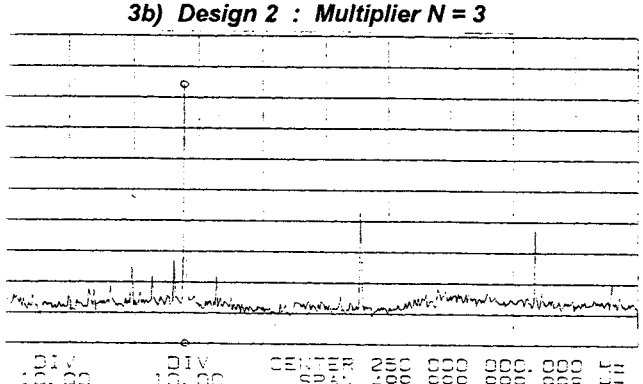
Such a characteristic is easy to obtain with low frequency fundamental mode or third overtone mode for crystal resonators, but stays difficult with high frequency fundamental mode.



3a) Design 1 : Direct



3b) Design 2 : Multiplier N = 3



3c) Design 3 : PLL $\frac{N}{M}$ = 8,5

FIGURE 3 - OUTPUT SPECTRUM (Fo = 139,264 MHz)

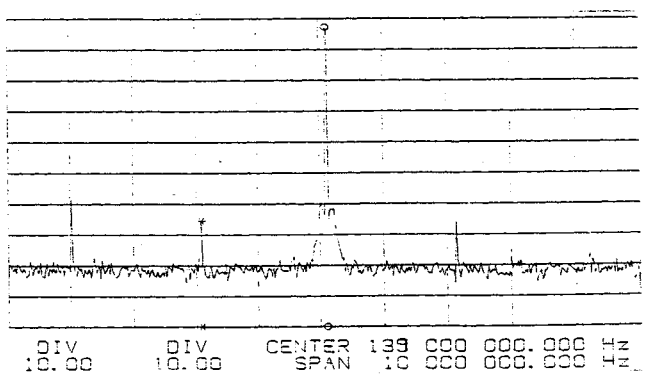
4./ OSCILLATOR OUTPUT FREQUENCY SPECTRUM

Now we will compare the frequency spectrum for the three designs.

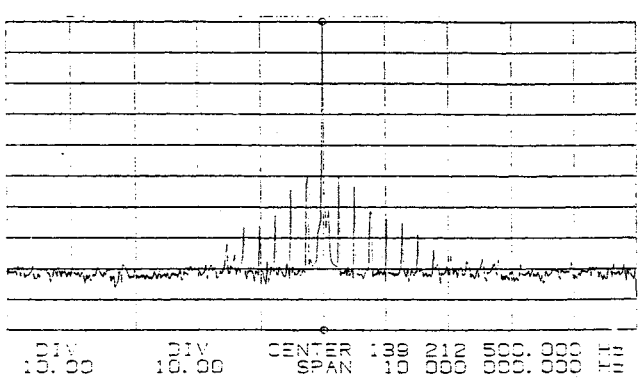
With the first design,we have at the output just the line at the fundamental frequency and harmonics. The harmonics level can be easily maintained under - 40 dBc (see figure 3a).

In the second design, with the use of a multiplier, we create sub-harmonics of the output frequency. A sub-harmonics level lower than - 30 dBc is difficult to achieve with a high level of multiplication and with a limited number of selective output amplifiers (see figure 3b).

For multiplication factors greater than three or four it is rather more advisable to use the design 3 (PLL). For this design, the wide band spectrum is shown in figure 3c. We have a harmonics rate of - 40 dBc and a sub-harmonics level better than - 50 dBc. Near the carrier, spurious are generated. The frequency between the spurious line and the carrier, are multiple of Fcomp. Fcomp is the comparison frequency of the PLL (Fin/M or Fo/N). Higher Fcomp is, more efficient PLL filtering is. The figure 4 presents the spectrum comparison for two Fcomp values : 2 MHz and 250 KHz. In spite of everything, the spurious level stays under - 50 dBc.



4a) FCOMP = 2 MHz

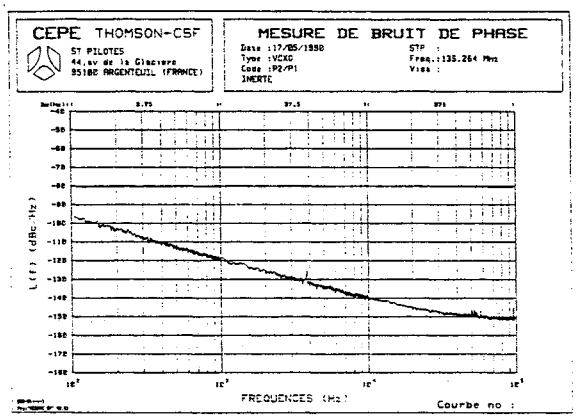


4b) FCOMP = 250 KHz

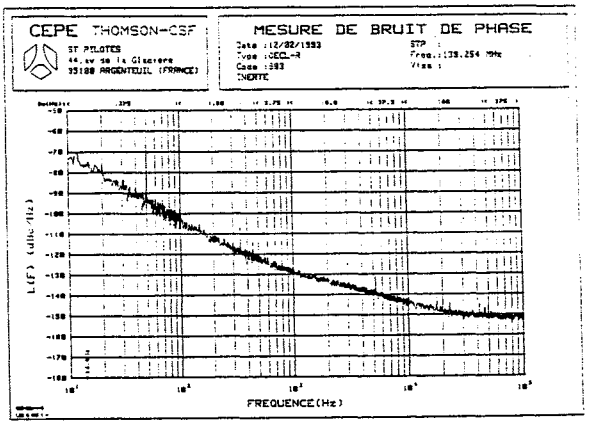
FIGURE 4 - COMPARISON OF SPECTRUM NEAR THE CARRIER AT THE OUTPUT OF THE DESIGN 3 : (PLL)

(PRESENCE OF SPURIOUS AT FREQUENCY FCOMP AND HARMONICS)

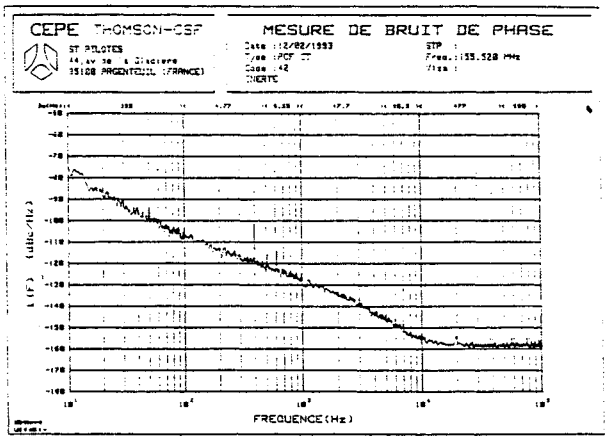
The last aspect, we view, is the two sided spectral density of phase noise, $\mathcal{L}(f)$. The curves of $\mathcal{L}(f)$ for designs 1 and 2 are presented in figure 5, for the three modes of quartz crystal noted in table 2. The frequency shift range which can affect the phase noise, is fixed, for these designs, at ± 30 ppm. We confirm that the lower the fundamental frequency of the crystal is, the lower the two sided spectral density of phase noise, at 100 Hz, is.



5a) Design 1 : Direct



5b) Design 2 : Multiplier N = 2



5c) Design 2 : Multiplier N = 3

FIGURE 5 - TWO-SIDED SPECTRAL DENSITY OF PHASE NOISE

In the last design with PLL, we have shown that the two-sided spectral density of phase noise detailed in figure 6. At the output, frequency of 622 MHz, is divided by four. Like this, we can compare with the last curves. At the difference of these, we observe three zones of slope for $\mathcal{L}(f)$ versus Fourier frequency :

Zone 1 : Lower than 200 Hz
Zone 2 : Between 200 Hz and 7 KHz
Zone 3 : Higher than 7 KHz

In zone 3, $\mathcal{L}(f)$ is due to the VCO. In zone 2, $\mathcal{L}(f)$ is due to the dividers N or M, and in zone 1, $\mathcal{L}(f)$ is equal to the $\mathcal{L}(f)$ of the quartz crystal oscillator at Fin, multiplied by N/M. In this example the frequency shift of Fin is ± 130 ppm, that is why the phase noise at 100 Hz is not better than the other examples.

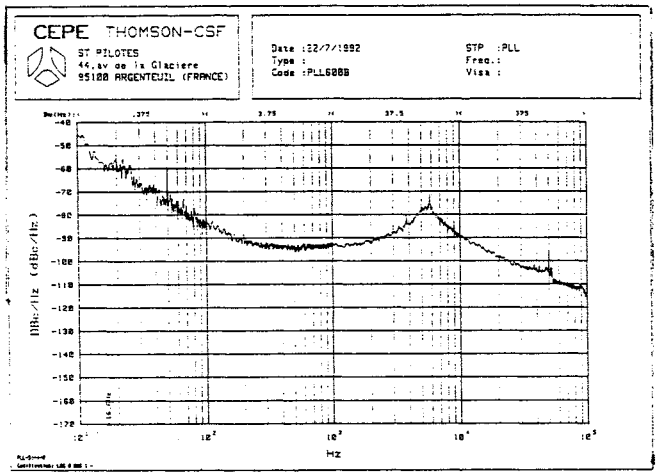


FIGURE 6 - TWO-SIDED SPECTRAL DENSITY OF PHASE NOISE AT 622 MHz/4 FOR DESIGN 3 (PLL)

5./ CONCLUSION

We have compared three designs of high frequency oscillators. In conclusion, for alarge frequency shift, a large bandwidth of modulation, and for a very low spurious level in spectrum, the design 1 is preferable. For high frequency stability versus temperature and versus aging without spurious in the spectrum, we would select design 2.

The new telecommunications systems need both performances and the trend is to increase the reference frequency. The PLL design is the only way to archieve such characteristics, but we have to control all the spurious due to phase comparison.

MINIATURE, LOW POWER CONSUMPTION OCXO OSCILLATOR

A. Masiukiewicz, B. Kalinowska

Tele and Radio Research Institute
Ratuszowa 11, 03-450, Warsaw, Poland

ABSTRACT

Basing on new technological possibilities, a miniature low power consumption OCXO oscillator has been designed at ITR. The 5 and 10 MHz SC-cut quartz crystal resonators of special construction, operating at fundamental mode, cdd-welded in TO-5 enclosures were used.

The resonator enclosure was used as an oven and its total volume is 0.4 cm^3.

Careful optimization of temperature sensor position assure good frequency stability. Electronic circuits employ SMD elements, which significantly reduced area necessary for printing boards.

1. INTRODUCTION

As a result of new application areas some new types of quartz crystal oscillators were developed. There were among them miniature oven controlled crystal oscillators (MCXO) (Ref. 1).

Typical parameters of these oscillators are as follow:

- frequency stability versus temperature 0.1 ppm
- volume less than 20 cm^3
- power consumption during warm-up less than 1 W
- power consumption at 25°C about 300 mW
- short term frequency stability $10^{-9} \div 10^{-10}$/1 s.

To obtain high frequency stability it is necessary to apply a high quality resonator and carry out a careful optimization of the oven. At ITR SC-cut resonators, operating at fundamental mode of 5 and 10 MHz frequency, cold-welded in TO-5 enclosures were developed.

The Q factor of $0.65 \cdot 10^6$ for the 10 MHz resonator and $1.2 \cdot 10^6$ for the 5 MHz resonator was achieved. The TO-5 enclosure was used as the oven. This led to dimensions and weigh reduction and as a consequence diminished power consumption.

Oven parameters optimization especially the oven Q factor and temperature overshoots, enabled finding the most advantageous position of the precision thermistor used as a temperature sensor.

Parameters of 5 MHz oscillator models were measured.

2. OSCILLATOR CONSTRUCTION

The chosen and realised solution bases on following principles:

- the oven is built directly on the resonator enclosure,
- the temperature control loop includes only the resonator and the varactor diode (two the most temperature - sensitive elements),
- polyurethane foam is used as thermal insulation,
- precision thermistor is used as a temperature sensor,
- heat is provided by a medium power SMD transistor in SOT 89 enclosure.

The block diagram of the oscillator is shown in figure 1 and the construction in figure 2. Electronic circuits including: oscillation circuit, oven temperature control circuit, frequency adjustment circuit and output circuit, were built using SMT elements. All circuits were placed on four printed boards.

The Colpitts configuration oscillation circuit could be applied due to very good suppression of the resonator B mode. The tuning circuit provides ±5 ppm frequency adjustment and if the resonator calibration tolerance is lower than ±2.5 ppm it doesn't require any variable elements.

The whole construction is placed in CO-08 enclosure (according to IEC specification) of 26x34.8x19.4 mm dimensions.

3. RESONATOR

The miniature SC-cut resonator was developed at ITR (Ref. 2).

The 5 and 10 MHz resonator operating at fundamental mode were built basing on plano-convex crystal blanks of 6.5mm diameter.

The special construction of vibrator suppressed the B mode intensity more than 7 times.

The vibrators were cold-welded in TO-5 vacuum enclosures.

Basic parameters of 5 and 10 MHz resonator are presented in Table 1, and the F-T characteristics close to the turn-over point in figure 3.

Table 1 Resonator parameters.

f	R_{1c}	R_{1b}	Q
MHz	Ω	Ω	10^6
5	25	250	1.2
10	15	110	0.65

4. OVEN OPTIMIZATION

The following features of the oven were considered:

- Q factor,
- dynamic behaviour,
- temperature stabilization time.

Optimization enabled finding the best location of the temperature sensor. The measurements of the oven Q factor were carried out on level similar to the crystal plate position (figure 4).

The precision thermistor was fixed in TO-5 enclosure and the oven was tested with the thermistor inside. Some results of the optimization process are presented in figure 5 and in Table 2.

Table 2. Oven optimization results

Characteristic number	ΔT_2 (K)	Oven quality	Warm up time	
			1K	0.1K
1	+0.25	+320	1min 50s	2min 30s
2	-1.2	-67	1min 50s	3min 10s
3	-0.58	-140	1min 50s	3min

Results show that there is a possibility of obtaining the high oven Q factor (about 300) and very small overshoots of temperature (0.07 K) in a very simple construction.

5. MEASUREMENTS

Measurements of 5 MHz oscillators parameters were carried out and results are presented in figures 6 to 11 and in Table 3.

Table 3. Oscillators parameters

Parameter	Osc. nr 1	Osc. nr 2
Frequency	5	
δ_f/20÷50°C	$9 \cdot 10^{-8}$	$8.5 \cdot 10^{-8}$
σ_{y_A}(1s)	$1.7 \cdot 10^{-11}$	$1.1 \cdot 10^{-11}$
δ_f(day)	$1.4 \cdot 10^{-9}$	
S_Φ(10kHz)	135dB/Hz	136dB/Hz
Supply voltage (V)	12	
Power consumption (mW)		
oven 25°C	160	170
oven -20°C	370	450
whole oscillator 25°C	340	370
whole oscillator -20°C	550	650
Dimensions (mm)	26x34.8x19.4	
Mass (g)	30	
Output	HCMOS or SIN	

In comparison with other oscillators of the same type this one has some advantages.

Aperiodic frequency stabilization characteristics and high short term frequency stability (due to the high quality resonator) are the features which can distinguish it (Ref. 3, 4, 5, 6, 7).

6. CONCLUSION

The further optimization of oscillator parameters will be continued. The achieving frequency stability is worse than it can be calculated from oven quality. It can be improved by heat loss through leads reduction.

Frequency stabilization time can be improved by two ways. First by increasing warm-up power (but it require applying more powerful heating transistor) or by using resonators with some helium inside the enclosure (such resonators were used previously at ITR).

7. REFERENCES

1. J. J. Gagnepain, Quartz Crystal Oscillators From Their Design To Their Performances, 4th EFTF, Neuchatel, 1990
2. W. Szulc, B. Gniewiñska, A. Masiukiewicz, B. Kalinowska, Miniature, Low Power Consumption OCXO Oscillator. Contract nr Z-532, ITR, 1992
3. L. Hurley, A Temperature Controlled Crystal Oscillator, 43rd ASFC, 1989
4. G. Marrianneau, D. Hauden, J. J. Gagnepaiu, Integrated Oven Controlled Quartz Crystal Oscillator, 40th ASFC, 1986
5. K. Schwaller, K. Schulthess, W. Zingg, Miniature Oscillators, 2nd EFTF, Nauchatel, 1988
6. B. Long, G. Wearer, Quartz Crystal Oscillators With Direct Resonators Heating, 45 ASFC, 1991
7. J. P. Aubry , G. Wagner, V. Candelier. Miniaturisation and Stability: A New Generation of Ovenized Oscillators, 6th EFTF, Noordwijk, 1992

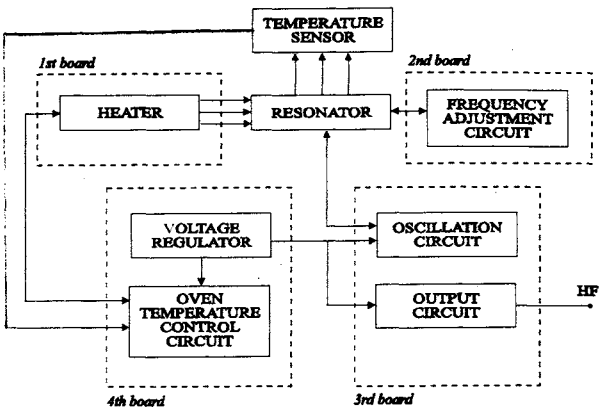


Fig. 1. Block diagram of the oscillator.

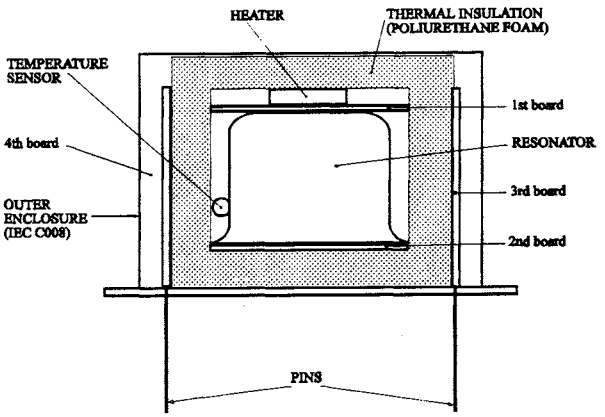


Fig. 2. Oscillator construction.

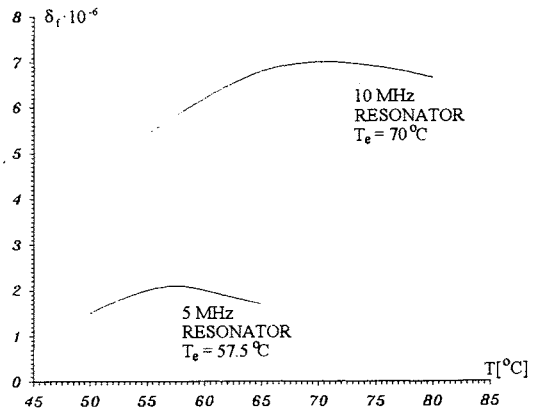


Fig. 3. F-T resonator characteristics.

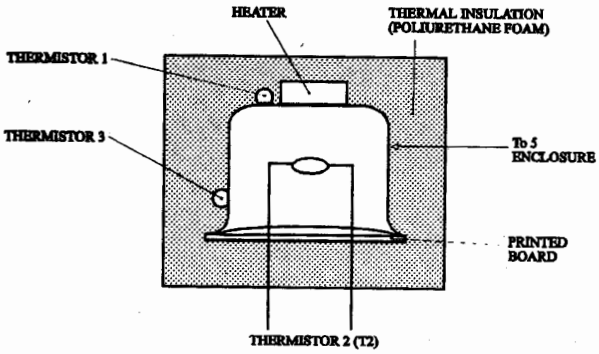


Fig. 4. Oven optimization.

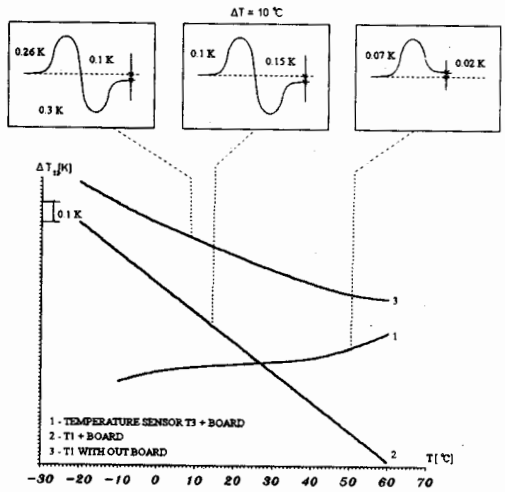


Fig. 5. Oven temperature characteristics.

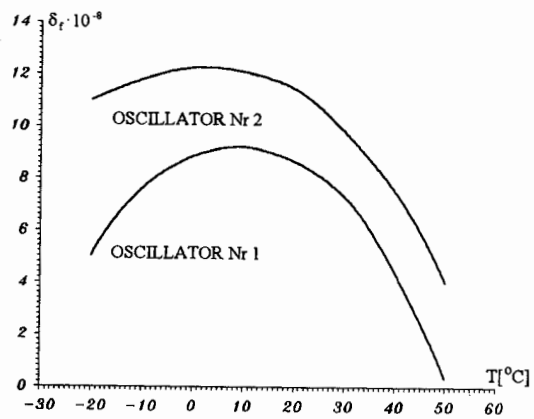


Fig. 6. F-T oscillators characteristics.

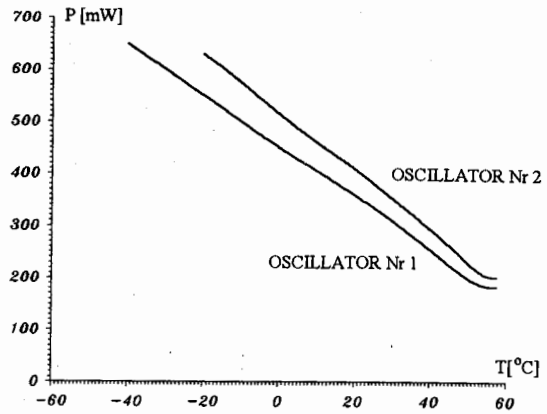


Fig. 7. Power consumption as temperature function.

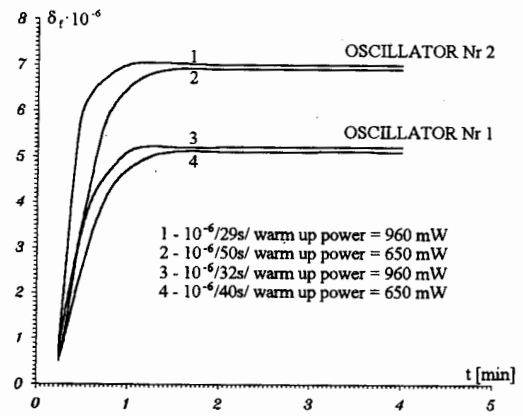


Fig. 8. Warm up time with 10^{-6} accuracy.

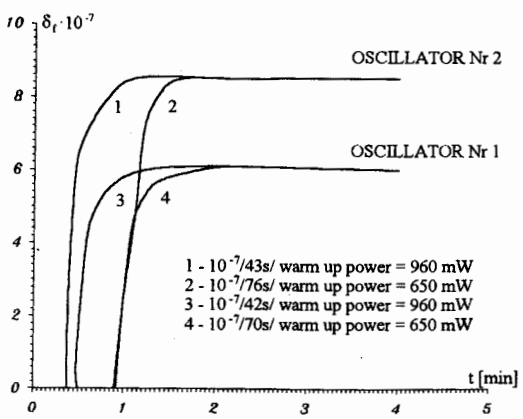


Fig. 9. Warm up time with 10^{-7} accuracy.

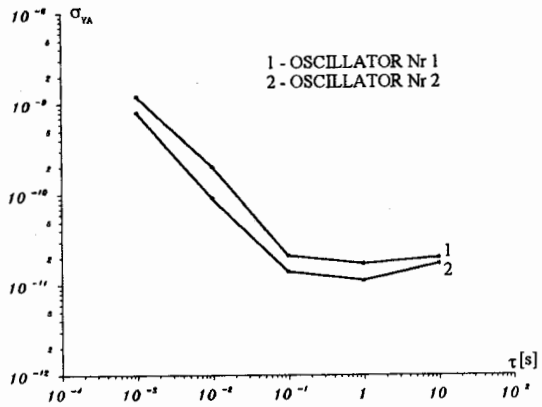


Fig. 10. Short-term frequency stability.

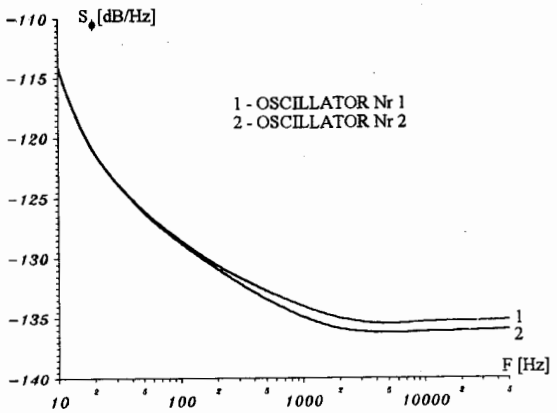


Fig. 11. Noise characteristics in frequency domain.

Quartz Frequency Control Products For Modern Telecommunications Systems.

M.P. Cracknell

C-MAC Quartz Crystals Ltd, Edinburgh Way, Harlow,
Essex, England, CM20 2DE.

ABSTRACT

This Paper describes a wide range of high quality, quartz frequency control products which have been designed for use in modern high bit rate telecommunication systems, such as;
SDH (Synchronous Digital Hierarchy) and SONET (Synchronous Optical Network), as well as non-synchronous systems. Data is presented on Voltage Controlled Oscillators (VCO) in the frequency range 155.52 MHz to 2488.32 MHz based on either bulk resonators or Surface Acoustic Wave (SAW) delay lines. One oscillator with temperature compensation is presented. Data is also provided on a family of SAW filters with centre frequencies in the range 278.528 MHz to 2592 MHz. Accelerated ageing tests on bulk resonators and SAW delay lines show that excellent long term stability is expected.

Keywords : Quartz, oscillators, filters, synchronous, SAW, STW

1. INTRODUCTION

Modern, high bit rate telecommunication systems, whether they are SDH, SONET or non-synchronous, use large quantities of quartz frequency control products. The main requirements are for VCOs, produced either with bulk resonators (VCXO) or SAW delay lines (VCSO) and for SAW filters (SAWF). Other product groups used are simple Crystal Oscillators (XO), Temperature Compensated Crystal Oscillators (TCXO), Oven Controlled Crystal Oscillators (OCXO) and Temperature Compensated Voltage Controlled Crystal Oscillators (TCVCXO). Typical system applications for VCOs are clock generation, clock synchronisation and clock extraction. SAW filters are used in clock extraction and data re-timing. C-MAC, until recently part of the major telecoms systems manufacturer Northern Telecom, and prior to that STC, has developed a full line of products designed specifically for these applications. Table 1 contains details of C-MAC's capability in products for these applications.

Frequency MHz	XO	VCXO	TCXO	TCVCXO	VCSO	SAWF
51.84 [1]	●	●	●	●		
68.736 [3]	●	●	●	●		
139.264 [3]	●	●	●	●	●	●
155.52 [1,2]	●	●	●	●	●	●
278.528 [3]	●	●			●	●
600.304 [3]					●	●
622.08 [1,2]					●	●
2488.32 [1,2]					●	●
2592.00					●	●
Application	1. SONET		2. SDH		3. Non-synchronous	

Table 1

2. OSCILLATORS

Oscillators for use in telecommunication systems have a number of key performance requirements. Amongst the more difficult requirements to meet are high carrier frequencies, low jitter, highly linear pulling and wide modulation bandwidths. Unlike some applications, temperature stability is not necessarily critical, as devices tend to be operated in phase locked loops. Providing that sufficient pulling is available to allow the unit to remain locked, despite variations in temperature, problems should not arise. For applications where temperature stability is critical, compensated oscillators are available. The following sections will give typical specifications and provide performance data.

2.1 TCVCXO

A TCVCXO for use in SONET applications has been produced. A summary of the specification follows;

Nominal Frequency	155.52 MHz
Supply Voltages	+12 V,-4.5 V
Operating Temperature Range	-40 to +85 °C
Frequency Change With Temperature And Control Voltage	±3.5 ppm
Frequency Control Range For Vc = ±8 V	±40 to ±55 ppm
Linearity (Ref.4)	±1.5 %
Output Type	Complimentary ECL
Package Type	24 Pin Dip

The device is realised using a High Frequency Fundamental (HFF) crystal at 51.84 MHz, and a three times frequency multiplier. This approach delivers the wide-range, linear pulling required by the application. High overtone crystals would not provide the necessary degree of pulling. The use of lower frequency fundamental crystals with high order frequency multiplication would demand complex filtering and yield low efficiency. Sub-harmonic content at the output is typically suppressed by 40 dB, thus minimising problems with jitter in the final application. Temperature compensation is performed using C-MAC's own analogue chip compensation technology (Refs.1,2).

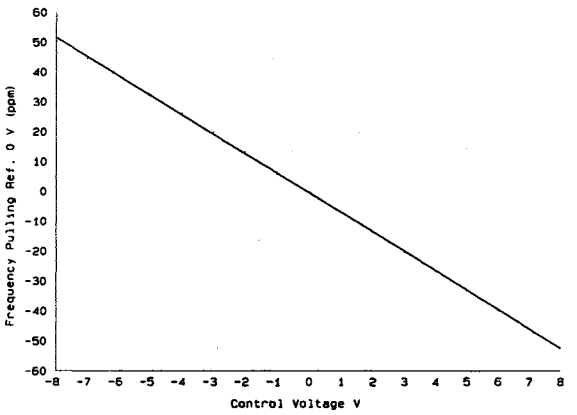


Figure 1

Figure 1 shows that a pulling of ±50 ppm and a linearity of +0.7/-1.3 % have been achieved.

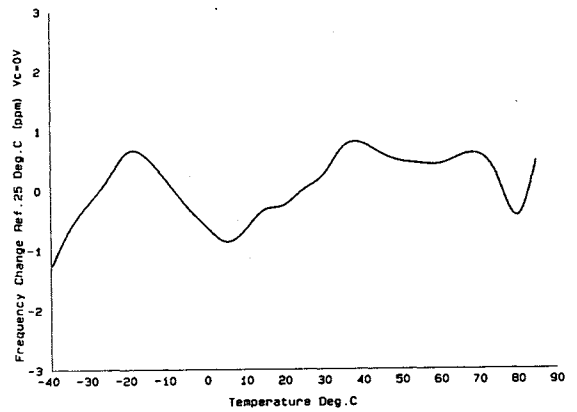


Figure 2

Figure 2 shows a frequency vs temperature performance of +0.8/-1.3 ppm Ref. 25 °C for the compensated device.

2.2 VCSO

VCSOs come into their own when there is a need to work at high frequencies such as 622.08 MHz or 2488.32 MHz. Several VCSOs for use in SDH/SONET applications have been produced. A summary of one specification follows;

Nominal Frequency	2488.32 MHz
Supply Voltages	+12 V
Operating Temperature Range	0 to +80 °C
Frequency Change With Temperature And Control Voltage	+20, -200 ppm
Frequency Control Range For Vc = ±5 V	+700 to -200 ppm
Linearity (Ref.4)	±2.5 %
1 dB Modulation Bandwidth	500 kHz
Output Type	Sine Wave, +16 dBm
Package Type	32 x 48 x 8 mm Hybrid Tub

The device designed to meet the above specification uses a SAW delay line in an oscillator at 622.08 MHz. The oscillator is followed by a four times frequency multiplier. The worst sub-harmonic content is typically suppressed by 45 dB. The major difficulty experienced with the design was in achieving the very wide frequency pulling and good linearity, whilst maintaining sufficient modulation bandwidth. The problem was resolved by using a multiple section phase shift network built around quadrature hybrids. Figure 3 shows that a pulling of 900 ppm and a linearity of +2.1/-1.4 % have been achieved.

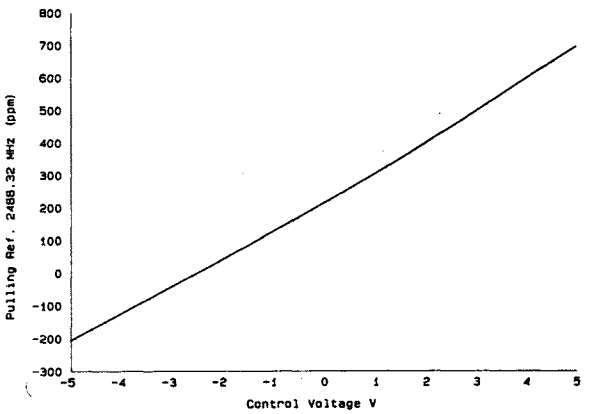


Figure 3

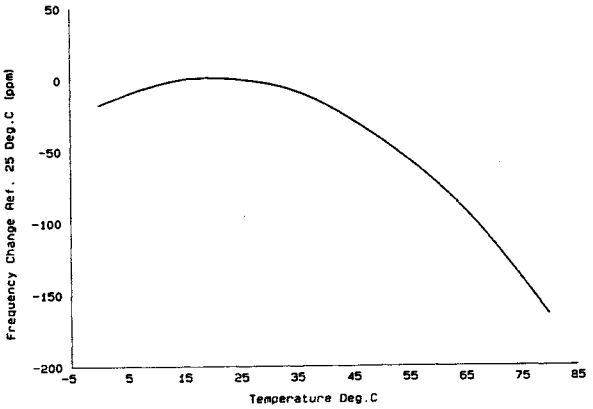


Figure 4

Figure 4 Shows the frequency vs temperature characteristics of the oscillator at the centre of the control voltage range. The results achieved at the extremes of the control voltage range typically track this result within 20 ppm.

The useful characteristics of the VCSO may be employed at lower frequencies by addition of an ECL divider to the O/P of the oscillator. A 155.52 MHz oscillator requiring in excess of 1000 ppm pulling has been produced in this way. VCSOs also lend themselves to temperature compensation. Further details of a device produced at 600.304 MHz can be found in Ref.3.

3. SAW FILTERS

3.1 Application

SAW Filters are most commonly used to extract the clock signal from an incoming data stream. The data received will have been degraded by jitter and noise acquired along the transmission path. Using the recovered clock signal, it is possible to re-time the received data, thus removing the uncertainties introduced during transmission. Figure 5 shows a simplified block diagram of a clock recovery system.

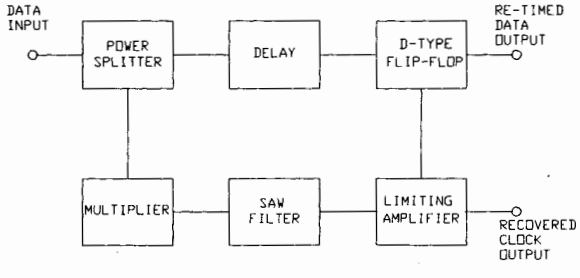


Figure 5

The incoming data stream is split into two parts. One part of the signal is applied to the input of a D-Type flip-flop, the other goes to a multiplier circuit. When viewed in the frequency domain, the incoming data stream does not contain a spectral line at the system clock frequency. The output from the multiplier circuit (commonly formed by a full-wave rectifier) however, does contain the appropriate spectral line, along with many others. The output is filtered by a SAW device leaving just the required clock frequency. To compensate for the loss of the SAW filter and to increase the steepness of the edges applied to the clock input of the D-Type flip-flop a limiting amplifier is used. The delay inserted in the data path ensures that the clock pulse is applied to the flip-flop at the centre of the bit period of the data to be re-timed.

3.2 Typical characteristics

SAW Filters used for clock recovery in SDH/SONET systems typically require a 3 dB bandwidth Q-value of 500-800. Earlier, non-synchronous systems worked with Q-values of around 150-500. The associated insertion loss is dependent on several factors but will usually be within the range 15 to 20 dB when measured in a 50 ohm system. Key areas of the device performance are passband shape, phase linearity, temperature coefficient and ageing. Figures 6, 7 and 8 show the performance of a device for use in an SDH clock recovery application.

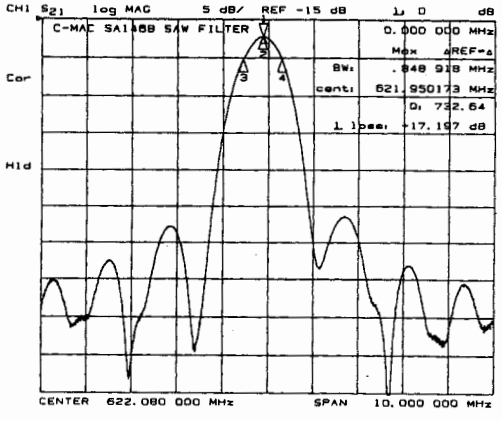


Figure 6

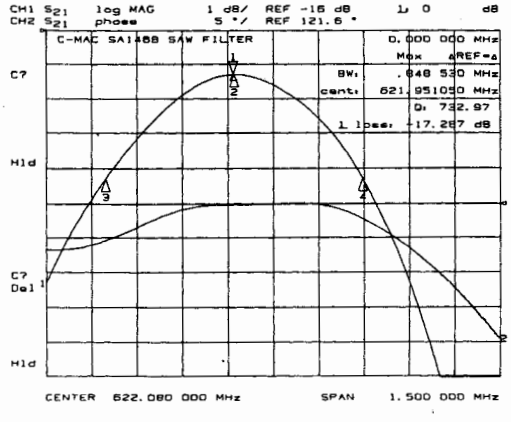


Figure 7

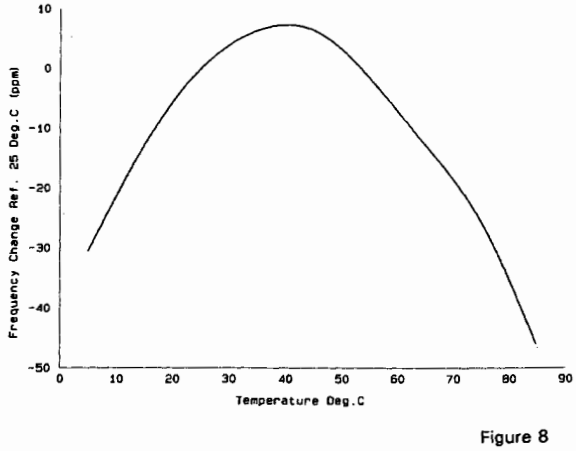


Figure 8

It can be seen from Figure 6 that the overall shape is very smooth with good sidelobe suppression. Figure 7 shows that the phase ripple over the region of the device's 3 dB bandwidth is 5 degrees peak to peak, while Figure 8 displays the typical, parabolic nature of the device's frequency vs temperature characteristic. It is common practice to place the inversion point at a temperature well away from the geometric centre of the operating temperature range. This can be used as a simple means of compensating for the temperature coefficient of the associated electronics.

The SAW filters C-MAC produces for use in the higher bit rate systems at 2.4 and 2.5 Gbits/sec do not rely on Rayleigh Wave propagation, but use Surface Transverse Waves (STW). The primary advantages of using the STW mode are the higher propagation velocity and lower propagation loss. The penalty that has to be paid with the STW mode is that the temperature coefficient increases from $\approx$-3.2 ppm/°C^2 to $\approx$-5.6 ppm/°C^2. In order to minimise the effects of direct breakthrough, the device is constructed using three identical transducers, one as the input and two for the output. There is a 180 degree phase shift between the two outputs at the clock frequency. When the outputs are combined using a 180 degree power splitter or differential amplifier, the desired signals add and the unwanted breakthrough signals cancel each other out.

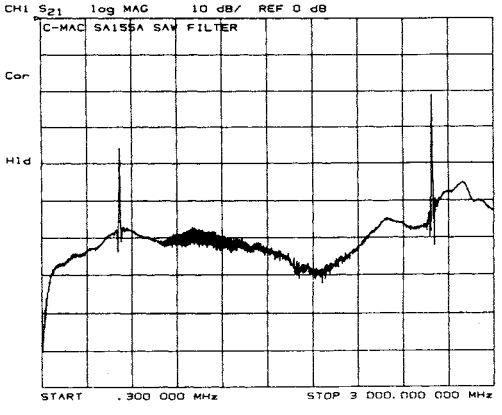


Figure 9

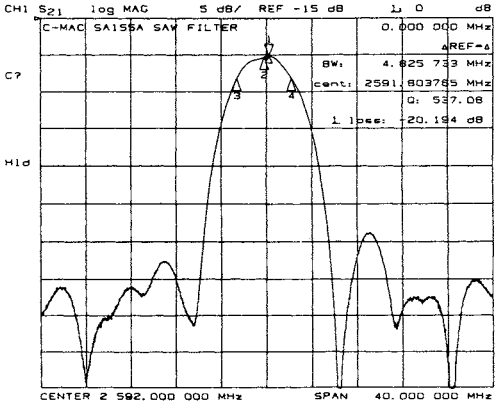


Figure 10

Figures 9 and 10 show the performance achieved by a 2592 MHz STW filter in a 50 ohms balanced test fixture. Figure 9 is the broad band response between 0.3 and 3000 MHz. The device operates on the fifth harmonic. Both the fundamental and the fifth harmonic can clearly be seen. Figure 10 shows an overall good shape with no apparent problems due to breakthrough.

4. AGEING

Data is provided on typical examples of bulk resonators and SAW filters housed in surface mount, seam welded, metal-ceramic packages. The intended application for these parts is in the next generation of submarine optical communication systems which will operate at 2.5 Gbits/s. The data has been gathered using automatic test equipment to monitor devices immersed in a bath of silicone oil at a temperature of 125 °C. Predicted ageing performance is based on a log fit as defined in Ref.4. The acceleration factors used in the predictions have been determined during previous extensive trials on similar products.

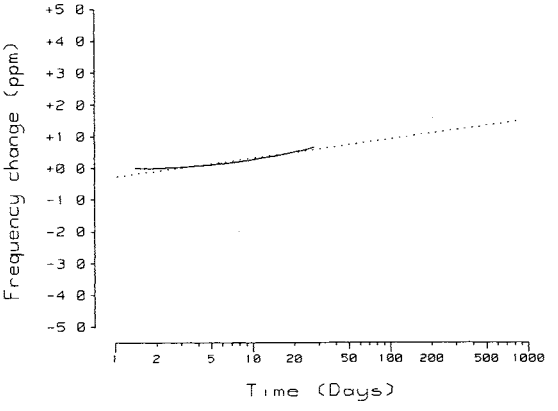


Figure 11

Figure 11 shows that an ageing of +0.67 ppm is observed on a 51.84 MHz bulk resonator over a period of 28 days at 125 °C. From this figure the predicted ageing for 25 years operation at 25 °C is +0.2 ppm. An acceleration factor of 12 has been used to arrive at this result.

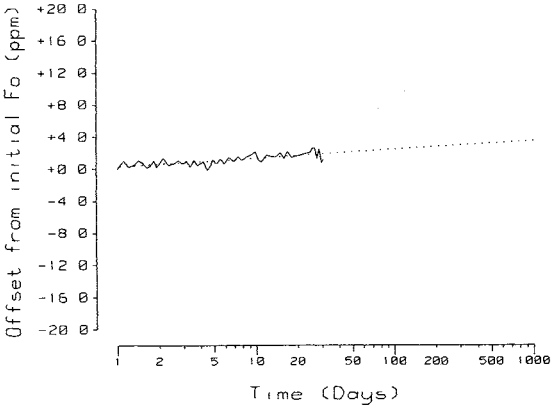


Figure 12

Figure 12 shows that an ageing of +1.6 ppm is observed on a 2592 MHz SAW filter over a period of 31 days at 125 °C. The predicted ageing during 25 years operation at 25 °C is +0.1 ppm. An acceleration factor of 35 has been used to arrive at this figure. These results are in close agreement with those obtained on other SAW devices which have been on long term ageing for periods in excess of 5000 hours (210 days).

It should be stressed that although the SAW filters which have been measured are going into probably one of most demanding of applications where long term performance is concerned, the materials and processing techniques used are common to those parts produced for commercial applications.

5. CONCLUSIONS

Data has been presented on a range of high quality, quartz frequency control products developed for use in modern high bit rate telecommunication systems. Results have shown that a high standard of performance has been achieved across a broad line of products, ranging from bulk resonators at 51.84 MHz through VCSOs at 2488.32 MHz to an STW filter at 2592 MHz. Examples of these products may be found in some of the most demanding of telecommunications applications.

6. ACKNOWLEDGEMENTS

The author wishes to acknowledge the assistance of the following colleagues in the preparation of this paper; Mr D.E. Bower, Mr N. Churchill, Mr M. Varcoe.

7. REFERENCES

1. James S Wilson "Precision Crystal Oscillator Design Thesis for Degree of Doctor of Philosophy". University of Aston, October 1983.

2. J.S. Wilson "An Improved Method for the Temperature Compensation of Crystal Oscillators". Proceedings of 32nd Annual Frequency Control Symposium (1983) p.442-447.

3. M.P. Cracknell, A.M. Harrison, D.J. Sharpe "A New Technique for the Temperature Compensation of SAW Oscillators". Proceedings of 3rd European Time and Frequency Forum (1989) p.169-172.

4. CECC 69000, Generic Specification: Quartz Crystal Controlled Oscillators.

POSTER 2 - FREQUENCY CONTROL AND TIMING SYSTEMS

Chairman: L. Prost

SPECTRUM CHARACTERISTICS OF GPS TIME COMPARISON AFFECTED BY SA EFFECT

Hu Jinlun Jian Shuguang

(SHANGHAI OBSERVATORY,CHINESE ACADEMIA OF SCIENCES)

ABSTRACT

At the beginning of 1992, we carried out some experiments in order to understand the characteristics of GPS time comparison affected by SA effect. The sampling time of data was 10 seconds and the tracking period of satellites was 1 hour in these experiments.

The results of spectral analyses in these experiments are shown as follows:

1. Under normal conditions, the spectrum characteristics of GPS time comparison are basically smooth. There are not obvious periodic phenomena and the noise process is a white noise PM.
2. When GPS time comparison affected by SA effect, the low spectrum are relatively obvious, and there are obvious periodic phenomena. The periodic signals of 5~15 minutes take up about 77% among these obvious periods. This is the reason which leads to a marked drop in the precision of GPS time comparison.

Keywords: GPS, Satellites, Time Comparison, SA, Spectrum, Period

INTRODUCTION

Until January 1992, there are 16 satellites which are operating in the sky for the implement of GPS plan. Block I are 5 satellites (PRN: 3,6,11,12,13) and Block II are 11 satellites (PRN: 2,14,15,16,17, 18,19,20,21,23,24). In order to make experiment or have other purpose at sometime, the signal characteristics of Block II satellites are turned from the normal conditions into SA conditions. The SA conditions means to add random modulation signal in the broadcast of GPS . Thus the timing results (or the positioning results) affected by SA effect have been much different from the results obtained under the normal conditions. Since november 15, 1991, the most satellites of Block II have been operated at SA conditions. Sometimes only one satellite or a few satellites of Block II operte in the normal conditions. However, it is not constant and can be changed at suitable times. But Block I satellites always operate at the normal conditions. How are the charcteristics of GPS time comparison affected by SA effect? it is special concerned by GPS time transfer users, because there are a marked drop in the accuracy and the precision of GPS time comparison affected by SA effect. Thus, it will seriously affect the results of GPS time synchronism in the long distance. Therefore, by means of comparisons and analyses from the receiving data under the two conditions, and trying to reduce the influence due to SA effect, it have already become an important problem for GPS time transfer.

At the beginning of 1992, we carried out some experimens and analyses in order to compare the results of GPS time comparison under the two conditions. The variance analyses in time domain are shown: the precision of GPS time comparison affected by SA effect (calculated from the receiving dsta which the sampling time is 10 seconds and the data length is 13 minutes) is reduced 5~10 times than the results under the normal conditions. Some analytic results in frequency domain for GPS time comparison affected by SA effect will be described in this paper.

CHARCTERISTICS ANALYSIS IN FREQUENCY DOMAIN

Fig. 1 is an example which expresses the changes of measurement values using GPS time comparison under the two conditions. In the figure, the comparison signal is [UTC(SO)-GPS(PRN J)], where J is PRN number of satellite, and the sampling time of data is 10 seconds, and the tracking period of satellites is 1 hour. The situations of other Block I satellites are similar with the situation of PRN3 satellite in fig. 1, in which their fluctuations of amplitude are not large. The situations of other Block II satellites under SA conditions are similar with the situation of PRN16 satellite in fig. 1, in which their fluctuations of amplitude present obvious change wave upon wave.

The objective of charcteristic analyses in frequency domain is to decide whether there are the obvious low frequency noise or periodic phenomena in the signal. It can be obtained by an analysis of function of power spectral density (PSD) and periodogram.

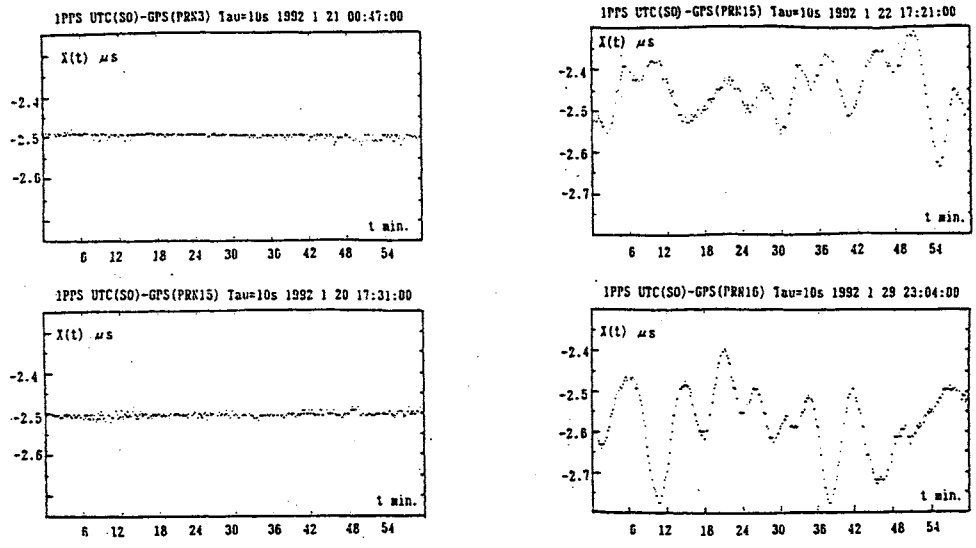


Fig. 1 The changes of time difference in GPS time comparision under the two conditions

The function of PSD is Fourier transform of autocorrelation function. The function of PSD Sx(f) can be calculated by the autocorrelation function Rx(k), which is obtained from the dispersed data of GPS time comparison. The expression is as follows:

$$Sx(f) = 2\,[Rx(0) + 2\sum_{k=1}^{N-1} Rx(k)\cdot COS2\pi kf] \qquad (1)$$

$$(\ f = j/T \quad j = 1,2,\ldots,N/2\ )$$

where N is the sampling numbers of data, and T is the time interval of finite data. The function of PSD Sx(f) can be directly calculated using FFT method with the time difference X(t) of GPS time comparison. Then the calculated results must be smoothed using the spectral window (Hamming or Hann), in order to reduce the dispersed degree. These details are omitted here.

Some charcteristics of the variable X(t) can be determined from the variations of function of PSD Sx(f) with frequency f: (1) If the variations are basically smooth or the fluctuations are small, X(t) belongs to white noise PM; (2) If the slope of the variations presents an exponential attenuation, X(t) belongs to low frequency noise process; (3) If there are obvious peak value in the variations, X(t) presents notable periodic phenomena.

It need to be determined notable periodic phenomena which can be obtained from the periodogram analysis by using Fourier transform. Fourier expression of the dispersed variable X(t) is:

$$X(t) = a_0 + \sum_{j=1}^{k}\left(a_j \cos\frac{2\pi jt}{N} + b_j \sin\frac{2\pi jt}{N}\right)$$

$$= a_0 + \sum_{j=1}^{k} A_j \sin\left(\frac{2\pi jt}{N} + \theta_j\right) \qquad (2)$$

$$k = \begin{cases} N/2 & N \text{ is even number} \\ (N-1)/2 & N \text{ is odd number} \end{cases}$$

$$t = 1,2,\ldots,N$$

$$A_j = \sqrt{a_j^2 + b_j^2} \qquad \theta_j = \arcsin(a_j/A_j)$$

The parameters a_0 、a_j 、b_j can be obtained by FFT method and are incorporated into the amplitude A_j , and the phase θ_j . Then it is derternined whether there are some notable periods using Fisher check. In Fisher chech, the notable level α is adoped 0.01, and choose three notable periods for comparison. The description of FFT and Fisher check are omitted here.

The results in frequency domain and time domain are correspondence. Their results are given together for comparison in following figure and tables. Fig. 2 is an example of the autocorreclogarm and the power spectra of GPS time comparison under two conditions. Table 1 and table 2 are the results of charcteristic analyses in frequency domain

and time domain for GPS time comperison of Block I and Block II satellites respectively In the two tables, the results in time domain are presented the charcteristic of autocorrelation function and the slope of noise process (fitting from two kind of Allan variance within different sampling time, the relative coefficient is most above 0.9) [1,2]. These figure and tables are shown: (1) The spectrum charcteristics of GPS time comparison under normal conditions are basically smooth (all Block I satellites and a few Block II satellites), its period charcteristics are not obvious, and the noise charcteristics of frequency are white noise PM; (2) The low spectrum of GPS time comparison affected by SA effect (most Block II satellites) are relatively obvious, its periodic phenomena are notable. The periodic signals of 5~10 minutes take up 77 % among the three notable periods. This is the reson why the precision of GPS time comparison affected by SA effect is quickly reduce. It is also shown that there are some statistical rule in GPS time comparison affected by SA effect.

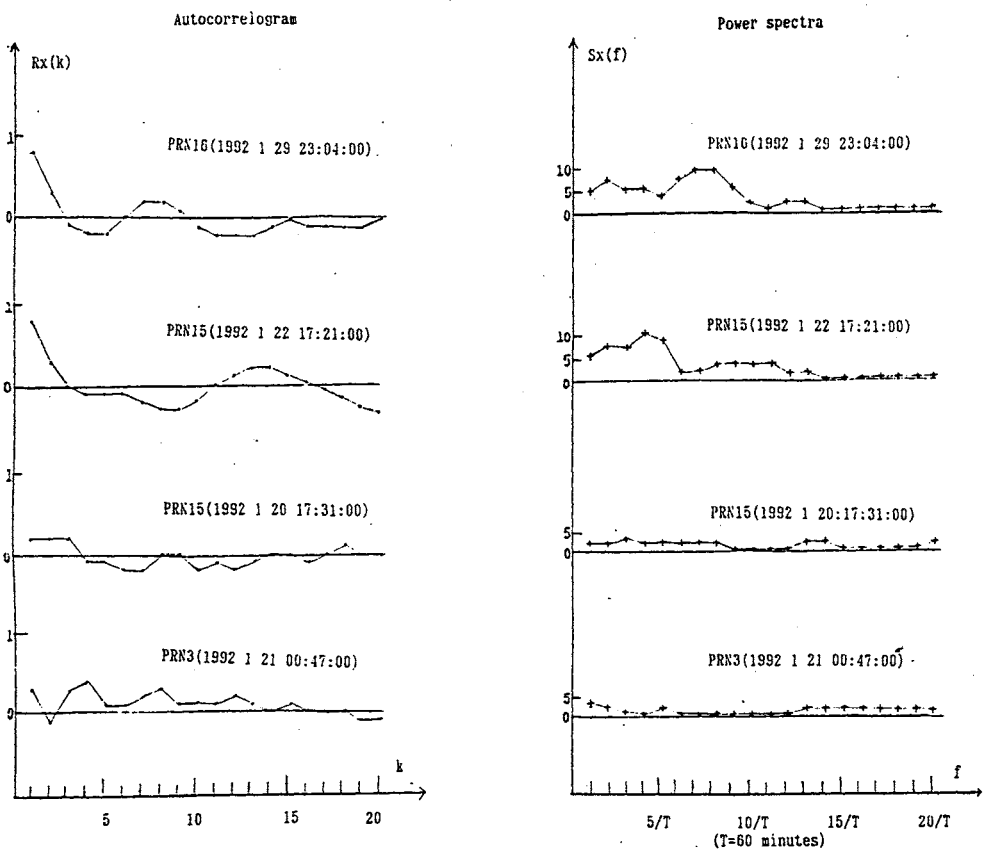


Fig. 2 Correlogram and spectrum of GPS time comparision

Table 1. The results of charcteristic analyses in frequency domain and time domain for GPS time comparison (BLOCK I satellites)

Satellite number	Receiving date and start time	PSD function	Periods	Autocorrelation function	Slope of noise process		
					τ s	$\sigma_y(\tau)$	Mod $\sigma_y(\tau)$
PRN 3	1992 1 21 00:47:00	A	B	A	10~500	-1.006	-1.467
PRN 6	1992 1 21 01:49:00	A	B	A	10~500	-0.998	-1.427
PRN11	1992 1 29 11:27:00	A	B	A	10~500	-0.995	-1.234
PRN12	1992 1 22 03:13:00	A	B	A	10~500	-1.027	-1.402
PRN13	1992 1 21 21:53:00	A	B	A	10~500	-0.978	-1.258

A:basically soomth, B:not obvious.

Table 2. The results of charcteristic analyses in frequency domain and time domain for GPS time comparison (BLOCK II satellites)

Satellite number	Receiving date and start time	PSD function	Periods(arrange in the notable order, time unit: minute)	autocorrelation function	Slope of noise process		
					τ s	$\sigma_y(\tau)$	Mod $\sigma_y(\tau)$
PRN 2	1992 1 20 16:07:00	C	8.6, 12.0, 15.0	D	10~ 40 50~100	-0.539 0.456	-0.725 0.472
PRN14	1992 1 20 19:27:00	C	10.0, 8.6, 60.0	D	10~ 40 50~100	-0.638 0.536	-0.812 0.472
PRN15	1992 1 22 17:21:00	C	15.0, 30.0, 12.0	D	10~ 40 50~100	-0.590 0.477	-0.749 0.443
PRN16	1992 1 29 23:04:00	C	6.7, 7.5, 8.6	D	10~ 40 50~100	-0.475 0.592	-0.577 0.564
PRN17	1992 1 20 02:11:00	C	20.0, 8.6, 6.7	D	10~ 40 50~100	-0.457 0.203	-0.614 0.245
PRN18	1992 1 19 22:31:00	C	7.5, 6.7, 30.0	D	10~ 40 50~100	-0.576 0.500	-0.735 0.469
PRN19	1992 1 20 13:21:00	C	4.6, 7.5, 10.0	D	10~ 40 50~100	-0.536 0.332	-0.681 0.315
PRN20	1992 1 19 06:41:00	C	20.0, 30.0, 12.0	D	10~ 40 50~100	-0.619 0.464	-0.790 0.435
PRN21	1992 1 19 07:44:00	C	12.0, 8.6, 15.0	D	10~ 40 50~100	-0.492 0.523	-0.612 0.481
PRN23	1992 1 19 04:31:00	C	30.0, 12.0, 10.0	D	10~ 40 50~100	-0.482 0.586	-0.587 0.517
PRN24	1992 1 30 18:25:00	C	5.4, 27.8, 17.6	D	10~ 40 50~100	-0.660 0.213	-0.871 0.216
PRN15	1992 1 20 17:31:00	A	B	A	10~500	-0.980	-1.445

A:basically smooth, B:not obvious, C:low spectra is relatively obvious, D:change wave upon wave.

CONCLUSION

By means of the charcteristic analyses in frequncy domain for GPS time comparison affected by SA effect, in comparison with the results obtained under the normal condition, some statistical substance of GPS time comparison affected by SA effect can be clearly understood. However, these results are still preliminary from this study.

REFERENCE

1. Allan,D.W.,Barnes,J.A., 35th Annu. Symp. on Freq. Cont., P.470～475 (1981).

2. Hu Jinlun, Navigation, No.2, P.101～111 (Sum 1990).

INSTRUMENTAL DELAY CALIBRATION WITH ZERO BASELINE INTERFEROMETER FOR VLBI TIME TRANSFER

H.Kiuchi, S.Hama, J.Amagai, T.Yoshino, M.Imae

Communications Research Laboratory
893-1, Hirai, Kashima, Ibaraki 314, Japan

ABSTRACT

In the time and frequency comparison, the instrumental delay is one of the most important factors for determining the accuracy. For the measurement of the instrumental delay in the VLBI time comparison method, Zero Baseline Interferometry (ZBI) with the directly controlled reference frequency system, which has high accuracy was developed in Communications Research Laboratory. Using ZBI method, the fundamental experiment was carried out and the accuracy better than 0.2 nsec was obtained.

Keywords
VLBI : Very Long Baseline Interferometer
ZBI : Zero Baseline Interferometer
Instrumental delay
Antenna structure delay
Fringe
Ref-Rx : Reference Receiver

1. INTRODUCTION

It is well known that VLBI is very effective in geodesy, it is possible to measure not only baseline vector but relative time difference. With any methods for time comparison, the instrumental delay is one of the most important factors which determines the accuracy of the time comparison. The detail of the instrumental delay of the VLBI system and the Zero Baseline Interferometry (ZBI) method for measuring the instrumental delay will be discussed. Using the ZBI, the fundamental experiment was carried out and the accuracy of the instrumental delay measurement is obtained to be 0.2 nsec.

2. INSTRUMENTAL DELAY OF THE VLBI RECEIVING SYSTEM

In order to analyze the instrumental delay, we consider the signal phase in the VLBI system. Figure 1, the upper part shows the acquisition system and lower part shows the phase and delay calibration system.

The phase calibration signal (Pcal) is injected at the Low noise amplifier input. Therefore, the calibration signal is added to the signal from the Quasar. The phase calibration signal is generated by a Step Recovery Diode (SRD) or a tunnel diode, and contains a narrow rectangular pulse train. SRD works as a comb generator. The pulse width is less than 40 psec, and period is 1 μsec. The Pcal signal has many Fourier components in each MHz step whose phase are coherent to the reference signal. The phase of the reference signal at the input of the SRD is given by Eq.(1).

$$\phi_1(t) = \omega_{ref}(t + \Delta t_L - \tau_{cable}) \quad (1)$$

where ω_{ref} : reference signal frequency of the station (nominal 5 or 10 MHz)
t : UT
Δt_L : local time difference from UT
τ_{cable} : delay between the delay calibrator and SRD

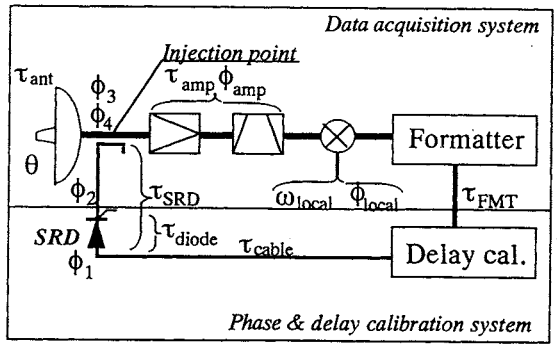


Fig. 1 The block diagram of the VLBI system

The output phase of SRD becomes as Eq.(2).

$$\phi_2(t) = \phi_1(t) \times n + \phi_{comb}$$
$$= \omega(t + \Delta t_L - \tau_{cable}) + \phi_{comb} \quad (2)$$

where $\omega = n\,\omega_{ref}$ (n is an integer)
ϕ_{comb} : $-\pi/2$ tunnel diode
: $\pi(n+(1/2))$ SRD

This signal is added to the signal from the Quasar at the injection point, at the front of the Low Noise Amplifier. Then it is assumed that the time delay by the cable from SRD output to the injection point is τ_{SRD}, which included delay of diode; τ_{diode}. The phase of the Pcal signal $\phi_3(t)$ at the injection point and the phase of the signal $\phi_4(t)$ from the Quasar are, respectively,

$$\phi_3(t) = \omega(t + \Delta t_L - \tau_{cable} - \tau_{SRD}) + \phi_{comb} \quad (3)$$

and

$$\phi_4(t) = \omega(t - \tau_{ant}) + \theta \quad (4)$$

where θ is the received signal phase from the Quasar and τ_{ant} is the antenna structure delay. It is supposed that the total phase of the frequency conversion in Down converter and Video converter is $\omega_{Local}\,t_L + \phi_{Local}$. And after frequency conversion,

the Pcal phase and the phase of the signal from the Quasar are as in Eq. (5) and Eq.(6).

$$\phi_5(t) = \phi_3(t) - \omega\tau_{amp} - \phi_{amp} - \omega_{Local}t_L - \phi_{Local} \quad (5)$$
$$\phi_6(t) = \phi_4(t) - \omega\tau_{amp} - \phi_{amp} - \omega_{Local}t_L - \phi_{Local} \quad (6)$$

We have to consider the delay of Formatter sampling clock. This delay affects the timing of the sampling clock. This delay, occurring between FMT and the phase and delay calibrator (Dcal), is expressed as τ_{FMT}. The phase of the Pcal signal at the sampling point and the phase of the signal are given in Eq.(7) and Eq. (8) respectively.

$$\phi_7(t) = \phi_5(t) - \omega_L\tau_{FMT} \quad (7)$$
$$\phi_8(t) = \phi_6(t) - \omega_L\tau_{FMT} \quad (8)$$

In the correlation processor, the cross correlation is made using the time on data. The time on Formatter, t_{FMT} or time on data, is delayed by τ_{FMT} from the phase and delay calibration system ground unit, because the origin of the calibration signal is the phase and delay calibration system ground unit. So one should consider t_{FMT} instead of t.

$$t_{FMT} = t + \Delta t_L - \tau_{FMT} \quad (9)$$

The phase of the sampled signal are given in Eq.(10) and (11) instead of Eqs.(7), (8) by using Eq. (9).

$$\phi_{Pcal}(t_{FMT}) = \omega(t + \Delta t_L - \tau_{cable} - \tau_{SRD}) + \phi_{comb} - \omega\tau_{amp} - \phi_{amp} - \omega_{Local}t_L - \phi_{Local} - \omega_L\tau_{FMT} \quad (10)$$

$$\phi_{signal}(t_{FMT}) = \omega(t - \tau_{ant}) + \theta - \omega\tau_{amp} - \phi_{amp} - \omega_{Local}t_L - \phi_{Local} - \omega_L\tau_{FMT} \quad (11)$$

The obtained delay τ corresponds to $(\Delta\phi_{signal} - \Delta\phi_{Pcal})$, bandwidth synthesis, where $\Delta\phi_{signal}$ and $\Delta\phi_{Pcal}$ are the difference of the ϕ_{signal} and ϕ_{Pcal} between the two stations, respectively. It is supposed that ω is the same in each station $(\omega=\omega_1=\omega_2)$. The suffix denotes each station.

$$\Delta\phi_{Pcal}(t_{FMT}) = \phi_{Pcal2}(t_{FMT}) - \phi_{Pcal1}(t_{FMT}) = \omega(\Delta t - \Delta\tau_{cable} - \Delta\tau_{SRD}) + [-\omega_2\tau_{amp2} - \phi_{amp2} - \omega_{Local2}t_{L2} - \phi_{Local2} - \omega_{L2}\tau_{FMT2}] - [-\omega_1\tau_{amp1} - \phi_{amp1} - \omega_{Local1}t_{L1} - \phi_{Local1} - \omega_{L1}\tau_{FMT1}] \quad (12)$$

$$\Delta\phi_{signal}(t_{FMT}) = \phi_{signal2}(t_{FMT}) - \phi_{signal1}(t_{FMT}) = \omega[-\Delta\tau_{ant}] + (\theta_2 - \theta_1) + [-\omega_2\tau_{amp2} - \phi_{amp2} - \omega_{Local2}t_{L2} - \phi_{Local2} - \omega_{L2}\tau_{FMT2}] - [-\omega_1\tau_{amp1} - \phi_{amp1} - \omega_{Local1}t_{L1} - \phi_{Local1} - \omega_{L1}\tau_{FMT1}] \quad (13)$$

where $\Delta t = \Delta t_{L2} - \Delta t_{L1}$

$\Delta\tau_{cable} = \tau_{cable2} - \tau_{cable1}$

$\Delta\tau_{SRD} = \tau_{SRD2} - \tau_{SRD1}$

$\Delta\tau_{ant} = \tau_{ant2} - \tau_{ant1}$

Then the obtained delay τ is as follows;

$$\omega\tau = \Delta\phi_{signal}(t_{FMT}) - \Delta\phi_{Pcal}(t_{FMT}) = \omega(-\Delta t - \Delta\tau_{ant} + \tau_g + \Delta\tau_{cable} + \Delta\tau_{SRD}) \quad (14)$$
$$\omega\tau_g = \theta_2 - \theta_1 \quad (15)$$

where τ_g is the geometrical delay.

In equation (14), Δt is the value of the clock difference, and τ_g is given by the VLBI experiment for geodesy. But we have to measure $\Delta\tau_{ant}$, $\Delta\tau_{cable}$ and $\Delta\tau_{SRD}$ to get the value of Δt. The time reference point of the station time in VLBI is not on Formatter but at the point of the phase and delay calibration system ground unit in Fig. 1, due to use the bandwidth synthesis method.

The delay τ_{ant} could be measured from the drawing, but the accuracy is limited to a few nsec. A solution to this problem is given in the next section. It is called **ZBI** (Zero Baseline Interferometry) method, and which is a kind of double difference method. Using this method, it is possible to get the value of $(-\Delta\tau_{ant} + \Delta\tau_{SRD})$.

We have discussed the ideal situation were each station has the same clock rate. In the truth, each station's clock is independent of others. But it is possible to compare the clock rate between the two stations if the change of the instrumental delay is supposed to be negligible. The differential phase is as follows;

$$\upsilon_F = \omega(d\tau/dt) = \omega(d\tau_{clk}/dt + d\tau_g/dt) = \{\delta_f + \alpha_f t + \omega_e(K_2\cos(\omega_e t) - K_1\sin(\omega_e t))\} \quad (16)$$

where;

υ_F : Fringe rate

δ_f : the frequency difference of the two stations

α_f : the aging rate difference of the two stations

ω_e : the earth rotation angular velocity

t : UT

α_f is negligible when the hydrogen maser is used. So it is possible to separate δ_f from other values because δ_f is independent of time. δ_f is the frequency difference between two stations. So after deciding the δ_f, in other words deciding the relative frequency rate $(\omega_2=\delta_f{}^*\omega_1)$, it is possible to apply the above given equations (Eq.(12) to Eq.(15)).

3.ZBI METHOD

It is clear that although a comparison of frequency between VLBI stations is possible without knowing $\Delta\tau_{ant}$, a comparison of time is impossible. Therefore, Zero Baseline Interferometry (**ZBI**) method is used in order to calibrate the instrumental delay.

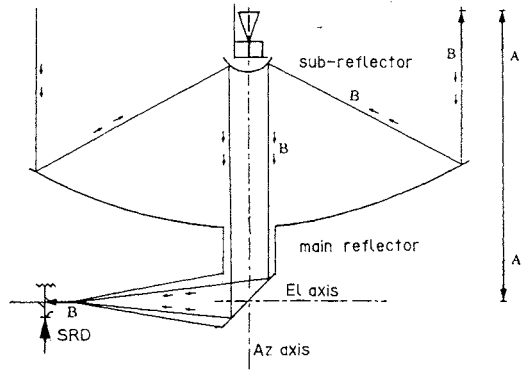


Fig. 2 Antenna structure delay and Zero Baseline Interferometer

In the geodetic VLBI, in principle, the geodetic reference point is the intersection point of the axis of azimuth and that of elevation, and is a stationary point. In other cases, such as X-Y mount etc., the geodetic reference point is the crossing of the

immovable axis and the projection of the movable axis. τ_{ant} is the delay between the geodetic reference point and the Pcal injection point from the star. In Figure 2, $\Delta\tau_{ant}$ means (the delay of the pass B) - (the delay of pass A). But it is difficult to measure the delay of the pass B from drawings because the pass B includes a tapered waveguide, a polarizer and so on.

In ZBI method, the Ref-Rx (reference receiver: which is a compact VLBI receiver) is mounted on the main antenna, and supply common reference frequency in both systems (main VLBI system and Ref-Rx system). The VLBI experiment was carried out between these systems. In this case the baseline length is zero and difference of the propagation media delay is zero, because the geodetic reference point is common for both systems. Therefore the estimated clock offset means the difference of the instrumental delays between the main antenna and the Ref-Rx (Eq.(14)).

So the Ref-Rx is used like a ruler. It is calculated that $\Delta\tau_{ant}$ is as follows;

$$\begin{aligned}&\Delta\tau_{ant}\\&= \{(\text{the delay of pass B}) - (\text{the delay of pass A})\}/c\\&= (\tau_{main} - \tau_{ref}) - (\text{the delay of pass A})/c \qquad (17)\end{aligned}$$

where τ_{main} : main antenna structure delay

τ_{ref} : Ref-Rx antenna structure delay

c : speed of light.

The length "pass A" from the setting point of the Ref-Rx to the geodetic reference point can be easily measured by using Laser ranging, because this pass is straight and the speed of light is as same as that in a vacuum, so the error is less than a few [mm].

ZBI is a kind of double difference method and the experiment must be carried out at both stations. In ZBI, Δt and τ_g are zero and the following two observation equations are obtained at both stations. Suffix means the each station.

$$\begin{aligned}&\Delta\tau_{station1}\\&= -(\tau_{main1} - \tau_{ref})\\&\quad + (\text{pass A}_{station1})/c + \Delta\tau_{cable1} + \Delta\tau_{SRD1} \qquad (18)\\&\Delta\tau_{station2}\\&= -(\tau_{main2} - \tau_{ref})\\&\quad + (\text{pass A}_{station2})/c + \Delta\tau_{cable2} + \Delta\tau_{SRD2} \qquad (19)\end{aligned}$$

where $\Delta\tau_{cable} = \tau_{cable} - \tau_{cable_ref}$

$\Delta\tau_{SRD} = \tau_{SRD} - \tau_{SRD_ref}$

τ_{cable_ref} is the cable delay of the Ref-Rx system which is the difference value in each station because it is difficult to use the same cable. τ_{SRD_ref} is the SRD delay of the Ref-Rx system which is common value in each station because it is the same SRD in Ref-Rx. To calculate the double difference, difference of the observed is as follows;

$$\begin{aligned}&\Delta\tau_{station1} - \Delta\tau_{station2}\\&= (-\tau_{ant1} + \tau_{ant2}) + (\text{pass A}_{station1} - \text{pass A}_{station2})/c\\&\quad + (\tau_{cable1} - \tau_{cable2}) - (\tau_{cable_ref1} - \tau_{cable_ref2})\\&\quad + (\tau_{SRD1} - \tau_{SRD2}) \qquad (20)\end{aligned}$$

then

$$\begin{aligned}&(-\tau_{ant1} + \tau_{ant2}) + (\tau_{SRD1} - \tau_{SRD2})\\&= (\Delta\tau_{station1} - \Delta\tau_{station2})\\&\quad - (\text{pass A}_{station1} - \text{pass A}_{station2})/c\\&\quad - (\tau_{cable1} - \tau_{cable2}) + (\tau_{cable_ref1} - \tau_{cable_ref2}). \qquad (21)\end{aligned}$$

where $(-\tau_{ant1} + \tau_{ant2}) + (\tau_{SRD1} - \tau_{SRD2})$

$= \Delta\tau_{ant} - \Delta\tau_{SRD}$ see Eq. (14).

τ_{cable_ref} is the delay of the cable used for installing the Ref-Rx in the ZBI experiment and can be measured by the delay calibrator. The term of (pass $A_{station1}$ - pass $A_{station2}$) can be measured by Laser ranging. Thus, using Eq. (21) and the above measured values, ($\Delta\tau_{SRD} - \Delta\tau_{ant}$) can be obtained and is used for the correction of the observation equation Eq.(14). So what we have to measure are simply two quantities for instrumental delay calibration, these been

$\Delta\tau_{cable}$ and (pass $A_{station1}$ - pass $A_{station2}$).

4.THE EFFECT OF THE PCAL SIGNAL IN CASE OF ZBI

The phase calibration signal is used to compensate for the instrumental delay and phase change from phase calibration signal injection point to Formatter, and is necessary for the bandwidth synthesis method. However, it may sometimes cause error in the VLBI correlation process. In fact, without fringe rotation, as in ZBI or for slow fringe rates in current VLBI, the output of the correlation processor has sine wave components consists of 10 kHz and 1 MHz + 10 kHz sine waves due to the phase calibration signal.

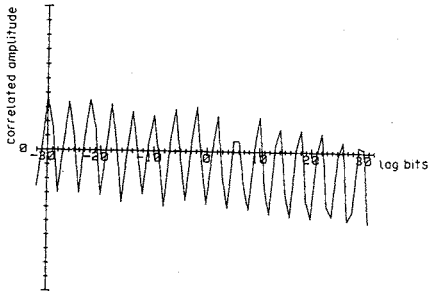


Fig. 3 The correlated raw data in case of ZBI

Figure 3 shows the correlated raw data and illustrates this effect in the case of ZBI. The abscissas are lag, and the ordinates are correlated amplitude. In VLBI observation with a long baseline, the signal from the star exhibits Doppler shift due to earth rotation, which has to be corrected by fringe rotation in order to obtain the fringe. In the process of fringe stopping, the phase calibration signal and DC offset due to the sampler are rotated by the fringe stopping (Doppler frequency canceling) and the coherence of the phase calibration signal is lost. As a result, the cross correlation is not affected by the phase calibration signal. But in VLBI observation with a zero or short baseline, the Doppler frequency between two stations is zero or very small. In such a situation, the results of cross correlation can be affected by the phase calibration signal.

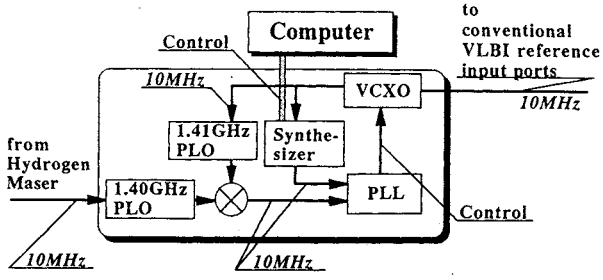


Fig. 4 The ZBI with the directly controlled reference frequency system

We can remove this effect by directly controlled reference frequency. This system is established between the hydrogen maser and the VLBI system, as shown in Fig. 4. The source of the frequency oscillator is a BVA crystal oscillator, which has high signal clarity and can be used as a VLBI signal reference. In this system, the BVA crystal is used as VCXO, which is phase-locked to the hydrogen maser. In order to obtain high frequency

resolution, the reference signal from the hydrogen maser is multiplied by PLO (phase-locked oscillator). The frequency of the PLO is selected as 1.400 GHz. The VCXO is controlled by a PLL circuit and the output frequency depends on the synthesizer frequency, which is under control of the host computer. With this method, frequency control is very accurate and frequency resolution reaches 7.1×10^{-13} with a mili-Hertz resolution synthesizer. This is a complicated system, but it has good signal purity.

We have adopted a system in which the rate of reference frequency used for both the frontend and backend of the VLBI data acquisition terminal, is controlled directly by a host computer. This method can be beneficial to us of conventional VLBI systems without any modification. In VLBI experiments using this system, pseudo fringe rotation is performed on all receiving frequencies simultaneously.

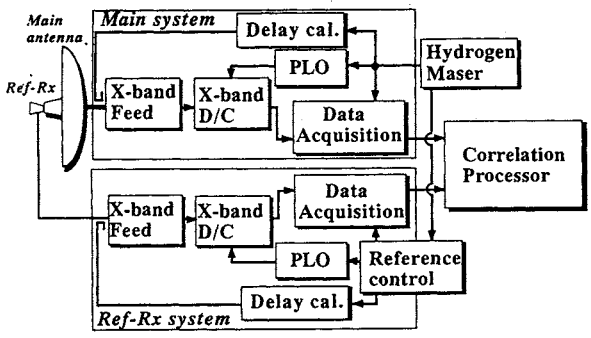


Fig. 5 Block diagram of ZBI experiment

A fixed local system as in conventional VLBI systems is used in main antenna, and a directly controlled reference frequency system is used for Ref-Rx, as shown in Fig. 5. The origin of the reference signal in both systems is supplied by the same hydrogen maser. Conventional correlation processing was performed between acquired data on both systems.

The effect due to the phase calibration signal and DC offset due to the sampler can be eliminated by this method.

5.RESULT OF THE FUNDAMENTAL EXPERIMENT

The purpose of the fundamental experiment is to detect fringe in order to check the overall system performance of ZBI. The observation was made for a few hours.

We used the 26 m Az-El type Radio telescope at CRL Kashima and the Ref-Rx system. The Ref-Rx is installed with an X-band polarizer, an X-band FET amplifier, a frequency converter with a local oscillator phase-locked to reference 10 MHz signal and a pulse generator for instrumental delay calibration. We adopted the septum type polarizer for Ref-Rx, which has no dielectronic components, in order to keep out of the disturbance of the external temperature change. The receiver is capable of receiving X-band signal for VLBI observation, but is not capable of receiving S-band signal for the ionospheric delay correction. In ZBI observation, the effect of the ionospheric and the tropospheric delay is common for both systems.

Hydrogen maser frequency standard supplied the reference signal for both systems. After the observations, the data on the tapes of both system were reproduced and cross correlated by K-3 correlation processor. The sources are Moon, Tau-A, Cas-A, Cyg-A, etc.. These sources are very strong. We can get good and high SNR fringes (Fig. 6) in spite of the very small diameter antenna (23 cm ϕ) of Ref-Rx.

In this experiment, we set different local frequencies (lst local and Video local frequency: Table 1)between main antenna system and Ref-Rx system, in order to keep out of influence of interference in IF band.

The length of "pass A" was measured using the Laser ranging method, which has an accuracy of better than a few [mm]. To summarize the result, we are confident that an accuracy of 0.2 ns or better has been reached.

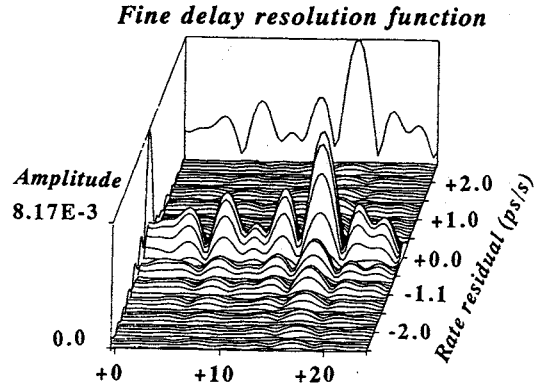


Fig. 6 Result of the ZBI experiment

Table 1 Local frequency configuration of the ZBI

Radio Frequency [MHz]	*Video frequency*	
	Main antenna local 8080 MHz	*Ref-Rx local 8120 MHz*
8220	140	100
8240	160	120
8300	220	180
8420	340	300
8460	380	340
8520	440	400
8560	480	440

6. DISCUSSION

We have discussed about Instrumental delay calibration. Our method is combined the ZBI method and the directly controlled reference frequency method. The Ref-Rx is used as a portable reference VLBI station, which is as same concept as the portable cesium clock for time comparison. To use ZBI method, it has some restrictions. The main antenna must be large enough to mount the Ref-Rx. The arrangement of the progress to mount the Ref-Rx on the other site is complicated job. We start to develop Short Baseline interferometer system, of which antenna must be larger than that of ZBI but easy to setting.

7. CONCLUSION

In the time and frequency comparison, the instrumental delay is one of the most important factors affecting the comparison accuracy. For the measurement of the instrumental delay, Zero Baseline Interferometry which has the highest accuracy was developed. Using ZBI with directly controlled reference frequency method, the fundamental experiment was made and the accuracy better than 0.2 nsec was obtained.

Acknowledgment

The work reported in this paper has been accomplished thanks to the previous work carried out by many VLBI researchers in Kashima. The authors are deeply indebted to Dr. Kawaguchi for helpful discussion and encouragement.

References

(1) W. Hurd.; "An analysis and demonstration of clock synchronization by VLBI", IEEE Trans., IM-23, 1, March 1974.

(2) W.Hurd,; "Demonstration of intercontinental DSN clock synchronization by VLBI", the deep space network progress report, Tech. Report 32-1526, XVII, p. 39, JPL, Oct. 1973.

(3) T. Clark, C. Counselman, L. Hanson, H. Hinteregger, W. Klepczynski, C. Knigth, A. Rogers, I. Shapiro, and A. Whitney; "Precise clock synchronization via very long base line interferometer", IEEE-IM.

(4) W. Cannon, W. Petrachenco, and R. Langley; "Clock rate comparison by long baseline interferometry", Proc. 9th PTTI Applications and Planning Meeting, 1977.

HIGH ACCURACY TWO-WAY TIME TRANSFER AND RANGING EXPERIMENTS USING SATRE MODEMS

Michael Dienert, Claus Baumer

TimeTech GmbH, Nobelstraße 15, 7000 Stuttgart 80, Germany

ABSTRACT

SATRE is a MITREX-compatible modem for two-way time synchronization and comparison of remote clocks via geosynchronous communication satellites /4/. The use of coherent spread spectrum techniques with truncated pseudo-noise codes offers an accuracy of better than 1 ns in time comparison. This paper summarizes some bench-test results of the SATRE -Modem.

1. TWO-WAY TIME TRANSFER

Time transfer with high precision and accuracy is needed for fast continuous comparison of atomic clocks (international atomic time, co-ordinated universal time) and to build-up national and world-wide time networks. This includes applications in the synchronization of nodes in communication networks, computer networks or electrical power networks. The basic system set-up is shown in Fig. 1-1. One of the main advantages of the two-way time transfer technique via geosynchronous communication satellites is, that path delays cancel out for reciprocal paths between the stations. Accuracies in time comparison below 1 ns have already been reported /1/. Further improvements in accuracy are possible using suitable calibration methods which account for path delay unsymmetries /2/.

The SATRE (SAtellite Time and Range Equipment)-modem uses a completely analogue, "non-coherent" delay lock-loop with a chip-rate of 2,5 MHz providing original MITREX-codes and modulation schemes. This kind of correlator is not susceptible to frequency and phase instabilities of the RF-link and ensures full compatibility to the original MITREX-modems. The frequency scheme of the RX-unit is carefully designed to avoid integral multiples of the chip-rate, keeping errors due to beat frequencies to negligible values.

2. TEST-RESULTS

Two SATRE modems were tested on the bench in laboratory environment. Fig. 2-1 shows the internal calibration-transponder delay measured after warm-up of the modem. The uncorrected instrument drift is about 100 ps in normal laboratory environment.

Figure 2-2 shows the performance of the modem over varying input-signal power and signal-to-noise density-ratios. The changes of the modem delay within the nominal range of input power (-40 ... -55 dBm) and link-quality (C/N_0 > 50 dBHz) can be kept below 200 ps, proper calibration provided.

A comparison of measured and calculated instrument jitter values for an integration time of 40 s is shown in Fig. 2-3.

3. EXTENDED CALIBRATION

The difference between the time scales (clocks) of station 1 and station 2 can be given as:

$$TA_1 - TA_2 = 1/2\,(TI_1 - TI_2) + 1/2\,E$$

where:

TA_K = time scale at station K

TI_K = time interval counter reading

E = Error term due to link unsymmetries /3/

The unsymmetry of the link is partly caused by the delay differences between transmit and receive path at each station. These delays have to be measured individually or they have to be calibrated by means of a suitable calibration set-up. Fig. 3-1 shows an overall set-up with an external calibration transponder, which is used for determining the total delay of station transmit path and receiving path including ground station TX- and RX-chain, antenna cables and feeds. The sum of delays in the transmitter and receiver chain of the complete station is thus covered. For calibration of the receive path only, a clean carrier is transmitted and the signal is than translated to the receiving frequency and PN-modulated directly in the calibration transponder. Hence the delay in the receiver chain can be measured if the cable feeding the PN-signal to the calibration transponder is well characterized in terms of its delay. The station transmit delay can then be calculated as well. The external calibration transponder is an optional unit. It operates with an additional carrier for the purpose of frequency translation and the TX-PN-signal to modulate the translated uplink-frequency. These two signals are generated within the modem and are fed to the outdoor transponder via cables.
By the use of the internal calibration transponder the delay of the transmit and receive path of the modem itself can be measured.

4. CONCLUSION

The results of numerous bench-tests obtained with two SATRE modems have been shown. The remaining uncertainty of the modem-delay can be kept well below 200 ps after calibration using the internal test-transponder. Extensive two-way time synchronization and ranging field-experiments are scheduled for the near

future.

REFERENCES

/1/ D. Kirchner, Two-Way Time Transfer via Communication Satellites, Proceedings of the IEEE Vol. 79 No. 7, July 1991.

/2/ G. de Jong, Prospects of Two-Way International Time Scale Comparisons, Proceedings of the 6th European Frequency and Time Forum, June 1992.

/3/ M. Dienert, A New Modem For Microwave Time Synchronization via Geosynchronous Telecommunication Satellites, Proceedings of the 6th European Frequency and Time Forum, June 1992.

/4/ Institute for Navigation University Stuttgart, "A Modem for Microwave Time and Ranging Experiments via Telecommunication Satellites", Stuttgart 1985.

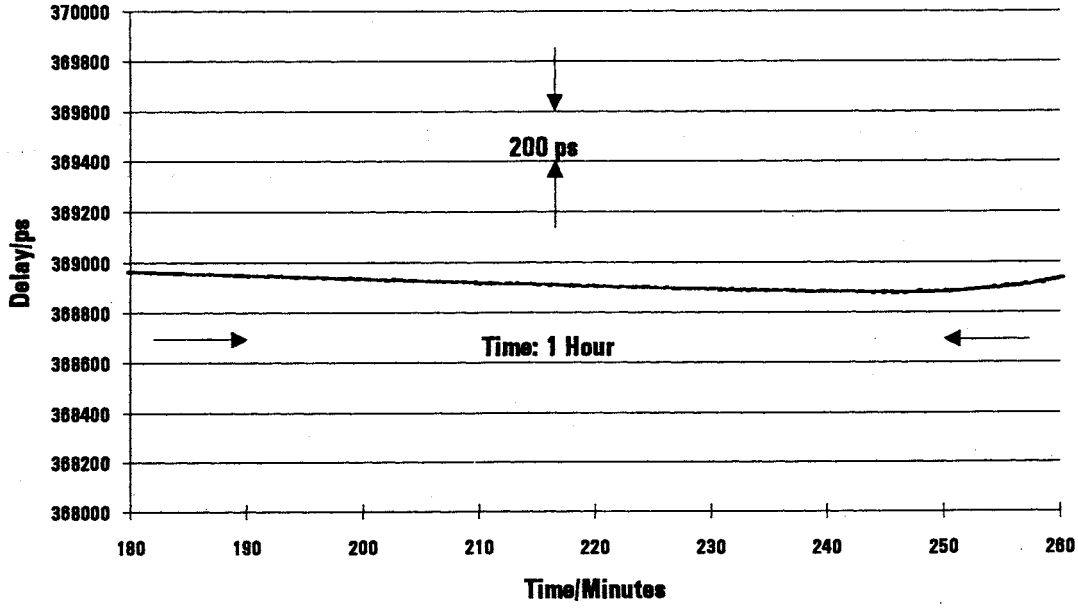
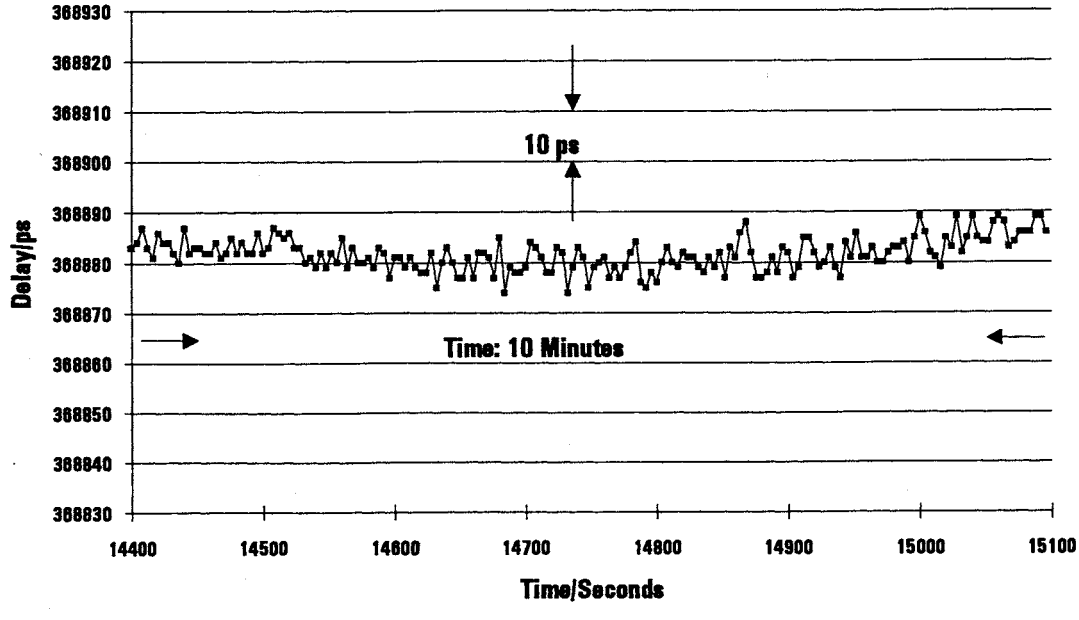


Figure 2-1 a,b: Uncorrected Stability of the Instrument Delay (Integration Time: 40 s, C/N_0 = 80 dBHz)



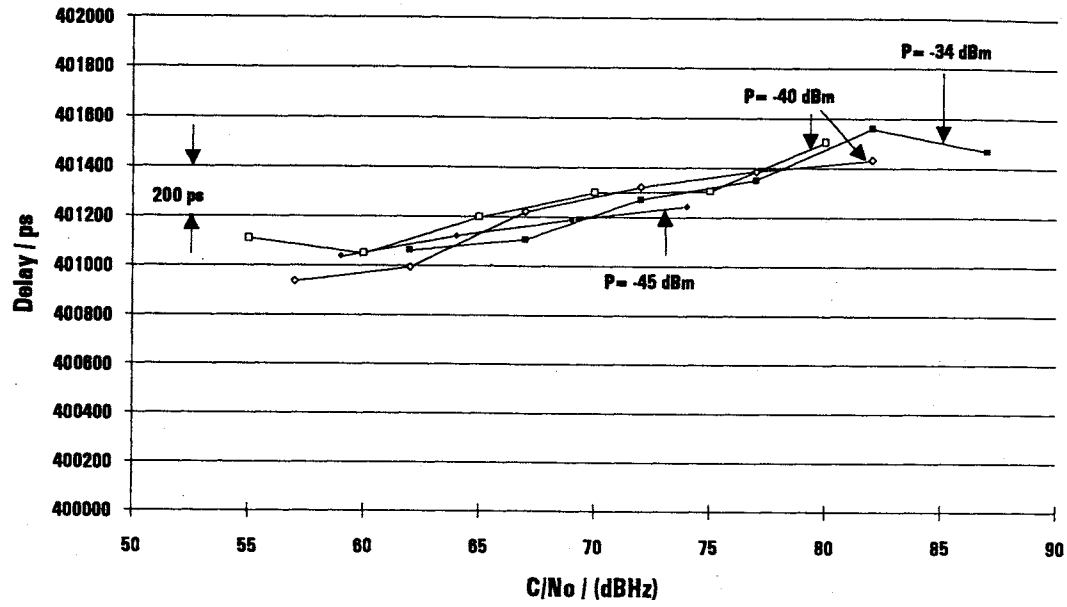


Figure 2-2: Typical instrument performance over varying input-signal power and link-quality; (Integration Time: 40s)

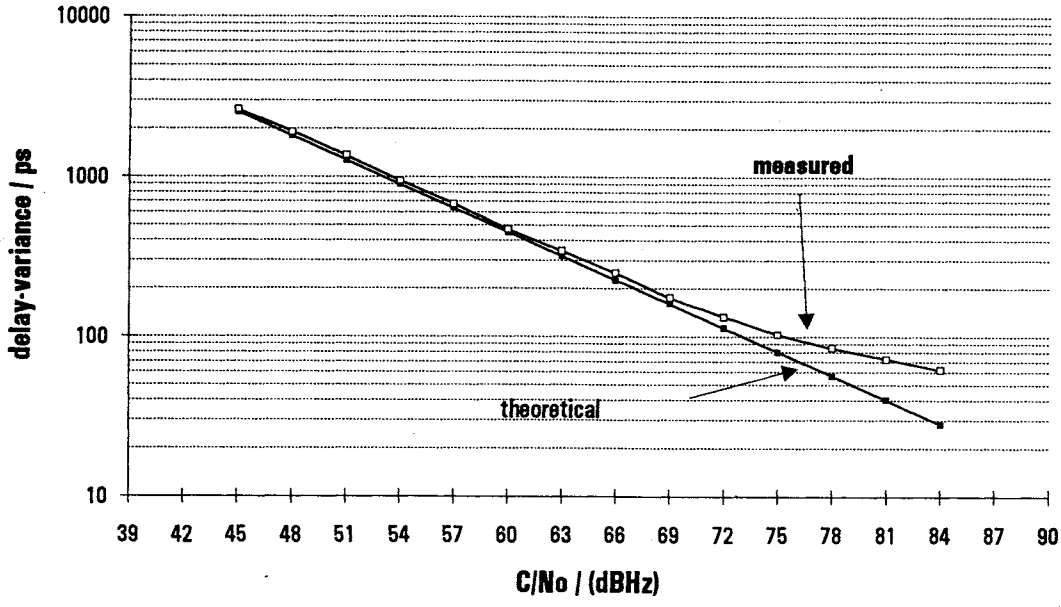


Figure 2-3: Instrument jitter of the 2.5 Mchip/s SATRE-modem versus signal-to-noise-density-ratio; (Integration Time: 40 s)

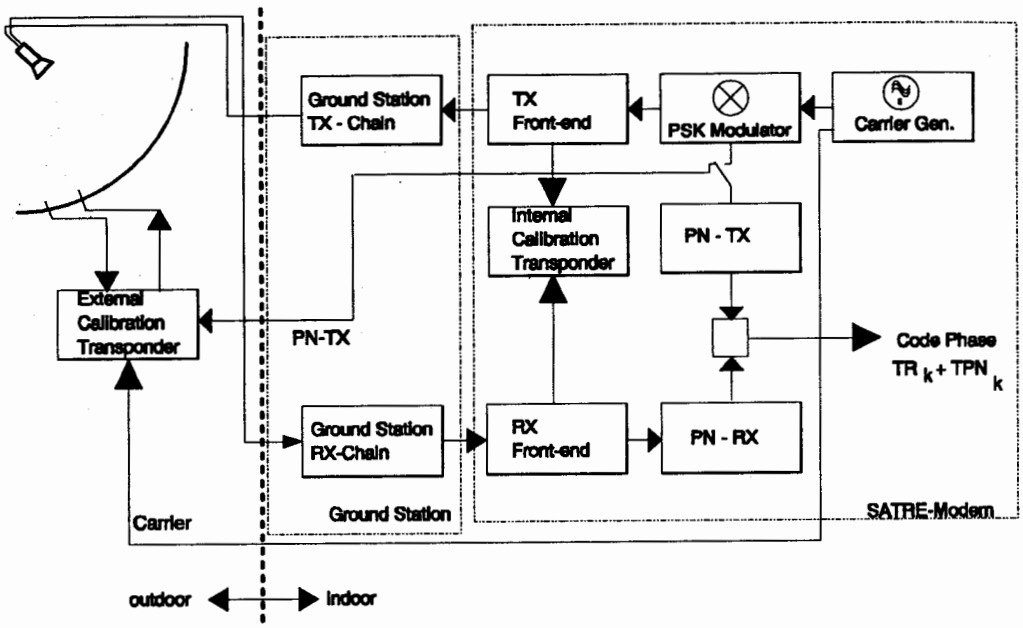


Figure 3-1: Set-up for determination of total station delays

MEASURES OF FREQUENCY LOCKING IN OSCILLATORS WITH APPLICATION TO ION COOLING IN A RADIO-FREQUENCY TRAP

M. 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 - 25044 Besançon Cedex - France

We have developped efficient tools of studying structurally stable solutions of oscillator differential equations. They are based on an accurate measure of frequency locked domains[1].

This strategy has been applied to a quasi-classical study of ion cooling in a radio-frequency trap, using the assumption of weak driving field and strong damping by spontaneous emission[2]. Following a suggestion by P. Toschek[3,4] we considered excitation by two lasers whose difference frequency is slightly detuned from the frequency of the harmonically bound ionic dipole.

The end result is the existence of frequency locked (cooled) domains in parameter space separated by chaotic (heated) gaps in qualitative agreement with recent experiments[3,4]. This emphasizes the interest of the method.

I - INTRODUCTION

Slowing and cooling of atoms and ions by traps and light is now a subject of wide interest in frequency metrology. This is due to the promise of greatly improving present time standards. One recent proposal has been to use the interaction of bichromatic light with a trapped ion. This has been motivated by the observation of a rich dynamics of a single Ba^+ ion when subjected to two incident laser fields [3, 4]. The experiment has revealed multistability of the ion's kinetic state, hysteresis loops and optical heating and cooling versus laser tuning and intensity. In this paper we develop a semi classical model of the ion cooling. Following ref [2] the radiation pressure is calculated from optical Bloch equations under the assumption of weak driving fields and strong damping. Assuming the ion is harmonically bound by the trap, the equation of motion becomes that of a forced oscillator and the cooling ressembles an injection locked oscillator. An accurate measure of the frequency locked domains is performed using the techniques derived recently [1] and a qualitative agreement with experiments is obtained.

II - THE INTERACTION OF THE TRAPPED PARTICLE WITH BICHROMATIC LIGHT

We study the influence of the light force on the trapped two level system. Optical Bloch equations govern the dynamics of a two level atom ; they are similar to those of a spin 1/2 magnetic dipole and may be derived from Heisenberg density matrix formulation in the rotating wave approximation [5].

Using parameters $\sigma = \rho_{12}$ (the transition probability between the two levels) and $D = \rho_{11}-\rho_{22}$ (the population difference between upper and lower levels), Bloch equations may be written as [6] :

$$\begin{aligned} \dot{\sigma} + \left(\frac{\Gamma}{2} - i\Delta \right) \sigma &= Dg \\ \dot{D} + \Gamma D &= -\Gamma - 2(g^*\sigma + g\sigma^*) \end{aligned} \tag{1}$$

where Δ is the detuning, Γ is the radiative decay rate of the atom to the ground state, μ_{12} is the atomic dipole moment connecting the two levels and $g = i\mu_{12} E/\hbar$ is proportional to the free external electric field E.

With the same approximations the radiation force has been derived from the atomic dipole moment **p** as follows [6]:

$$\boldsymbol{F} = -\boldsymbol{grad}\,(\boldsymbol{p}\,\boldsymbol{E}) = 2\hbar\ Im\,(\sigma^*\,\boldsymbol{grad}\,g) \tag{2}$$

For an incident bichromatic field the radiation force may be calculated by solving Bloch equations with the forcing :

$$g = i \sum_{n=1}^{2} \mu_n \exp\left\{ -i \left(\boldsymbol{k}_n . \boldsymbol{x} - \Delta_n t + \phi_n \right) \right\} \tag{3}$$

where μ_n (n = 1,2) are the Rabi frequencies, Δ_n the laser detunings and ϕ_n the initial phases.

At large times $t > 1/\gamma$, Bloch equations have solutions in the form :

$$\begin{aligned} \sigma &\sim \int_0^t g(t')\,D(t')\,e^{\gamma(t'-t)}\,dt' \\ D &\sim -1 - 2\int_0^t (g\sigma^* + g^*\sigma) e^{\Gamma(t'-t)}\,dt' \end{aligned} \tag{4}$$

To further simplify the problem Joos [2] uses the low amplitude approximation $D\sim$-1 and takes an expansion around the upper integration limit to get :

$$\sigma = -\frac{g}{\gamma}+\frac{\dot{g}}{\gamma^2}-\frac{\ddot{g}}{2\gamma^3} \qquad (5)$$

By using (3)-(5) the radiation force (2) may be calculated. The two subcases are considered. Either the two waves are parallel $k_2\sim k_1 = k$ or they are antiparallel $k_2 = -k_1 = -k$. Neglecting the effect of detuning in (1), i.e.

$$\gamma = \frac{\Gamma}{2}-i\Delta\sim\frac{\Gamma}{2}$$

the radiation force along these directions is calculated as follows :

$$\frac{F}{m} = \varepsilon_o+\varepsilon_1\cos\psi-\varepsilon_2\sin\psi$$
$$+\dot{x}(a_1+\beta_1\cos\psi)-\dot{x}^2(a_2+\beta_2\cos\psi) \qquad (6)$$

$$with\ \psi = \delta t-\chi x+\phi o\ ;\chi = k_1-k_2\ ;\delta = \Delta_1-\Delta_2$$

The parameters at the r.h.s. of (6) depends on the length $X_o = 2\hbar k/(\Gamma m)$, on the spontaneous decay rate $\gamma = \Gamma/2$ and of detunings Δ_i and amplitudes μ_i of the incident wave.

In this approximation the motion of the trapped particle is that of a forced oscillator :

$$\ddot{x}+\omega^2 x = F(x,t)/m \qquad (7)$$

Problem (7) will now be studied using an averaging technique developped before [8, 9]. The end result will be frequency locked states at some resonances $\Omega = \omega/\delta$ corresponding to cooling of the particle.

III - SOLVING THE PARTICLE-WAVE INTERACTION PROBLEM

Equ. (7) is a particle wave-interaction problem. In its simplest prototype form it is written as :

$$\ddot{x}+\omega^2 x = \varepsilon\sin(\chi x-\delta t-\phi_o) \qquad (8)$$

Equ. (8) has been studied before as representing the motion of a charged particle in the presence of a plane wave that propagates orthogonally to a magnetic field. In the particular case where the r.h.s. of equ. (8) is a quasimonochromatic wave packet with $k\sim k_0$ and has n resolved side bands $\omega\sim\omega_o+n\ \Delta\omega$, an approximate mapping has been derived and stochastic webs have been found at rational ratios of the cycloton to the Larmor frequency $\Delta\omega$ [7].

In contrast to the a forementioned case we are interested here in a purely monochromatic excitation and look at conditions for frequency locking.

We recently presented an averaging technique to get a Poincaré mapping at resonances of forced differential equations [8, 9]. Rewritting equ. (8) in polar coordinates $x = \rho\sin\theta$; $\dot{x} = \rho\cos\theta$, the resonance condition is $\rho = \rho_n$ (n integer) and $\theta = \delta t$; this should be valid over time intervalls of length Γ^{-1} provided the quality factor $Q = \delta/2\pi\Gamma\geq 1$. With these assumptions, the cross section of trajectories reads [9].

$$r_{n+1} = r_n-c\sin\theta_n\ J_1(r_n)/r_n$$
$$\theta_{n+1} = \theta_n+2\pi\Omega Q+c\cos\theta_n\ J_1'(r_n+1)/(r_n+1) \qquad (9)$$

where $\Omega = \omega/\delta$ is the resonance ratio, $r_n = \chi\rho_n$, J_1 is the Bessel function and $c = \varepsilon\chi/(\omega\Gamma)$.

The analysis of fixed points of the linearized problem reveals that non dissipative dynamics occurs at phases o or π at zeros of Bessel function J_1. Non linear analysis reveals frequency locking around rational ratios Ω. More complicated results with subbands are obtained for initial conditions away from a zero of Bessel function J_1 [9].

No new technical problem happens in solving equ. (7) despite the length of calculations. We have studied the two subcases where the incident waves are parallel or antiparallel and have closed wavenumbers ($\chi\ll k$) closed frequencies ($\delta\ll\Gamma$) and equal amplitudes ($\mu_1 = \mu_2 = \mu$).

Antiparallel case :

In such a situation $a_2 = \beta_2 = 0$. Introducing the extra parameter $\eta = a_1\pi Q\ \delta^{-1}\sim k\ X_o\ \mu^2\Delta/\Gamma^3$, we have $|\eta/c|\sim\Omega\ k/\chi\gg 1$ so that terms in J_1 may be neglected. The Poincaré mapping takes the simple form :

$$r_{n+1} = r_n(1+\eta)-2\eta\cos(\theta_n+\theta_o)\ J_2(r_n)$$
$$\theta_{n+1} = \theta_n+2\pi\Omega Q-2\eta\sin(\theta_n+\theta_o)\ J_2(r_{n+1})/r_{n+1} \qquad (10)$$

with $\theta o = tg^{-1}(\pi Q)$

Parallel case :

In the parallel case $a_2\sim\beta_2\neq 0$ and the ratio $|\eta/c|\sim 1$. All terms contribute the same amount and cannot be neglected. The Poincaré mapping is :

$$r_{n+1} = r_n(1+\eta)-c\sin\theta_n\ J_1(r_n)/r_n$$
$$-\eta\cos(\theta_n+\theta_o)\ J_2(r_n)$$
$$\theta_{n+1} = \theta_n+2\pi\Omega Q+c\cos\theta_n\ J_1'(r_{n+1})/r_{n+1} \qquad (11)$$
$$+2\eta\sin(\theta_n+\theta_o)\left\{\frac{1}{8}J_1'(r_{n+1})\right.$$
$$\left.+\frac{3}{8}\frac{J_1(r_{n+1})}{r_{n+1}}+J_2(r_{n+1})\left[\frac{1}{8}+\frac{1}{r_{n+1}}-\frac{3}{2}\frac{1}{r_{n+1}^2}\right]\right\}$$

with $c\sim-\eta$ and $\theta_o\sim-\pi/2$.

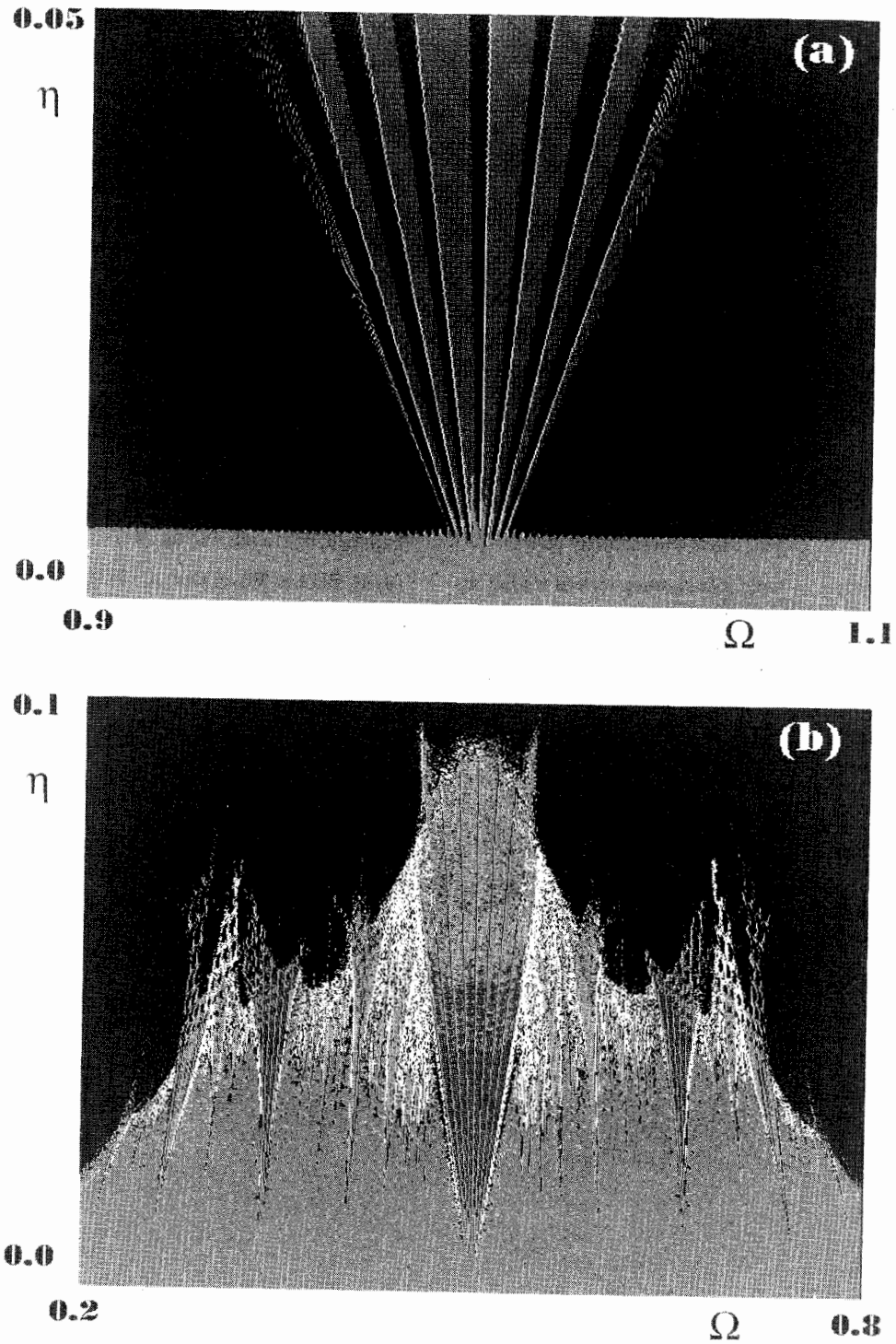


Fig. 1 : Diagram of the frequency locked states from the α exponent
(initial condition $\theta_0 = 0, r_0 = 0{,}5$)
(a) antiparallel case
(b) parallel case

IV - THE FREQUENCY LOCKED STATES

Mapping (10) and (11) are rotation mappings, similar to those studied in our recent publications [1, 8, 9]. They are essentially two parameters, one for the rotation (the resonance ratio $\Omega = \omega/\delta$, the second one η for the strengh of the perturbation.

At a given point $x(\Omega, \eta)$ of parameter space P the perturbed resonance ratio ν is given by :

$$\nu : x \to \nu(x) = (\theta_n - \theta_o)/(2\pi n) \tag{12}$$

Frequency locked regions are generally characterized by constant rational resonance ratios [10]. We recently found that the Hölder exponent a defined by

$$|\nu(x) - \nu(y)| = \|x - y\|^a \quad ; \quad x, y \in P \tag{13}$$

provides an accurate measure of the stability [1, 9].

According to our approach a frequency locked region in P should satisfy the two conditions :

$a > 1$ and Lipschitz continuous, i.e. $|a(x) - a(y)| < \|x - y\|$.

A plot of the frequency locked regions is given in Fig. 1 for the counterpropagating case (a) and for the parallel case (b). A set of alternating cooled and heated regions is clearly seen closed the 1/1 ratio for case (a), and closed to 1/2 ratio for case (b). The maximum width of the resonance is for $\eta \sim 0.05$.

Strictly speaking these calculations cannot be used to interpret the Ba^+ experiments [3, 4] since the low amplitude assumption $\mu < \Gamma$ is not satisfied .

However using a set of parameters for Ba^+: Q = 1 ; $k \sim 10^7$ m^{-1} ; $\Gamma \sim 50$ MHz ; $X_0 = 2\hbar k/(\Gamma m) \sim 4$Å ; $\Delta \sim 400$ MHz we obtain $\mu \sim 50$ MHz.
These data should provide a motivation for further work using the same approach.

REFERENCES

[1] M. Planat, EFTF 1991, ESA publication division, p. 333.

[2] E. Joos, A. Lindner, preprint.

[3] H. Gilhaus, Th. Sauter, W. Neuhauser, R. Blatt and P. E. Toschek, Optics Communications, 69, 25 (1988).

[4] Th. Sauter, H. Gilhaus, W. Neuhauser, R. Blatt and P. E. Toschek, Europhysics Letters, 7, 317 (1988).

[5] M. Sargent III, M. O. Scully and W. E. Lamb, "Laser Physics", Addison-Wesley, (1974), pp. 79-95.

[6] J. P. Gordon and A. Ashkin, Phys. Rev. A, 21, 1606 (1980).

[7] G. M. Zaslavskii, M. Yu Zakharov, R. Z. Sagdeev, D.A Usikov and A. A. Chernikov. Sov. Phys. JETP, 64, 294 (1986).

[8] M. Planat "Non Linear Coherent Structures in Physics and Biology" edited by M. Remoissenet and M. Peyrard, Springer Verlag, N-Y (1991), p. 292.

[9] M. Planat, Phys. Rev. A 15 (1993), in press.

[10] A. J. Lichtenberg, M. A. Lieberman, "Regular and chaotic dynamics" Second Edition, Springer Verlag (1991).

PASSIVE HYDROGEN MASER FREQUENCY STABILITY AND ACCURACY INVESTIGATIONS

N.A.Demidov, A.A.Belyaev, B.A.Sakharov, A.A.Uljanov
Institute of electronic measurements KVARZ Nizhny Novgorod Russia

F.L.Walls
National Institute of Standards and Technology Boudler USA

Abstract

Frequency stability and frequency accuracy of passive hydrogen maser CH1-76 are investigated. It is noted that together with already investigated frequency maser errors electronic part of the instrument contributes essentially to passive hydrogen maser instability. The influence of electronic units on output frequency of the instrument is analysed.

By the results of investigations of the series of CH1-76 passive hydrogen masers the possible error limits and frequency reproducibility inserted by electronic units of the instrument are experimentally defined CH1-76 passive maser instability is obtained experimentally and is described by the expression:

$\sigma(2,\tau) = (5\text{-}9)\cdot 10^{-13}\tau^{-0,5}$ $1<\tau<10^4$s
and $(4\text{-}7)\cdot 10^{-15}$ per day.

Keywords: Hydrogen maser, frequency stability.

1. Introduction

Passive hydrogen masers, developed during last ten years, successfully combine the highest frequency stabillity and small dimentions of table instrument. Passive hydrogen maser frequency stability is better 10^{-14} per day. This parameter is only a little bit worse than in active hydrogen maser.

High stability of passive maser depends first of all on the qu- antum physical part (hydrogen discriminator) having a very high quality hydrogen emission line.The main sources of passive maser frequency instability are the same as in an active hydrogen maser and namely: cavity frequency detuning, storage bulb coat ageing, magnetic field and beam flux system instability. However, the detailed passive hydrogen maser investigation showed essential influence of electronic units of instrument on the output signal frequency instability and accuracy. Among them it is necessary to note the discriminator test signal spectrum distortions, irregularity of the receiver amplitude-frequency response and instability of receiver input circuit voltage stand- ing-wave ratio (VSWR).

The investigation results of the CH1-76 passive hydrogen ma- sers produced by IEM KVARZ (Russia) are given in this article. Experimental research was conductes by specialists of IEM KVARZ and independently by the scientists of NIST (USA) in 1991-92 years. [1]

2. Design And Construction of CH1-76 Passive Hydrogen Maser. Quantum Hydrogen Discriminator.

The design of a physical part (quantum discriminator) of the passive hydrogen maser is given in Fig.1.

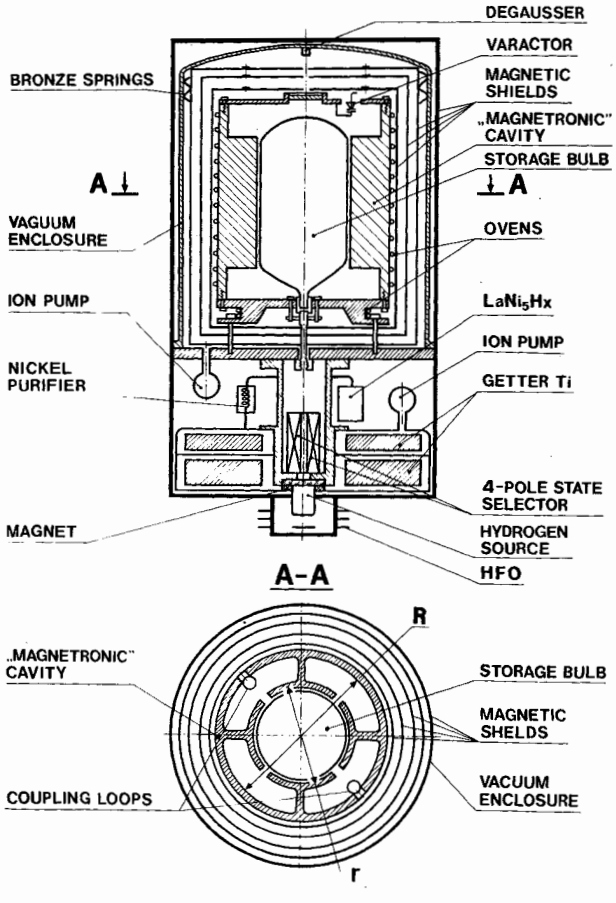


Fig.1. CH1-76 Quantum discriminator

Compaund $LaNi_5H_x$ (about 250g), placed into the metal bulb, is used as molecular hydrogen source. It comprises 20-25 litres of hydrogen at normal pressure that is quite enough for a 5 years continuous maser operation. The bulb with molecular hydrogen is vacuum-tightly connected to purifier. Nickel tube of 0.5 mm diameter and 0.05 mm wall-thickness coiled for compactness is used as purifier. The hydrogen flux is controlled by purifier temperature. Purifier power consumption is about 1W.

Atomic hydrogen source is made of quartz glass in the form of cylinder of 30 mm diameter and 40 mm height. The source is connected to the vacuum pump by indium gasket. Output source channel (collimator) is a set of 200-250 channels of 13μm diameter and 0.65 mm length each. It provides for hydrogen atom beam flux.

Hydrogen dissociation is performed by molecular impact ioni- zation in electromagnetic field of RF-generator coil. RF-generator tor power is 12W.The magnetic state selection is performed by means of quadripole magnet. The bore diameter of the magnet is 1.6 mm, its length is 75 mm and outer diameter is 30 mm.The magnetic induction at the pole tips is 1T.

The maser has two separated vacuum chambers. The first one includes storage bulb, state selector magnet and atomic hydrogen source, the second one includes cavity ovens and magnetic shields. Such a construction permits to achieve excellent vacuum- conditions in the storage bulb. Hydrogen pumping is performed by means of absorption pump with titanium as getter material. The getter is activated at 800^0C by an internal tungsten heater. The weight of getter is 1kg which is sufficient for more then 5-7 years of continuous operation. Residual gases are pumped in both vacuum chambers by a 2 l/s ion pumps.

The storage bulb is made of quartz glass in the form of a cylin- der of 0.45 l volume. The inner surface of the bulb is coated with 3 layers of fluroplastic, type F-10. Among other known materials F-10 ensures the smallest wall-shift and the largest emission line quality ($Q_1 \sim 10^9$).

The storage bulb is located in the middle of "magnetronic" ca- vity [2] Fig.1.Such cavity construction ensures high mechanical rigidity and reliability, high Q-factor ($Q \sim 10^4$) and a good filling factor value η. Variations of the ηQ_c -parameter versus the internal cavity plates radius is shown in Fig.2.

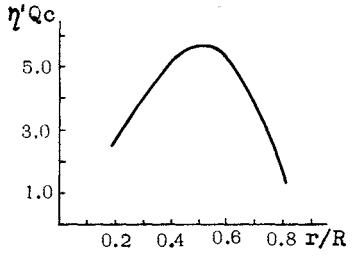


Fig.2. Variation of the filling factor versus the internal plate magnetronic cavity relative radius.

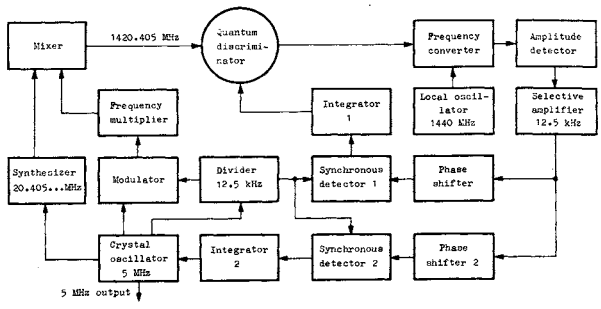
These results have been received theoretically [3] and proved by a real maser parameter measurements. Fig.2 shows that maser operates near the self-excitation point and ensures high signal amplification and discriminator characteristic steepness.

Thermal cavity stabilization is insured by the two-zone oven, placed in vacuum. The temperature is controlled to 1 10^{-2o}C at the 5-40^0C environmental temperature range. CH1-76 maser real warm-up time is about 6 hours at 20^0C. Oven power consumption is <1W.

Maser has a four layers permalloy magnetic shield system. Three of them are located in vacuum and made of 81 NMA permalloy of 0.5 mm thickness. The fourth shield made of 79 NMA permalloy covers all discriminator construction with vacuum pumps and hydrogen beam source. To prevent undesirable mixing of Zeeman sub-levels in the drift region, a ring magnet is placed on the discharge bulb of the hydrogen beam source to assure a smooth transition between the quadripole magnet and the apertures of the magnetic shields.

Discriminator weight is 19.5 kg, diameter 224 mm, height 520 mm.

Block-diagram of CH1-76 is given in Fig.3.



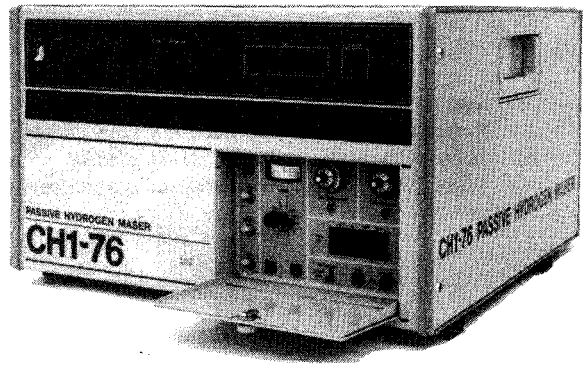


Fig.3. CH1-76 Passive hydrogen Maser: a) block-diagram b) photo.

It comprises two frequency lock-loops: 5 MHz crystal oscillator frequency lock to the hydrogen emission line and discriminator RF-cavity frequency adjustment loop to the crystal oscillator frequency. The frequency modulation auto-tuning method with high modulation frequency in both loops was used in CH1-76 passive hydrogen maser [4]. 5 MHz crystal oscillator signal phase is modulated by the 12.5 kHz signal, multiplied up to 1400 MHz frequency, mixed with the synthesizes 20.4057 MHz signal and put in the RF-cavity of quantum hydrogen discriminator. The power of this test signal is $\sim 1 \cdot 10^{-12}$W and frequency modulation factor is 1.5.

Error signals (information about the test signal frequency detuning relative to the hydrogen emission line and RF-cavity) are formed as a result of test signal spectrum linear distortions which appear as a result of interaction of the emission line, RF-cavity and test signal in the discriminator. Error signals are contained in the 12.5 kHz amplitude modulation of the output discriminator signal. Low noise (NF~1.5dB) RF-converter (receiver) provides sufficient signal amplification for amplitude detection and the 1440MHz local oscillator eliminates the influence of synchronous disturbances (5 MHz harmonics and 20.405MHz signal) influence in the intermediate frequency amplifier [5].Error signals from crystal oscillator and RF- cavity are divided into two ways by the phaze- shifters setting at the 90^0 angle and entered to the frequency control circuits.

Open-loop proportional-action coefficients are $1 \cdot 10^7$ for the crystal oscillator loop and $1 \cdot 10^4$ for the RF-cavity loop. Total frequency statical tuning error for 5 MHz output signal is less than $1 \cdot 10^{-15}$.

3. Frequency Accuracy and Instability of Passive Hydrogen Maser.

Quantum Discriminator

Physical effects causing hydrogen discriminator frequency variations are the same as in an active hydrogen maser and well known [6,7]. They are: cavity pulling, magnetic fields, storage bulb aging, second-order Doppler shift, spin-exchange shift. Frequency shift values and their instabilities are shown in Table 1.

Table 1

Physical effect	Shift value	Shift instability	Note
1. Cavity pulling	$-(2\text{-}5)\cdot10^{-12}$	$2\cdot10^{-15}$	Due to used method of tuning and spin-exchange shift
2. Magnetic fields	$2.5\cdot10^{-13}$	$-2\cdot10^{-14}/0e$	
3. Wall-shift of storage bulb coating	$6.7\cdot10^{-12}$	$6\cdot10^{-13}/^0C$	Possible changing up to $+1.5\cdot10^{-13}$/ year
4. Second-order Doppler-shift	$-4.46\cdot10^{-11}$	$-1.4\cdot10^{-13}/^0C$	

The passive hydrogen maser feature is the relatively large cavity pulling frequency shift arising due to used method of tuning and spin-exchange emission line shift.Cavity-pulling frequency shift may be significantly decreased by special test signal spectrum pre-distortions.

Electronics

Passive hydrogen maser short-term frequency stability depends on the quantum discriminator characteristic steepness and the noise figure of preamplifier in the receiver, which define the low frequency (< 10 Hz) white noise level of the output 5 MHz signal phase noise spectrum.

In practice it is very convenient to estimate discriminator quality in combination with electronic part of the instrument. Normal FM test signal is entered to the discriminator and the amplitude detector output voltage change is measured at the calibrated test signal carrier frequency shift by a half of hydrogen amission line- width. As a result of measurement, the discriminator quality factor is calculated:

$$R=Us/Un\Delta f \qquad (1)$$

where Us - is the 12.5 kHz voltage change at the amplitude detector output

Un - noise voltage at the amplitude detector output without frequency detuning between test signal carrier and emission line in the 1 Hz bandwidth [$V/\sqrt{Hz}$]

Δf - hydrogen emission line-width

For the CH1-76 passive hydrogen maser the quality factor value is within the limits 300:400 and ensures 5 MHz output signal short-term frequency stability (5-9) $10^{-13}\tau^{-0.5}$ for the time intervals $1<\tau<1\ 10^4$ s

Due to the used modulation method of tuning in the passive maser and detuning information contained in the weak amplitude modulation signal, the electronic part influence on the long-term maser stability and accuracy will be manifest itself as an additional amplitude modulation at the same modulation frequency 12.5 kHz. This additional modulation signal can appear as a result of FM test signal spectrum distortions in several units of instrument: modulator, frequency multiplier and receiver.

Modulator

The main requirements to the phase-modulator of the instrument are the highest linearity of modulation characteristic and the reduction of the even harmonics in the modulation signal. The presence of even harmonics of the modulation signal [8] leads to the test signal spectral components distortions and gives an additional amplitude modulation signal and frequency shift of output 5 MHz signal. The estimation of this shift for CH1-76 passive maser can be done by the next formula [8]

$$\Delta f/f=3m_1/4Q_l, \qquad (2)$$

where: m_1 is the 2-nd harmonic modulation index

Q_l is the emission line quality factor $1\cdot10^9$

According to expression (2) the m_1 value should be less then $1\cdot10^{-4}$ to achieve 5 MHz output frequency shift less $1\cdot10^{-13}$. And, consequently, the second harmonic of modulation signal reduction of the test signal should be more than 80 dB. In CH1-76 passive hydrogen maser these values are achieved by a special type of a varactor with very high linearity of capacity-voltage characteristic and usage of even harmonic reduction modulation signal former.

Frequency Multiplier

Additional amplitude modulation FM signal arises in the frequency multiplier due to FM-spectrum distortions by the multiplier steps resonant systems. This effect leads to the RF discriminator cavity adjustment frequency shift and therefore to the emission line frequency shift (cavity pulling). Maser output 5 MHz signal frequency shift value in this case is proportional to the multiplier resonant system detuning and quality factor and can be estimated by the expression

$$\Delta f/f=[Q_s/Q_l\,Q_c]\cdot[J_1(m_s)/J_1(m_c)]\cdot[\Delta f_s/f_s]\cdot[f_c/f_s] \qquad (3)$$

In expression (3) f_s, Q_s, Δf_s are the frequency, quality factor and detuning of multiplier step resonant system; Q_l, Q_c - are the quality-factors of emission line and RF-cavity in the discriminator; $J_1(m_s)$ and $J_1(m_c)$ are the first modulation harmonic component FM-spectrum amplitudes in the multiplier step and in the RF-cavity.

Frequency multiplier contribution to long term passive maser frequency stability depends on the temperature stability and ageing of multiplier steps resonant circuits and can reach a few part $1\cdot10^{-13}$. For supression of this effect multiplier has several amplitude-limiters and automatic output voltage control system providing 12.5 kHz amplitude modulation suppression up to -110 dB and eliminating the multiplier influence on the output frequency stability.

Reciever (frequency converter)

Reciever's circuit influence on the output standard frequency arise the same way as in frequency multiplier due to the AFC unlinearity. However it is impossible to use the amplitude modulation suppression here because the reciever's signal envelope contains useful information. To reduce an additional amplitude modulation the reciever's intermediate amplifier bandwidth is made as wide as possible provided that signal/noise ratio is > 1. IF amplifier in CH1-76 has ~20 MHz bandwidth at 19.6 MHz carrier and the output standard frequency shift accoding to expression (3) less then $3.5\cdot10^{-13}$ for $\Delta f_s/f_s=5\%$ and the temperature stability of this shift less $5\cdot10^{-15}/^oC$

Experimentaly defined that the RF-discriminator cavity tuning point in the passive hydrogen maser cavity tuning system being described depends on the RF-cavity and reciever coupling

and output standard frequency changes with the VSWR of the reciever input and coaxial cable length (Fig.4.)

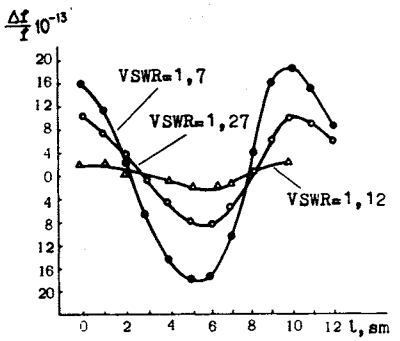


Fig.4. 5 MHz output frequency variations versus coaxial cable length coupling guantum discriminator and the reciever.

This effect was not investigated theoretically and one of the probable reasons is the RF-cavity resonant curver distortions by the load reactivity.

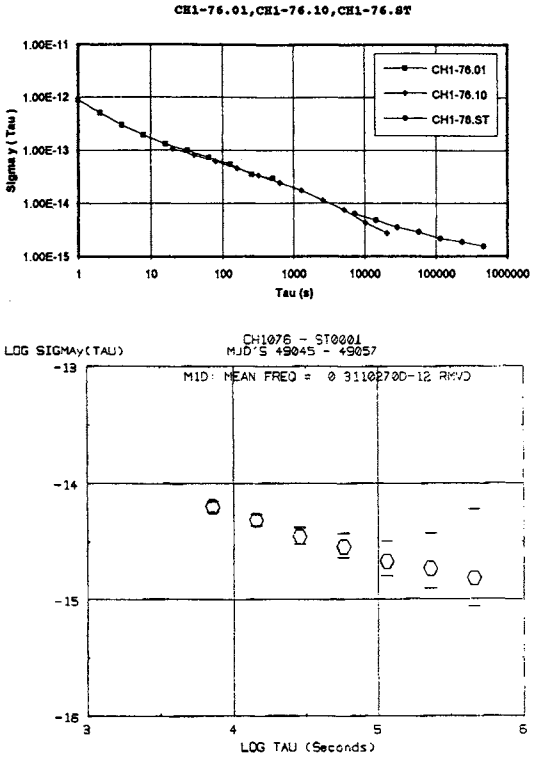
4. Experimental Results

Frequency stability measurement tests of the CH1-76 type passive hydrogen masers were conducted at IEM KVARZ and independently at NIST (USA). Frequency accuracy, frequency drift, time-domain and long term frequency stability, frequency temperature coefficient and phase noise spectrum are estimated.

Short term frequency stability (Allan deviation$\sigma(2,\tau)$ (Fig.5) for most of masers is within the limits $(5\div9)\cdot10^{-13}$ for a 1s and for the time intervals $1<\tau<1E+4$ s may be expressed as

$$\sigma(2,\tau)=(5\div9)\cdot10^{-13}\tau^{-0,5}$$

For a long time intervals $\sigma(2,\tau)$ decreases slowly and is estimated as $(4\div7)\cdot10^{-15}$ per day and depends on the environmental conditions.



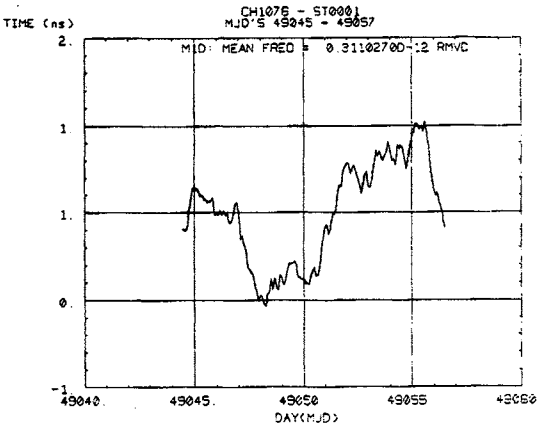


Fig.5 CH1-76 5 MHz output signal freguency stability (Allan deviation)

Typical frequency drift estimations are less then $1\cdot10^{-15}$ per day, but CH1-76 serial number 023 accoding to NISTestimation showed the frequency drift value $-1.82\pm0.06\cdot10^{-15}$ per day for the 100 days time interval. Long term CH1-7 6 frequency behaviour is represented in Fig.6 with respect to environmental temperature. Frequency temperature coefficient experimental estimations for several samples of CH1-76 is $(0.5\div1.5)\cdot10^{-14}$/C.

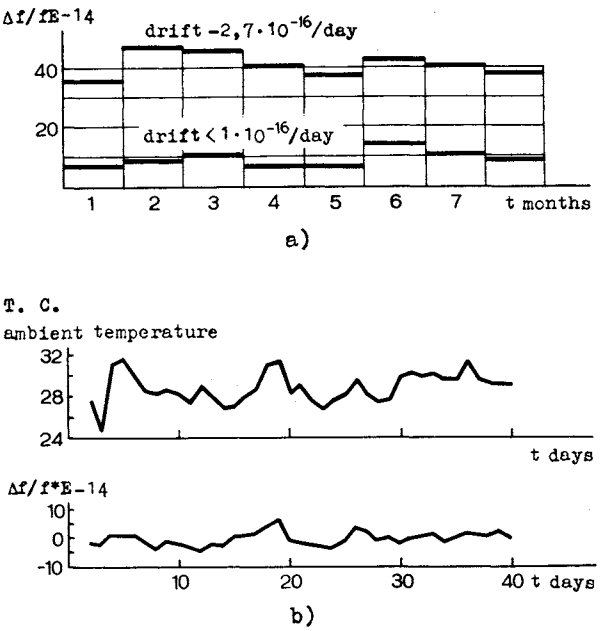


Fig.6 Long term frequency stability of the CH1-76 passive hydrogen masers
a) averaging per month
b) averaging per day with respect the ambient temperature

Phase noise spectrum of the CH1-76 5 MHz output signal is shown at the Fig.7.

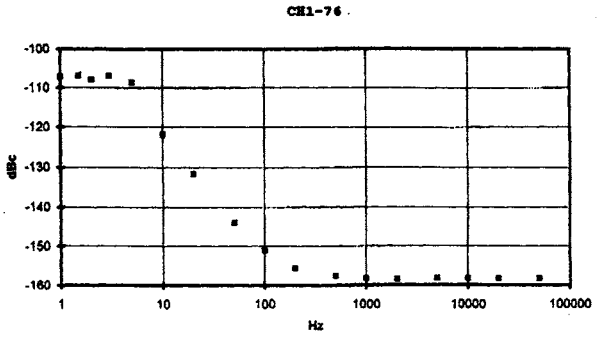


Fig.7 Phase noise spectrum CH1-76 5 MHz output signal

References

1.F.Walls, D.Allan, J.Gray, C.Hackman. Calibration of KVARZ Passive hydrogen maser CH1-76/2391 for Efratom Division Ball Corporation. NIST, 1992

2.A.A.Belyaev, N.A.Demidov, S.A.Makarychev. Hydrogen maser. A.s. No1459568 USSR Otkrytiya Izobreteniya No 6, 1989

3.A.A.Belyaev, N.A.Demidov, V.A.Savin Hydrogen maser SAS- cavity energetic characteristics investigation Technica sredstv svjasy RIT 1987 No 2 p.23

4.G.Busca,H.Brandenberger Passive H-maser Proc. 33-rd Ann.Freq.Cont.Symp, 1979 p.563-568

5.N.A.Demidov, B.A.Sakharov Passive H-maser A.s. No1.258.212 USSR Otkrytiya Izobreteniya No 34 1986

6.E.M.Mattison Physics of Systematic Frequency Variations in Hydrogen Masers IEEE Transacytions on ulrasonics, ferroelectrics and frequency control vol 39 No 2 March 1992

7.F.Walls Characteristics and Perfomance of miniature NBS Passive Hydrogen maser. IEEE Transactions on instrumentation and Measurement vol IM-36 No 2 June 1987

8.V.A.Logachev Investigation of passive hydrogen maser frequency stability Technica sredstv svjazy, RIT, No 1 1986 .

HIGH PERFORMANCE HYDROGEN FLOW REGULATOR FOR HYDROGEN MASERS

L. C. Giurgiu, M. P. Dinca

University of Bucharest, Bucharest-Magurele,
PO BOX MG-11, R-76900, Romania

O. C. Gheorghiu

Institute of Atomic Physics, Bucharest-Magurele, Romania

ABSTRACT

This paper reports an improved system for controlling the hydrogen flow in hydrogen masers. The main features of its different parts are presented, as well as the achieved performances. Rise time less than 2 s, setlling time less than 3 s and over-shoots below 1% have been obtained.

Keywords: Atomic frequency standard, Hydrogen maser, Flow regulator, PID controller.

1. INTRODUCTION

For a proper operation of hydrogen maser the hydrogen flow must be accurately measured and maintained constant at a programmed level. Moreover, in the cavity auto-tuning process the flow level is periodically switched between two different values and the system response must have low setlling time and reasonable over-shoots.

Here is described the system which measures and controls the hydrogen flow, built for the third generation H masers performed at the Institute for Physics and Technology of Radiation Devices, Bucharest (Ref. 1). In order to overcome the drawbacks of the earlier reported versions (Refs. 2,3) several improvements have been made.

2. SYSTEM CONFIGURATION

The system designed to measure and control the hydrogen pressure in the atomic hydrogen source, and consequently the atomic hydrogen flow, is sketched in Fig. 1. Molecular hydrogen from a FeTi hydrure bottle, B, under a pressure arround 0.2MPa enters the palladium leak, heated by the electrical current Ipd,and flows to the atomic hydrogen source which operates at low pressure, in the range of 10^{-1} Torr ($\simeq$ 10 Pa). From here the atoms pass into a high vacuum chamber where the pressure value is below 10^{-7} Torr ($\simeq 10^{-5}$ Pa). The system measures and controls the pressure in the atomic hydrogen source and, in this way, controls the intensity of the atomic beam flow.

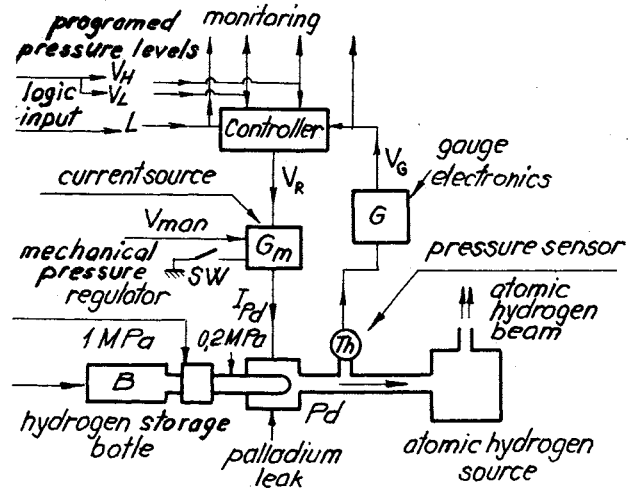


Fig. 1. Block diagram of the flow regulator

The pressure in the atomic H source is measured by the transducer Th and the gauge electronics G (Ref. 4). The output voltage V_G, having a known dependence on the pressure, is compared by the PID controller with one of the programmed levels V_L or V_H . The selection between these levels is done according to the value of the logic input L. Then, the comparison result is processed by the PID controller with an adequate transfer function providing an output voltage V_R which drives the controlled current source G_m. The heating current I_{Pd} supplied by this source controls the temperature of the palladium leak and, consequently, the intensity of the molecular hydrogen flow entering the atomic H source.

3. PRESSURE GAUGE

3.1 Basic operation

A miniature NTC thermistor suspended in a small cavity is the pressure transducer. Its resistance may be expressed as

$$R_m = A_m \exp(B_m/T_m) \qquad (1)$$

where A_m and B_m are the thermistor con-

stants and T_m its operating temperature. In the previously specified range of pressure involved the dissipated power may be approximated as

$$P_d = 1/R_{th}\ (1+p/p_o)\ (T_m - T_o) \qquad (2)$$

where p is the pressure to be measured, T_o is the temperature of the cavity (ambiant) and the constants R_{th} and p_o are depending on the thermistor and gauge design.

In order to obtain a fast response for the pressure gauge the thermistor is connected in a self-ballancing bridge, as shown in Fig. 2. The DC gain of the bridge loop exceeds 120 dB resulting an accuracy for the control of the thermistor temperature better than $5 \cdot 10^{-4}$ K.

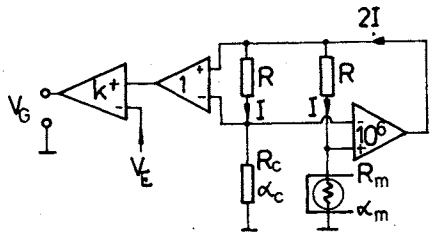


Fig. 2. Block diagram of the pressure gauge electronics

The bridge current may be computed using (1) and (2) giving

$$I = (1+p/p_o)\ (T_m - T_o)^{1/2} \ /\ (R_{th} R_c). \qquad (3)$$

The experimental dependence $I^2(p)$ is plotted in Fig. 3 for two samples of the realized gauges, proving the validity of the approximation made. Moreover, every gauge presented the same value of the constant p_o, namely $p_o = 0.22$ Torr.

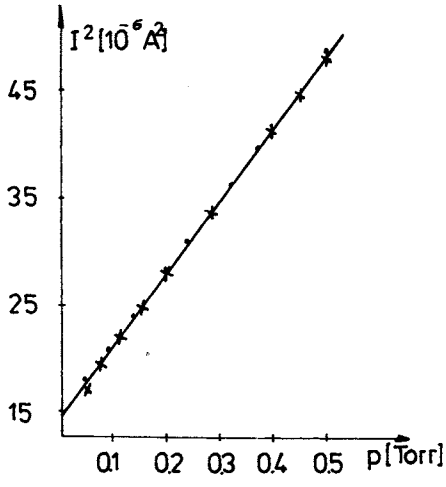


Fig. 3. $I^2(p)$ versus p for two of the realized gauges

Therefore, by trimering the values of V_E and K for each gauge resulted for all gauges the same calibrating curve:

$$V_G(p) = \frac{V_{cal}\ p/p_o}{1+(1+p/p_o)^{1/2}} \qquad (4)$$

with V_{cal} a convenient constant voltage. The measured time constant of the gauge is less than 10^{-2} s allowing a proper operation for the whole system.

3.2 Thermal compensation

Assuming that the other resistors of the bridge have no temperature dependence, one can obtain the temperature coefficient of the current I in respect with ambiental temperature as

$$\alpha_I = \frac{1}{I}\frac{\partial I}{\partial T_o} = -\frac{1}{2}\frac{1}{T_m - T_o}\ . \qquad (5)$$

Thermal compensation is achieved by allowing the dipole R_c in the bridge to have a proper temperature dependence. Thus, (5) is replaced by

$$\alpha_I = \frac{1}{2}\left[\left(\frac{\alpha_c}{\alpha_m} - 1\right)\frac{1}{T_m - T_o} - \alpha_c\right] \qquad (6)$$

where α_c and α_m are the temperature coefficients of R_c and pressure transducer, respectively. The coefficient α_I vanishes at the nominal temperature of the ambient T_{oo} if between the values of the temperature cofficients at this temperature the following condition is satisfied:

$$\alpha_{co} = \frac{\alpha_{mo}}{1-(T_{mo} - T_{oo})\,\alpha_{mo}}\ . \qquad (7)$$

The dipol R_c was syntetized by the network presented in Fig. 4, where R_o is a NTC thermistor mounted in very close thermal contact with the gauge body.

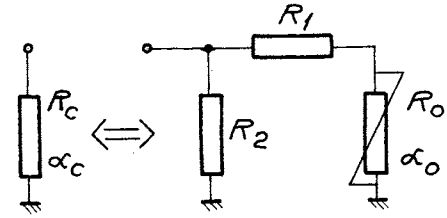


Fig. 4. Compensation network

The dependence of the residual coefficient α_I versus the ambiental temperature T_o is presented in Fig. 5 (solid line). With dashed line is drawn the dependence of α_I for an uncompensated gauge, as given by (5).

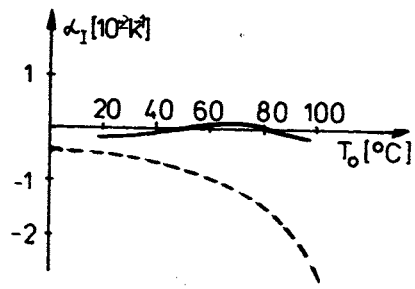


Fig. 5. Dependence of the coefficient α_I versus ambiental temperature for compensated gauge (solid line) and uncompensated gauge (dashed line).

4. PID CONTROLLER

The difference between the programmed level and the output of the electronic gauge V_G is processed by the PID controller. Its transfer function is the product of $H_1(s)$ and $H_2(s)$, where s is the Laplace variable. $H_1(s)$ is a proportional-derivative one and is expressed as

$$H_1(s)=\frac{\omega_2+\omega_3}{\omega_1}\,\frac{s+\omega_1}{(s+\omega_2)(s+\omega_3)} \quad . \qquad (8)$$

The poles lying at $-\omega_2$ and $-\omega_3$ have been introduced from realizability and noise bandwidth considerations. The second transfer function is a proportional-integral one

$$H_2(s)=K_I\,\frac{s+\omega_4}{s}\,\frac{1}{\omega_4} \quad , \qquad (9)$$

the pole in origin providing the elimination of the steady-state loop error for the whole system.

In fact the derivative effect acts on the voltage V_G before the comparison in order to avoid the devices saturation due to the step variation of the programmed level. Although this modification does not affect the system stability, it allows the achieving of better dynamic performances by a proper tuning of the PID controller.

5. CURRENT SOURCE AND PALLADIUM LEAK

Because the resistance of the cables and conectors is comparable with the palladium tube resistance and is subject of variations the power for palladium heating is delivered by a current source controlled by the output voltage V_R of the PID controller. The efficiency of this source is greater than 80% thanks to its switched mode operation, and the response time is less than $5\cdot10^{-2}$ s.

The palladium leak is heated by the electrical current flowing directly through the tube in order to decrease the thermal inertia. However, the time necesary for heating and cooling of the palladium tube is the main limitation for the fastness of the system transitions between the two programmed levels.

6. DYNAMIC PERFORMANCES OF THE SYSTEM

The behaviour of the system was optimized for pressure transitions between 0.10 Torr and 0.15 Torr. By measuring the rate of pressure variation when the palladium tube was supplied with the maximum available heating current (5A) the minimum duration of the "up" transition was estimated. For the "down" transition the rate of pressure variation was determined in the condition $I_{pd}=0$. From these measurements, the following values for the minimum obtenable duration of the transition resulted:

$$t^{up}\geqslant 1.3 \text{ s}; \quad t^{down}\geqslant 3.1 \text{ s}. \qquad (10)$$

In the case of very small variations a linear analysis apply. For the system without controller, in open loop mode, the transfer function was experimentally investigated by analysis of the response to a step input, yielding

$$H_o(s)=H_o(0)\,\frac{\omega_n^2\omega_p}{(s^2+2\zeta\omega_n s+\omega_n^2)(s+\omega_p)} \qquad (11)$$

with $\zeta=0.66$ and $\omega_n=0.22$ rad/s. The location of the real pole was estimated from the frequency of the oscillation obtained by closing the loop with a proportional controller and trimering its gain to reach the limit of stability. Thus, a value of about 70 rad/s resulted for ω_p. It must be emphasized that a pair of complex poles appears in the transfer function (11) due to the palladium leak operation, a result different from the other reported (Ref. 2).

By using the notation

$$K_t=H_o(0)K_I\,\frac{\omega_2\omega_3\omega_n^2\omega_p}{\omega_1\omega_4} \qquad (12)$$

the closed loop transfer function can be expressed as

$$H_{CL}(s)=K_t\omega_1/(\omega_2\omega_3)\;(s+\omega_2)(s+\omega_3)(s+\omega_4)/$$
$$\left[(s^2+2\zeta\omega_n s+\omega_n^2)(s+\omega_p)(s+\omega_2)\cdot\right.$$
$$\left.(s+\omega_3)s+K_t(s+\omega_1)(s+\omega_4)\right] \quad . \qquad (13)$$

Here the programmed level is considered as input variable while the output variable is the output voltage of the pressure gauge.

The system behaviour was studied by the root locus method. For the proper selected values of ω_1 to ω_4 the root-locus is sketched in Fig. 6, the locus parameter being $K'=H_o(0)K_I$.

In the closed-loop transfer function (13) the zeros located at $-\omega_3$ and $-\omega_4$, and the poles lying on the branches starting from $-\omega_3$ and $-\omega_p$ are far enough from the origin to be neglected. The pole on the branch starting from the origin becomes close to $-\omega_4$ for the value K' selected and practically is cancelled by the zero located at $-\omega_4$ (this is not apparent

on the figure due to the scale used).

Therefore, the system behaviour is practically given only by the real pole on the branch starting from $-\omega_2$ and the dominant pair of complex poles. For a value of K' in the range 60-90 rad/s the pair of complex poles is well damped ($\zeta > 0.7$) and the location of these three dominant poles allows a response time less 0.25 s. Thus, for pressure variations between 0.10 and 0.15 Torr, for which the linear analysis do not apply, the response fastness is affected only by the limitation stated in (10).

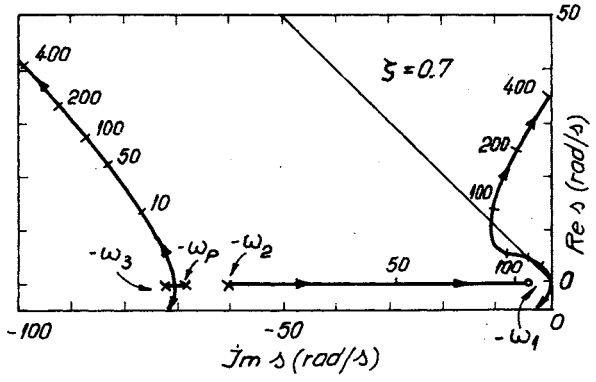


Fig. 6. Root-locus of the closed-loop transfer function.

7. MEASURED PERFORMANCES

The proper tuning of the PID controller was achieved for ω_1=3.9 rad/s, ω_2=61 rad/s ω_3=73 rad/s, ω_4=0.1 rad/s and K=80 rad/s. In Fig. 7 is presented the system response to step transitions between 0.10 Torr and 0.15 Torr. Defining the rise (decay) time t_r (t_d) as the time necessary to pass from 10% to 90% of the whole variation, we obtain

$$t_r=1.85 \text{ s}, \quad t_d=3.20 \text{ s} \qquad (14)$$

in agreement with (10).

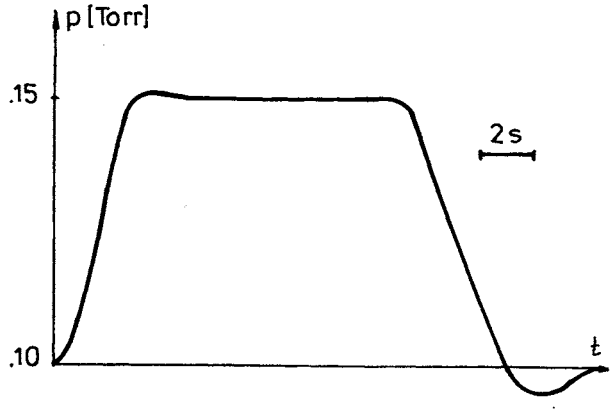


Fig. 7. System response to step variations of the programmed pressure between 0.10 Torr and 0.15 Torr.

Another merit figure is the setlling time, the time required for the response setlling to within 5% of the whole variation. For the transition which are discussed

$$t_s^{up}=2.70 \text{ s}, \quad t_s^{down}=6.25 \text{ s}. \qquad (15)$$

If the over-shoots are defined as

$$S=2\,(p_{ext}-p_f)/(p_f-p_i) \qquad (16)$$

where p_{ext} is the extreme pressure reached during the transition, p_i and p_f are the initial and final levels, respectively, one can calculate

$$S^{up}=0.94\%\ , \quad S^{down}=4.72\%\ . \qquad (17)$$

Fig. 8 presents the system response to repeated equal steps from p=0 to p=0.20 Torr

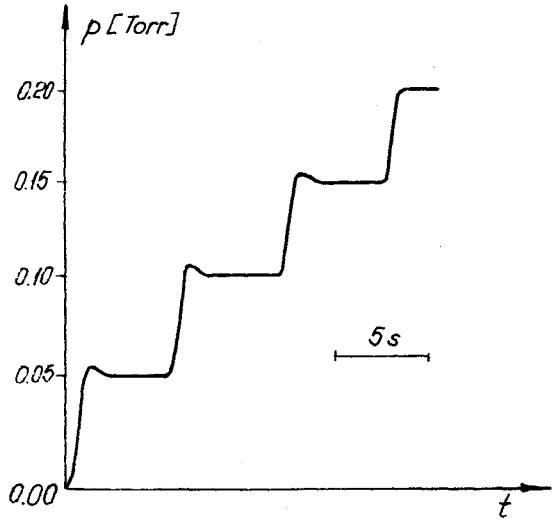


Fig. 8. System response to repeated equal steps from p=0 to p=0.20 Torr.

REFERENCES

1. O.C.Gheorghiu,L.C.Giurgiu,M.Dinca,Cs. Szekely and C.Mandache,"H maser and Rb work in Romania", in Frequency Standards and Metrology, Editor A. De Marchi, Springer-Verlag Berlin,Heidelberg 1989,p.414.
2. J.Viennet,P.Petit and C.Audoin,J.Phys. E,Sci.Instrum.,p.257,1973.
3. P.Petit,R.Barillet,J.Viennet and C. Audoin, Congrés International de Chronométrie,Besançon,1984.
4. L.C.Giurgiu and M.P.Dinca, Revue Roumaine de Physique, in press.

POSSIBILITIES AND LIMITS OF PERFORMANCE IMPROVEMENTS FOR THE DIODE-LASER-PUMPED Rb MASER

Zhu Xiwen, Deng Jinquan, Liu Jinting, An Shaofeng and Tan Yongfang

Laboratory of Magnetic Resonance and Atomic and Molecular Physics, Wuhan Institute of Physics, The Chinese Academy of Sciences, Wuhan P.R.China

ABSTRACT

In this paper we discussed the possibilities and limits of performance improvements for the diode-laser-pumped Rb maser based on the experimental and theoretical results. Some performance improvements might be expected .

Keywords: Rb maser, Laser pumping.

1. INTRODUCTION

Rb maser has been required for a variety of special purposes owing to its outstanding short-term frequency stability when the average time is less than 1s. The short-term frequency stability of the maser can be improved by increasing the output power. A feasible way to do this is to operate a maser with a diode laser. Several groups[1,2,3] obtained the self-sustained oscillations of a diode-laser-pumped Rb maser, but the output powers are not higher than that with a lamp pumping. Our analyses indicated that we may reach the limit of the short-term frequency stability of 10^{-14} by using a high-power(20mW) diode laser.

We measured light shifts in a diode-laser-pumped Rb maser and found that the light shift could be minimized when the laser frequency was adjusted to near the center frequency of the resonance cell. Therefore, the light shift is no longer the main effect on the medium- and long-term stabilities. Due to a low atomic line Q's resulting from high-power laser pumping, these performances will be limited by the cavity pulling shift, which is determined by the the thermal stability of the cavity and the precision of the temperature controller.

A disadvantage with the laser pumping is the laser's residual high frequency phase noise which limits the short-term frequency stability of the maser. In addition, the laser frequency modulation which enlarges the pumping bandwidth to cover as most atom velocity as possible results in additional contributions to the short-term frequency stability. So, there is a considerable interest in whether the performance can be improved further, whether the medium-term and long-term performances can be improved in the meantime and what the limits are.

2. SHORT-TERM STABILITY IMPROVEMENTS

There are two ways to improve short-term frequency stability of the maser. One is increasing the maser output power, and the other increasing atomic resonance line Q, which are discussed respectively as follows.

Increasing maser output power: We studied the variations of the diode-laser-pumped Rb maser output power vs. cell temperature and pumping light intensity. The calculated results and their comparison with experiment[3] are given in Tab.1. From Tab.1 we see that the theoretical results are smaller than experimental ones. The reason is that the present calculation did not take into account of some positive effects for pumping, e.g., velocity-changing-collision which occurs in reality and will increase the pumping efficiency. We proposed a special lamp whose light intensity is about 2 times higher than that of the traditional lamp. Fig.1 shows the variations of the output power of the maser with the lamp pumping vs. light intensity under various cell temperature. The maximum output power is 1×10^{-9}W at cell

temperature of 68°C. The short-term frequency stability of the maser is evaluated as $2\times10^{-14}\tau^{-1}$(1ms≤τ≤200ms) by comparison with another same type maser. The equivalent bandwidth of measurement setup is 1kHz. Therefore, we may expect same order of magnitude improvement in short-term frequency stability upon incorporation of high-power and high-quality diode laser.

Tab.1 The calculated (T) and experimental(E) values of the output power

CT(°C)		55	56	57	58	60	65	70
OP(10^{-10}W)	T	0.41	---	—	—	0.84	1.70	3.0
	E	0.61	0.70	0.85	1.0	—	—	—

** CT:cell temperature; OP:output power

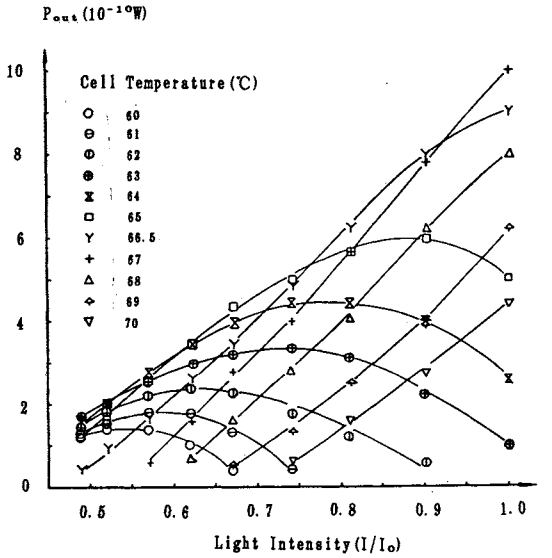


Fig.1 Variations of the output power vs. light intensity under various cell temperature.

Increasing atomic resonance line Q: In order to increase atomic line Q, we must operate a Rb maser in a smaller atomic vapor density and lower pumping light intensity. Both are bad for increasing the output power of the maser. At a given light intensity we can operate a maser at higher cell temperature to reduce the total linewidth by means of the effect of density-narrowing of atomic linewidth in laser pumping[4].

3. MEDIUM- AND LONG-TERM STABILITY IMPROVEMENTS

Decreasing light shift: One of the main reasons motivated the study of diode-laser-pumped Rb maser is that this brings the possibility of better control of the light fluctuations which were supposed to be a major cause for long-term frequency instabilities. We measured the light shifts under condition of various laser intensity and frequency detunings(Fig.2). The smallest light shift is 8.8Hz when the laser frequency is adjusted to near the frequency of the resonance cell(100MHz low). The light shift might be minimized by setting the laser frequency near somewhere of the center frequency of the resonance cell. The relative maser frequency pulling from the laser was found to be about 4.1×10^{-3}, which is fairly agreement with the result shown in Ref.[1]. Therefore, if one designs a maser with a long-term stability of 1×10^{-14}, the stability of the laser frequency must be better than 2×10^{-12}, which can be obtained by using an atomic absorption technique.

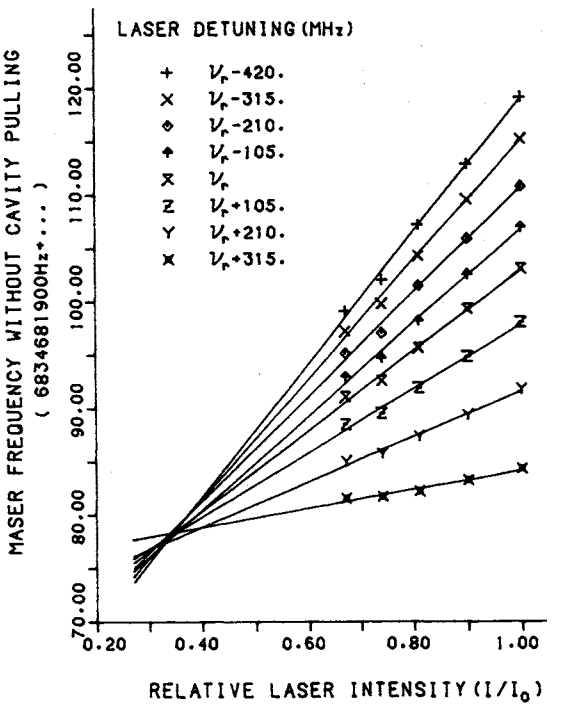


Fig.2 Variations of the light shift vs. incident laser intensity for eight laser detunings.

Reducing cavity pulling: Besides the light shift, the cavity pulling shift is another major cause for long-term frequency instabilities. We can reduce this shift by means of using higher thermal stability of the cavity, better temperature

controller and more precise compensation design. So far, the frequency temperature characteristics of the cavity has been improved by an order of magnitude and to be 0.36kHz/°C [2]. There is no difficult to stable thermally a cavity within 10^{-2} °C. Therefore the relative cavity pulling shift can be reduced to be order of 10^{-13} and the maser frequency temperature characteristics might be improved to be 1×10^{-14} with compensation by filling appropriate kinds and pressures of buffer gases.

4.OVERALL PERFORMANCE OF A LASER-PUMPED MASER

From the results discussed above, we see that high intensity and good frequency tunability of the diode laser can both increase the output power and drastically reduce the light shift effect. The former is good to improve short-term frequency stability and the latter is good to improve medium- and long-term frequency stability. However, the laser's residual high frequency phase noise may degrade the short-term frequency stability. And the high pumping light will both degrade the short-, medium- and long-term frequency stability. So there will exist some optimum operation parameters. Further study of diode-laser-pumped Rb maser is necessary to provide insights into choice of the optimal pumping conditions. In conclusion, we may expect performance improvements upon incorporation of high-power and high-quality diode lasers.

ACKNOWLEDGEMENTS

One of the authors (X.Zhu) gratefully acknowledges the support of K.C. Wong Education Foundation, Hong Kong.

REFERENCES

[1] A.Michaud et al., IEEE trans. IM40, 170(1991).

[2] J.Deng et al., Proc. 6th EFTF, Noordwijk, 319(1992).

[3] N.Oura et al., Proc. 6th EFTF, Noordwijk, 527(1992).

[4] T. McClelland, L. K. Lam and P. R. Berman, Phys.Rev., A33, 1687(1986).

A HIGH UNIFORMITY MAGNET FOR A HIGH C-FIELD CESIUM BEAM RESONATOR

Andrea De Marchi*, Giovanni Costanzo*, Carlo Lombardi*, Antonio Del Casale*, Franco Maddaleno*, Enrico Rubiola*, Maurizio Repetto#, Mario Chiampi#

* Dip. di Elettronica, # Dip. di Ingegneria Elettrica Industriale,
Politecnico di Torino, Torino, Italy

Abstract

The magnetic field uniformity and stability required in a high field Cesium beam resonator [1], for an accuracy of 10^{-14}, are both of the order of 10^{-6} over the whole interaction region. This paper reports on realization and preliminary measurements on a resistive solenoidal magnet designed for this application, complete of a low ripple high stability power supply for its excitation.

1 Introduction

The magnetic field uniformity required in a high field (82 mT) Cesium beam resonator, for an accuracy of 10^{-14}, is of the order of 10^{-6} over the whole interaction region. This goal is clearly very difficult to attain because of the tight mechanical tolerances which it imposes on any realization.

In particular, a gap width uniformity of 10^{-6} would be necessary in a configuration with the field transversal to a Cesium beam. Some consideration was given to this approach, mainly because the possibility that it offers to realize the desired field with permanent magnets appeared very attractive in view of the power consumption foreseen for the operation of an electromagnet at the relatively high target field. Unfortunately it appears very hard to meet specifications in this way, because an unearthly 30 nm tolerance on flatness and parallelism of the magnet's polar expansions would be required.

High uniformity seems to be more easily obtainable in a longitudinal field configuration realized with a multi-layer solenoid. In fact, statistical compensation of imperfections in different winding layers can be expected to average out variations in turn pitch and area, and these are the relevant quantities in determining the field. Unfortunately this kind of statistical smoothing cannot instead be expected to reduce the effect of variations in the section of the pipe on which the coil is wound. Therefore also in this case high mechanical precision on the pipe dimensions is required. However, two factors make life easier in this configuration: one is that a section area is less hard to keep uniform than a gap; the other is that correction with shim coils or ferromagnetic rings is a viable path to improvement in this case. On the other hand a longitudinal field is more convenient from another point of view also. In fact it matches well microwave cavities in which beam holes can be opened up wide without excessively disturbing the rf field distribution. For these reasons, and because (superconducting) magnets with the desired characteristics are commercially available, the choice was made to design the whole beam machine around a high uniformity solenoidal electromagnet.

This paper reports on design, realization, and preliminary uniformity measurements for a 0.7 m long resistive solenoidal magnet with a useful bore diameter of 60 mm.

2 Design and realization of the magnet

The philosophy of magnet design was obviously based on the uniformity target. The latter was 10^{-6} rms in a cylindrical volume 400 mm long and 8 mm in diameter.

The choice was made to realize it by winding a 3 mm × 1 mm rectangular cross section copper wire on a calibrated copper pipe, which also serves as the vacuum envelope. The length of the solenoid was decided on the basis of finite element method calculations (FEM) on a geometrically ideal model, with the idea of achieving the uniformity target with minimal shim coil correction of the effect of the extrem-

This work is partially funded by:
- Consiglio Nazionale delle Ricerche
- Istituto Nazionale di Fisica della Materia

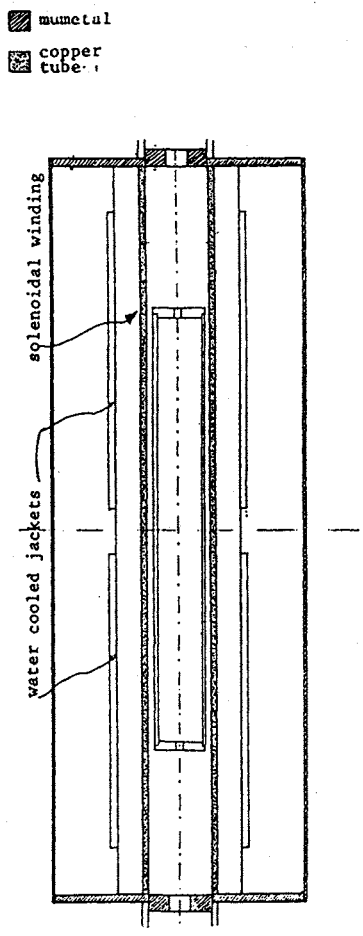


Figure 1: Overview of the magnet.

ities. It was decided after preliminary FEM evaluations that 700 mm would be a reasonable length.

An overview of the magnet is shown in fig. 1, in which different parts of the magnetic circuit can be recognized: the copper pipe, the winding, the mumetal end plates, and the external mumetal flux return tube. For the material of the ferromagnetic part of the circuit, mumetal was chosen in order to maximize μ_r. The thickness of the flux return tube was also designed for maximum μ_r. This is obtained by imposing that the magnetic induction within the metal be about 0.4 T, at which field μ_r is of the order of $3{\cdot}10^5$. The return tube diameter necessary to make the field on the axis independent of its shape and local μ_r was checked by numerical calculation to be over 200 mm. A diameter of 240 mm was then designed. This imposes a thickness of 2 mm to channel at 0.4 T the flux of 82 mT inside the 70 mm bore. At the μ_r value reached in this situation

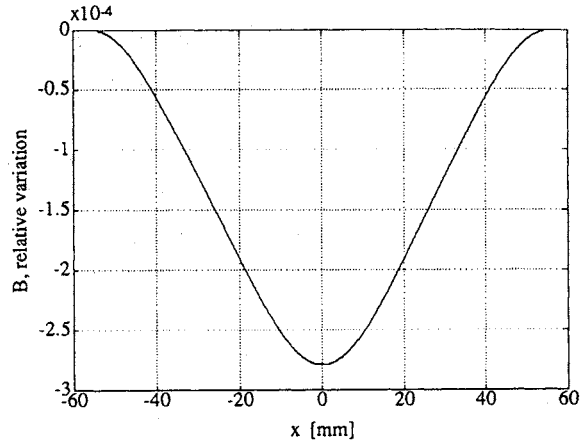


Figure 2: Calculated effect on the field of a bulge of 10 μm in the copper pipe on which the magnet is wound. The defect profile is triangular and its overall extension is 100 mm.

this is equivalent to a 10 mm thick mumetal shield at low field.

The external diameter of the copper pipe was in fact set at 70 mm, both to conform to standard (NW63) vacuum equipment, and to find a compromise between power consumption in the resistive magnet (which increases with bore diameter) and the availability of space in the bore for a suitable microwave cavity. This diameter was checked to be uniform within a measurement resolution of 10 μm.

Sensitivities to fabrication tolerances were also checked with the same axisymmetric FEM code. In fig. 2 the relative variation of the field on the axis is shown as a function of position calculated for an imperfection of the external pipe radius having a triangular profile with a maximum displacement of 10μm and a full development of 100 mm. Writing theoretical formulas which describe the variation of this field perturbation with maximum displacement and with its wavelength is not difficult, and shows, in agreement with fig. 2, that most dangerous are disuniformities with a wavelength of the order of the pipe diameter or more. More difficult is to check on actual manufacturing tolerances, beyond the mentioned 10 μm limit. The approach was taken of asking the machine shop for the "best" diameter calibration obtainable on a limited budget, being as careful as possible in winding the coil, and then resort to trimming the field after assembly.

Mumetal end caps 18 mm thick with a 20 mm hole in the center terminate the bore at the two extremities inside the vacuum enclosure, for better field uniformity. The effect of the reluctance introduced in the magnetic circuit by the thin gap through which these are coupled to the 8 mm thick end plates was investigated with the same FEM code, and is shown in fig. 3 with a grey scale map. The conclusion was

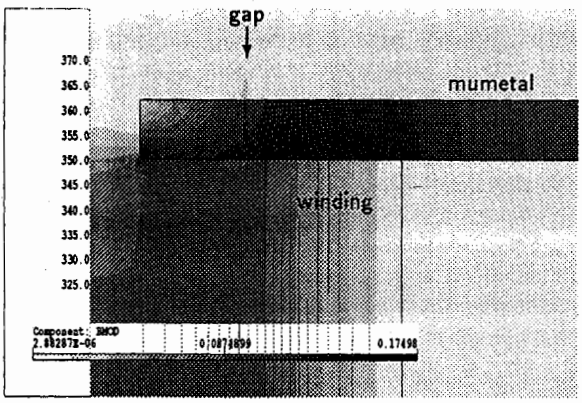


Figure 3: Grey scale map of the field intensity in the region of the magnetic vacuum feedthrough.

made that the gap was acceptable, enabling in this way the design of a solenoid which is all external to the vacuum enclosure.

About 5400 turns were wound in 24 superimposed layers of 224 turns each by the aid of a low speed lathe. A silicon thermally conductive paste was smeared between layers during assembly to make heat removal easier and to fill the gaps between the turns. The length of the wire is 1.5 km, its total weight is 41 kg; the resistance of the winding is 9.6 Ω at 20 °C. The 8.6 A current necessary to excite the desired field of 82 mT (65 kA/m) develops less than 0.8 kW at the operating temperature of 50 °C.

The impedance of the solenoid was measured to have a low frequency inductive part of 250 mH, and resonate at 7 kHz with a Q factor of 4.6. These parameters have been used in designing the power supply.

3 Power supply design

A specification on field stability similar to the one mentioned for uniformity must also be observed not only in long term, but also in short term, since the atomic flight time may be of a few ms. In the long term this limit can be met by servo loop stabilization on a field dependent atomic transition, but for the short term it is the C-field power supply which must guarantee the necessary stability.

Specifications on the latter are of three types:

1. Long term stability must be good enough to guarantee comfortable dynamic range margins for the servo control based on the frequency of a field dependent atomic transition.

2. Short term stability up to tens of ms must be adequate to provide field stability in the 10^{-7} range. The frequency dependent shielding effect of the 5 mm thick copper pipe on which the coil is wound is a great help. The low frequency end of the spectrum is most critical because such shielding factor becomes small ($\approx$10 dB at 50 Hz). For this reason power line related ripple was considered to be a problem.

3. High free running stability for several hours is required in order to allow meaningful high resolution field uniformity measurements.

Most severe are the last two specifications.

In order to fulfill the last requirement, a control system can be built for current stabilization on a stable external voltage reference. However, control bandwidth and residual ground loop problems make it very difficult to follow the same approach for the reduction of low frequency ripple. The latter must be very low in the supply itself. The second requirement appears therefore to be the most critical.

No commercial equipment was readily found that would meet the power line ripple specification of 100 μV peak to peak on 100 V. For this reason, the decision was made to build a switching supply, working at 100 kHz. By suitable internal feedback employing pulse width modulation (PWM) techniques, ripple in these supplies can in fact in principle be reduced below any desired level with proper trimming. On the other hand, at the used switching frequency, the shielding factor of the copper pipe is so high (in excess of 200 dB) that no further filtering is really needed, in this respect, at the output of the supply.

In order to reduce ripple specifications two further precautions were taken. One was to feed line power to the switcher through a three-phase transformer and a six-diode rectifier network. Ideally, this design moves up the ripple frequency to 300 Hz, where the shielding factor of the pipe is 30 dB, and reduces its amplitude to roughly 10% of the DC voltage, in the absence of transformer dispersed flux. The other one was to build the supply in a voltage output mode. In this way the filtering effect of the inductive part of the magnet impedance introduces an additional 30 dB rejection at 300 Hz. The specification for the ripple at the output of the supply is reduced in this way to 100 mV_{pp}.

This specification was met by employing a buck converter with a current-mode control within the supply, which is specifically aimed at reducing the ripple at 300 Hz.

As for the last requirement, random walk-like long term behavior of the field was identified by Allan variance analysis of the signal frequency from an NMR probe placed in the center of the magnet. An instability of 10^{-4} was reached after 300 s of averaging, whereas the short term stability is about 10^{-6}. The conclusion was made that the long term stability of the free running power supply is not adequate for the acquisition of field uniformity data with the desired precision. In order to support the design of

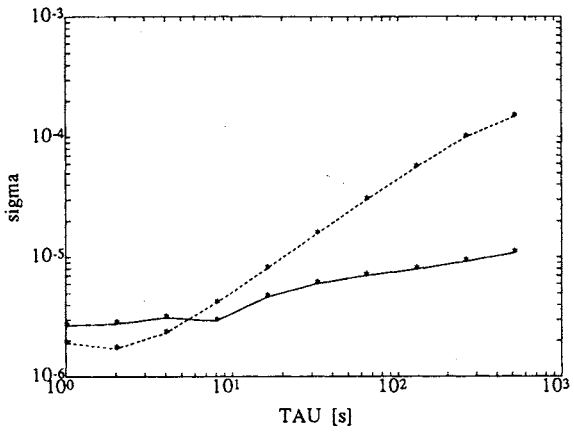


Figure 4: Field stability measured in open loop (dashed line) and closed loop (solid line) conditions.

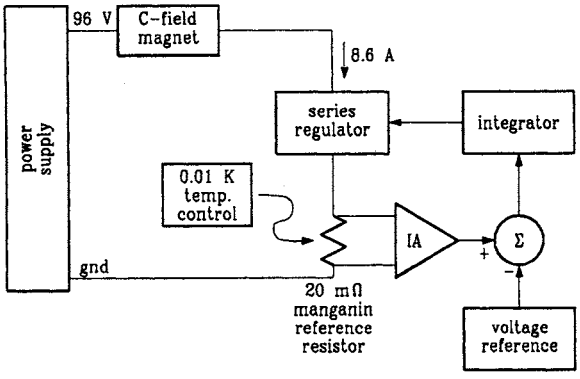


Figure 5: Schematics of current control electronics.

shims for a 10^{-6} uniformity, consistency within the same limit is in fact required of field measurements throughout the longitudinal scan of the magnet.

A current stabilization system was then realized, which is illustrated in the following paragraph. Plots of the square root of the Allan variance of the field measured in the same way for open loop (dashed line) and closed loop (solid line) conditions are shown in fig. 4.

4 Current stabilization

Suitable control of the magnet exciting current is achieved by series load regulation according to the scheme of fig. 5. A signal proportional to current is sensed at the voltage terminals of a 20 mΩ manganin resistor. This .17 V signal is amplified by 40 in a differential amplifier to about 7 V, and then compared with a reference in a current summing node at the input of another amplifier. The output of the latter is the error signal. A simple integrator follows, which drives the regulator.

As instabilities generated inside the loop cannot be corrected, all such circuit variations which are above 100 nV when reported to the input of the first amplifier must be considered unacceptable. In this respect it must be admitted that it would have been better to build a sensing manganin resistor of 100 mΩ, which would have relaxed by a factor of five this limit. However, a five times greater power dissipation on the resistor (7 W) was thought at the time to be unacceptable for precision temperature stabilization with the available Peltier module.

Because of the amplification by 40 right up front, only the sense resistor and the differential amplifier are very critical in this configuration. The stability of their offsets and transfer functions must be analyzed closely.

The sensing resistor is built with two manganin wires in parallel, whose section is 1 mm^2. The wires are enclosed in a copper block, to which they are thermally coupled. The latter is temperature controlled to a stability of 10 mK within 1°C of the point at which the manganin resistor turned out to be maximized. Such point was found experimentally to be about 35 °C. This is fortunate because it happens to be within 2 °C of the equilibrium temperature reached in the environment by self heating alone at the nominal current of 8.6 A. Such coincidence simplifies regulation by taking the load off the Peltier module.

The long term stability of the sense resistor is limited to $5 \cdot 10^{-8}$ by the temperature control.

As for the differential amplifier, a selected version of OP177 was chosen, which features a gain of 130 dB and a thermal drift smaller than 100 nV/°C over a wide temperature range. Gain precision and stability are ensured by a network of four thermally coupled 1 ppm/°C Vishay resistors. By careful resistance matching, a common mode rejection ratio in excess of 110 dB was obtained, which reduces the effective input common mode from above 100 mV to below 300 nV. Variations of the common mode become then unimportant. The amplifier's performance is limited by the Vishay resistors. Further improvement is expected by the use of carefully built manganin resistors.

The three resistors of the current summing node are also thermally coupled high stability Vishay resistors, and will be perfectly adequate even after the improvements on the differential amplifier are made.

The voltage reference is an LTZ1000, which exhibits a temperature coefficient of $2 \cdot 10^{-8}$/°C and a long term stability of $3 \cdot 10^{-7}$ in the proper circuit.

The series regulator is made of a IRFP254 power mosfet with a 500 mΩ resistor connected between drain and source. With this arrangement, the drain circuit Thevenin equivalent is a 4 V generator in series with a 450 mΩ resistance. A 3 V dynamic range is then obtained for the control, which should guarantee locked operation for hours.

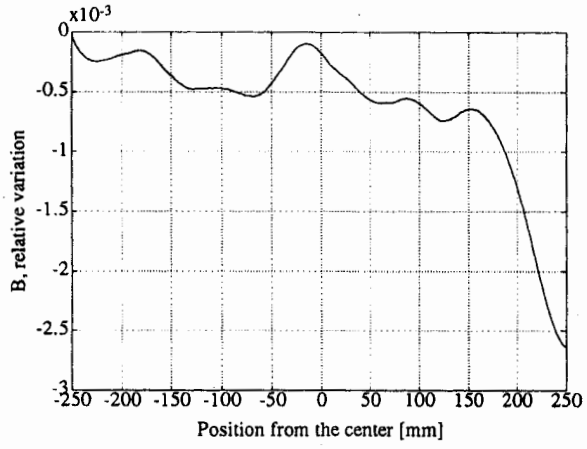


Figure 6: Field uniformity in the central section of the magnet before corrections are applied.

5 Field uniformity

Commercial NMR measurement equipment with a transfer function of 43 MHz/T and a resolution of 1 Hz was used for field uniformity assessment. A resolution of $3{\cdot}10^{-7}$ is in this way realized at the operating field of 82 mT, where the NMR frequency is 3.5 MHz. As a result, uniformity measurements are still limited by the long term stability of the magnet's power supply, even when locked to the voltage reference. The present uncertainty on the complete scan of the magnet in fact is not better than $3{\cdot}10^{-6}$, due to current instability. This is not yet adequate for the design of a shimming configuration apt to take uniformity down to 10^{-6}.

On the other hand, this level is probably hard to reach with the existing magnet, for which the starting point is a uniformity of 10^{-4} rms, as shown in fig. 6. It appears in fact possible to improve this figure by an order of magnitude by shimming. How much better than this it may be possible to improve the existing magnet is not clear at this time. The obtained current stabilization appears then adequate for the goals that are considered reasonable with the present prototype.

The measured spatial disuniformity shown in fig. 6 is thought to be mainly caused by variations of the outer diameter of the copper pipe, as already mentioned. It is to be expected that a possible second prototype magnet, wound on a pipe machined to 1μm, may yield a uniformity as good as 10^{-5} rms before shimming. It may then be possible to apply corrections and push it down to the 10^{-6} level.

Both active shimming by trimmed currents in precisely positioned additional turns placed on the solenoid, and passive shimming by ferromagnetic rings placed in the bore, are considered for the improvement of uniformity in the present magnet. The first technique is probably more stable with time and temperature and is more flexible. In fact shim positions and currents can be easily changed and corrections can be made in either direction. The second offers finer control, as the diameter of the iron shim is much smaller than the diameter of the shim turn. In fact, the range of action of a shim depends linearly on its diameter. A coil placed around the winding has necessarily a minimum diameter of 130 mm, whereas the iron shim diameter would be 60 mm. The bell shaped correction curves of the two ones are accordingly different in width. A maximum effect of 10^{-4} at the top of the bell is obtained with .75 A in the shim turn or a section of 0.03 mm^2 of the iron ring.

In either case the optimization problem is nonlinear in the position of the shim, and linear in current values or ring section. It is therefore easier to fix positions and solve for currents or sections. This procedure is actually much better for active shimming because currents are easily adjusted. Position is a more relevant parameter for the iron rings technique, which leads therefore to a nonlinear problem.

Least square minimization criteria were used with limits on the number of shims and on currents or ring sections. Computer simulations suggest feasibility of an rms uniformity better than 10^{-5} with the available structure. The first implementations of actual shim configurations have realized an rms uniformity of only slightly better than $3{\cdot}10^{-5}$.

6 Conclusions

This paper is a progress report on the realization of a highly uniform magnet for a high C-field cesium beam resonator. The target spatial rms uniformity is 10^{-6}. An rms uniformity better then 10^{-5} appears to be obtainable with the existing prototype.

Acknowledgments: The authors wish to express their deep appreciation for the friendly help of S. Parodi and his employer Ansaldo Componenti, Genova, for the generous loan of the NMR magnetic field measurement equipment.

References

[1] A. De Marchi, "The high C-field concept for an accurate cesium beam resonator," *Proc. 7th EFTF*, Neuchâtel (1993).

Optimization of the Escape Rate for Best Frequency Stability in the Atomic Hydrogen Maser

L. G. Bernier

Neuchâtel Observatory
2000 Neuchâtel, Switzerland

Abstract

Optimization of the frequency stability of the atomic hydrogen maser is analysed in terms of the basic operating parameters that are controllable by the designer, i.e. the flux and the escape rate of the storage bulb. It is shown that for a given flux, escape rates can be found that optimise either the atomic linewidth, the output power or the frequency stability. It is also shown that if the optimum escape rate is maintained while the flux is increased, then the frequency stability improves continuously.

Hydrogen Maser, Frequency Stability, Optimisation, Escape Rate

1. Introduction

The problem of the optimisation of the frequency stability of the hydrogen maser was considered by a number of authors. In most studies [1], [2] and [3] section 6.7.3, the optimisation is performed in terms of the parameters q and I_{th}, as defined by Kleppner et al. [4], and the escape rate of the storage bulb is considered as a constant.

The present study, on the other hand, restricts the optimisation problem to the basic operating parameters that are actually controllable by the designer. We consider that while the storage volume, the wall relaxation rate, the quality and filling factors of the cavity and the efficiency of the state selector are imposed by available technology and by size, weight and complexity considerations, the only operating parameters that are left free for optimisation are the operating flux and the escape rate of the storage bulb.

We consider only the white frequency noise contribution in the frequency stability optimisation. The white phase noise contribution, on the other hand, depends only on the maser output power, i.e. on the flux, and is not the result of a compromise between power and linewidth as the latter.

2. Theoretical Background

We use the basic equations from ref. [3] with the same notations which are mainly those of the original Kleppner et al. article [4]. We mention here only the supplementary notations and hypotheses.

We note P the atomic signal power inside the cavity and P_0 the fraction of this power that is sampled from the cavity and that goes into the receiver.

$$P_0 = \frac{\beta}{1 + \beta} P \qquad (1)$$

β is the coupling coefficient of the cavity.

Assuming that the transverse and longitudinal wall relaxation rates are equal, $\gamma_{1w} = \gamma_{2w}$, and that the magnetic relaxation rate is negligible, [3] equation (6.3.34) and (6.3.37), the theoretical threshold flux I_{th} is given by

$$I_{th} = \frac{h\, V_b}{2\pi\, \mu_0\, \mu_B^2\, Q_c\, \eta'} (\gamma_b + \gamma_{2w})^2 \qquad (2)$$

where h is Planck's constant, V_b the volume of the storage bulb Q_c the loaded quality factor of the microwave cavity and η' the filling factor. γ_b is the escape rate of the storage bulb.

The atomic oscillation signal, [3] equation (6.3.40), is given by

$$\frac{I_0}{I_{th}} = \frac{P}{P_c} = \frac{P_0}{P_0^{th}} = (1 - 3q)\phi - 2\, q^2\, \phi^2 - 1 \qquad (3)$$

where q is the maser figure of merit defined below and ϕ the ratio between the flux I in [atoms/s] and the threshold flux I_{th}. The atomic oscillation signal may be expressed in terms of an excess flux I_0, of a signal power P inside the cavity or in terms of an output power P_0 to the receiver.

In the same manner the maser threshold may be expressed in terms of a flux I_{th}, as defined

above, of a signal power P_c inside the cavity, [3] equation (6.3.41), or in terms of an output power to the receiver.

$$P_0^{th} = \frac{\beta}{1+\beta} P_c \tag{4}$$

Neglecting again the contribution of magnetic relaxation, the total transverse relaxation rate including the density dependent spin exchange contribution is given by [3] equation (6.3.30)

$$\gamma_2 = \gamma_b + \gamma_{2w} + \frac{1}{2}\left(\frac{I_{tot}}{I}\right)\left(\frac{\sigma_{ex}\, v_r}{V_b}\right)\frac{I}{\gamma_b} \tag{5}$$

where I_{tot}/I is the efficiency of the state selector, σ_{ex} the spin exchange cross-section and v_r the mean relative velocity of the hydrogen atoms.

Under the above assumptions the maser figure of merit q is given by

$$q = \frac{\sigma_{ex}\, v_r\, h}{4\pi\, \mu_0\, \mu_B^2\, \eta'\, Q_c}\left(\frac{I_{tot}}{I}\right)\left(\frac{\gamma_b + \gamma_{2w}}{\gamma_b}\right) \tag{6}$$

following [3] equation (6.3.33).

From equation (5) it can be deduced that the relaxation rate extrapolated to zero flux, i.e. without the spin exchange broadening effect, is

$$\gamma_2(0) = \gamma_b + \gamma_{2w} \tag{7}$$

and from equation (3) it can be deduced that when the ratio

$$r = \frac{P_0}{P_0^{th}} \tag{8}$$

and the figure of merit q are known from experimental measurements, the ratio ϕ can be computed.

3. Optimisation of γ_2

For a given flux I the escape rate that minimizes the atomic linewidth is determined by finding the root of the derivative of γ_2, equation (5), with respect to γ_b. The result is :

$$\gamma_b^{opt(\gamma_2)} = \sqrt{\frac{1}{2}\left(\frac{I_{tot}}{I}\right)\left(\frac{\sigma_{ex}\, v_r}{V_b}\right) I} \tag{9}$$

4. Optimisation of Output Power

The output power P_0 to the receiver is proportional to the excess flux I_0 which is given by

$$I_0 = \left(1 - 3\, k_1 \left(\frac{\gamma_b + \gamma_{2w}}{\gamma_b}\right)\right) I - \left(\frac{2\, k_1^2}{k_2\, V_b\, \gamma_b^2}\right) I^2 - k_2\, V_b\, (\gamma_b + \gamma_{2w})^2 \tag{10}$$

according to equations (2), (3) and (6). The constants k_1 and k_2 are defined below.

$$k_1 = \frac{\sigma_{ex}\, v_r\, h}{4\pi\, \mu_0\, \mu_B^2\, \eta'\, Q_c}\left(\frac{I_{tot}}{I}\right) \tag{11}$$

and

$$k_2 = \frac{h}{2\pi\, \mu_0\, \mu_B^2\, \eta'\, Q_c} \tag{12}$$

For a given flux I the escape rate $\gamma_b^{opt(I_0)}$ that maximizes the output power is determined by finding the root of the derivative of I_0 with respect to γ_b. The analytic result is not tractable. Therefore it is better to compute this optimum escape rate by numerical optimisation.

5. Optimisation of Frequency Stability

It can be shown, [3] equation (6.7.31), that the stability limit of the hydrogen maser is determined by the white frequency noise component which may be expressed as the following Allan standard deviation

$$\sigma_y(t) = \frac{1}{Q}\sqrt{\frac{kT}{2P}}\,\tau^{-1/2} \tag{13}$$

where $Q = \pi\nu_0/\gamma_2$ is the operating atomic quality factor and P the maser signal power in the cavity as defined above. The frequency stability of the maser is therefore optimised for the minimum of the function

$$s = \frac{\gamma_2}{\sqrt{I_0}} \tag{14}$$

For a given flux I the escape rate that optimises the frequency stability is given by the root $\gamma_b^{opt(s)}$ of the derivative of s with respect to γ_b. The root cannot be found analytically but the optimum escape rate can still be computed numerically. We will show in the 2 examples below that the escape rates that optimise I_0 and γ_2 are very close to each other so that the escape rate that optimises the frequency stability is well approximated by taking the mean value of the other optima.

6. Influence of the Flux and of the Wall Relaxation Rate

It is common practice to chose the storage bulb escape rate approximately equal to the wall relaxation rate on the rule of thumb that the hydrogen atoms must stay long enough in the storage bulb so that the linewidth is limited by the wall relaxation rate but not too long because relaxed atoms participate to the spin exchange broadening effect.

However the escape rate that yields the minimum linewidth is independent of the wall relaxation rate and depends only on the flux, equation (9). On the other hand we have computed the escape rate that optimises the output power for 3 different values of the wall relaxation rate : 50% 100% and 200% of the 0.75 s^{-1} relaxation rate assumed in the compact maser example below. The variation of the optimum escape rate for these largely different wall relaxation rates was less than 1%. Therefore it appears that the escape rate that optimises the output power is also a function of the flux only.

Figure 1 shows the 3 different optimum escape rates computed as a function of the flux.

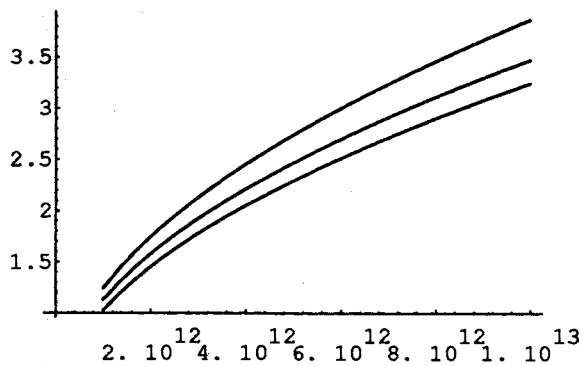


Figure 1
Optimum γ_b versus the flux I

The lower curve shows the escape rate that optimises the linewidth, the middle curve the escape rate that optimises the frequency stability and the upper curve the escape rate that optimises the output power. The operating parameters are those of the compact maser example below. It appears that all three optima increase with the square root of the flux. The optimum region of the s vs γ_b curve (see figure 4) is broad and the three optima close to each other.

Therefore, in practice, choosing the escape rate that optimises the linewidth is a good compromise. The resulting frequency stability is close enough to the optimum with the advantages that the escape rate for optimum linewidth has an analytic expression and that the frequency pulling effect is minimum.

7. Example : Optimisation of EFOS-14

7.1. Comparison of Experimental and Theoretical Parameters

EFOS-14 is a standard maser produced by Neuchâtel Observatory for use in VLBI radiotelescopes. The estimated operating parameters are shown in table 1.

p_0^{th} was estimated by measuring the maser signal output power versus the hydrogen pressure in the dissociator and by extrapolating to zero hydrogen pressure. The quality factor Q(0) of the atomic resonance extrapolated to zero flux, i.e. without the spin exchange contribution, was estimated by extrapolating the atomic linewidth to zero flux. q was estimated by plotting the linewidth versus output power curve using a method similar to that used in [5]. γ_{2w} was estimated by removing the contribution of the storage bulb escape rate from the atomic linewidth extrapolated to zero flux.

Parameter	Measured	Theoretical
I_{th} [s^{-1}]		3.43×10^{11}
q	0.1	0.0086
p_0^{th} [W]	3.94×10^{-14}	3.23×10^{-14}
Q(0)	3.89×10^{9}	3.75×10^{9}
γ_{2w} [s^{-1}]	0.44	
Q_c	34×10^{3}	
β	0.25	
η'		0.45
V_b [litre]	3.8	
I_{tot}/I		2
r	0.8 †	
ϕ	2.80 †	
I [s^{-1}]	9.60×10^{11} †	

Table 1
EFOS-14 Operating Parameters

The standard EFOS collimator is made of solid PTFE and has a length of 51 mm and an internal radius of 3.15 mm yielding $\gamma_b = 0.74$ s^{-1}. The state selector is of the quadrupole type and the efficiency I_{tot}/I is assumed to be 2. The parameters marked with † were estimated for the standard -105 dBm output power to the receiver.

The theoretical value of the threshold flux was computed using (2). The theoretical value of q was computed using (6). The above figures are reported in table 1 and show that the match between estimated and theoretical parameters is good.

7.2. Optimisation of γ_b

Assuming a flux I = 9.60×10^{11} s^{-1} , i.e. the flux that yields the standard -105 dBm output, the escape rate that optimises the linewidth is computed using (9) and the escape rate that optimises the output power is computed numerically from (10).

$$\gamma_b^{opt(\gamma_2)} = 0.46 \ s^{-1} \qquad \gamma_b^{opt(I_0)} = 0.55 \ s^{-1}$$

The escape rate that optimises the frequency stability is

$$\gamma_b^{opt(s)} = 0.52 \ s^{-1}$$

which is very close to the average of the 2 other optima. The relevant parameters were computed and are reported in table 2 for both the standard and optimum collimators. Q is the operating quality factor. The results show that the standard EFOS escape rate is already close enough to the optimum so that the expected improvement in frequency stability is very small.

relaxation rate is, according to equation (7), $\gamma_2 = 1.86\ s^{-1}$ therefore $Q(0) = 2.40 \times 10^9$.

The same flux as used in the full size EFOS maser will yield a slightly higher output power in the compact maser. The threshold flux is now, according to equation (2) and using the optimum escape rate, $I_{th} = 1.76 \times 10^{11}\ s^{-1}$ therefore $\phi = 5.43$ and, after equation (4), $P_0^{th} = 1.66 \times 10^{-14}$ W. After equation (6) we also have $q = 0.09$. Therefore, after (3), we have $r = 2.48$ and $P_0 = 4.13 \times 10^{-14}$ W (-104 dBm). The same result is obtained directly from (10). Using equation (5) the operating relaxation rate is found to be $\gamma_2 = 2.77\ s^{-1}$ therefore $Q = 1.61 \times 10^9$.

8.4. Frequency Stability with Optimum Collimator

The output power P_0 = -104 dBm yields $P = 2.07 \times 10^{-13}$ W and from equation (13) we obtain $\sigma_y(t) = 6.22 \times 10^{-14}\ \tau^{-1/2}$ i.e. 2.0×10^{-15} at 1000 s averaging time.

8.5. Extrapolation to Higher Fluxes

We have done the same optimisation process under the assumption that the dissociator is able to yield 10 times the original flux, i.e. $I = 9.60 \times 10^{12}\ s^{-1}$ all other parameters being the same. Under these conditions the escape rate that optimises the frequency stability is $3.4\ s^{-1}$, the signal output power is $P_0 = 5.57 \times 10^{-13}$ W (-93 dBm) and the operating atomic quality factor is $Q = 6.26 \times 10^8$. Therefore $\sigma_y(t) = 6.22 \times 10^{-14}\ t^{-1/2}$ i.e. 1.4×10^{-15} at 1000 s averaging time.

9. Conclusion

In this paper we have shown that since the escape rate of the storage bulb is a free parameter, very easy to control by design, it should be optimised either for the maximum output power, for the minimum atomic linewidth or for the best frequency stability. We have shown that the 3 optima are close to each other, increase with the square root of the flux and are independent of the wall relaxation rate.

In 2 examples we have shown that the degradation of the Allan deviation is in a factor of 1.8 between a full size maser design (3.8 litre storage) and a compact maser design (0.8 litre storage) operated at the same flux.

On the other hand we have also shown that the frequency stability improves continuously when the flux is increased but the improvement is very slow, a factor of 1.4 for a tenfold increase of the flux. This is a small improvement of frequency stability at the price of a very large increase of the hydrogen consumption and of an increased sensitivity to the cavity pulling effect since the operating atomic quality factor is sensibly degraded.

10. References

[1] Vessot R.F.C., Levine M.W., Mattison E.M., "Comparison of Theoretical and Observed Hydrogen Maser Stability Limitation due to Thermal Noise and the Prospect for Improvement by Low-Temperature Operation, " *Proc. 9th PTTI*, Greenbelt, March 1978, pp. 549-569.

[2] Mattison E.M., Vessot R.F.C., "Analysis of Fundamental and Systematic Effects Limiting Hydrogen Maser Frequency Stability, " *Proc. 21st PTTI*, Redondo Beach, California, November 1989, pp. 433-444.

[3] Vanier J., Audoin C., *The Quantum Physics of Atomic Frequency Standards*, Adam Hilger, IOP Publishing, 1989.

[4] Kleppner D., Berg H.C., Crampton S.B., Ramsey N.F., Vessot R.F.C., Peters H.E., Vanier J., "Hydrogen-Maser Principles and Techniques, " Physical review, vol. 138, no. 4A, May 1965, pp. A972-A983.

[5] Mattison E.M., Shen W., Vessot R.F.C., "Measurement and Interpretation of Hydrogen Maser Quality Parameters, " *Proc. 39th Annual Symposium on Frequency Control*, Philadelphia, May 1985, pp. 72-74.

[6] Gaygerov B.A., Rusin F.S., Sysoev V.P., "Portable Atomic Clock on the Basis of an Active Hydrogen Maser "Sapphir"", *Proc. 5th EFTF*, Besançon, March 1991, pp. 293-294.

[7] Sphicopoulos T., Bernier L.G., Gardiol F., "Theoretical Basis for the Design of the Radially Stratified Dielectric Loaded Cavities Used in Miniaturized Atomic Frequency Standards", *IEE Proceedings*, vol. 131, part H, no. 2, April 1984, pp. 94-98.

Parameter	Standard escape rate	Optimum escape rate
γ_b [s^{-1}]	0.74	0.52
P_0 [W]	-105 dBm	-105 dBm
I_{th} [s^{-1}]	3.43×10^{11}	2.25×10^{11}
r	0.80	1.6
ϕ	2.80	4.26
I [s^{-1}]	9.60×10^{11}	
$\gamma_2(0)$ [s^{-1}]	1.19	0.96
$Q(0)$	3.75×10^{9}	4.63×10^{9}
γ_2 [s^{-1}]	1.48	1.38
Q	3.02×10^{9}	3.24×10^{9}
P [W]	1.58×10^{-13}	1.71×10^{-13}
$\sigma_y(1000s)$	1.2×10^{-15}	1.1×10^{-15}

Table 2
EFOS-14 Operating Parameters for Standard & Optimum Escape Rate

8. Example : Optimisation of a Compact Maser

8.1. Introduction

A compact maser based on the Russian concept of a sapphire loaded microwave cavity [6] design would have a V_b = 0.8 litre storage volume. For a given relaxation probability per collision, if the storage volume is reduced while keeping constant the shape of the storage bulb, the wall relaxation rate is expected to increase proportionally to the inverse of the radius, [3] equations (3.4.27) and (3.4.28).

Under this assumption the relaxation rate measured in the 3.8 litre EFOS storage bulb can be extrapolated to a 0.8 litre storage volume yielding $\gamma_{2w} = 0.750\ s^{-1}$. Moreover we assume that the flux I, the cavity coupling β and the filling factor η' are the same as in the EFOS-14 example above. As a matter of fact the filling factor of a dielectric loaded cavity is expected to be at least as high as in an empty cavity [7].

8.2. Optimisation of γ_b

Using the above parameters the function $I_0(\gamma_b)$ was computed from equation (10) and is plotted on figure 2.

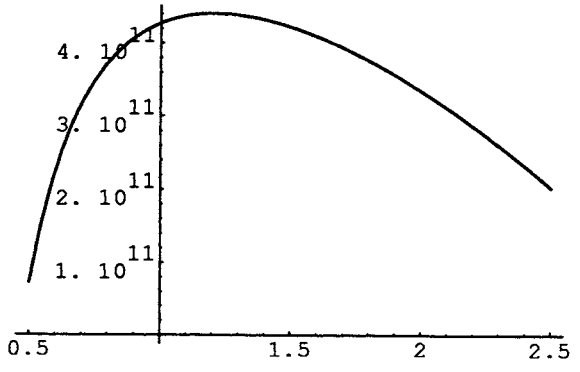


Figure 2
I_0 versus γ_b plotted with Compact Maser Parameters

The escape rate that yields the maximum output power was found by numerical optimisation.

$$\gamma_b^{opt(I_0)} = 1.206\ s^{-1}$$

The function $\gamma_2(\gamma_b)$ was also computed from equations (5) and is plotted on figure 3.

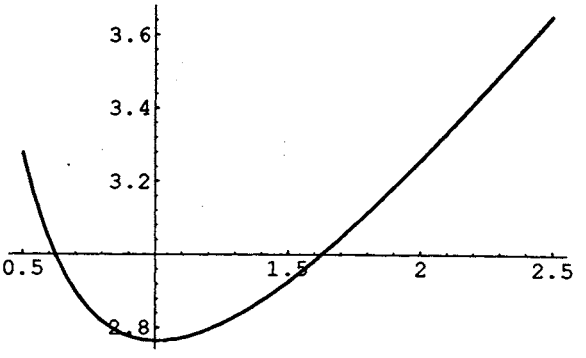


Figure 3
γ_2 versus γ_b plotted with Compact Maser Parameters

The escape rate that optimises the linewidth is given by equation (9).

$$\gamma_b^{opt(\gamma_2)} = 1.007\ s^{-1}$$

The frequency stability function s is shown on figure 4.

The escape rate that optimises the frequency stability function s was computed numerically and is found to be

$$\gamma_b^{opt(s)} = 1.112\ s^{-1}$$

Again the above result shows that the escape rates that maximize the output power and the operating Q are very close together so that the mean value of the latter yields the escape rate that optimises the frequency stability.

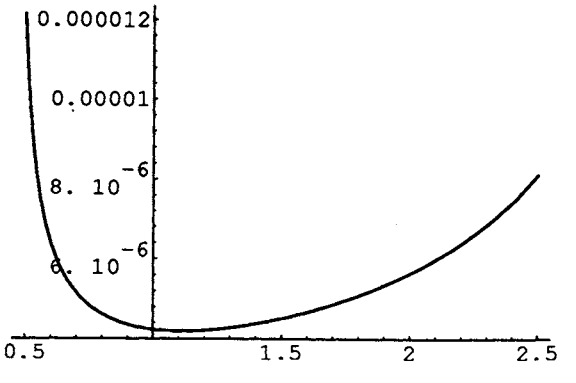


Figure 4
s versus γ_b plotted with Compact Maser Parameters

8.3. Atomic Quality Factor with Optimum Collimator

Using the above defined optimum escape rate and assuming that the flux is the same as in the previous example, the zero flux

CHARACTERISTICS OF FIBER-OPTIC FREQUENCY REFERENCE DISTRIBUTION SYSTEMS

Richard A. Dragonette and Joseph J. Suter
The Johns Hopkins University Applied Physics Laboratory
Laurel, Maryland 20723-6099 USA
Telephone (301) 953-5826

ABSTRACT

Precision time and frequency reference signals have been effectively disseminated using high-quality intricate fiber-optic distribution systems. The quality of signals distributed by such systems is excellent, but the cost of these systems makes them unavailable to many potential users. In this paper, a study of signal quality maintained using inexpensive commercial transmitter/receiver pairs is undertaken. Seven different transmitter/receiver pairs obtained from four different manufacturers have been thoroughly tested using a 5-MHz sinusoid derived from a precision, temperature-controlled, crystal-controlled oscillator. The electrical signal output from each fiber-optic receiver was tested for spectral purity, single-sideband phase noise, and AM noise, and the results are tabulated and discussed without identification of the manufacturer or the equipment model number.

INTRODUCTION

One advantage of maintaining atomic frequency standards is that their reference frequency signals can be distributed to a number of users. One of the difficulties associated with a remotely located atomic frequency standard is that its signals tend to be degraded by transmission over long distances. Long interbuilding connections typically share cable trays and underground conduits with digital computer-network cables and 60-Hz power cables. Sinusoidal frequency-reference signals are often badly distorted by induced electrical interference as they are distributed via coaxial cables.

One effective way to minimize electrical interference and distortion is the use of multichannel fiber-optic transmission networks. Monetarily, it is impractical to implement a large number of interbuilding connections with customized analog fiber-optic links, which would cost thousands of dollars per channel. We therefore began testing inexpensive commercial fiber-optic transmitters and receivers that were designed for video signal transmission as a means of disseminating frequency reference signals.

In this study, seven different fiber-optic transmitter/receiver pairs (TRP's) from four different manufacturers were tested to verify the extent to which they degrade the single-sideband phase noise, harmonic distortion, spurious noise, and AM noise of a 5-MHz sinusoid derived from a precision, temperature-controlled, crystal-controlled oscillator. The test results are summarized here without identification of the manufacturer or the model number. The cost per TRP of the systems were as follows: System A, $600; System B, $700; System C, $700; System D, $2850; System E, $3800; System F, $10,000; System G, $700.

APL currently maintains 14 fiber-optic distribution lines, consisting exclusively of equipment from this study, that cost less than $1000 per TRP. In each case, the fiber-optic systems have improved the reliability and noise quality of the signals being transmitted.

EXPERIMENTAL PROCEDURE

The seven systems tested here are designated System A through System G. All tests were performed using the TRP under test to transmit a low-phase-noise 5-MHz sinusoid derived from the same precision, temperature-controlled, crystal-controlled oscillator. The output of each system was tested for single-sideband phase noise performance, harmonic content, spurious noise, and AM noise.

For the tests, each TRP was connected by a short fiber-optic cable (less than 50 ft long). For systems A, B, C, and F, the tests were repeated using a 3200-ft, 250/62.5-μm-diameter multimode fiber.

Phase-noise tests were conducted over the frequency range of 0.1 Hz to 100 kHz from the carrier frequency, using an HP 3048A phase-noise test system, which was calibrated (traceable to the National Institute of Standards and Technology) to be accurate to within ± 2 dB over the frequency range of interest. It is clear from Figures 1 through 8 that the fiber-optic systems measurably degraded the phase noise of the reference oscillator at frequencies greater than 1 Hz from the carrier.

AM noise tests were also performed using the HP 3048A phase-noise system over the frequency range of 1 Hz to 100 kHz from the carrier frequency. The spectral purity of the reference oscillator and the systems under test was verified with an HP 8560A spectrum analyzer that was set for a frequency range of 1 to 51 MHz, with a video bandwidth and resolution bandwidth of 10 kHz.

DESCRIPTION OF THE SYSTEMS

Systems B, C, and G are rack-mounted systems in which the transmitters and receivers are plug-in cards that plug in to a common rack chassis. Each card houses up to two transmitters or receivers, and the rack chassis provides power and cable connectors. The remaining systems consist of stand-alone transmitters and receivers.

Three identical pairs of System A equipment were available and tested. The system uses an LED light

source at 850 nm for light transmission, and has an input frequency range of DC to 60 MHz. Its electrical output can be adjusted from 50 mV to 2 V p-p.

System B has an operational frequency range of near-DC to 20 MHz, with an electrical output amplitude that is adjustable up to 3 V p-p. This system uses an LED operating at 850 nm as its light source. The performance of this system was partially limited by the presence of noise added by a 1-pps signal being distributed by a transmitter on the same chassis, although the manufacturer claims that this system is capable of distributing both sinusoidal and pulsed signals from the same chassis.

System C consists of the same equipment as System B, but with the transmitter distributing the 1-pps signal removed from the chassis. Comparing this system with and without the 1-pps lines active is important because it is not uncommon for a timing station to distribute both 5-MHz frequency references and 1-pps timing references.

System D has an operating frequency range of 5 to 125 MHz, and uses an 1300-nm LED as its light source. Its electrical output signal level is approximately -10 dBm.

System E uses a 1350-nm laser transmitter and single-mode optical fiber for interconnection. Its operating frequency range is 5 to 300 MHz, with an electrical output power near -25 dBm.

System F uses a 1350-nm laser light source and single-mode optical fiber interconnection, with an operating frequency range of 5 MHz to 3 GHz.

System G has an operating frequency range of DC to 10 MHz, and uses an 850-nm LED as a light source.

EXPERIMENTAL RESULTS

The results of the single-sideband phase-noise tests of each individual system are summarized in Table 1. Perhaps the most significant result of this study is found in Table 1; each of these vastly different systems exhibit a single-sideband phase noise floor between -115 and -125 dBc/Hz for frequencies greater than 10 Hz from the carrier. Figure 1 shows the single-sideband phase noise of the reference oscillator. Figure 2 shows the single-sideband phase noise of System A's output when transmitting the reference oscillator through a short optical fiber. Figures 3 through 8 show the phase noise test results for Systems B through G, respectively.

Table 2 summarizes the AM noise test results. Table 3 describes the harmonic content observed in the outputs of the reference oscillator and in each of the fiber-optic links.

Three pairs of System A equipment were available for testing. Each one was tested with both a 10-ft and a 3200-ft multimode fiber. The single-sideband phase-noise measured in the outputs of all three receivers exhibited a noise floor of 115 $\pm$ 4 dBc/Hz for frequencies greater than 10 Hz from the carrier frequency.

The excessive spurious signals seen in the phase-noise characteristic of System B (Figure 3) were caused by the presence of a 1-pps transmitter in the rack chassis used to supply power to System B. Similar noise spurs were seen in System B's AM noise response between 5 and 100 Hz from the carrier.

The 1-pps transmitter was removed and the tests were repeated. The results of the retests were designated as System C tests to differentiate the new results from the original System B tests. As is indicated by Figure 4, System C transmitter/receiver pairs achieved a phase noise floor as great as 10 dBc/Hz below the phase noise floor of System B.

System G has 75-Ω electrical inputs and outputs, but both the reference oscillator and the HP3048A are configured for 50-Ω systems, requiring impedance-matching networks to interconnect the devices. In is possible that those impedance-matching networks may have limited the test results of System G, although its measured performance compares favorably with the other systems.

As can be seen in Table 2, each of the TRP's significantly degraded the AM noise characteristics of the reference oscillator. With the exception of System B, each system achieved an AM noise floor in the neighborhood of -120 dBc/Hz. The reference oscillator's AM noise floor was closer to -150 dBc/Hz. No AM noise measurements were made on System A.

As can be seen in Table 3, each of the fiber optic transmission systems added to the harmonic content of the reference oscillator. The first harmonics of the systems ranged between -32 and -43 dB below the carrier, and the second harmonics between -37 and -57 dB below the carrier. In Table 3, the word "noise" means that the specified harmonic was hidden by the receiver's noise floor.

CONCLUSIONS

The most interesting result of testing these fiber-optic systems is that they all achieved a single-sideband phase-noise floor of -115 to -125 dBc/Hz. This noise floor was amazingly consistent across the tests even though the systems used different light sources, operated over different frequency spans, varied widely in cost, and were manufactured by different manufacturers.

Since a number of examples representative of Systems A and C were available, they were tested in the full permutation of configurations; that is, System A transmitter 1 was tested with System A receiver 1, 2, and 3, etc. From these tests, it is clear that, at least for these systems, the phase noise in the output signal is not completely limited by either the transmitter or the receiver, but rather by a combination of both elements.

It is also interesting to note that optical cable length can significantly affect the system's performance. For example, in the three tests run on System A the phase noise floor was 2 to 3 dBc/Hz lower when a 3200-ft cable was used to connect the transmitter and receiver than it was when a 50-ft cable was used.

It is clear that all the fiber-optic systems significantly degrade the single-sideband phase noise and AM noise, and add to the harmonic content of a quality oscillator. The levels of performance achieved by these systems is acceptable for reference-frequency distribution if the transmitted signals are being used as a common time base for test equipment, or as a reference for a system that phase locks its own internal voltage-controlled oscillator to the transmitted reference.

BIBLIOGRAPHY

1. R. Simons, *Optical Control of Microwave Devices*, Artech House, Boston (1990).

2. E. Udd, *Fiber Optic Sensors*, John Wiley, New York (1991).

3. H. W. Yen et al., "High-Speed Optical Modulation Techniques," *Optical Technology for Microwave Applications II*, SPIE, 645, 2-9 (1985).

4. G. Keiser, *Optical Fiber Communications*, McGraw-Hill, New York (1983).

Table 1. Single-sideband phase-noise performance of the fiber-optic systems.

	Single-sideband phase noise (dBc/Hz)							
Freq. (Hz)	Ref.	Sys. A	Sys. B	Sys. C	Sys. D	Sys. E	Sys. F	Sys. G
0.1	- 90	- 85	- 90	- 90	- 90	- 90	- 90	- 80
1	-120	-115	-115	-115	-115	-110	-117	-113
10	-145	-115	-116	-125	-120	-118	-124	-118
100	-152	-115	-116	-125	-120	-118	-124	-119
1k	-153	-115	-116	-125	-119	-119	-124	-119
10k	-153	-115	-116	-125	-120	-119	-124	-119
100k	-153	-115	-116	-125	-118	-119	-124	-119

Table 2. AM noise response of the fiber-optic systems.

	AM noise response (dBc/Hz)						
Freq. (Hz)	Ref.	Sys. B	Sys. C	Sys. D	Sys. E	Sys. F	Sys. G
1	-125	-112	-115	-108	-90	-120	-110
10	-133	-115	-121	-110	-100	-120	-115
100	-140	-112	-121	-115	-105	-120	-117
1k	-148	-113	-120	-118	-115	-120	-117
10k	-153	-112	-119	-118	-120	-120	-117
100k	-154	-111	-117	-118	-120	-120	-117

Table 3. Harmonic Content of the fiber optic systems outputs.

	Harmonic Content (dBc)							
Harmonic	Ref.	Sys. A	Sys. B	Sys. C	Sys. D	Sys. E	Sys. F	Sys. G
1st	-53	-39	-43	-42	-40	-37	-38	-32
2nd	-55	-43	-52	-51	-57	-37	-50	-48
3rd	-69	-62	-66	noise	-72	-47	-55	noise
4th	-67	noise	-64	noise	-70	-57	-57	noise

NOTE: Harmonics higher than 4th order were insignificant except for system G.

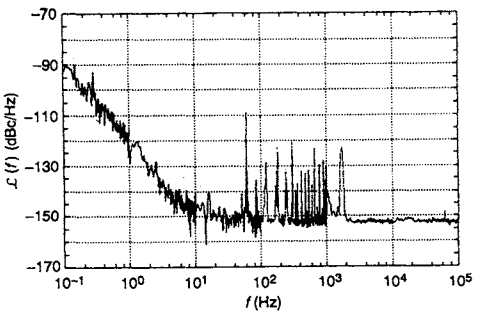


Figure 1. Reference oscillator single-sideband phase noise.

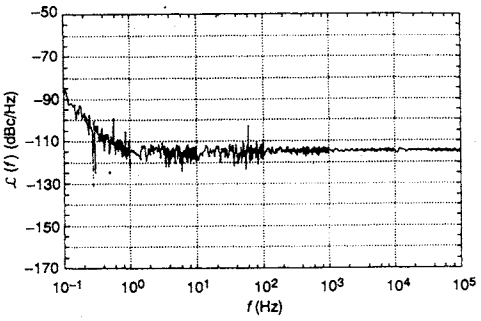


Figure 2. System A single-sideband phase noise.

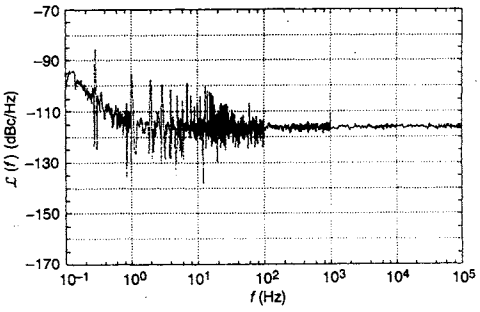


Figure 3. System B single-sideband phase noise.

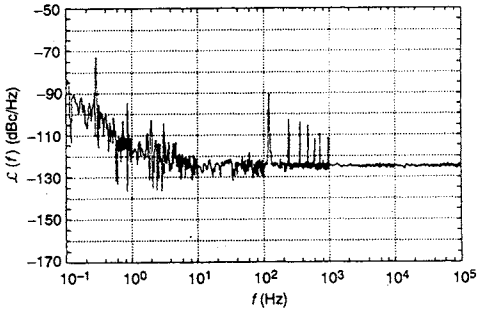


Figure 4. System C single-sideband phase noise.

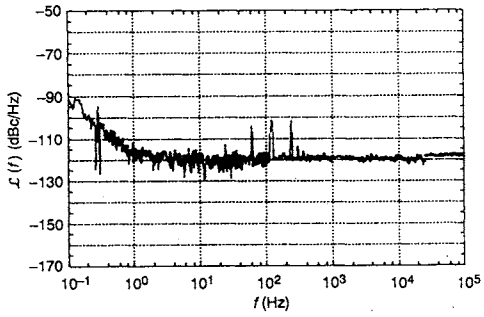


Figure 5. System D single-sideband phase noise.

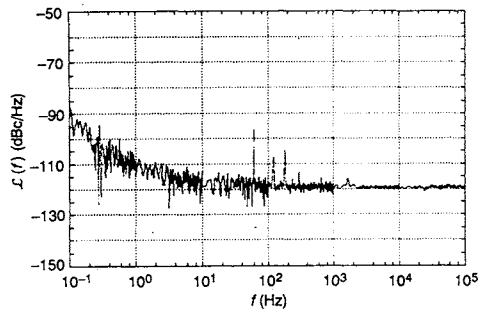


Figure 6. System E single-sideband phase noise.

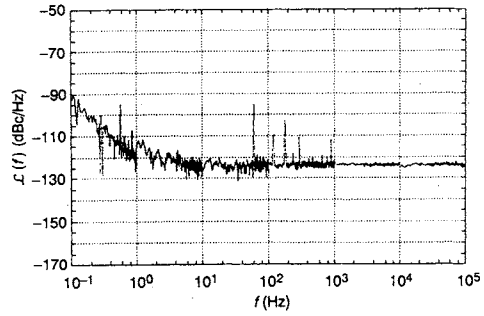


Figure 7. System F single-sideband phase noise.

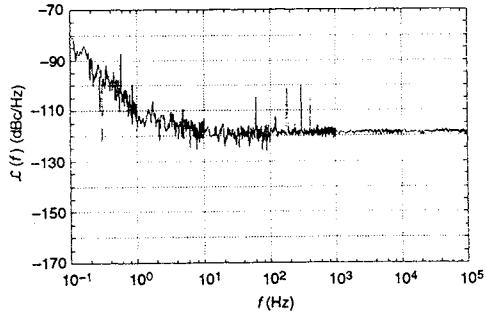


Figure 8. System G single-sideband phase noise.

TV-FREQUENCY STANDARD AS A TRACEABILITY SOURCE IN ACCREDITED CALIBRATION LABORATORIES

Kalevi Kalliomäki (1,2), Tapio Mansten (1)

(1) Technical Research Centre of Finland (VTT)
(2) University of Oulu, Oulu, Finland

Abstract

In Finland TV-sync pulses of our national broadcasting company (YLE) are locked to a rubidium clock. TV-phase is monitored at VTT. The short term stability of the phase is about 10 ns. The potential of TV-frequency dissemination method has led to an product called TV-frequency standard. In Finland there are seven accredited frequency calibration laboratories ALL of them are equipped with the TV-frequency standard as a traceability source. To check the measurement methods and instruments of each calibration laboratory, audit measurements are performed annually. The typical laboratory error is 1.8*10exp-11 (transfer error included) demonstrating the potential of TV-sync. pulse frequency transfer. Now most of our customers calibrate their counters themselves, the accuracy level is 1*10exp-11 at best.

Introduction

Frequency and time dissemination by using the frame-synchronization pulses of TV-network is an old and wellknown method. TV-synchronization pulses in many countries have been locked to atomic frequency standards to allow easy frequency and time comparisons. For example in Finland TV-sync pulses of our national broadcasting company were locked, first to a Loran-C receiver in 1984 via telephone line and then to a rubidium clock since 1987.

The requirements for better quality control, broadening of European common market, the Western European Calibration Cooperation (WECC) and ISO 9000 standard are some reasons to improve traceability of all electrical quantities.

Earlier the traceability of frequency in Finland was realized by calibrating the counters of our customers annually at VTT, which is the national measurement laboratory for electrical quantities. The inaccuracy thus obtained was about 1*10exp-8. Some customers claimed better accuracy without bying expensive atomic frequency standards. Our colleques in Sweden had the same problem. Then the study of the potential of TV-frequency dissemination method was started leading to an industrial product called TV-frequency standard. Now most of our customers calibrate themselves their counters, the accuracy level being 1*10exp-11 at the best.

Stability of TV-synchronization pulse

The TV-synchronization unit of Finnish broadcasting company (YLE) is locked to a rubidium frequency standard (Rohde & Schwarz), which in turn is monitored at VTT by continuously measuring the phase of the frame

synchronization pulse.
VTT uses GPS and Loran-C (Sylt) as references.
In addition this phase is monitored at the University of Oulu, 600 km to the North of Helsinki. Here the reference is Loran-C (Bo). Thus the possible effect of long microwave link chain on the phase may be seen. Some stability data has been published in 1990 EFTF, Neuchatel (J.Mannermaa). The short term (minutes) stability of the phase is about 10 ns in both monitoring sites.
The R&S rubidium oscillator of YLE seem to work very well, its drift is less than 1*10exp-11 per year. Daily variation of the frequency is about 1*10exp-12 (SD) obviously depending on temperature and pressure variations. The daily frequency deviation from nominal is published in the TV-frequency information letters of VTT, which are sent free of charge to all interested.
Since september 1992 the rubidium oscillator of YLE is phaselocked to 1 kHz reference, which is obtained from VTT Loran-C receiver via telephone line.

Principle of the TV-frequency standard

The TV-frequency standard contains a 10 MHz crystal oscillator, which is phaselocked to TV-line frequency (15625 Hz). In addition it contains an intelligent, analog phase detector for external phase comparisons and buffers to deliver standard frequencies (0.1, 1, 5 and 10 MHz).
To avoid galvanic connection to TV-set we have utilized an inductive coupling from line output stage of TV-set to a small, tuned antenna coil inside the frequency standard. The drawback of this method is that computer monitors has to keep far away from the standard.
The intelligent phase detector has 10 us full scale reading and it accepts all standard frequencies (fin = 20 MHz / N, where N is an integer) without changing the full scale reading. To facilitate the documentation work, the phase detector calibrates time and phase scales automatically by producing a 1.00 us calibration mark every half an hour.

Results

Stability of rubidium oscillator

Because the reference signal (1 kHz) is supplied to YLE via normal, lossy 20 km long telephone line, it is subject to delay variation caused by ambient temperature. It seems that diurnal variation of the electrical length of the cable is around 20 us. Thus the effect of this variation to the frequency of rubidium oscillator could be a 2..3 *10exp-11.

Audit results using TV-frequency standard

In Finland there are around twenty accredited calibration laboratories (K1..K20) seven of which have frequency calibration facilities. All of them are equipped with the TV-frequency standard as a traceability source. Naturally they have frequency counters with oven controlled oscillators and often other precision oscillators too to quarantee reliable results even in case of TV-synchronization failure.

To check the measurement methods and instruments and the ability of the personnel of each calibration laboratory, audit measurements are performed annually. As an audit oscillator we have used an old HP105B crystal oscillator, which in very carefully controlled conditions gives a transfer accuracy up to 1*10exp-11.
The transfer accuracy (2*SD) is estimated from three frequency average readings, two hours before (1) and after (3) the audit trip and during (2) the whole trip calculated from the phase shift of the audit oscillator. Transportation lowers the frequency of HP105B systematically by 5*10exp-11 (df/f range +1..-17 * 10exp-11). Frequency comparison is performed using the intelligent phase detector mentioned before. Comparison time is as long as possible but at least 30 minutes. Table 1 shows the essential results, the unit of relative frequency error is 1*10exp-11.

Round	Accredited Laboratory	Laboratory error	Transfer error
	mean	0.3	4.0
	SD	2.7	2.3
1	F	0.0	4.9
3	G	0.5	1.6
2	G	-0.5	4.9
1	E	-1.0	3.3
1	A	1.3	0.9
1	D	1.5	4.3
1	G	1.8	2.5
2	E	-3.2	3.4
1	C	-3.3	9.8
3	E	-3.5	4.9
2	B	4.0	5.7
2	B	5.5	1.6

Table 1: The observed and transfer uncertainties during the last frequency audits 1990..92 (unit 1*10exp-11).

Discussion

The average error is 0.3*10exp-11 and the variation (SD) 2.7*10exp-11. As a null hypotesis the average error is zero and the result of t-test supports it.
The transfer error is as a rule higher than the observed accredited laboratory error indicating the need for a better transfer (audit) oscillator. In spite of that the typical calibration error is 1.8*10exp-11 only demonstrating the potential of TV-sync. pulse frequency transfer.

A PRECISE GPS-BASED TIME AND FREQUENCY SYSTEM

Jack Mc Nabb & Earl Fossler

TRAK Systems, Division of TRAK Microwave Corp, Tampa, Florida, USA

ABSTRACT

This paper describes an approach to implementing a compact, highly reliable and precise Master Time and Frequency subsystem usable in a variety of applications. These applications include, among others, Satellite Ground Terminals, Range Timing Stations, Communications Terminals, and Power Station Timing subsystems. All time and frequency output signals are locked to Universal Time via the GPS Satellite system. The system provides for continued output of precise signals in the event of GPS signal interruption from antenna or lead-in breakage or other causes. Cost/performance tradeoffs affecting system accuracy over the short, medium and long term are discussed. A unique approach to redundant system design provides an architecture with the reliability advantage of triple-redundant majority voting and the cost advantages of dual-redundant elements. The system can be configured to output a variety of precise time and frequency signals and the design can be tailored to output as few, or as many, types and quantities of signals as are required by the application .

I. INTRODUCTION

Overview

This paper first presents a background of precise time and frequency generation and briefly lists current-day applications. The GPS system concept is then introduced and the needs in precise frequency and timekeeping that are filled by the GPS system are discussed. Various accuracy alternatives are presented. Next, a practical Station Clock that meets these needs is described. The discussions then proceed to the implementation of a highly reliable, relatively low cost redundant time and frequency system.

Background

The art of precise frequency generation and timekeeping using electronic means has been with us for several decades, with achieved precision steadily improving as new frequency standards and synchronization methods have been developed. Since 1967, the length of a second has been defined in terms of the resonance of the cesium atom, and since the 1970's, cesium beam standards have been used in all applications requiring the most precise frequencies and precise timekeeping. Maintenance of frequency calibration and precise time setting have been a matter of some difficulty and great expense, requiring the use of standard frequency radio broadcasts, portable clocks, and extensive manpower allocation. With the advent of the GPS system, ultra-precise frequency generation and timekeeping is with us at immeasurably reduced effort and at a fraction of the costs of previous methods. *When GPS synchronization is used, all generated frequency and time outputs, worldwide, are precisely locked to a common source.*

The GPS system is making most older methods of precise frequency and time generation obsolete and has, with the help of CPU-based instruments, brought precise frequency and timekeeping to a myriad of new applications that could not have afforded the old methods. Present day applications include, among others, the following:

SOME GPS TIME & FREQUENCY APPLICATIONS

* Satellite ground terminals
* Communications terminals
* TV and radio networks
* Astronomical observatories
* Electrical power system monitoring
* Telecommunications systems
* Cellular radio
* Frequency-hopping systems
* Telemetry data acquisition stations

II. GPS CLOCK BASICS

Definitions

The word "clock" has many definitions and connotations. In this paper, we use the term to denote a self-contained electronic instrument that provides a user with precise frequency and (usually) time signals. A GPS Clock is an instrument that synchronizes to GPS to produce precise frequency and time signals. Often, when such an instrument provides all of the signals required at a given site, it is called a GPS Station Clock. Add redundancy, backup power supplies and/or multiple output signal units and we have a GPS Station Clock System (sometimes the word "system" is left off).

Clock Elements

The basic Clock has several necessary elements, as listed below. When only reference frequency outputs are required, the elements related to time accumulation, correction, and generation are not required.

* A reference frequency source
* Time correction mechanisms
* Frequency counters and time accumulators
* Output signal generators
* A synchronizing source

Reference Frequency Source

Being self-contained, the Clock has an internal frequency source (oscillator) and, since this paper addresses a GPS-based system, GPS disciplining of the oscillator is assumed. Crystal or rubidium based oscillators are normally used. A disciplined cesium frequency standard can provide a clock that eliminates almost all the time inaccuracies due to SA. With GPS disciplining, using properly-designed disciplining hardware and software, the long term accuracy, i.e.; 24 hours, of the Clock is equal to that of the core GPS system standard. This results from the fact that long-term phase-locked disciplining locks the local oscillator to the GPS standard over the long term and the fact that the core GPS standard does not vary. As long as oscillator-inherent (free running) drift is not excessive, loss of GPS disciplining for periods of even several hours does not degrade long term accuracy. The length of time that GPS disciplining can be cut off is proportional to oscillator stability.

Now, "long term" must be defined. With SA active, the resulting signal distortions last from a few seconds to a few hours. The most visible indication of these distortions is a plot of the 1 PPS output of a GPS receiver against a local cesium standard over a period of several days. Experience has indicated that 1 PPS phase returns to "neutral" at intervals not exceeding two-to-three hours and that the summation of deviations integrates to zero over a 16 to 24 hour period. The conclusion is that the accuracy will be on the order of one part in 10^{12} to 10^{13} when measured over a period of 10 to 20 hours.

Next we come to the situation of expected intermediate term (one second to several minutes) frequency accuracy while tracking satellites. The figure here is determined by the oscillator characteristics. Typically, in this application, voltage control is used to discipline the oscillator, and the disciplining process is one of continually varying the oscillator voltage to track the inherent drifts (primarily aging and temperature stability). The poorer the oscillator stability, the poorer the intermediate term accuracy will be. Ovenized crystal oscillators typically have accuracies in the range of 1×10^{-9} thru 5×10^{-11}. Rubidium oscillator accuracy is approximately 1×10^{-11}. Table 1 lists the characteristics for a representative set of oscillators.

Short term stability is totally determined by the oscillator characteristics. Many different suitable oscillators for this application with a great variation of characteristics are available. Data on typical oscillators are given in Table 1.

Knowing the coasting drift is important in some applications. Coasting drift is defined as the total 1 PPS phase drift at the Clock output during those periods when the oscillator is not being actively disciplined by GPS. Since there is worldwide coverage by at least one satellite 24 hours a day, this situation arises only under unusual circumstances, but it is useful information for those cases. Two possible situations where satellite data may not be available are (1) the temporary disconnection of an antenna and (2) the placement of the antenna at a location where all satellites are obscured for periods of time. The coasting drift for various types of oscillators is given in Table 1.

Table 1 gives data for oscillators available from TRAK Systems Division of TRAK Microwave Corporation. If an instrument has accumulated an offset of less than 10 microseconds while drifting, when a satellite is reacquired corrections are applied smoothly to return the phase difference to zero and to reestablish long term accuracy. Conversely, if, at the time of satellite reacquisition, accumulated offset is greater than 10 microseconds, all counters are reset to zero and a new baseline for long term accuracy measurements is established. As can be seen from Table 1, even the least-precise of the oscillators

Table 1. GPS Clock Oscillator Performance Comparisons

CHARACTERISTICS	STD CRYSTAL	LO DRIFT	LO DRIFT/ REDUCED NOISE	RUBIDIUM
Accuracy while tracking	1 x 10-9	1 x 10-10	1 x 10-10	1 x 10-11
Typical coasting drift (first 2 hours)	1.5 us/hr	250 ns/hr	250 ns/hr	40 ns/hr
100 sec Allan variance	5 x 10-11	1 x 10-12	1 x 10-12	1 x 10-11
10 sec Allan variance	1 x 10-10	1 x 10-12	1 x 10-12	4 x 10-11
1 sec Allan variance	N/C	5 x 10-12	5 x 10-12	1 x 10-10
Phase noise @ 1 Hz offset	-80 dBc	-80 dBc	-100 dBc	-70 dBc
Phase noise @ 10 Hz offset	-85 dBc	-85 dBc	-120 dBc	-90 dBc
Phase noise @ 100 Hz offset	-100 dBc	-100 dBc	-130 dBc	-110 dBc
Phase noise @ 1 KHz offset	-105 dBc	-105 dBc	-145 dBc	-130 dbc
Phase noise @ 10 KHz offset	-115 dBc	-115 dBc	-150 dBc	-130 dBc
Harmonic distortion	-30 dBc	-30 dBc	-30 dBc	-40 dBc
Non-harmonic distortion	-40 dBc	-50 dBc	-50 dBc	-65 dBc

TYPICAL COASTING DRIFT - Oscillator drift when no satellites are in view following two to four days stabilization. Since there is currently full coverage worldwide, this information covers only very unusual conditions.

CHARACTERISTICS	STD CRYSTAL	LO DRIFT	LO DRIFT/REDUCED NOISE	RUBIDIUM
1 hour	1.5 ms	250 ns	250 ns	40 ns
2 hours	3 ms	500 ns	500 ns	80 ns
4 hours	7 ms	1.5 ms	1.5 ms	200 ns
8 hours	20 ms	3 ms	3 ms	400 ns
16 hours	50 ms	8 ms	8 ms	800 ns

offered can withstand greater than four hours without disciplining and still maintain integrity of long term accuracy.

Frequency Counters and Time Accumulators
Clocks of the type described herein contain, as a minimum, the frequency counters to produce a 1 PPS output. Producing a 1 PPS by countdown from the oscillator is required to implement the disciplining process. Figure 1 is a simplified block diagram showing the disciplining circuits.

Clocks producing time of year outputs also contain time accumulators for seconds through years. Modern time Clocks use software accumulators.

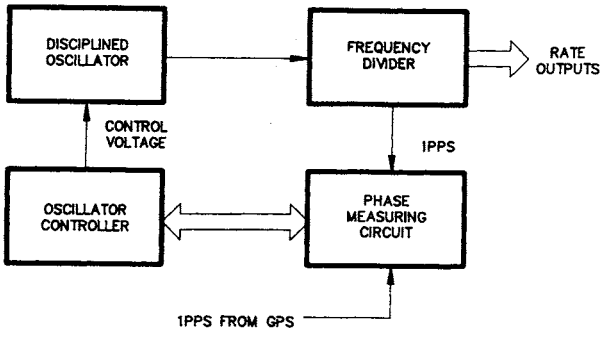


Figure 1. Block Diagram of Disciplining Circuits

Synchronizing Source
As described above, the internal oscillator and 1 PPS output are synchronized to the GPS received 1 PPS signal. The GPS receiver also outputs the data to initially synchronize the time accumulators. It is important to note that, once synchronized, the Clock and its time accumulators can run for periods of many hours without resynchronization. In other words, the Clock does not merely process received GPS data and output it -- the Clock uses the GPS received signals for synchronization and, without loss of long term accuracy, bridges periods of no GPS input.

Time Correction Mechanisms
The various time correction factors are Leap Year, Leap Second, Time Zone, and Summer Time. Leap Year for the lifetime of the equipment is preprogrammed in ROM. Leap Second is automatically entered at the precise rollover time, and provisions are made to enter the other required corrections.

Output Signal Generators
A Clock containing all of the elements described thus far can output a variety of rates and encoded time signals merely by adding circuits that use the basic clock signals to produce the desired outputs. A Clock designed only as a frequency source can output a variety of optional digital and sinewave frequencies. The next section of this paper illustrates a typical GPS Station Clock containing a great variety of outputs.

III. HARDWARE IMPLEMENTATION

A GPS Station Clock
A functional block diagram for a typical production GPS Station Clock incorporating all of the features outlined in the previous paragraphs is illustrated in Figure 2.

On the front panel of a typical GPS Station Clock, keypad and alphanumeric display implement setup of the unit and status display. All desired setups can also be entered by remote terminal or computer. The setup system provides for entering such data as time zone and other optional data differing from default values. Status screens include satellite status, current modes, time and date, latest navigation solution, GPS-lock status, and many others.

Referring to Figure 2, the instrument contains a six-parallel-channel GPS Receiver for frequency disciplining and time presetting, a Disciplined Frequency Subsystem including stable oscillator (crystal or rubidium), a System Processor including receiver interface, executive controller, frequency and time countdown circuits, computer I/O, and other functions. A great variety of signals, including digital rates, sinewave rates, serial time codes, and parallel time codes can be generated and supplied as outputs.

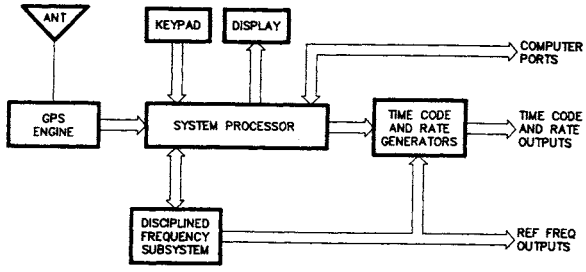


Figure 2. Block Diagram, GPS Station Clock

GPS Antenna System
Three primary choices of antenna system exist: (1) passive, (2) active with preamp at antenna and (optionally) in line, and (3) with downconverter at antenna. These are described in the following paragraphs. For lower-accuracy applications, it is possible to mount the GPS receiver in the antenna, but this approach is outside the scope of this paper.

PASSIVE: The GPS frequency of 1.5 GHZ precludes using a passive antenna with more than just a few feet of lead-in cable, and this type of system is restricted to field and vehicular applications.

ACTIVE WITH PREAMP: This is the most common type of system. It is not uncommon to multiplex a dc voltage up the antenna lead-in conductor to power preamps in the line and at the antenna. A well-designed system can use inexpensive coaxial cable to lengths of 200 feet with a preamp in the antenna and to 400 feet with the addition of a line amplifier at midpoint. This distance is more than adequate for most applications. A view of two typical antennas is shown in Figure 4. Lead-in distance can be extended to over 1000 feet using low-loss (expensive and cumbersome) cable and/or additional line amplifiers, but dynamic range decreases with added amplifiers.

DOWNCONVERTER: Lead-in distance can be extended to a few thousand feet if a downconverter is used at the antenna. Typically, the 1.5 GHZ signal is converted to the range of 10-to-80 MHz to greatly reduce cable losses. An upconverter can then be used at the receiver if it has a standard 1.5 GHz front end, or a special receiver made to accept the downconverted frequency can be used. Some systems send an LO signal up to the antenna on a separate cable.

Design and Performance of the NIST Satellite Two-Way Time Transfer Modem

D. Davis, D. Howe, M. Lombardi, and M. Weiss

Time and Frequency Division
National Institute of Standards and Technology
325 Broadway
Boulder, Colorado 80303

This paper describes the design and performance of the NIST-designed direct sequence spread spectrum (DSSS) modem. The modem is intended to be used in two-way time synchronization between two locations via a geostationary satellite operating typically at Ku-band. The modem has the capability for variable PN sequence chip rate from 102.3 Kchips/s to 2.5 Mchips/s. Two classes of codes can be selected. (1) Gold code of length 1023 bits, and (2) truncated maximal length sequence (MLS) code of length 10,000 bits (Mitrex-compatible).

The NIST Modem is controlled by a 486 PC type computer operating at 33 MHz. The software, designed specifically for the modem, is written in assembly language and C, and essentially replaces the DOS operating system. The modem software allows four interrupt driven programs to run simultaneously, as well as numerous interruptible ones. The interrupt driven programs are the transmit or TX program, the receive or RX program, the counter program, and a spare. The interruptible programs include a monitor demon, a menu system, programs to store and reduce the measurements, and a disk drive demon.

The transmit and receive signals are at 70 MHz. Enough frequency agility is built into the modem to allow its use with low-cost VSAT earth terminals. Acquisition of lock is possible with C/N_0 as low as 50 dB-Hz. Various data will be presented which indicate that accuracy of time synchronization approaching 1 ns will be possible with suitable earth terminal calibrations.

TEMPORAL AND SPATIAL TEMPERATURE VARIATIONS IN THE SURROUNDINGS OF A CLOCK

Jari Mannermaa (1), Kalevi Kalliomäki (1,2), Tapio Mansten (2)

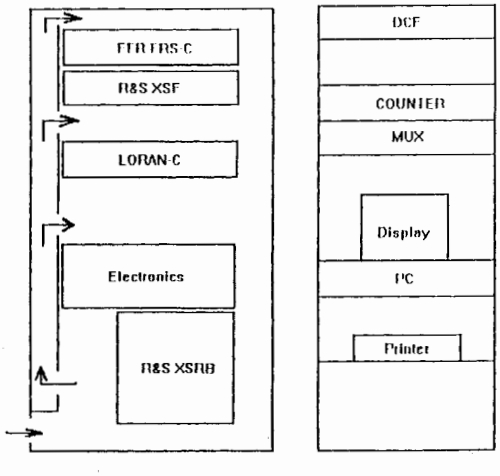
(1) University of Oulu, Oulu, Finland
(2) Technical Research Centre of Finland (VTT)

Abstract

The temperature variations inside and outside of a clock and a clock cabinet have been studied. The problem was to determine suitable measurement points in order to ascertain the temperature distribution around a clock.
The results showed that the clock significantly perturbed the temperature distribution in its vicinity. The sphere of influences reached up to 50 cm in the upward direction and about 30 cm in the horizontal direction inside a large laboratory.
The local temperature changes in the cabinet varied from 0 °C to 1 °C within a few minutes, even under stabilised conditions.

Introduction

An often encountered problem with frequency standards is their sensitivity to fluctuations of ambient factors. The most significant of them is temperature and its non-homogeneous distribution in the neighbourhood of devices. Energy supplies need to be used to operate devices. When clocks are operated with the help of electric power, energy is released as heat. There are at least two important compartments, the rack where the clocks are housed and the cases of the clocks where there can be a temperature gradient formed.
Heat is removed by three mechanisms from the clock into the surroudings, namely conduction, convection and radiation. In this instance, convection, especially forced, is the dominating mechanism and thus the others have been neclected. One of the important quantities which influences the temperature behaviour is the thickness of the boundary layer. The objects that were measured in this instance were a single clock placed on a table in the middle of a large laboratory (Case A1 and B1) and several clocks placed inside a stabilised clock cabinet (Case A2 and B2). The flow of air was either caused by the natural heat of the clocks or it was enhanced by a fan which mixed the air and tried to equalise the temperature.
The clock cabinets are presented below:



The rack on the left is insulated by thermal shields. The arrow denotes the directions of the flow of air. There were two fans in the rack, one which took in cool air from the room while the other circulated it through the air channel. The rack on the right did not have thermal shields.

There were two temperature sensor distributions used to study the profile in the clock cabinet. The first (Case A2):

Sensor	Explanation	Sensor	Explanation
0	Room, on the ceiling, above the racks	8	Between racks level with MUX
1	In front of EFR, outside the rack	9	On the bottom of Loran-C, in the middle
2	In front of DCF, outside the rack	10	On the bottom of the rack, in the middle
3	On the floor, near right rack	11	In the air channel, level with Loran-C
4	On the right side, level with R&S XSFrack,	12	In the air channel of the near the top hole
5	On top of R&S XSRB, in the middle	13	On the front panel of Loran-C, in the middle
6	On top of Loran-C, in the middle	14	Inside Loran-C, beside its crystal oven
7	In the middle of the rack	15	On the left side of the rack, outside

The second configuration was (Case B2)

Sensor	Explanation	Sensor	Explanation
0	On top of R&S XSRB, in the middle	8	On the bottom of the counter, in the middle
1	In front of R&S XSRB, in the middle	9	On top of DCF, in the middle
2	On the left of R&S XSRB, in the middle	10	Behind R&S XSRB, in the middle
3	On the right of R&S XSRB, in the middle	11	In front of R&S XSRB, 5 cm away
4	On the floor under R&S XSRB, in the middle	12	On the left of R&S XSRB, 10 cm away
5	Under the printer	13	On the right side of R&S XSRB, 10 cm away
6	Under the PC	14	Behind R&S XSRB, 10 cm away
7	On the bottom of MUX, in the middle	15	5 cm above the top of R&S XSRB, in the middle

The total power dissipated by the devices in the closed rack was about 100 W.
The measurements were conducted using a computer which had either 12 or 16 analog channels. The sampling intervals were half, one or five minutes. Static and dynamic conditions were applied by varying the temperature inside the laboratory or the cabinet both stepwise and sinusoidally.

Method

Initially, the theoretical calculations were carried out. In convection, heat is transferred from the sample into the moving gas (like air) or liquid.
The dissipated power during convection is

$$P = kA(T_s - T_{inf})$$

Where

P is the dissipated power,
k is the heat transfer coefficient,
A is the total area of the rack,
T_s is the temperature of the surface and
T_{inf} is the temperature remote from the surface

The value of k can approximated from the formulae for vertical homoisothermic surfaces in the free convection

Laminary flow:

$$k = 1.3(dT/L)^{0.25}$$

Turbulent flow:

$$k = 1.3(dT)^{0.33}$$

and for the forced convection

Laminary flow:

$$k = 4\sqrt{v/L}$$

Turbulent flow:

$$k = 5v^{0.8}L^{-0.2}$$

where

the dimension of k is $[W/(m^2\ ^oC)]$,
dT is temperature difference,
L is the length of the surface parallel to the flow,
v is the flow speed.

For air, k is less than 10 $W/(m^2\ ^oC)$ and in free convection and greater than 10 $W/(m^2\ ^oC)$ in forced convection.

Near the surface is the boundary layer, which can approximated by

$$dl = 0.1(x/dT)^{0.25}$$

where

x is the distance from the initial point of flow.

The temperature gradient of the lower and upper levels can be approximated using the formula

$$T_u = T_l + P/(1200q) \quad [\ m^3/s]$$

where

T_u is the temperature at the top of the rack,
T_l is the temperature at the bottom of the rack,
P is the total dissipated power of the devices and
q is the performance of the fan.

In the case of the single clock, two sensor distribution patterns were used. In the first, there were only 12 sensors 4 of them being placed on the axis which penetrated the midpoint of the front and rear panel horizontally, and 4 on the respective axis of side panels, while 4 were on the axis from bottom upwards (Case A1). Subsequently 4 sensors were added to the horizontal axes, one at each of the four ends (Case B1). The sensors on the horizontal axes were classified as follows:
S1 = 4 sensors that were on the surface panels,
S2 = 4 sensors at a distance of 5 cm and

S3 = 4 sensors at a distance of 15 cm.
S4 denoted the sensors which were fixed to the bottom surface, fixed to the top, and the two others which were fixed 15 cm and 200 cm above the top.
Three experiments were conducted in both cases.
In Case A1 the following measurements were conducted

A11:
The temperature distribution was measured under stable conditions when the clock was operated and with S1+S2+S4.

A12:
As with A11 but with sensors S1+S3+S4.

A13:
As with A11 and A12 but with sensors S1+S2+S3+S4.

These measurements revealed not only the rough distribution of the temperature but also the repeatability of the experiments.
In Case B1 the following measurements were conducted

B11:
The clock was disturbed by periodic temperature steps, each lasting 20 minutes after an interval of 2 hours.

B12:
B11 was repeated.

B13:
As in B12 but with only one step which lasted 1.5 hours.

In the other experiments, the temperature changes of the two racks in the laboratory were studied. The Studies included three arrengements in the Case A2 and five in the Case B2:

A21:
The closed rack was disturbed by stopping the circulating fan for 3 hours.
A22:

The front door of the closed rack was opened periodically for half an hour after every 2 hours.

A23:
Periodic temperature steps of 1 ^{o}C disturbed the whole laboratory for half an hour after every 2 hours.

B21:
The circulating fan was stopped for 2 hours.

B22:
The thermally stabilised cabinet was heated with the maximum power for a long time and then the heater and the both fans were stopped for several hours, until finally the fans were switched on again.

B23:
The front door was opened and the fans were stopped for 13 hours.

B24:
The temperature step of 1 ^{o}C disturbed the system for 3 hours.

B25:
A pulse train of the temperature steps of 2 ^{o}C was applied to the system four times, each state lasting for 12 hours.

The temperature difference between the room and the closed rack was about 5 ^{o}C. The amplitudes of the fluctuations in temperature at different points around the clocks were compared in order to determine the temperature coefficients correctly. The effect of air flow on the volume of the relevant space surrounding the clocks was approximated.

Exponential fits were applied to determine the time constants of the heating and cooling systems.

The experiments were divided into two series. Firstly, a single clock on the table of the large laboratory room was studied. Secondly, the study concerned the whole clock cabinet.

Results

Based on theoretical considerations, the results can be explained as follows. The natural cooling ability of a device case is about 7 to 10 $W/(m^2\ ^oC)$. The heat transfer coefficient, k, in this was about 1.3, the area A of the surface of the rack was about 2 m^2 and the dissipated power about 120 W. The vertical temperature gradient in the rack was about 1.5 $^oC/m$.

The temperature gradients were significant even inside carefully stabilised cabinets, being of the order of 1...2 $^oC/m$.

The amplitudes of the temperature fluctuation varied in a ratio of more than 1:2 depending on the point at which it was measured. The temperature was also found to vary by as much as 0.2 ^oC in a half hour just beside the surface of the crystal oven inside the clock case.

Experiments A11 & A12 & A13 & B13:
The results from the single clock point out that it disturbed the other clocks if they were situated closer than 30 cm in the horizontal direction and 50 cm in the vertical direction above the clock. Moreover, asymmetry due to the power supply was clearly observed e.g. the temperature in the rear was 7 ^oC higher than in the front panel, 2 ^oC when distance was 5 cm and 0.3 ^oC when distance was 15 cm. If there was any convection, the asymmetry further emphasised the temperature difference.

Experiments B11 & B12:
In the cyclic experiment in which temperature steps of 1 ^oC was applied, the change of temperature was about 0.3 ^oC on the surface of the clock, 0.6 ^oC at a distance of 5 cm and 0.9 ^oC at 15 cm. A high variation of about 4 ^oC was observed 15 cm above the clock. Asymmetry remained and preserved the relative differences between the measuring points.
The results of the experiments from the clock cabinets were found to be as follows:

Case A2:

A21:
When the circulating fan was stopped, the temperature decreased 1.8 to 3.1 ^oC inside the cabinet, especially in the middle. The influence of the upper heating (EFR and R&S XSF) was probably eliminated by stopping the circulation. The heat controller was at its maximum. The changes measured inside the cabinet were of the order 1:2 and thus the temperature coefficient would have the similar error.

A22:
When the rack door was opened periodically, the temperature in the top decreased 2 ^oC but at the bottom it remained the same. The largest decrease was in the middle of the rack. The time constant of the heating and cooling systems were 40 and 17 minutes, respectively.

A23:
The temperature step in the room air was damped by the ratio 1:10 when compared to the changes in the closed rack.

Case B2:

B21:
The temperature inside the closed rack rose about 2 ^{o}C and decreased by about 2 ^{o}C outside when the circulating fan was stopped. The surface temperature of the rubidium changed by between about 0 to 0.4 ^{o}C during the test.

B22:
The results pointed out that the room temperature changed by less than 0.1 ^{o}C during the test. The heater increased the temperature inside the rack by 0.8 ^{o}C and the fan decreased it by 2.1 ^{o}C. The control range of the temperature controller was thus about 2.9 ^{o}C.

B23:
When the fans were stopped and the door was opened, the temperature increased by 1.1 ^{o}C. The highest change was at the top of the XSRB.

B24 & B25:
A23 was repeated with longer term disturbances because more stable surroudings were achieved between the temperature steps. These experiments ascertained that both temporal and spatial thermal asymmetry existed in the stabilised rack.

Conclusions

The temperature distribution in the cabinet was asymmetric, both horizontally and vertically, despite effective mixing of air and the symmetric placement of the clocks. This was due to the location of the power supplies and crystal ovens of the clocks. Moreover, the maximum theoretical temperature gradient should have been about 1.5 ^{o}C/m in the closed rack, but between the surfaces of the clocks there could have been differences of more than 5 ^{o}C. The determination of the time constants of the heating and cooling systems required the use of a multiexponential model, at least 2-exponential fitting.

It was surprising that the temperature increased in the closed rack when the front door was opened. The possible reasons for this phenomenon were the change in the air flow pattern and the overreaction of the temperature controller. Moreover, the results pointed out that the circulation created by fans in the rack is not effective because of the free convection to the opposite direction.

The study revealed that the temperature coefficients were more or less academic concept because of the ambiguous definition of the surroundings. The "surroudings" were not constant or stable but the function of time, location and the temporal and spatial temperature distribution. Thus it is not possible to define the "surroundings" in a tightly packed cabinet.

Manufacturers often describe the behaviour of their devices with specifications and statistics. One of those most frequently used is the temperature coefficient. In general, the surroundings of a clock cannot be defined unambiguously. This means that, for example, the temperature coefficient loses its meaning, in practise, even inside a stabilised oven.

PRECISE FREQUENCY DISTRIBUTION USING FIBER OPTICS

Dr. Richard L. Sydnor and Dr. Malcolm Calhoun
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California 91109 U.S.A.

This paper descibes several frequency distribution systems developed at the Jet Propulsion Laboratory. They cover distances from a few meters to several kilometers and performance levels ranging from Cæsium performance to Hydrogen Maser performance amd beyond.

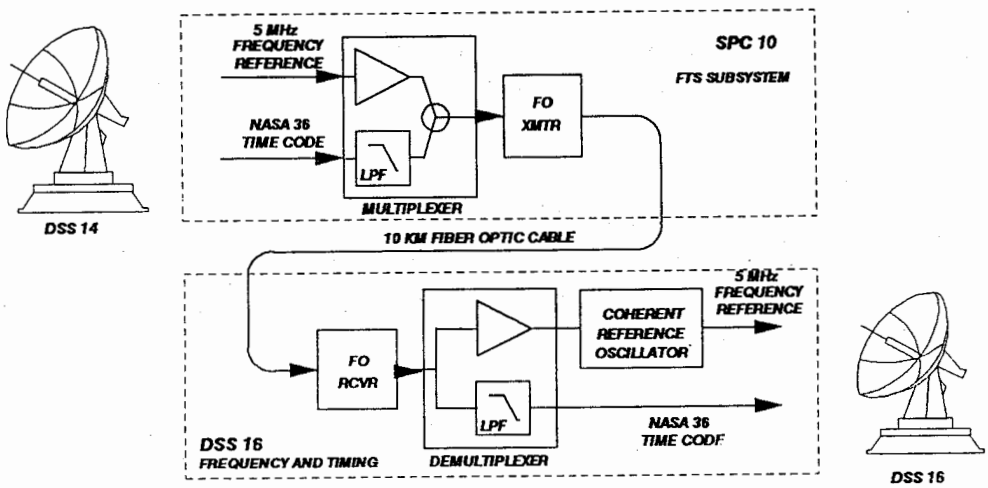
This work represents one phase of research carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract sponsored by the National Aeronautics and Space Administration.

I. Cesium Quality, Long Distance Distribution

- Figure 1A shows a block diagram of the system used in order to achieve stability adequate for support of cesium quality frequency and time code distribution over a 10 kilometer path.
- A standard single mode fiber was used, buried 1.6 meters in the earth.
- The multiplexer was a commercial unit, modified to improve its stability.
- The receiver/transmitter combination were obtained from the Grass Valley Group.
- Figure IB shows the stability obtained with this link, better than the Cesium which it replaced.
- Figure IC shows the Allan Deviation for the link, also better than that of the Cesium standard.
- Figure ID shows the system for measuring the Phase Noise of the system.

JPL

FIGURE 1A

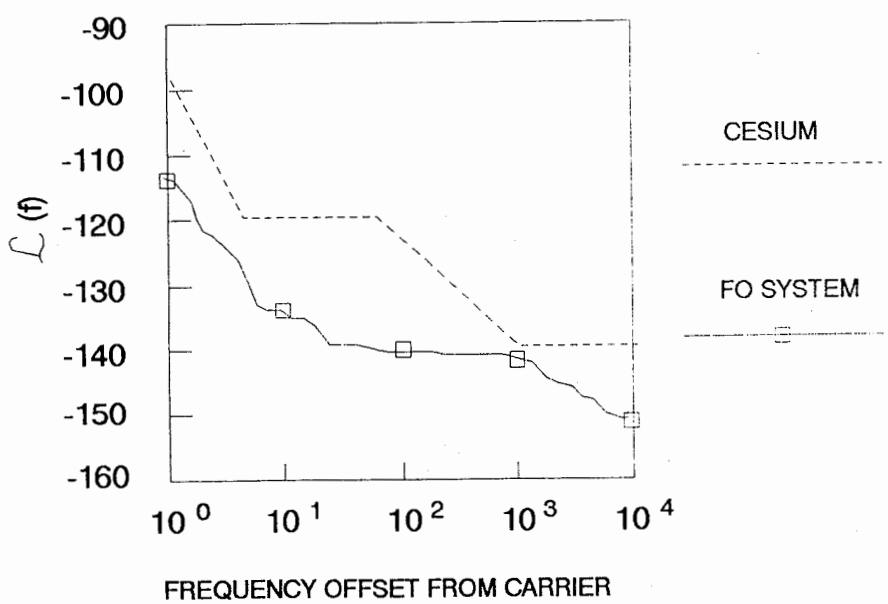


FIBER OPTIC REFERENCE FREQUENCY AND TIMING DISTRIBUTION ASSEMBLY SPC 10 TO DSS 16

JPL

PHASE NOISE OF THE FIBER OPTIC REFERENCE FREQUENCY AND TIMING DISTRIBUTION ASSEMBLY

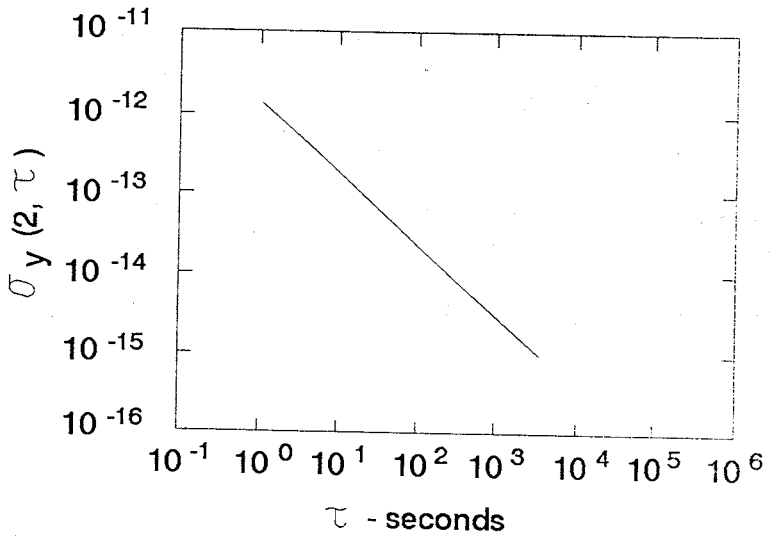
FIGURE 1B



JPL

ALLAN DEVIATION OF THE REFERENCE FREQUENCY AND TIMING DISTRIBUTION SYSTEM

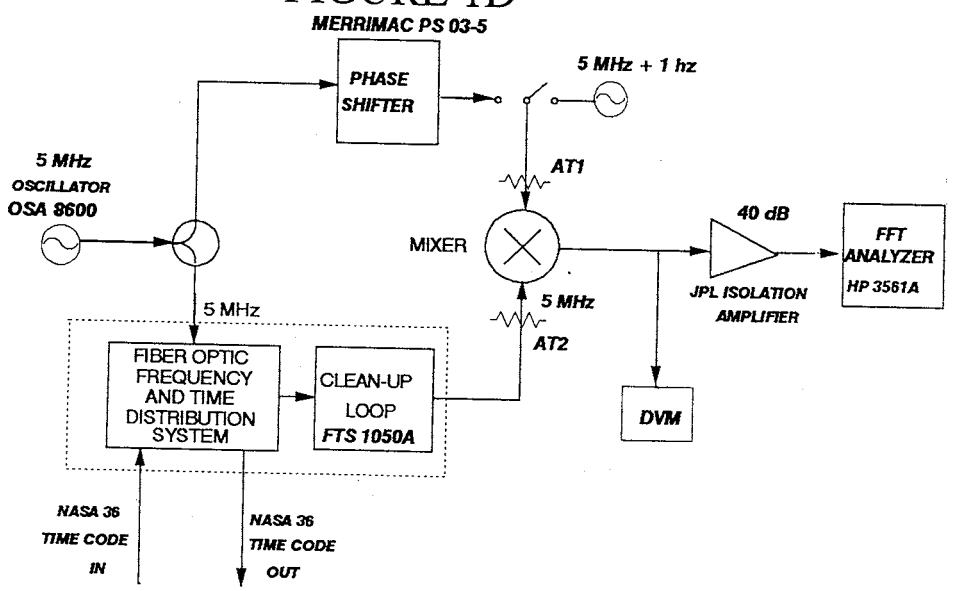
FIGURE 1C



JPL

PHASE NOISE TEST SETUP

FIGURE 1D



II. High Quality Medium Distance Frequency Distribution

- Figure 2A is a block diagram of this system. This is a Low-Cost System which distributes Hydrogen Maser Signals within the Jet Propulsion Laboratory
- Low-Temperature-Coefficient-of Delay fiber was used because the fiber was installed in existing duct-work which was not well insulated from environmental changes.
- Grass Valley Group transmitter and receivers were used.
- Figure 2B show the phase noise of the system. There is approximately 7 dB of degradation of the Hydrogen maser signal due to the electronics.
- Figure 2C shows the Allan deviation and the various residuals for the system.

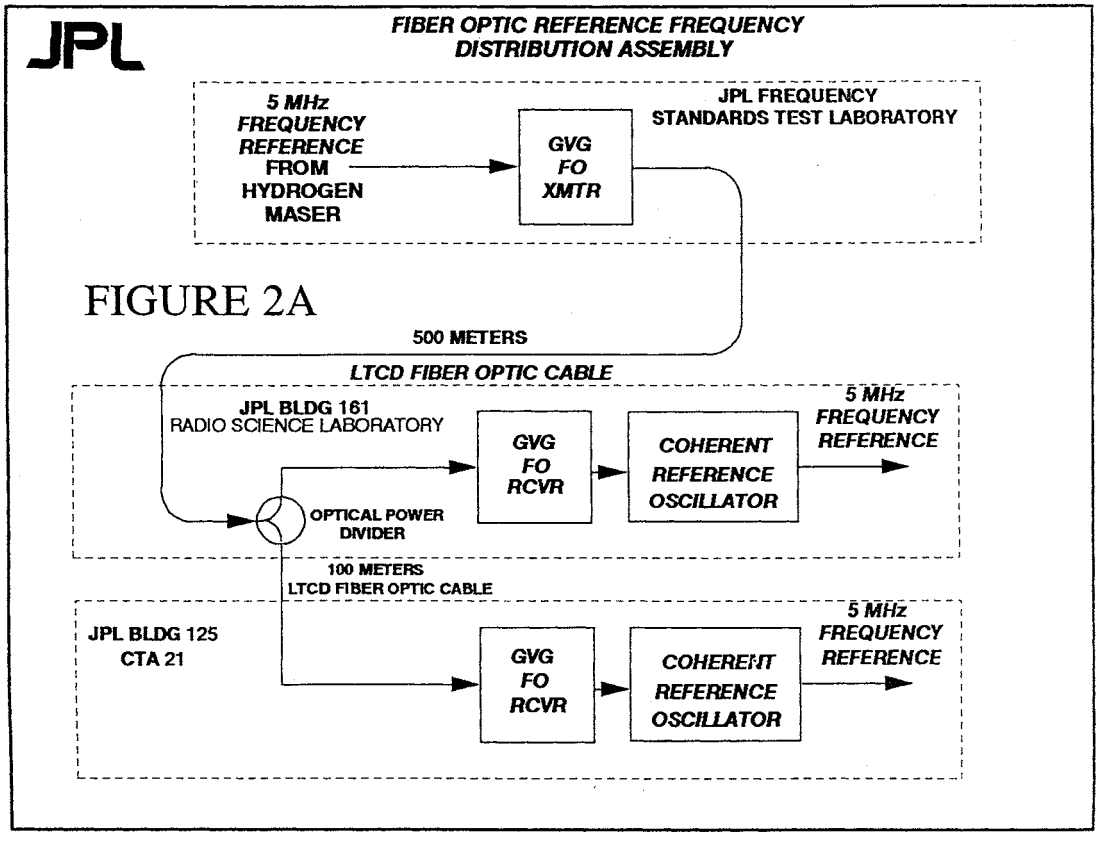
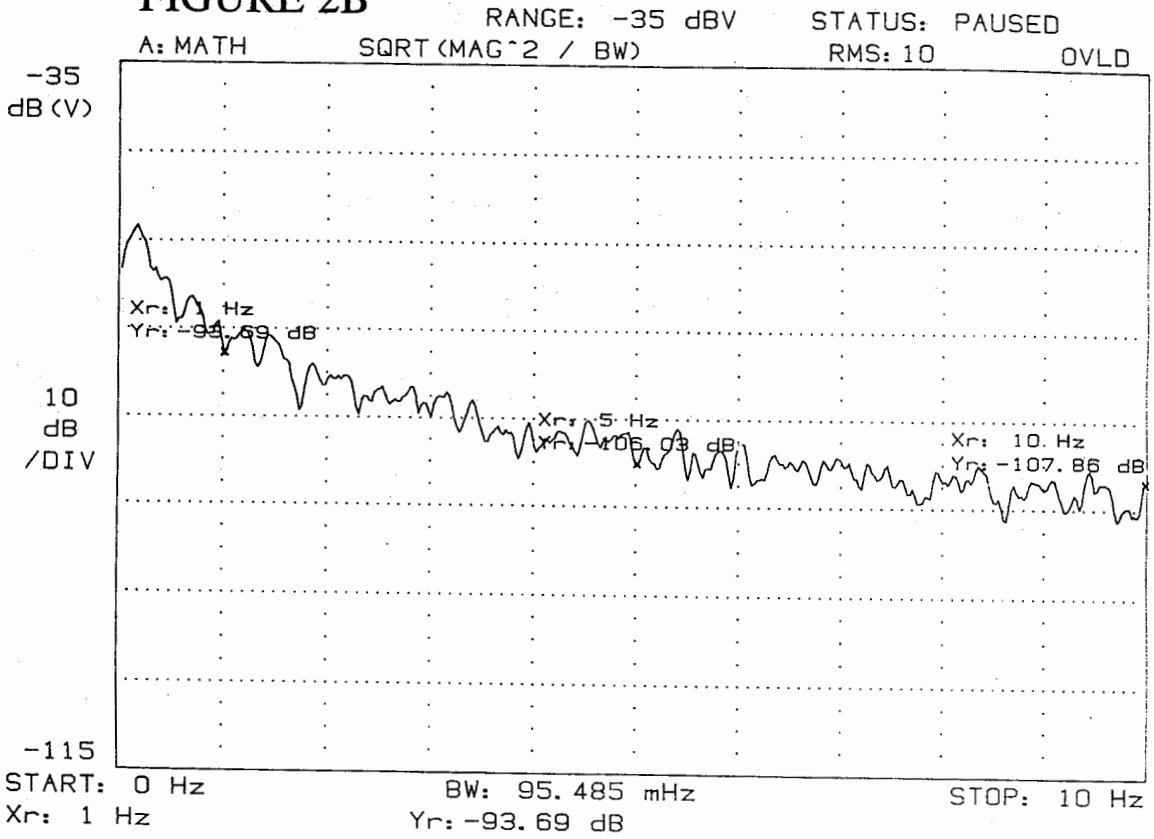


FIGURE 2A

FIGURE 2B



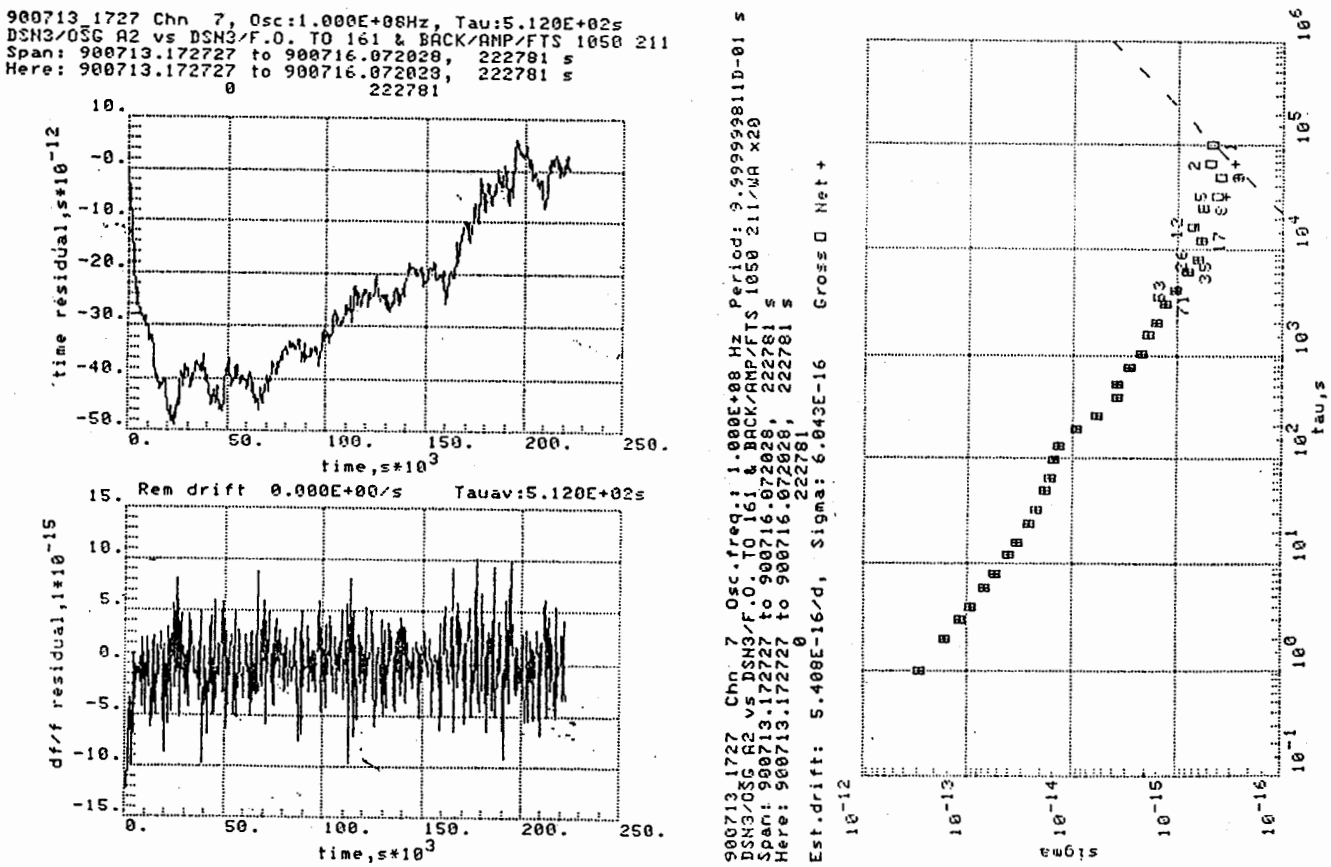


FIGURE 2C

III. Field Station Distribution of Hydrogen Maser Signals

- **Figure 3A shows the block diagram of this system. The length of the path is approximately 150 meters. This system includes complete monitoring of its operational health.**
- **Figure 3B shows the performance of this system versus the Radio Science specifications.**
- **Figure 3C shows the Allan deviation of the system. This measurement used a dual link in order to bring the signal back down the antenna, so the actual performance is probably 3 dB better than shown.**
- **Figures 3D and 3E show the phase noise performance of the system. It is approximately 20 dB better than the maser, so it has a negligible effect on the stability of the signal.**

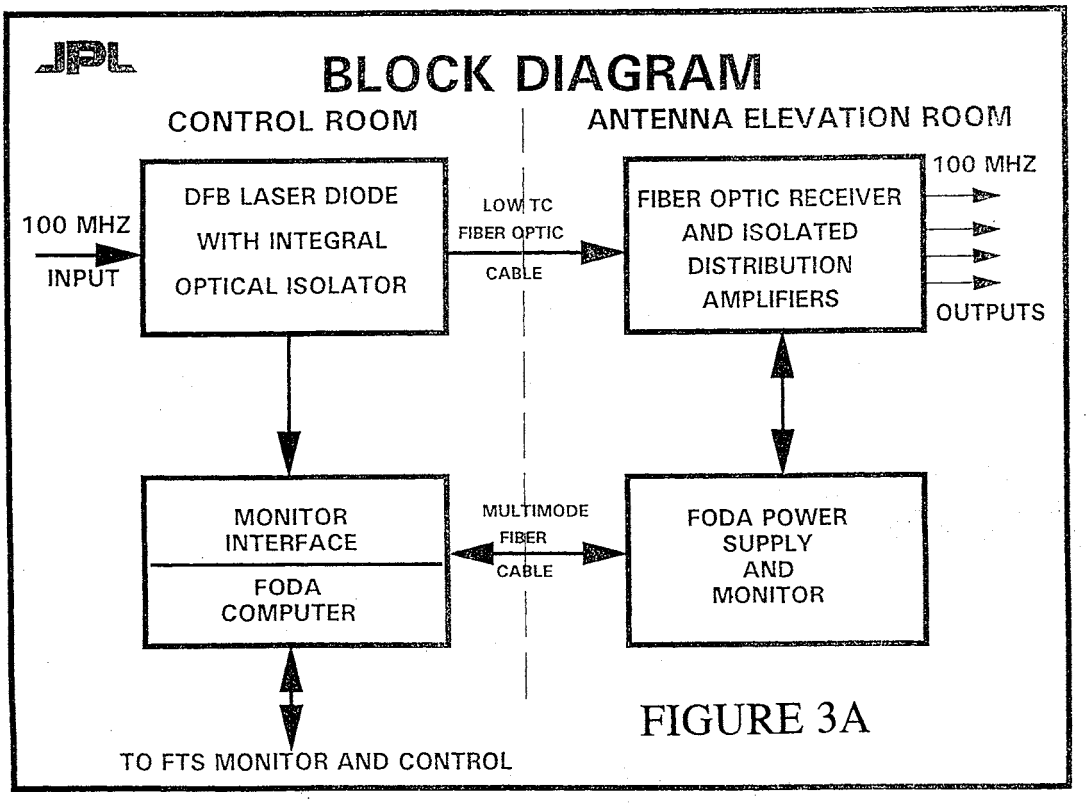


FIGURE 3A

PERFORMANCE vs SPECIFICATIONS

L(f) dBc AT 100 MHz

OFFSET FREQUENCY, Hz	FTS SPEC.	LINK PERFORMANCE
1	-92	-119
10	-103	-128
100	-115	-134
1000	-115	-140
10,000	-115	-140

SPURIOUS OUTPUTS

NONE MEASUREABLE ABOVE NOISE LEVEL IN 1 Hz BANDWIDTH

ALLAN DEVIATION

TAU, SECONDS	FTS SPEC.	LINK PERFORMANCE
1	2E-13	2E-14
10	4E-14	4E-15
100	----	1E-15
1000	1.5E-15	5E-16
3000	1.5E-15	3E-16
10,000	----	2E-16

FIGURE 3B

FIGURE 3C

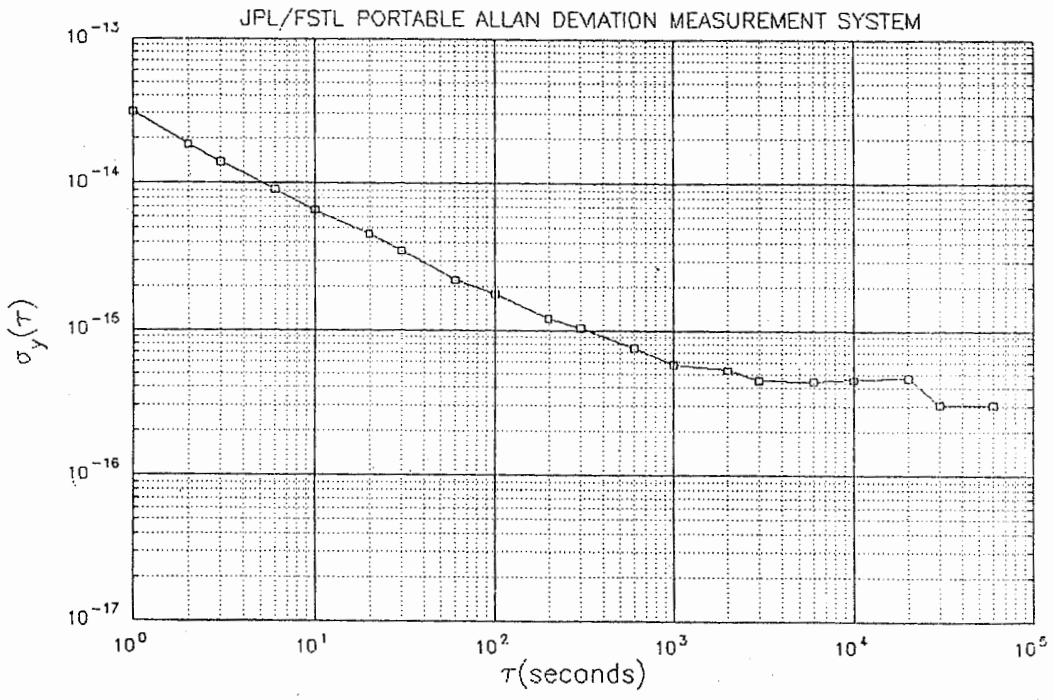


FIGURE 3D

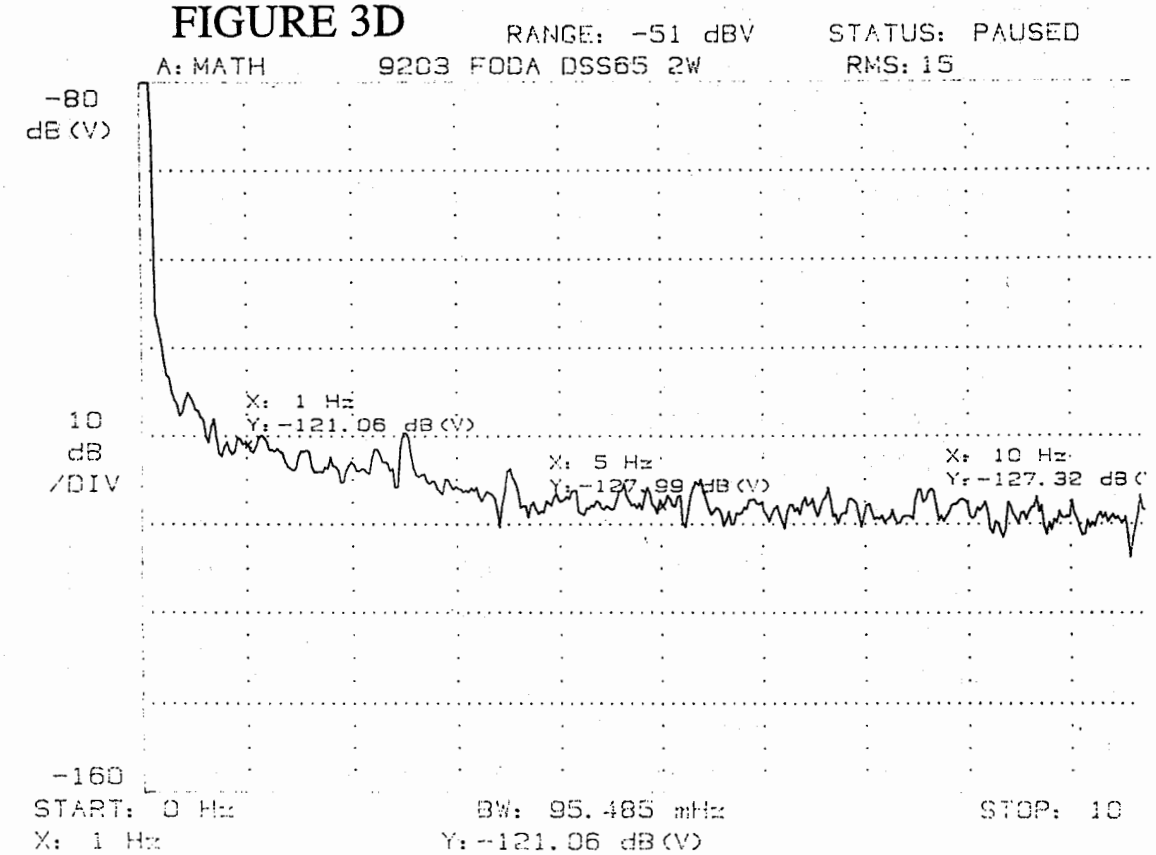
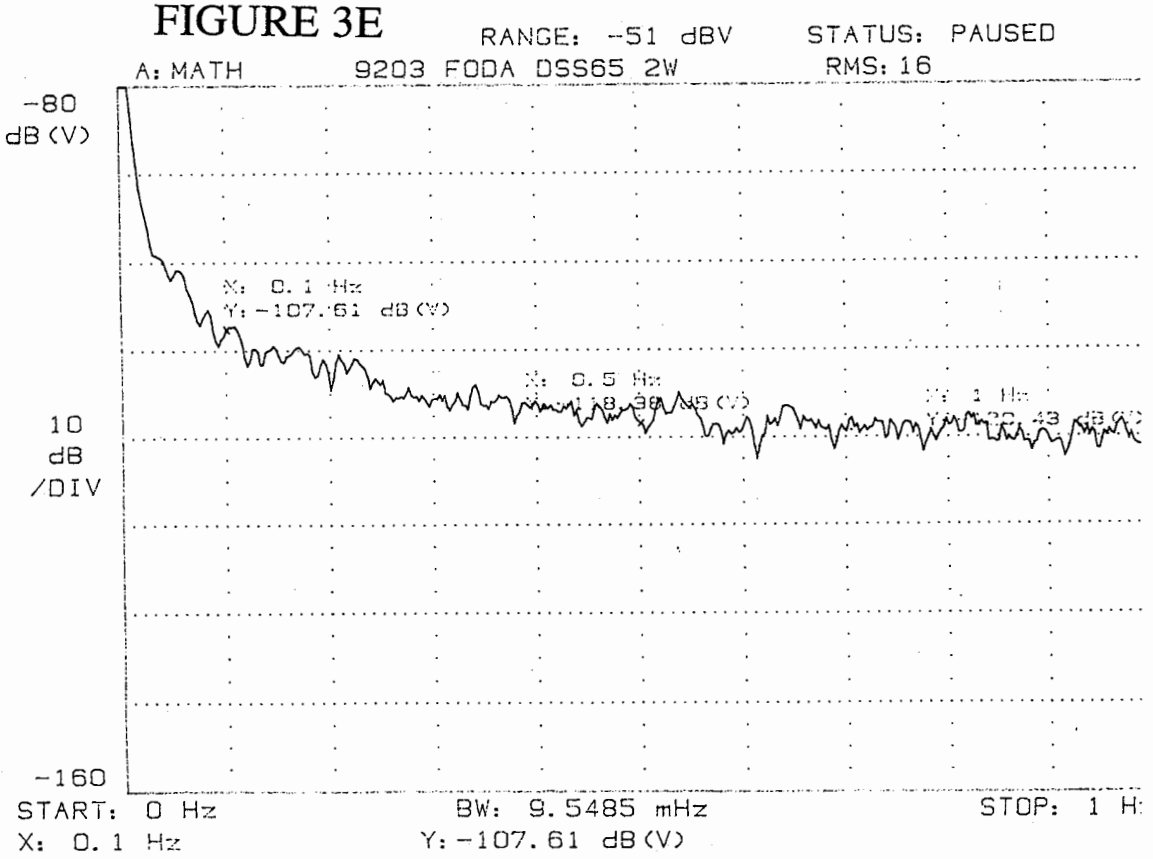


FIGURE 3E



IV. Summary

- **Figure 4 is a summary of the performance of the various links and the specifications for them.**
- **JPL is developing an active stabilizer which will reduce the contributions of the distribution system to well below 10^{-17} for future frequency standards and Radio Science requirements.**

FIGURE 4

FIBER OPTIC
REFERENCE FREQUENCY DISTRIBUTION

	824-13 DSCC FTS RQMNT	EXISTING LINK SPC 10/DSS 15 5MHz w/code	EXISTING LINK FSL to TDL 5MHz w/code	EXISTING LINK SPC 10/DSS 15 100MHz/no code
τ(seconds)	$\sigma(\tau)$	$\sigma(\tau)$	$\sigma(\tau)$	$\sigma(\tau)$
1	2E-13	1.5E-12	3E-13	1.5E-14
10	4E-14	1.5E-13	4E-14	2E-15
100		2E-14	1.5E-14	7E-16
1000	1.5E-15	2E-15	2E-15	4E-16
3600	1.5E-15	2E-15	1E-15	2E-16

f_f(Hz)	L(f), dBc		L(f),dBc		L(f), dBc		L(f), dBc	
	5 MHz	100MHz	5 MHz	100MHz	5MHz	100MHz	5 MHz	100MHz
1	-118	-92	-113	-87	-117	-91	-134	-108
10	-129	-103	-133	-107	-136	-110	-144	-118
100	-141	-115	-141	-115	-144	-118	-150	-124
1000	-141	-115	-141	-115	-146	-120	-152	-126
10,000	-141	-115	-150	-124	-150	-124	-152	-126

CALIBRATION EXPERIMENTS OF GPS TIME TRANSFER RECEIVERS AND IONOSPHERIC DELAY MEASURING RECEIVERS BETWEEN JAPAN AND FRANCE

M. Imae[*1], M. Aida[*1], T. Goto[*1], C. Miki[*1], W. Lewandowski[*2], C. Thomas[*2], G. Petit[*2] and M. Granveaud[*3]

*1: Communications Research Laboratory, *2: Bureau International des Poids et Mesures
*3: Observatoire de Paris

ABSTRACT

We performed calibration experiment of GPS dual frequency ionospheric delay measuring receivers between Japan and France to establish a accurate ionospheric delay correction for the international GPS time transfer link by the portable receiver method. By using the several days simultaneous receiving data of the portable receiver and the working receiver at each site, the relative receiver delay between CRL and BIPM was estimated within 2 ns.
At the same time, GPS time transfer receivers between CRL and OP was also performed. The calibration result was not so accurate compared with a such kind of calibration by some reasons, but we could estimated the relative receiver delay within about 5ns uncertainty.

Keywords/GPS, Time transfer, Calibration, Ionospheric delay

1. INTRODUCTION

Ionospheric delay measurements have been performed at BIPM starting from October 1988 by using prototype one of GTR-2 receiver(Ref.1) which use the dual frequency ranging signals transmitted from GPS satellite, and also have been performed at CRL starting from May 1989 using the same type of GTR-2 in regular base. They are used for the correction of ionospheric delay of the daily time transfer using conventional L1 C/A code GPS time receivers(Ref.2). The GTR-2 receiver has the measurement precision of ionospheric delay of L1 signal of about 1 or 2 ns and the receiver delays were individually estimated within a few ns level at each site. But the relative receiver delay has not been estimated.
The common-view GPS time transfer(Ref.3) is widely used for daily time transfer at major time-keeping stations and the calibration experiments or campaigns have been performed among European countries and USA by several times. But such kind of calibration experiment between Japanese time keeping laboratories and the other countries was made only once by NIST several years ago.
Due to the situations described above, we conducted the calibration experiments of GPS ionospheric delay measuring receivers and time transfer receivers between Japan and France.

2. CALIBRATION EXPERIMENTS OF GPS RECEIVERS BETWEEN JAPAN AND FRANCE

Two type of GPS receivers were transported from Japan to France to perform the calibration experiments, one is L1 C/A code time transfer receiver and the other ionospheric delay measuring receiver. At Japanese site, the portable receivers made simultaneous receiving according to the time transfer schedule for East Asian laboratories with the working receiver at CRL Koganei headquarters, Tokyo Japan. Then the time receiver was placed at Observatoire de Paris(OP, Paris France) and the ionospheric delay measuring receiver was placed at Bureau International des Poids et Mesures(BIPM, Sevres, France) and they made the simultaneous receiving with working receivers at each site according to the time transfer schedule for European laboratories.
Table 1 denotes the list of the receivers which were used for the calibration experiments, and calibration periods at each site.

Table 1 GPS receivers which are used for the calibration experiments

Japanese site;
at CRL(Koganei headquarters, Tokyo, Japan)
Ionospheric delay measuring receiver:
Prototype one of GTR-2
Time transfer receiver:
made by Allen Osborne Associates
Model TTR-5A S/N 0184(L1, C/A code only)

French site;
at OP(Paris, France)
Time transfer receiver:
made by Allen Osborne Associates
Model TTR-5 S/N 051(L1, C/A code only)
at BIPM(Sevres, France)
Ionospheric delay measuring receiver:
Prototype of one of GTR-2

Portable receivers:
Ionospheric delay measuring receiver:
made by Nihon Tsushinki
Real Time TECMETER Model 7633
(commercial version of GTR-2)
Time transfer receiver:
made by SONY
Model GTT-3000(L1, C/A code only)

Calibration periods:
at CRL: Feb. 4 to Feb. 13, 1992
at OP and BIPM: March 4 to March 8, 1992

3. CALIBRATION RESULTS AND DISCUSSION

3.1 Calibration of GTR-2 receivers

Figure 1 (a) and (b) shows the measurement results of the differences between the working GTR-2 receiver and the portable receiver(real time TEC meter) measured at CRL and BIPM respectively. And figure 2 (a) and (b) show the correlation graphs of the measured values of GTR-2 receiver and portable one at each site. In these figures the bad data which were estimated to be caused by the multi-path effect were omitted.
If we use the mean offset from figure 1 and the correlation coefficient from figure 2, we can roughly obtain the relation between the working receivers at CRL and BIPM by following equation.

$$(\text{GTR-2})_{\text{CRL}} = -0.78 + 1.16*(\text{GTR-2})_{\text{BIPM}} +/- 2.48\ (\text{ns}) \quad [1]$$

There is a possibility that the mean offset for small ionospheric delay case is not same value for large ionospheric delay case. Figure 2 (b), correlation graph at BIPM, indicates this possibility. Then we divided the obtained data into two groups, one has the ionospheric delays less than 20ns and the other grater than 20ns.
Figures 3-4 and figures 5-6 denote the two case. The results of the CRL receiver show very little difference between the two cases but for the BIPM receiver there exists a few ns of difference of offset value. Following equations relations for the two case.

$$(\text{GTR-2})_{\text{CRL}} = 0.03 + 1.15*(\text{GTR-2})_{\text{BIPM}} +/- 2.16(\text{ns}) \quad [2]$$

$$\text{for } (\text{GTR-2})_{\text{BIPM}} < 20 \text{ ns}$$

and

$$(\text{GTR-2})_{\text{CRL}} = 2.84 + 1.09*(\text{GTR-2})_{\text{BIPM}} +/- 2.15(\text{ns}) \quad [3]$$

$$\text{for } (\text{GTR-2})_{\text{BIPM}} > 20 \text{ ns}$$

For precise correction, it is better to use the equations [2] and [3].

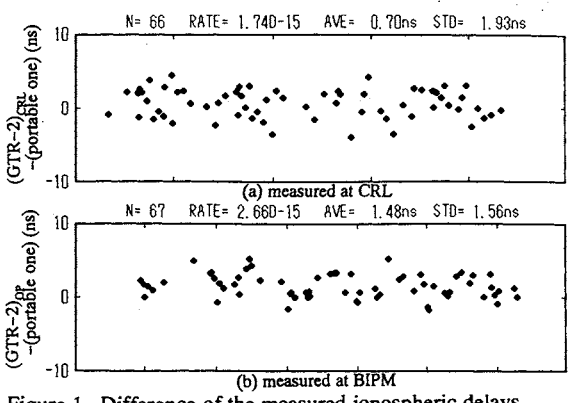


Figure 1. Difference of the measured ionospheric delays between GTR-2 receiver and portable one.

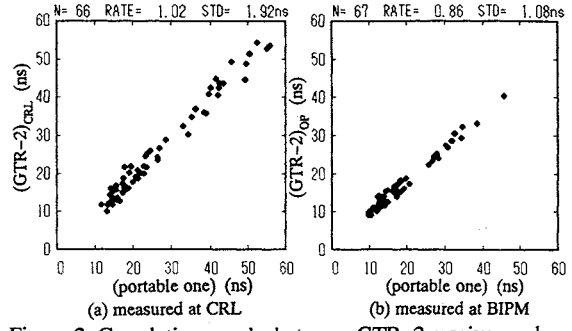


Figure 2. Correlation graphs between GTR-2 receiver and portable one.

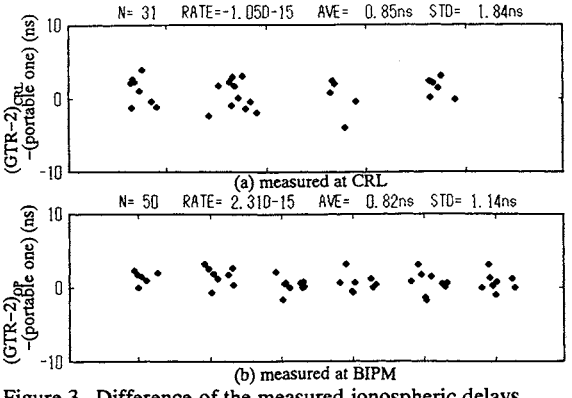


Figure 3. Difference of the measured ionospheric delays between GTR-2 receiver and portable one when measured ionospheric delays were less than 20ns.

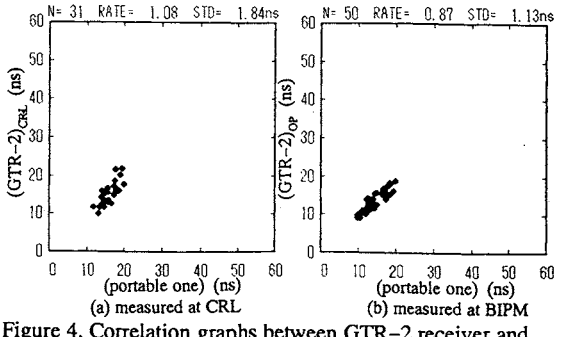


Figure 4. Correlation graphs between GTR-2 receiver and portable one when measured ionospheric delays were less than 20ns.

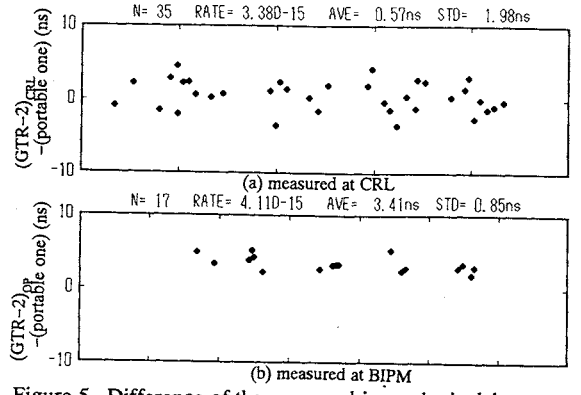


Figure 5. Difference of the measured ionospheric delays between GTR-2 receiver and portable one when measured ionospheric delays were greater than 20ns.

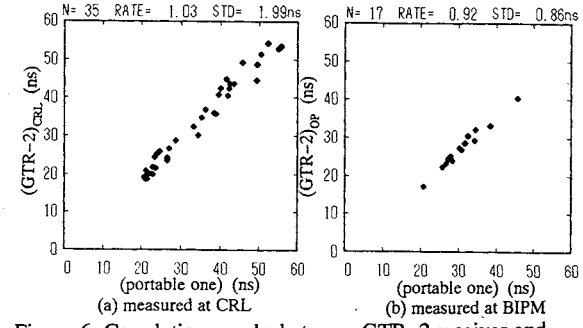


Figure 6. Correlation graphs between GTR-2 receiver and portable one when measured ionospheric delays were greater than 20ns.

3.2 Calibration results of the time transfer receivers

Calibration results of the GPS time transfer receivers at CRL and OP are summarized in table 2. During these calibration experiments, the Selective Availability(SA) was operating for Block II satellites of GPS, so we should make the common-view tracking exactly. And we omitted the tracking data which have the elevation angle of less than 20 degrees to avoid the multi path effects.

The numbers of the measured tracks at CRL largely changed depending on date because the SONY receiver couldn't receive the Block II satellite when the Anti Spoofing(AS) was operated.

Table 2. Calibration results of the time transfer receivers. ($[\text{Ref-GPS}]_{\text{LAB}} - [\text{Ref-GPS}]_{\text{PORTABLE}}$)

	Date (MJD)	Number tracks	daily mean offset (ns)	standard deviation (ns)
at CRL				
	Feb. 4(48656)	5	-7.20	3.87
	Feb. 5(48657)	7	-5.57	8.88
	Feb. 6(48658)	9	-1.56	4.81
	Feb. 7(48659)	9	-9.33	8.38
	Feb. 8(48660)	8	-1.75	4.79
	Feb. 9(48661)	21	-1.24	10.49
	Feb.10(48662)	18	-7.44	18.58
	Feb.11(48663)	23	+2.52	9.51
	Feb.12(48664)	22	-2.82	13.55
	Feb.13(48665)	22	-4.41	10.80
at OP				
	Mar. 4(48685)	20	+1.25	8.29
	Mar. 5(48686)	34	+2.18	13.10
	Mar. 6(48687)	33	+2.36	14.35
	Mar. 7(48688)	32	+1.84	12.08
	Mar. 8(48689)	14	-1.93	7.61
	Mar. 9(48690)	12	-1.17	7.57

The daily mean offsets have large discrepancy of about 12ns at CRL and about 4ns at OP. One of the reason of this discrepancy can be estimated as the problem of AOA receiver at CRL and the other is the problem of the SONY receiver.

For the first one, the CRL's AOA receiver could be considered that it was not in normal condition. Because it was modified to be able to receive Block-II satellites by manufacturer just before this experiment and it was estimated that it had a electronics problem during this experiment. In fact just after this experiment, it became out of order and needed to repair. It is the reason we could not make the calibration after the calibration trip. The second one is difference of the signal processing software in the receivers, AOA(or NIST) type and SONY one. The SONY one is not opened the calculation algorithms, such as the tropospheric delay or other calculation method of time transfer results.

Table 3. Mean value of daily offset.

Laboratory	Number of date	Mean offset (ns)	Standard deviation (ns)
CRL	10	-3.88	3.40
OP	6	0.76	1.68

Therefore, unfortunately, this calibration experiment of GPS time transfer receivers has not good precision compared with usual such kind of calibration(Ref.4), and table 3 denotes the mean value of daily offset between the working receiver and the portable receiver at each site. Finally we got the following result for the receiver mean offset between CRL and OP:

$$(AOA)_{CRL} - (AOA)_{OP} = -4.64ns +/- 5.09 (ns)$$

4. CONCLUSION

The calibration experiments of the GPS ionospheric delay measuring receivers between Japan and France was performed within about 2ns uncertainty and also time transfer receiver was performed within 5ns uncertainty. The precision of the calibration results of the time transfer receiver was not good compared with the usual such kind of calibration experiments because of a few reasons. We must repeat such kind of calibration experiments more frequently.

5. ACKNOWLEDGEMENTS

We would like to acknowledge to the colleagues of OP and BIPM for their collaborations to the calibration experiments. And the calibration experiments were financially supported by the Science and Technology Agency of Japan.

REFERENCES

1.M. Imae et. al.,"A dual frequency codeless GPS receiver measuring ionospheric effects and its application to time comparison between Europe and USA," in Proc. 3rd European Frequency and Time Forum, pp. 89–93, Mar. 1989.

2. M. Imae, E. Kawai and M. Aida,"Long term and long distance GPS time transfer Results corrected by measured ionospheric delay," IEEE trans. on IM, April, 1993(to be published).

3. D. W. Allan and M. Weiss,"Accurate time and frequency transfer during common-view of a GPS satellite," in Proc. 34th Ann. Symp. on Frequency Control, pp.334–346, May 1980.

4. W. Lewandowski,"Determination of differential time correction between the GPS time receivers located at the Observatoire de la Cote d'Azur and the Technical University of Graz," BIPM Report 91/6, Nov. 1991.

APPLICATION OF METEOR BURST EQUIPMENT FOR HIGH PRECISION COMPARISONS OF TIME & FREQUENCY STANDARDS

Epictetov L.A., Merzakreev R.R. and Sidorov V.V

Applied Radiophysics Laboratory, Kazan State University, Kazan, 420008, Tatarstan, Russia

ABSTRACT/RESUME

The meteor burst equipment is successfully utilized for two-way time transfer in Russia for a long time. The time transfer accuracy of utilized broadband-pulse equipment is about 20-30 ns now, and might be improved up to better than 10 ns by equipment modernization. The estimates of the main components of delay non-reciprocity in meteor burst propagation and the time transfer experiments made by Kazan State University are presented in this paper. They show that the two-way time transfer error via meteor burst equipment may be reduced down to less than 0.5 ns if the multiple-frequency phase method of time transfer is used.

Keywords: two-way time transfer, meteor burst, phase method, non-reciprocity.

INTRODUCTION

The meteor burst (MB) communications uses, as a means of radio signal propagation, reflecting radio waves from the ionized meteor trails. The height of arising the meteor trails (about 100 km) determines the maximum range of MB communication up to 2200 km. Now the meteor burst communications is a proven medium for reliable cost-effective data transfer in data acquisition, position monitoring and communications systems (Ref.1). In spite of limited range a meteor burst equipment is advantageous to use for precise two-way time transfer in navigation and time-keeping systems too. This is due to high accuracy and reliability of time transfer on the one hand, and self-content, low cost and simplicity of the equipment on the other hand.

The meteor burst equipment is successfully utilized for two-way time transfer in Russia for a long time. First in Russia the experimental radio link for time scale comparisons via MB radio was done by Kazan State University (KSU, Kazan, Russia) and NPO"VNIIFTRI" (Moscow region, Russia) in 1967 (Ref.2). Since that time the detail theoretical and experimental researches of MB time transfer characteristics were provided, and more than ten different experimental models of MB equipment for two-way time transfer were designed by KSU and Institute of Radioelectronics (KIRE, Kharkov, Ukraine). We took the traditional meteor-radar approaches to design the models of first generations. As the time signals we used the broadband radio pulses with the bandwidth up to 0.5-1.0 MHz and the emitted power up to 10-40 kW. The Russian National Time & Frequency Service still utilizes one of such models, the METKA-6 model developed by KIRE, for scale comparisons of remote time/frequency standards (Ref.3). The random error and residual systematic error of point-to-point time comparisons with the METKA-6 equipment are both about 20-30 ns now (Ref.3,4). Within the framework of such approach it is feasible to reduce the time transfer error less than 10 ns (Ref.5).

But it is enough easy to enhance of time transfer accuracy better than 0.5 ns if we use the multiple-frequency phase method of time transfer. Besides in this way we can use the traditional MB communications equipment with the bandwidth in range 8-32 kHz and the output power up to 300-500 W. The problems of using the carrier phase measurements and the multiple-frequency phase method for two-way time transfer via the MB radio channel have been developed in KSU for a long time. The experimental models of the MB time transfer equipment of new generation have been created there. The main results obtained in KSU concerning the investigations of MB propagation phase reciprocity and achievable accuracy of two-way MB time transfer are presented in this paper.

1. NON-RECIPROCITY IN METEOR BURST PROPAGATION

The top precision of time transfer is achieved with using the carrier phase measurements. With these measurements on the frequencies of MB band (40-60 MHz) the noise component of time transfer error may be easily decreased to 0.1 ns and more. However, when such accuracy is achieved the errors connected with the delay non- reciprocity in meteor burst propagation become major. It is known, if there is a deference between delays of propagation in contrary directions of radio link, there is an error of two-way time transfer, the value of which is equal to a half of this difference. The phase delay difference for radio waves propagated in contrary MB link directions we call the phase non-reciprocity. Several possible reasons exist for appearance of phase non-reciprocity in MB radio channel used for time transfer. The all main components of non-reciprocity may be conventionally divided into three groups according to mechanism of their appearance: the dynamic non-reciprocity, the polarization non-reciprocity and the relativistic non-reciprocity.

1.1 Dynamic phase non-reciprocity

The first obvious reason of phase non-reciprocity is the dynamic change of trail-reflection conditions during the time between the trail-reflections of the opposing radio signals. The value of phase dynamic non-reciprocity depends on 1) the phase change velocity in reflected signal which caused by several physical factors, and 2) the quantity of time interval between the trail-reflections of signals passing in forward and back directions of radio link. If the compared scales are roughly synchronized and time signals of the master and slave clock stations are emitted simultaneously, the trail-reflection time difference is minimum and caused by possible asymmetry of meteor trail location about the middle of path.

The main reasons of relatively fast propagation delay changes for meteor scattering are, the first, the radio wave diffraction from forming meteor trail during a meteor particle pass through the main Fresnel zone ("heading echo"), and the second, the spatial movement and deformation of meteor trail by the turbulent wind. The diffraction and wind phase nonstability characteristics have been investigated detail in the KSU (Ref.6,7).

The phase of heading echo varies parabolically and very fast because the meteors pass with the cosmic velocities (up to 72 km/s). However, the time interval, when phase changes so fast, is enough small part of the meteor burst duration. Therefore the initial part of meteor burst is not used for high-precise measuring usually.

The results of wind measurements, which are provided by KSU for many years, were used for computer simulation of the wind phase non-stability and non-reciprocity for several MB links. The estimates with the Monte Carlo computer model of MB propagation developed by KSU (Ref.8) show that the maximal RMS value of the wind non-reciprocity error is about 0.2 ns on the links of 300-600 km length (Ref.9). If the link is shorter or longer, the wind dynamic non-reciprocity is smaller. The wind varying of propagation delay is practically linear, so that the wind non-reciprocity is practically constant during the one meteor burst. But its values differ from one meteor trail to another as their location and the turbulent wind velocity differ.

1.2 Polarization phase non-reciprocity

In the meteor burst radio the antennas with linear polarization are usually used. The influence of earth's magnetic field on propagating and scattering the polarized radio waves in the plasma of ionosphere and meteor trail leads to arising the polarization phase non-reciprocity. There are some mechanisms of polarization phase non-reciprocity in meteoric propagation.

The first mechanism of polarization non-reciprocity is caused by that Lorentz force, acting on electrons of meteor-trail plasma in geomagnetic field, changes the complex diagram of scattering from a meteor trail of small volume electron density (Ref.10). The numeric solution of general diffraction problem for this case shows that, when the volume electron density of meteor trail is small enough, the phase non-reciprocity value does not exceed 5-10 degrees under the worst circumstances. If the volume electron density is greater, the non-reciprocity is smaller. Therefore the value of this kind error is not more than 0.1-0.2 ns, and in most cases it is far smaller.

The second mechanism of polarization non-reciprocity connects with the difference of polarization angles for the opposing waves, incident on meteor trail. This difference arises from both the non-reciprocity of ionospheric Faraday rotation and the asymmetric location of meteor trail about the middle of path. As calculations (Ref.11) show the maximum value of this non-reciprocity does not exceed 60-90 degrees (time transfer error is less than 1.5-2.5 ns) in the worst. But this large value takes place only in initial part of meteor burst where the large dynamic diffraction

error also exists. The value of this polarization non-reciprocity quickly decreases with the diffusion expansion of meteor trail.

The third non-reciprocity mechanism connects with the influence of the earth's magnetic field on propagation of the polarized radio waves in ionosphere. The calculations (Ref.12) show that under worst circumstances the maximal values of the phase non-reciprocity do not exceed 10-20 degrees on the frequencies of meteor band. Accordingly the maximal value of the two-way time transfer error caused by this non- reciprocity is not more than 0.25-0.5 ns.

1.3. Relativistic phase non-reciprocity

The general fundamental cause of non-reciprocity for synchronization channels of time scales, placed on the Earth surface, is the anisotropy of wave speed in Earth-axes non- inertial coordinate system (the Sagnac effect). If the geocentric coordinates of synchronization points and trail-reflection point are known, the relativistic non-reciprocity value may be calculated from general equation (Ref.13). As trail-reflection point coordinates change from one meteor trail to another in some limits, there is a variation of the relativistic non-reciprocity around some mean value. We used the Monte Carlo computer model of MB propagation to estimate the relativistic non-reciprocity values for some real MB links. For example, the calculated daily average value and the RMS burst-to-burst variation value of the relativistic non-reciprocity are about 2.1 ns and 0.05 ns on the Moscow-Kazan latitudeous MB link of about 720 km length. Thus if the coordinates of synchronization points and the antenna orientations are known, the average relativistic time comparison error may be calculated with the computer modeling and taken into account for scale comparison correction. Both the residual systematic error and the burst-to-burst error variation are small enough.

2. MB TIME TRANSFER EXPERIMENTS WITH PHASE METHOD

We can conventionally divide the phase non-reciprocity on two components: 1) systematic or slow-changing component (it may be caused by relativistic and ionospheric propagation components of non-reciprocity); 2) random or variable component which is distinctly change both during one meteor burst and from one burst to another. As mentioned above the calculated values of slow-changing components of phase non-reciprocity on MB radio frequencies are not more than 0.5 ns. As the experimental detecting of such small systematic component is very difficult now, we have not done such experiments yet.In the series of our expirements we have done the attempts to appreciate the total variable MB phase non-reciprocity component.

We performed three phase measurement experiments on different MB links: St.Petersburg-Kazan in 1979 (Ref.14), Moscow-St.Petersburg in 1988 (Ref.15), and Moscow-Kazan in 1992 (Ref.16). The most accuracy and representative experiment was performed on the 720 km Moscow-Kazan MB link in June of 1992. There was used the KAMA-5 maltiple-frequency phase MB equipment developed by KSU. The instrumental errors were minimized by regular automatic calibration of phase delay in receiver and transmitter sections on all frequencies. The RMS random component value of the instrumental error of phase single-measurement was not more than 4 degrees (0.2 ns). The half-difference of opposing phase measured on stations in each transmission period was taken as the single top precision estimate of time scale shift, and the burst-average value of single estimates was taken as the result of time scale comparison.

It was calculated the RMS deviations of single estimates from the burst-average values for all long bursts with the single measurement numbers greater 10-20. The average value of these RMS deviations was about 0.8 ns (without excluding the initial parts of meteor bursts). It exceeds the average value of instrumental single-measurement error (0.2 ns) and indicates the existence of random component phase non-reciprocity in meteor burst radio.

In view of the random locations and orientations of meteor trails we can consider that the random errors of the burst-average scale comparison results are not correlated for any different meteor trails. Therefore we can use the two-sampling variance of the scale comparison results with small intervals between meteor bursts as the simple and convenient sample statistic of the random errors (Ref.17). In examined case the two-sample variance is determined as the half of the mean square of the difference between the burst-average scale comparison results obtained on any two meteor bursts. For the small intervals between meteor bursts when the mutual scale drifts within these intervals are neglected in comparison with the time transfer errors, the square root of the two-sample variance was found to be about 0.25 ns for the last experiment data. We can consider this value as the random error value of time transfer with KAMA-5 meteor burst

equipment if we resolve the ambiguity of these carrier phase measurements.

In the previous experiment, which was performed on Moscow-St. Petersburg MB link in 1988 with the KAMA-3 pulse MB equipment, the analogous estimate was about 0.45 ns (Ref.15). It may be explained by the less accurate measurements mainly and the other experiment- performance conditions also. In any case we can consider the time transfer error with carrier phase measurements is less than 0.5 ns.

The period of ambiguity of two-way time transfer with carrier phase method is in range 8-12 ns for the MB radio frequencies (40-60 MHz). To resolve this ambiguity, the unambiguous measurements of maximal error less than +-4-6 ns are necessary. The KAMA-5 model with the 0.5 MHz maximal frequency shift can ensure such small error long averaging interval only. The new meteor burst equipment model with maximal frequency shift about 10 MHz has been created by KSU. Such frequency shift ensures the resolving of ambiguity of the single phase measurements.

CONCLUSION

The measurement noise floor of two-way time transfer with the meteor burst equipment may be reduced down to 0.1 ns and more if the multiple-frequency phase method of time transfer is used. In this case the errors connected with the delay non-reciprocity of meteor burst propagation become major. The calculations and the time transfer experiments provided by Kazan State University are shown that the total random error caused by non-reciprocity is not more than 0.25-0.5 ns. Thus the phase meteor burst equipment ensures reliable point-to-point time transfer to the sub-nanosecond accuracy for the distances up to 1500-2000 km. We hope that in spite of the limited range the meteor burst systems might find application for the precise time scale comparisons in navigation and timekeeping systems (at least as an independent supplementary metrological mean). This is due to the high accuracy and reliability of the time and data transfer on the one hand, and the self-content, low cost and simplicity of the meteor burst equipment on the other hand.

ACKNOWLEDGMENTS

The authors thank all their colleagues from the Applied Radiophysics Laboratory of KSU who in either degree contributed to this work. The authors are also grateful to S.B.Pushkin and N.B. Koshelyaevsky from NPO"VNIIFTRI" for the support of the meteor burst experiment performed in 1992.

REFERENCE

1. W.E. Day. - Electronics, 1982, vol.55, N 20, pp.71-75.
2. Минуллин Р.Г., Палий Г.М., Сидоров В.В. и др. - Измерительная техника, 1971, N 1.
3. Пушкин С.Б., Кащеев Б.Л., Дудник Б.С. и др. - Измерительная техника, 1989, N 4, с.22-24.
4. Дудник Б.С., Кащеев Б.Л., Коваль Ю.А. и др. - Измерительная техника, 1986, N 4, с.15-16.
5. Кащеев Б.Л., Дудник Б.С., Коваль Ю.А. и др. - Тезисы докл. науч.-техн. конференции "Метрологическое обеспечение частотно-временных измерений при испытаниях и эксплуатации техники", Москва, 15 октября 1991 г., - М.: 1991. с.55-56.
6. Андрианов Н.С. - Метеорное распространение радиоволн. - Казань: Изд-во КГУ, 1980, вып.16, с.63-68.
7. Тептин Г.М. - Изв. АН СССР. Физика атмосферы и океана, Т. 8, N 3, 1972.
8. Карпов А.В., Сидоров В.В. - Метеорное распространение радиоволн. - Казань: Издат-во Казанского университета, 1978, Вып.15, С.52-60.
9. Курганов А.Р., Сидоров В.В. - Материалы итоговой науч.конф. Казанского университета за 1987 г., - Казань, Издательство Казанского университета, 1989, с.65-67.
10. Сидоров В.В., Хузяшев Р.Г. - Известия вузов. Радиофизика, 1984, т.27, N 5, с.661-663.
11. Сидоров В.В., Хузяшев Р.Г., Плеухов А.Н. - Изв.вузов.Радиофизика, 1984, т.27, N 8, с.1075 - 1077.
12. Плеухов А.Н. - Изв.вузов. Радиофизика, 1988, т.31, с.545-549. 12. Плеухов А.Н.-Изв.вузов. Радиофизика. 1988, т.31, с.545-549.
13. Ландау Л.Д., Лифшиц Е.М. Теория поля. - Москва: Наука, 1967.
14. Сидоров В.В., Курганов А.Р., Плеухов А.Н., Овчинников В.В., Хузяшев Р.Г. - Метеорное распространение радиоволн. -Казань: Изд-во КГУ, 1981, вып.17, с.30-39.
15. Сидоров В.В., Базлов А.Е., Мерзакреев Р.Р., Эпиктетов Л.А. - Тезисы докл. XVI Всесоюз. конф. по распространению радиоволн. -Харьков, 1990. с.83.
16. Мерзакреев Р.Р., Сазонов М.С., Эпиктетов Л.А. - Тезисы докл. XVII конф. по распространению радиоволн. - Ульяновск, 1993.
17. Сидоров В.В., Базлов А.Е., Мерзакреев Р.Р., Эпиктетов Л.А. - Материалы итоговой науч. конф. Казанского университета за 1988 г. - Казань, Изд-во КГУ.

PRECISE GPS EPHEMERIDES FOR TIME TRANSFER

G. Petit, W. Lewandowski, C. Thomas

Bureau International des Poids et Mesures
Pavillon de Breteuil, 92312 Sèvres Cedex, France

ABSTRACT/RESUME

The quality of GPS ephemerides is a key factor in the GPS time transfer technique. In the common-view method, the error can reach 10-20 ns for a single measurement, when using broadcast ephemerides, and several times these values with the full implementation of Selective Availability. The measurements can be corrected in post-processing, if precise ephemerides are made available to the user. Recently, international GPS tracking networks have been set up for civilian geodetic purposes. They provide post-processed ephemerides which are estimated to be accurate at the metre level, and the corresponding effect on GPS time transfer should, on average, be down to 1 ns. The accuracy of GPS ephemerides provided by numerous agencies is tested by verifying the closure condition on the sum of three time links around the world. It is found that the precision of GPS time transfer, when corrected using such ephemerides and measured ionospheric delays, reaches 3-4 ns. Uncertainties of several nanoseconds remain in this test of the accuracy of time transfer.

Keywords: GPS time transfer, ephemerides, reference frames.

1. INTRODUCTION

The excellence of worldwide time unification depends on the quality of the clocks kept by the national timing centres and on the means of time comparison. With the Global Positioning System, time comparisons can easily be performed with an accuracy of 10 to 20 nanoseconds over intercontinental distances (Ref. 1). It is possible to improve this performance to 4 to 5 nanoseconds when using measured ionospheric delays and post-processed precise ephemerides (Ref. 2). Until recently, such ephemerides have been available on a regular basis only from the US Defense Mapping Agency and from the US National Geodetic Survey, with a delay of about 6 weeks.

In June 1992, the International Association of Geodesy has set up the International GPS Geodynamics Service (IGS) to support geodetic and geophysical research activities (Ref. 3). Through a number of observatories equipped with GPS receivers, data centres archiving the collected data, and analysis centres, it makes its data and products available to users. In particular, GPS ephemerides computed by seven analysis centres are available on a regular basis with a delay of about one week. This service is intended to be permanent.

The purpose of this paper is to study the accuracy and precision of GPS time transfer, when using measured ionospheric delays and precise ephemerides provided by several agencies. The test of accuracy is the closure condition provided by the sum of three independent links around the world (Ref. 4). Section 2 presents the ephemerides used in the tests. Section 3 describes the closure experiment and the possible interpretations of a misclosure. Section 4 reviews the results obtained on several closure experiments conducted over the last three years.

2. GPS EPHEMERIDES

2.1 DMA ephemerides

The US Defense Mapping Agency produces precise ephemerides from its own 10-station network. They have been provided to the BIPM on a regular basis since 1990. The description of the computation process can be found in (Ref. 5). The ephemerides represent an Earth-centered Earth-fixed trajectory, referenced to the satellite centre of mass, expressed in the WGS84 reference frame, the time scale being GPS time.

2.2 NGS ephemerides

The US National Geodetic Survey produces precise ephemerides using data from the Cooperative International GPS Network (CIGNET). Since 1991, these have been made available on a free-access computer. The procedure of orbit determination is described in (Ref. 6). The ephemerides represent an Earth-centered Earth-fixed trajectory, expressed in the International Earth Rotation Service (IERS) Terrestrial Reference Frame (ITRF), the time scale being GPS time.

2.3 IGS ephemerides

Seven IGS analysis centres have produced precise ephemerides from the data of the IGS network since June 1992. They are the Center for Orbit Determination in Europe, University of Bern (CODE), the Geodetic Survey of Canada, Ottawa (EMR), the European Space Agency, Darmstadt (ESA), the GeoForschungsZentrum, Potsdam (GFZ), the Jet Propulsion Laboratory, Pasadena, California (JPL), the Scripps Institution of Oceanography, La Jolla, California (SIO), the University of Texas, Austin (UTX). The ephemerides are made available on free-access computers at the IGS data centres. The procedures of orbit determination are described in messages available on the IGS computers (Ref. 7). The ephemerides represent an Earth-centered Earth-fixed trajectory, expressed in the ITRF reference frame, the time scale being GPS time.

3. TEST OF ACCURACY

3.1 Description of the test

All closure experiments reported in this paper use the three links between the Paris Observatory (OP), the National Institute of Standards and Technology in Boulder, Colorado (NIST), and the Communications Research Laboratory in Tokyo (CRL). The GPS data taken at the three sites correspond to the international schedule issued by the Bureau International des Poids et Mesures for the establishment of TAI. Each link is computed independently using the common-view method, the results being passed through a low-pass filter with a cut-off period of about three days. The scatter of the unfiltered data points relative to the filtered ones provides an estimation of the short term measurement noise. The filtered data are then interpolated to one point per day, and these daily values for the three links are summed to provide one estimation of the closure per each day.

At each site ionospheric delay measurements are performed by dedicated dual-frequency codeless GPS receivers. These measurements are used to compute the ionospheric correction to the raw time-transfer data. A correction for precise ephemerides is also computed for all the data points. In the following, "raw data" means common-view data without corrections (other than those transmitted by the satellites), while "corrected data" refers to data to which have been applied the corrections for ionospheric measurements and precise ephemerides.

In the ideal case, where there is no measurement noise and no systematic error of any kind, the sum of the three time links, although they are completely independent from one another, must be zero. In the real world there is measurement noise, in the time measurements themselves, and in the ionospheric delay measurements. In addition systematic errors are probably present in the ionospheric measurements. Errors in the relative positions of stations and satellites will also cause systematic effects. In the next sections, we consider this particular point in more detail. Finally, although the closure procedure provides a cancellation of calibration errors, it is still possible that short term variations of the calibrations cause systematic effects due to the regularity of the observing schedule over durations of months.

3.2 Realization of terrestrial reference frames

The station coordinates on the one hand, and the satellite coordinates on the other, are two realizations of a terrestrial reference frame, which can be totally independent. In a first approach, one can state that a realization by a set of station coordinates is a geometric one, while a realization by the ephemerides of satellites is a dynamic one. In general, apart from local effects, the differences between two realizations of a terrestrial reference frame can be represented by a 7-parameter transformation which includes a translation of the origin (3 parameters), a scale factor (1 parameter), and a global rotation of the axes (3 parameters).

When choosing a set of station coordinates and a set of satellite ephemerides to compute the common-view time transfer data, one implicitly assumes that the two realizations of the terrestrial reference frames by the station coordinates and by the ephemerides are the same. If this assumption is not justified, errors will be caused in the time links, which will return a non-zero value when computing the sum of the three links (deviation from closure). However, due to the use of the common-view method and to the long distance of the links, the overall geometry of our network is quite particular, with stations at about the same latitude, regularly spaced in longitude, and with each station observing towards the east for one link, and towards the west for the other. This geometry has the important consequence that the closure is, to first order, relatively insensitive to a global transformation between the frame of the station coordinates and that of the ephemerides, with one exception: a global rotation about the Z axis (the axis of the poles) fully affects the closure. This can be seen in figure 1, which represents the effect of such a rotation on one link. In this case, the link computed as H(A)-H(B) has a higher numerical value for the geometry A'B' than for the geometry AB. As the geometry of all three links is equivalent regarding a rotation about the Z axis, the effect of such a rotation on the closure is three times its effect on one link. The deviation from closure is estimated to be about 1 ns per 0.004" of rotation about the Z axis (Such a rotation represents about 40 cm at satellite level).

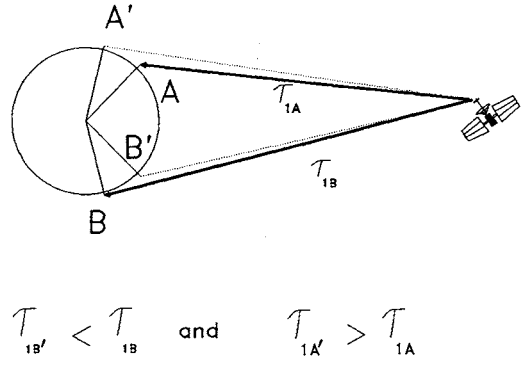


FIGURE 1: *Effect of a rotation of the station frame relative to the satellite frame on the GPS time transfer.*

3.3 Effect of an error in station coordinates

In addition to a global difference decribed in the previous section, an error in the coordinates of one station can be viewed as a local difference that affects the time links in which this station participates. However, due to the more or less constant geometry of the observations (due east for one link, due west for the other), the effect of an error in the coordinates of one station is a constant misclosure. Based on the geometry of the network, table 1 lists the average effect of errors in station coordinates on the closure. Incidentally one can verify (by summing the values in each column of table 1) that the effect of a global translation, ie. an error of the same value in a coordinate for all stations, is about zero.

TABLE 1: *Misclosure/ns caused by 1 metre of error for each station coordinate*

	X	Y	Z
OP	+2	-4	0
NIST	-4	+1	0.5
CRL	+2	+3	0

4. RESULTS

4.1 Closure experiments

The closure has been computed for three periods of time between mid-1990 and end-1992. To account for the ephemerides available at different times, the results are presented in four groups.

The first represents an extensive study that has been conducted over the 393 days from 1990 June 16 (MJD 48058) to 1991 July 13 (MJD 48450). The results have been presented in (Refs. 2,4). This test only uses DMA ephemerides and checks the long term consistency of such a test.

The second is included in the first, but consists in a 3-week period early in 1991 where a global geodetic experiment named GIG'91 (Ref. 8) was conducted. This ran from 1991 January 22 (MJD 48278) to February 13 (MJD 48300). Available ephemerides for this period were of DMA origin, together with a set produced by the JPL from the GIG'91 data.

The third group covers a 2-week period at the beginning of the operation of the IGS network in the summer of 1992. It starts on July 5 (MJD 48808) and ends on July 17 (MJD 48820). Available ephemerides were the ones from the DMA and the NGS, with others from the IGS centres CODE, ESA, JPL and SIO (others were available from IGS, but have not been considered in this analysis).

The fourth group covers a 4-week period of operation of the IGS network in the autumn of 1992. It runs from October 13 (MJD 48908) to November 7 (MJD 48933). Ephemerides from DMA and CODE were used in this analysis.

In each case the same processing scheme was applied, following the description in section 3.1. Figures 2 to 11 present the daily values of the deviation from closure obtained. The results from raw data are shown in figures 2, 4, 6 and 10 for groups 1 to 4, respectively. The results for corrected data are given in figures 3 and 5 for groups 1 and 2, figures 7 to 9 for group 3, and figure 11 for group 4.

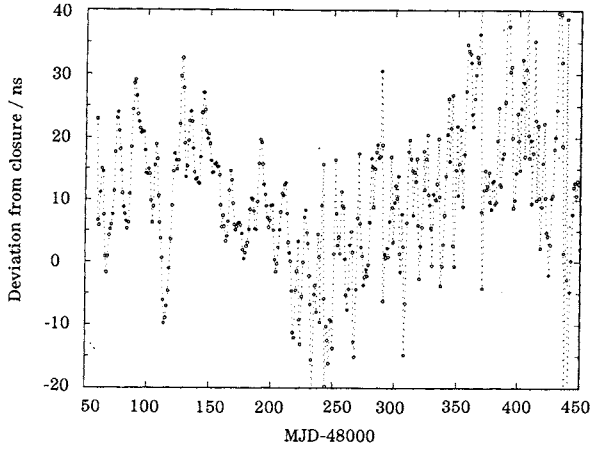


FIGURE 2: *Deviation from closure around the world via OP, NIST and CRL with non-corrected data.*

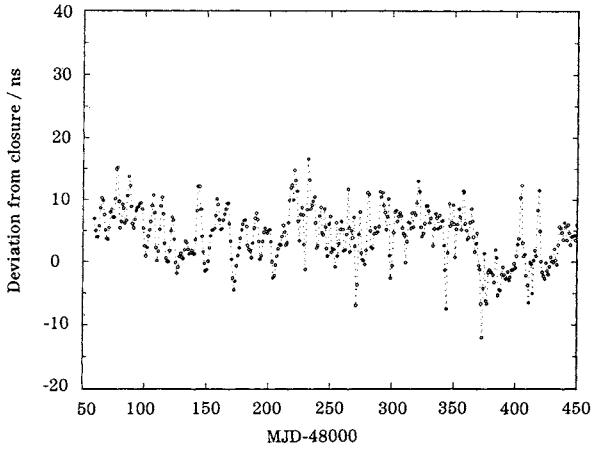


FIGURE 3: *Deviation from closure around the world via OP, NIST and CRL with data corrected for measured ionospheric delay and DMA precise ephemerides.*

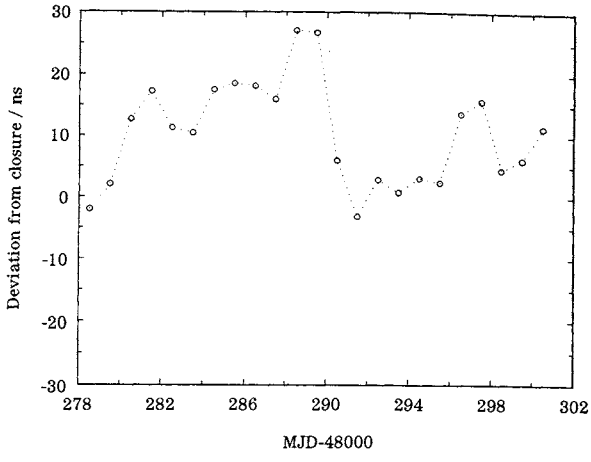


FIGURE 4: *Deviation from closure around the world via OP, NIST and CRL with non-corrected data.*

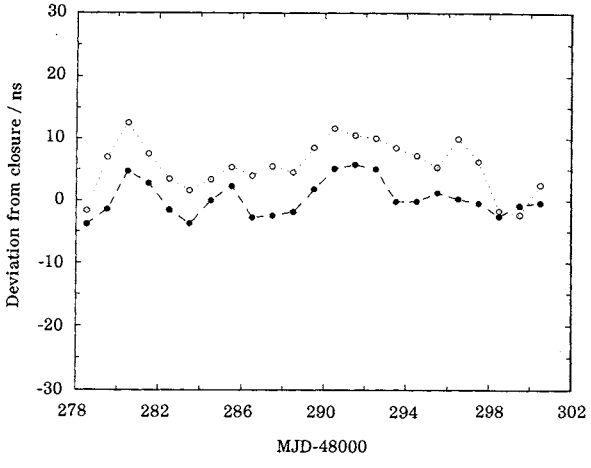


FIGURE 5: *Deviation from closure around the world via OP, NIST and CRL with data corrected for measured ionospheric delay and precise ephemerides from DMA (open circles) or JPL (filled circles).*

4.2 Discussion of the results

4.2.1 Precision of the time transfer

The precision of the time transfer, and thus the effect of ephemerides on it, can be estimated from the scatter of the difference between unfiltered and filtered data, as stated in section 3.1. This procedure yields values that range from 3 ns to 6 ns, depending on the length of the link and on the ephemerides used. The best result (2.4 ns) is for the OP-NIST link (the shortest one) using the JPL ephemerides of the GIG'91 campaign. Such values are estimates of the precision of one single 13 min measurement. Given that there is measurement noise on the time measurements and on the ionspheric delays, at the level of a few nanoseconds, this implies that the contribution of the ephemerides is generally down to 1 ns, or even less, when using any of the precise ephemerides. This, in turn, implies that the precision of the ephemerides is at the level of one metre.

4.2.2 Differences between reference frames

From the figures presenting corrected data, it can be seen that the sum of the three links is generally not zero, and the deviation from closure can reach 15 ns. As noted in section 3, this can have a number of causes and it is not obvious how to account for the absolute value of the deviation from closure. But, in all cases, the differences between the results using different ephemerides can easily be interpreted.

It is evident from the figures that, in general, results using different ephemerides follow the same pattern with a more or less constant offset of a few nanoseconds (The sole exception to this is the result with NGS ephemerides obtained in the first few days of the series shown in figure 7). This behavior can be related directly to a global rotation between the frames of the ephemerides, as explained in section 3.2. To check this, we have determined the best-fitting 7-parameter transformation between several of our satellite ephemerides.

From this, for example, it is found that the transformation from CODE to DMA, computed for 7 days during the July 92 period and 7 days during the November 92 period, consists mainly of a scale factor of 5×10^{-8}, which has almost no effect on the closure, and of a rotation of -0.015" about the Z axis, which exactly accounts for the offset of 4 ns between the results with CODE and DMA ephemerides (figures 7-8 and 11). Incidentally, the scale factor is certainly due to CODE and DMA using different values for the gravitational constant GM of the Earth. In a similar analysis, it is found that the transformation from the GIG'91 JPL ephemerides to the DMA ones consists also in the same scale factor, and a rotation about the Z axis of -0.013". This accounts for the offset clearly present between the two results of figure 5.

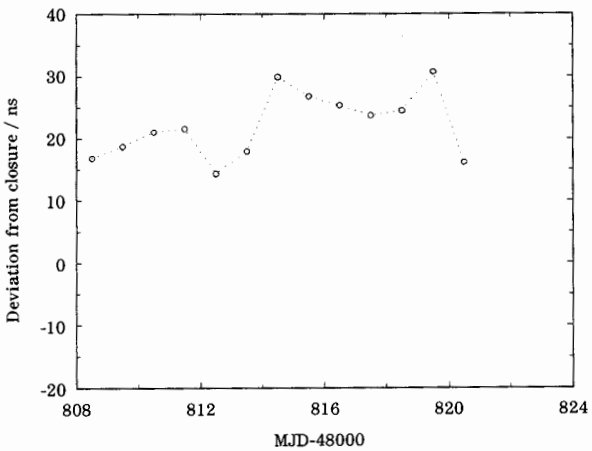


FIGURE 6: *Deviation from closure around the world via OP, NIST and CRL with non-corrected data.*

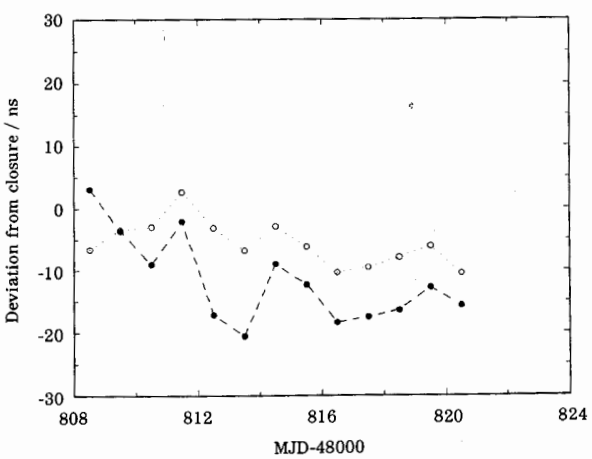


FIGURE 7: *Deviation from closure around the world via OP, NIST and CRL with data corrected for measured ionospheric delay and precise ephemerides from DMA (open circles) or NGS (filled circles).*

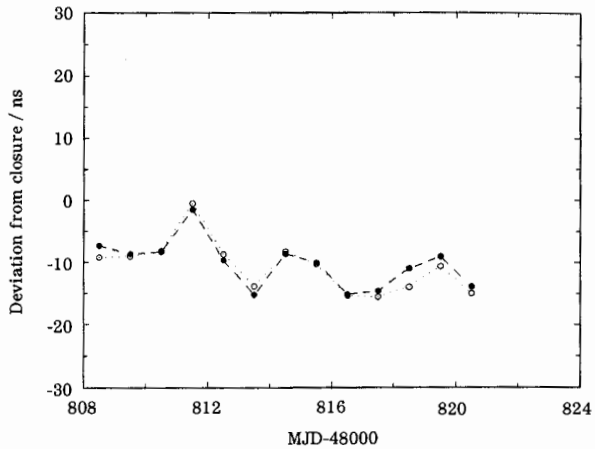


FIGURE 8: *Deviation from closure around the world via OP, NIST and CRL with data corrected for measured ionospheric delay and precise ephemerides from CODE (open circles) or ESA (filled circles).*

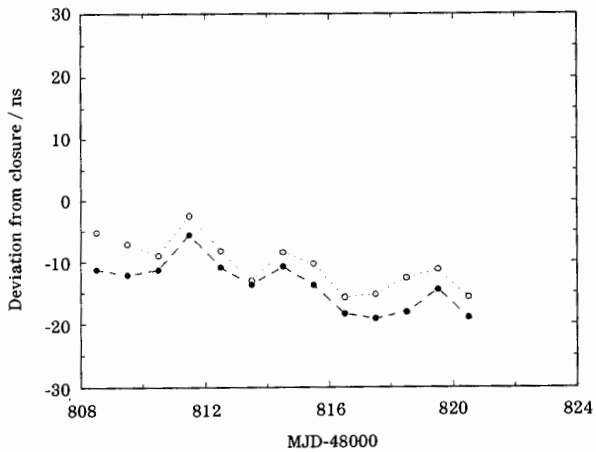


FIGURE 9: *Deviation from closure around the world via OP, NIST and CRL with data corrected for measured ionospheric delay and precise ephemerides from JPL (open circles) or SIO (filled circles).*

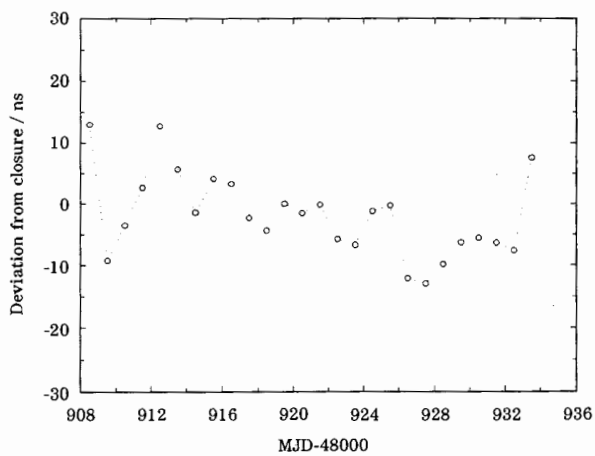


FIGURE 10: *Deviation from closure around the world via OP, NIST and CRL with non-corrected data.*

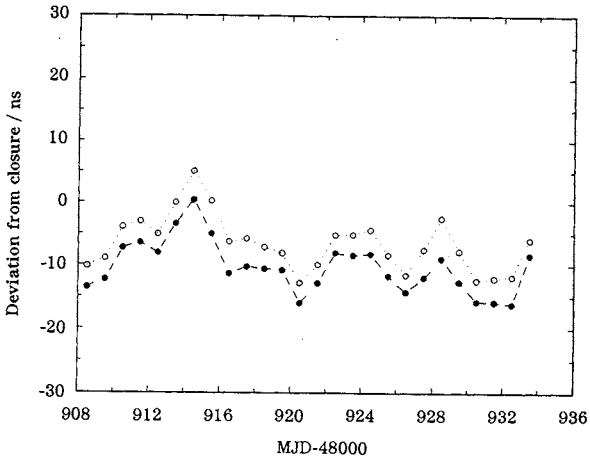


FIGURE 11: *Deviation from closure around the world via OP, NIST and CRL with data corrected for measured ionospheric delay and precise ephemerides from DMA (open circles) or CODE (filled circles).*

4.2.3 Accuracy of the time transfer

If the station coordinates and the satellite ephemerides were actually consistent realizations of the same terrestrial reference frames, it would be expected that the closure condition be verified with an uncertainty of a few nanoseconds. It is clear from all the results that the closure condition is not verified at this level: the deviation from closure can reach 15 ns. But it is also clear that all ephemerides give rise to a deviation from closure which remains fairly constant with time. The RMS of the daily values over several weeks range from 3 ns to 5 ns, depending on the ephemerides used. In our case, the station coordinates have been chosen to be consistent with the ITRF reference frame (Ref. 9), so it is to be expected that the closure condition will be verified for those ephemerides that are a realization of ITRF, ie. the NGS ephemerides and those produced by the IGS centres. We find that the results for these ephemerides are in good agreement, but show a deviation from closure of about -10 ns.

One possible explanation of this misclosure is an error in the coordinates of one station. From table 1, it can be seen that an error of 2 metres, in some of the coordinates, could account for much of the effect that is seen. However, the uncertainties in station coordinates are thought to be 10 cm to 50 cm (Ref. 9), which means that such an effect would come from an unnoticeable mistake. If this explanation is eventually proven to be correct, it will have the consequence that the closure is verified to within a few nanoseconds, and thus that the accuracy of GPS time transfer is validated at the level of about two nanoseconds.

Another possibility is that systematic effects, as yet unidentified, remain in the data. These could have their origin in the ionospheric measurements or in the time receivers themselves, and manifest themselves as biases, with variations showing over many months. In this case, it would be necessary to identify those effects before reaching the level of a few nanoseconds in the accuracy of GPS time transfer.

4.3 Comparison of GPS ephemerides

The precision of GPS ephemerides is generally estimated by the groups that produce them. The NGS reported an internal consistency of one part in 10^7 (Ref. 6). The IGS centre CODE regularly produces a comparison of all IGS ephemerides (Ref. 7), which states that the RMS of the differences between any two sets ranges from a few decimetres to slightly more than one metre on daily averages, after adjustment of a 7-parameter transformation. Finally the comparisons that were done for this study (see section 4.2.2) show that the RMS of the differences between DMA and IGS ephemerides ranges from 1.3 to 1.8 metre on daily averages, after adjustment of a 7-parameter transformation.

These estimations all yield values from 0.5 m to 2 m for the internal consistency of the ephemerides used. This turns into an average error of 1 ns or less in the time transfer over long distances (the worst case error in time transfer over 10000 km is about half the error on the satellite position, ie. 3 ns of time transfer error for 2 m of satellite position error). This is perfectly consistent with the observed precision of time transfer, as indicated in section 4.2.1.

5. CONCLUSIONS

Numerous precise ephemerides are now available for the post-processing of GPS time transfer data, some of them being available with a delay of one week or less. They claim to be precise at the level of about one metre, which affects time transfer by one nanosecond or less, on average. This claim is fully supported by the results of several closure experiments, conducted over a period of two years. For the deferred-time user, such ephemerides allow to overcome the effects of Selective Availability (Ref. 1), and contribute negligeably to the error budget of the time transfer. If ionospheric delay measurements are also available, the precision of one 13 min time transfer reaches 3-4 ns.

The GPS ephemerides are a realization of a terrestrial reference frame. It is to be checked that the station coordinates which are used to process the time data are consistent with the reference frame of the ephemerides. It is advised that the IERS Terrestrial Reference Frame (ITRF) be used whenever possible. If the WGS84 reference frame is used, it should be realized that, by its definition, it may differ by up to one metre from the ITRF.

Uncertainties of several nanoseconds remain in the accuracy of time transfer, as tested by the closure condition of the links OP-NIST-CRL. It is possible that the source of uncertainty is in the test itself (error in station coordinates), but it is also possible that unknown biases remain at the level of several nanoseconds.

Acknowledgements

We thank the staff of the three laboratories, the OP, the NIST and the CRL, who, in addition to the time data, provided the ionospheric measurements on a regular basis. We are indebted to G. Beutler and T. Springer at the University of Bern for providing the software used for the comparison of ephemerides, and for fruitful discussions.

References

1. Lewandowski W., Thomas C., GPS time transfer, Proc. IEEE, 79, 5, p. 991, 1991.
2. Lewandowski W., Petit G., Thomas C., Precision and accuracy of GPS time transfer, IEEE Trans. I&M, Special issue for CPEM'92, 1993, in press.
3. Mueller I.I., Beutler G., The international GPS service for geodynamics: development and current structure, Proc. 6th International geodetic symp. on satellite positioning, Columbus, p. 823, 1992.
4. Lewandowski W., Petit G., Thomas C., Accuracy of GPS time transfer verified by the closure around the world, Proc. 23rd Annual PTTI meeting, p. 331, 1991.
5. Swift E.R., GPS orbit/clock estimation based on smoothed pseudorange data from a ten-station network, Proc. IAG Symp. G2, Vienna, 1991, in press.
6. Schenewerk M.S., GPS orbit determination at the National Geodetic Survey, Proc. 23rd Annual PTTI meeting, p. 59, 1991.
7. IGS reports, computer files, Internet address 130.92.4.10.
8. Melbourne W., Fisher S., Neilan R., Yunck T., Engen B., Reigber C., Tatavjan S., The first GPS IERS and geodynamics experiment 1991, Proc. IGGU meeting, Vienna, 1991, in press.
9. Lewandowski W., High accuracy ground antenna coordinates for GPS time transfer, Proc. IAG Symp. G2, Vienna, 1991, in press.

USE OF LORAN C SIGNALS FOR TIME/FREQUENCY APPLICATIONS

D. SENNEDOT, Yves MINGUY and B. THOMAS

DCN BREST - FRANCE

ABSTRACT

A large number of means are currently used for time/frequency applications, in order to compare or generate stable times and frequencies. These means include the LORAN C radio navigation system, which should assume an important role in the future. The creation in Europe of a modern network synchronized using the "TOE Control" method with UTC(B) time, which is the local representation of Brest atomic time (TAB).

Keywords: Time, Frequency, LORAN C, Radio navigation, Synchronization, Clocks, Cesium.

1 - SUBJECT OF THE DOCUMENT

1.1 HISTORY

The time references used throughout the world are the TAI or UTC, calculated by the Paris Observatory for the BIPM. The link between these two references is:
TUC - TAI = n.seconds
Today n = 27 seconds.

Each country generates its own time references called: TA(i) and creates a physical representation of UTC time: UTC(i).

In France, the Paris Observatory is responsible for calculating the TAF and generating UTC(OP).

For these times to be usable, they must be "transportable", or comparable. To enable comparison, several means exist:

- clock transport,
- television signal,
- radio transmitters,
- telecommunication satellites,
- radio navigation satellites,
- ground based radio navigation resources.

1.1.1 Clock transport

This means of comparison requires moving a clock (cesium or rubidium frequency standard) from one place to another and then, after studying the relative behaviours, measuring and calculating the time deviation between a reference and a laboratory. The measurement accuracy is mainly linked to the time deviation between the measurements on each site. This accuracy can generally be estimated at +/- 20 ns.

1.1.2 Television signal

1.1.2.1 Microwave

In France, the signals broadcast by TDF contain a time reference, which is a specially coded line in the television frame. The Paris Observatory calibrates and measures the time deviation between a special mark transmitted on the TV frame and the UTC (OP). Using an appropriate receiver, the user can also measure the deviation between his time reference and reception of the television signal. After the Paris Observatory laboratory analyses the results, and corrects TV signal propagation time delays, it distributes the time deviations for the clocks measured. When a number of laboratories are connected via the same television repeaters, the accuracy of the results is generally greater than 10 ns (this is chiefly the case in the Paris region, where accuracy can reach 2 ns).

1.1.2.2 Television satellite

Experiments are in progress to examine the possible uses of these satellites for time measurement. The Besançon Observatory is at the present time carrying out work, for which the expected time transfer results will be about 10 ns after data processing, to take account of satellite trajectory calculation corrections.

1.1.3 Radio transmitters

In France, the France Inter longwave transmitter (162 kHz) is phase modulated to transmit time information: day, hour, minute and second, and the time signal (representing one second). The use of receivers requires initial setting in order to correct these various delays. After setting, it is estimated that the time can be monitored with an accuracy of +/- 200 ns, provided that the measurements are averaged over at least 10 consecutive acquisitions.

1.1.4 Telecommunication satellites

The use of these satellites gives extremely good results. Nonetheless, this means of transferring or measuring the time deviation requires the use of special equipment which is hard to transport from one site to another. The EUTELSAT and INMARSAT satellites have been used for experiments between various laboratories and institutes. The results so far obtained experimentally give an accuracy of up to 10 ns.

1.1.5 Radio navigation satellites

The TRANSIT System
This system transmits time data with the navigation message. The accuracy of this system is often inadequate, and is given at 150 s.
This system is reaching the end of its useful life.

The GPS system
The GPS system was initially dedicated to radio navigation. Satellites transmit the information needed to calculate a position, altitude and time. The appropriate receivers output a representative UTC(USNO) signal as defined in the ICD 060 specification (GPS system time interface). The absolute accuracies expected for this system are as follows:

- code P: 176 ns at 2 σ (95%)
- code C/A: 363 ns at 2 σ (95%)

1.2 REQUIREMENTS

These days, it is frequently common to have highly stable and accurate time references, whether or not permanent. To be usable, they must be interconnected either physically, or by calculation.

The easiest means at present is the GPS system, but it belongs to the USA Department of Defense (D0D). Furthermore the signal degradation (SA mode) of block II and IIR satellites will also affect the time measurement results.

Current measurements show that with time information degradation it is possible to obtain correct times in "common views" mode with errors of less than 10 ns. However, with the possible degradation of ephemeris data in SA mode, no figures are given.

1.3 TIME/FREQUENCY USE OF LORAN C

All the LORAN C signals have been linked to a common time frequency by means of Times Of Coincidence (TOC), since 1st January 1958. Analysis of the times of arrival of these signals thus gives a reference time.

The "TOE CONTROL" synchronization mode (time of emission) enables relatively "tight" slaving of the LORAN C time of emission to the reference time.

For the French system (GRI: 8940), the time reference is generated by a high-accuracy clock using four cesium frequency standards, corrected by a calculation algorithm.

The behaviour of the UTC(B) time scale is compared with the other time scales in figure 1 below.

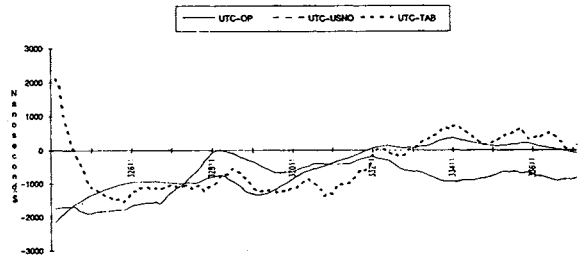


Figure 1 - time scale comparison in relation to UTC

The fixed point measurements thus taken indicate very good stability of the LORAN C signals. The LORAN C control centre in fact uses two computers to ensure perfect synchronization of the signals transmitted with the time reference, according to the following specifications:

Lessay - UTC(BREST) < 50 ns at 1 sigma (66%)
Lessay - Soustons < 30 ns at 1 sigma (66%)

1.4 PROSPECTS

Following the decision by the American DOD to cease operations in stations outside America, and to hand them over to the host countries as of 1st January 1995, a number of countries formed a group to take over and develop the LORAN C system in Europe. An agreement protocol was signed by 6 European countries on 6th August 1992. The project drawn up by the working group led to a structure of 9 transmission stations split into 4 LORAN C systems. These stations will all be synchronized using the "TOE Control" mode with the same time reference: UTC(B), which will be the physical representation of the UTC time in Brest.

Given the announced specifications, the transmissions of each station will therefore be linked to the time reference with the following accuracy:
Station - UTC(B) < 100 ns at 1 σ (66%)

With an appropriate receiver, it will therefore be possible throughout the coverage area to be linked to the same reference. This coverage area will be far greater than the navigation area, as it is enough to be able to receive a single transmitter in the network (see enclosed coverage area).

2 - RECAP OF LORAN C

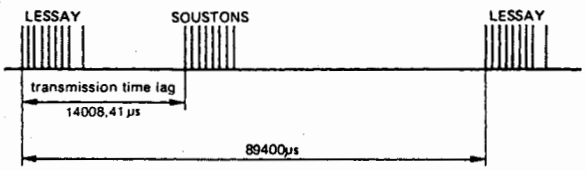
LORAN C is a low frequency radio navigation system (LOng RAnge Navigation). The carrier frequency is: 100 kHz.

This is a land-based system organized into chains. Each chain usually comprises a master transmission station and 2 to 4 secondary stations.

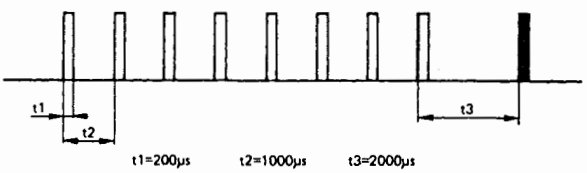
The user of this system can, with an appropriate receiver, obtain his position on the Earth. The operating modes can differ depending on whether the receiver uses the signals from 2 transmitters (circular mode still called "rho-rho-mode" or 3 transmitters in the same chain (normal hyperbolic mode) or 4 transmitters in 2 different chains(cross-rate hyperbolic mode) to calculate its position.

2.1 LORAN C SIGNALS

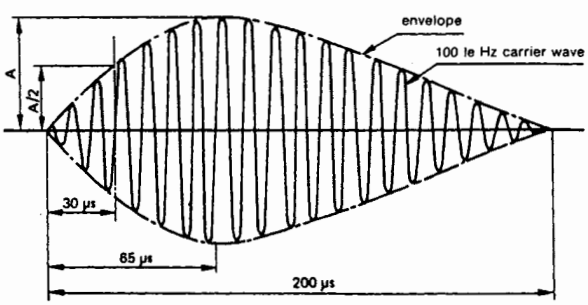
The LORAN C signals transmitted comply with a basic specification drafted by the USCG: M 16542-4 July 1981.



2a - Sequencing



2b - Pulse group



2c - LORAN C pulse

Figure 2 - LORAN C signals

Reminder of main characteristics:

- Carrier frequency: 100 kHz
- Pulsed signals: (see figure 2)
  - master station: 9 pulses
  - secondary stations: 8 pulses
- Pulse form such that 99% of the energy is within the band: 90 to 100 kHz.
- Chain recognition: GRI (Group Repetition Interval).
- **Phase Code**: this is the sign of the first half-wave of the LORAN C signal. This code also allows recognition of a master station from a secondary station.
- The **blink** consists in blinking either the 9th pulse of a master station, or the first two pulses for a secondary station.
- **Range**: this depends on the power of the transmitter and the sensitivity of the receiver. For a 250 kW transmitter, it is about 1000 NM for -10dB reception at sea.

2.2 SYNCHRONIZATION

Synchronization of a LORAN C chain follows two principles:

The **SAM mode** (System Area Monitor). This mode allows synchronization of the secondary stations with the master stations. An SAM station located in a triad allows measurement and checking of the hyperbolae intersection point in relation to the reference point. As a general rule, this mode is manual. It is today used in most chains around the world.

The **TOE Control mode** (time of emission control): this mode is used to control the time of emission from the stations in relation to the master station and in relation to a reference time. This control is therefore per station and per chain, in order to maintain the EDs (emission delays) constant for the given chain. The French chain is synchronized on this principle, as will be the case, from 1995, for the Northern Europe chain.

2.3 TIME OF EMISSION CONTROL

The time of emission control is performed by a LORAN C **control centre** (figure 3). This dialogues permanently with the transmission stations via communication resources (leased lines or packet transmission).

Each site (transmission station and control centre) supplies the computer with the results of measurements made using appropriate receivers. Simulators, synchronized with the transmission stations and reference time, transmit reference LORAN C signals to each site. The control centre automatically calculates the phase corrections to be made on each transmitter in order to keep the network synchronized with the reference time.

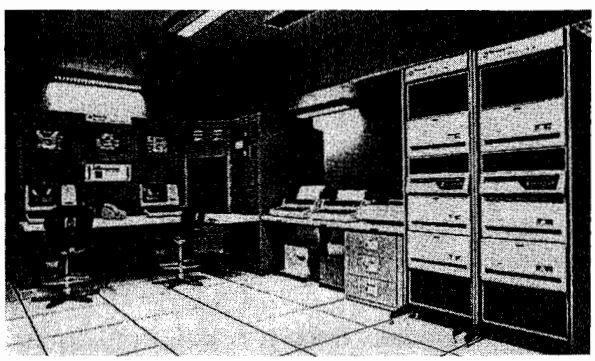


Figure 3 - French chain control centre GRI = 8940

For the French chain (GRI 8940), the specifications are as follows:

- The reference time is the TAB (Brest atomic time) whose physical representation is UTC(B).
- The error tolerated on master station transmission (LESSAY) in relation to TAB is < 50 ns at 1 σ.
- SOUSTONS synchronization in relation to LESSAY must be better than 30 ns at 1 σ.

2.4 USES

Each transmission station is locally synchronized by two cesium clocks, which give extremely stable transmitted signals. The use of these cesium clocks, together with automatic control by the control centre, guarantees transmission of reference time synchronous signals. The phase corrections made to the transmission stations enable the LORAN C transmissions to be kept synchronized with the TOC (Time Of Coincidence). For stations synchronized in TOE mode, the fine phase corrections are of 10 ns.

Using an appropriate receiver, it is possible to make full use of this system. The use of ground wave stability (100 kHz) over a distance of 1000 NM with a sea path gives a frequency of stability better than $1*10^{-12}$ for 1 day, if an average is calculated.

The LORAN C time/frequency receiver would therefore enable a reference time to be obtained in a known geographical site of better than one microsecond, right from startup. Residual errors would be mainly due to LORAN C signal propagation errors between the transmitter and the reception site. The use of calculation methods or the performance of initial calibration would correct this propagation error. There would then only remain a measurement noise linked to the specific nature of the wave propagation phenomena (seasonal effects, day/night transition, interference, etc.).

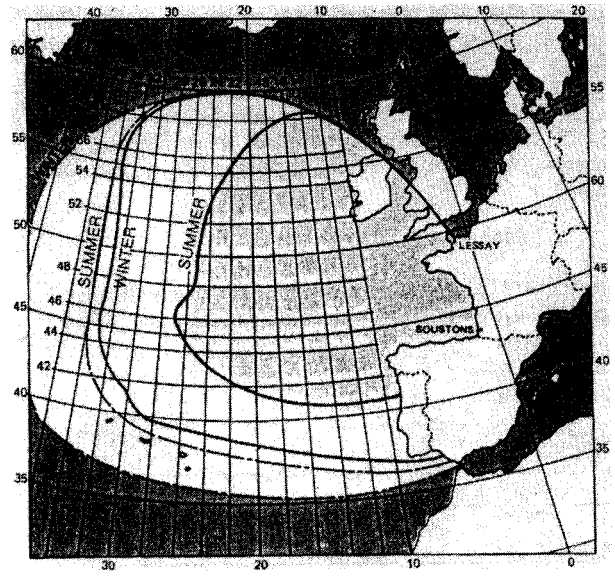
A receiver of this type would therefore give:

- a stable frequency (dependent on the stability of the LORAN C signals on the reception site),
- a time reference (based on the time reference of the LORAN C chain followed):
  - for the USCG chains: UTC(USNO)
  - for the French chain: UTC(B)
  - tomorrow for the European network: UTC(B)

2.5 COVERAGE

This type of receiver is generally used in a laboratory or fixed station. It only uses a single transmitter for time or frequency slaving.

For the North Atlantic, coverage is today that given by the American stations (reference: UTC(USNO)) and the two French stations (reference: TAB) (figure 4).



Signal/ambient noise ratio = - 10 dB (conventional receiver)
Signal/ambient noise ration = - 25 dB (improved receiver)

Figure 4 - Coverage of the French chain

With the "Northwest European Loran C System", the coverage is given in figure 5. The NELS reference time will then be UTC(B).

The position of the UTC(B) reference time will be observed and controlled by the control centre. A server will later inform users of deviations from the other time scales on a daily basis (UTC(USNO) in particular).

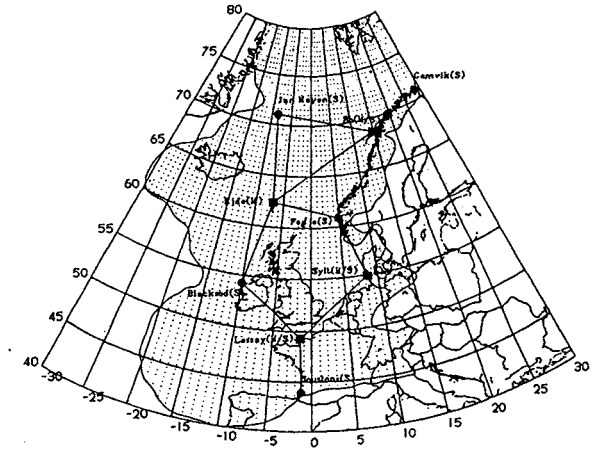


Figure 5 - Coverage of the LORAN C European System

2.6 LIMITATIONS/POTENTIAL

2.6.1 Time/frequency receivers

Specific time/frequency LORAN C receivers are at present few in number:

AUSTRON (2100)
Stanford Research (FS 700)
Locus (Lad LORAN LRS)

2.6.2 Availability

The availability of the LORAN C signals is given at 99.9% per transmitter and 99.7% for a chain. The same applies to the system reference time (UTC(B)): 99.9%.

3 - USE OF LORAN C SIGNALS FOR TIME/FREQUENCY APPLICATIONS

3.1 PROPAGATION OF LORAN C SIGNALS

The LORAN C wave with a frequency of 100 kHz is influenced by the radioelectric properties of the earth and the ionosphere. In LORAN C propagation, the sky wave must be differentiated from the ground wave.

The first is the result of reflection from the ionosphere, and is of low stability owing to ionospheric variations. It is therefore not generally used.

The ground wave comprises a space wave and a surface wave. The space wave is the combination of the direct wave (direct path between the transmitter and the receiver) and the reflected wave (wave received by the receiver reflected from the ground). The surface wave propagates over the surface of the earth in an interface zone consisting of an insulating medium (air) and a conducting medium (surface). Figures 6 and 7 show the different waves. In LORAN C, when the transmitting and receiving antennas are located close to the surface of the earth, the components (direct and reflected) of the space wave may cancel each other out and the LORAN C signal is then transmitted by surface wave.

Furthermore, to differentiate the sky wave from the ground wave, the LORAN C system uses pulse coding. This ensures that it is the ground wave that is measured.

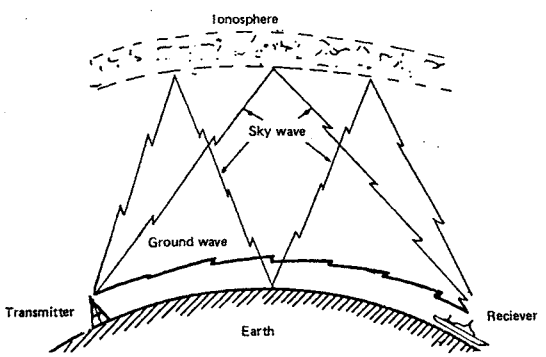


Figure 6 - Presentation of direct waves

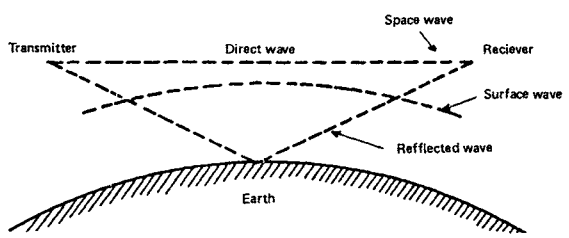


Figure 7 - Presentation of reflected waves

3.2 MEASUREMENT METHODS

Measurement of LORAN C signal propagation times requires slaving to the received signals. The reference point for time measurement is the rising edge of the sixth zero transition for a positive phase code pulse.

The LORAN C measurement is made on the ground wave phase. The measurement accuracy is therefore linked to knowledge of its properties.

If we consider an arc length, D, in metres, for which we wish to know the path time TP in microseconds, the problem entails measurement of the total phase (FT) of the LORAN C wave, which is the sum of three terms:

$\Phi T = PF + SF + ASF$
where: PF = Primary phase factor
SF = Secondary phase factor
ASF = Additional secondary phase factor

The path time (TP) of the wave is then given by:

$$TP = \frac{\Phi T}{2\pi f} \text{ in s}$$

or:

$$TP = \frac{\Phi T}{2\pi} \text{ in } \mu s$$

where f = LORAN C frequency: 100 kHz

We therefore obtain: $TP = T_{PF} + T_{SF} + T_{ASF}$

$$T_{PF} = d/C$$

where d = distance between transmitter and receiver

$$C = Co/n$$

where n = index of refraction in air

T_{SF} is the path time in the air-seawater interface.
The T_{PF} and T_{SF} terms are generally calculated by a model called the "SALT Model". This model takes account of the radioelectric properties of the air-seawater medium. When the LORAN C wave propagates in a different medium, an additional phase correction is required: T_{ASF}.

The T_{ASF} term can be defined as the correction to be made to a homogeneous path for the LORAN C wave propagating in a medium other than seawater. The T_{ASF} time can be determined in various ways:

- integral equation method,
- equivalent impedance method,
- Millington method,
- practical method.

The Millington method is that most commonly used and in practice proves to be accurate enough. The calculation of T_{ASF} is generally limited by a lack of information on the media crossed. In practice, this term can be measured by a time transfer between the LORAN C time reference and the measurement site.

STABILITY OF THE LORAN C WAVE

At a given point, the stability of the LORAN C wave is linked to the characteristic impedance of the media crossed. The ground is not only characterized by its nature but also by its humidity, temperature, general geological formation and equivalent wave penetration depth.

In practice, the variation in all these parameters leads to seasonal effects whose amplitude can reach several hundred nanoseconds. Figure 8 presents the LORAN C wave variations observed between the LESSAY station and the Paris Observatory, for a period of 1 year.

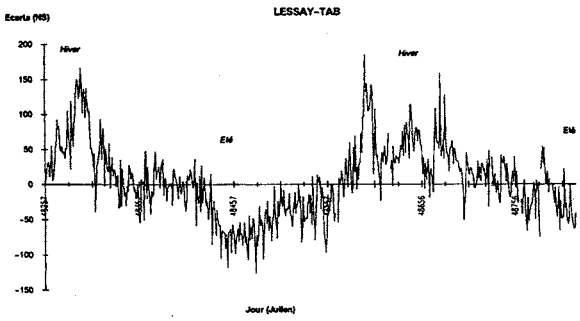


Figure 8 - Observation of seasonal ASF variations between LESSAY and PARIS (deviation less than 100 ns)

3.3 MEASUREMENT RESULTS

The DCN BREST has performed a large number of maritime and shore- based measurements of LORAN synchronization in the various North Atlantic and Mediterranean chains, for a variety of requirements, from the French Navy to a number of working groups.

The time of emission from each transmission station is measured by a LORAN C receiver in circular mode synchronized on UTC(USNO) time via the GPS system. The path times measured are compared with the times calculated from a point supplied by the GPS receiver, into which should be integrated the various calibration delays. To calculate the exact times of transmission from the USCG stations, the time variations supplied by UTC(USNO) in the "series 4" type publications, should be included.

Figure 9 presents the existing bias on the North Atlantic systems which should be added to the "series 4" file values to obtain the exact times of emission in relation to UTC(USNO). The deviations observed are due on the one hand to the control mode of these LORAN C stations (SAM mode) and on the other to the "series 4" file production method.

3.4 USABLE CHAINS

The use of one or other chain will depend on the one hand on the reception in the zone, and on the other on the ability to obtain the time corrections to be made, in order to lock onto a known time scale.

RESULTS 1991	9980M SANDU	9980X EDJE	9980W ANGIS	7970M EDJE	7970Y SANDU	7930M FOX H	7930X ANGIS				5930Y CAP R	
MAX-MIN	1,6	0,5	0,1	0,8	0,5	0,3	0,3				0,5	
AVERAGE	-2,1	-2,6	-2,3	-3,3	-3,4	-2,7	-2,9				-7,1	

RESULTS 1992	9980M SANDU	9980X EDJE	9980W ANGIS	7970M EDJE	7970Y SANDU	7930M FOX H	7930X ANGIS	7930W CAP R	5930M CARIB	5930X NANTU	5930Y CAP R	9960X NANTU
MAX-MIN	1,0	0,4	0,4	0,6	0,8	0,4	1,0	0,5	0,6	0,2	0,2	0,7
AVERAGE	-2,6	-2,4	-2,8	-4,8	-5,0	-2,6	-2,4	-2,0	1,8	2,1	3,0	2,1

COMPARAISON	9980M SANDU	9980X EDJE	9980W ANGIS	7970M EDJE	7970Y SANDU	7930M FOX H	7930X ANGIS				5930Y CAP R	
AVE.92 - AVE.91	-0,5	0,2	-0,5	-1,6	-1,6	0,1	0,4				10,1	

Figure 9 - Time deviations in the "series 4" UTC(USNO) files in relation to UTC(USNO) in 1991 and 1992

3.5 COMPARISON WITH GPS

Evaluations are in progress at DCN BREST. Our aim is to show that the LORAN C system can give the same time/frequency accuracy as GPS, to which it could become a useful complement.

4 - POTENTIAL USERS

4.1 FRENCH CHAIN

Today, many DGA laboratories already use LORAN C signals as a reference.

4.2 FUTURE EUROPEAN NETWORK

When the European network is commissioned, on 1st January 1995, LORAN C coverage will enable numerous laboratories to connect to a common reference time. These applications will be possible in the civil, scientific and military fields.

5 - CONCLUSIONS

The land coverage area of LORAN C is interesting. Western Europe is in effect fully covered and could be extended North-East by linking up with the CHAYKA network (the Russian equivalent of LORAN C). Thus LORAN C can offer any user the same information as GPS, admittedly with poorer location accuracy, but with comparable Time/Frequency accuracy.

LORAN C should not be presented as an alternative to GPS, and the two means should complement each other. There is no doubt that dual capacity receivers will soon be marketed, offering the user improved or even total integrity.

While LORAN C is hard to deceive, the same is not true of GPS, as was intended by the United States. LORAN C is jammable (at atmospheric noise level), while GPS is less so. LORAN C is sensitive to reception conditions (mountains, forests, etc.) while GPS requires direct view between satellites and receivers. By setting up the European network, LORAN C will be under the control of the European countries.

In the Time/Frequency field, LORAN C can be used to good advantage by the scientific community, but also by all those (including the military) who need a precise time reference of a few hundred nanoseconds.

THE FIRST MONTHS OPERATION OF A MASER CH1-75 AND TWO HP 5071A CESIUM.

F. COLLIN, P. PAQUET

Royal Observatory of Belgium

Abstract

Since July 1992 a russian Maser CH1-75 is operating at the Royal Observatory of Belgium
It is continuously compared to the local time scale UTC(ORB) and atomic clocks.
This paper will report about its behaviour and stability during the 8 first months of operation.

In July 1992 a Maser CH1-75 produced by KVARZ (Russian) has been turned in continuous operation. To get the best performance of this unit KVARZ recommends to use, at the same site, two masers to perform an automatic auto-tuning of the cavity.
As only one maser is in operation at the Royal Observatory of Belgium, in principle the auto-tuning is not operational. However, in the same laboratory, a BVA quartz oscillator is in operation since several years. Its short term stability is of the order $5\,10^{-13}$ over time interval of 30 seconds; for this reason, KVARZ suggested to experiment to tune the CH-75 with an external frequency provided by the BVA. In consequence once a week the tuning has been applied for about 10 hours. Typical frequency changes were of the order of $1.2\,10^{-13}$.
During the 8 months of operation and with respect to GPS (fig1), the normalized frequency of the CH1-75 can be subdived in three periods:

- MJD = 48820 to 48865, $\frac{\Delta f}{f} = -7.6\,10^{-14}$
- MJD = 48865 to 49042, $\frac{\Delta f}{f} = 1.6\,10^{-13}$
- on MJD = 49040 the auto-tuning mode was full-scale. In that case a manual frequency readjustement is mandatory; it took few days to readjust properly the frequency.

In a next period of experiment we intend to leave the maser in free runing mode; we expect to reach a better short and long term stability.

The Allan variance is the following (fig1):

1. with respect to GPS (MJD = 48820 to 49042), $\sigma = 1.6\,10^{-13}$, $\tau = 1$ day.
2. with respect to UTC (MJD = 48820 to 48999), $\sigma = 0.9\,10^{-13}$, $\tau = 10$ days.

It is interesting to compare the results above with the performances of two new Cesium HP-5071A. Their drift with respect to GPS is given in Figure 2; their Allan variances are respectively $1.27\,10^{-13}$ and $1.36\,10^{-13}$ for $\tau = 1$ day.
Figure 3 is the evolution of one HP 5071A with respect to the other one; the figure on the top reports one measurement per day ($\sigma = 0.95\,10^{-13}$) while the figure at the bottom refers to one measurement per hour ($\sigma = 5.7\,10^{-13}$).
Comparing the Allan variances of the maser CH1-75 and the two cesiums we expect that the BVA quartz oscillator did not match the performance to allow a correct tuning of the maser. This is the reason why we shall experiment, for few months, a free runing mode of the maser.

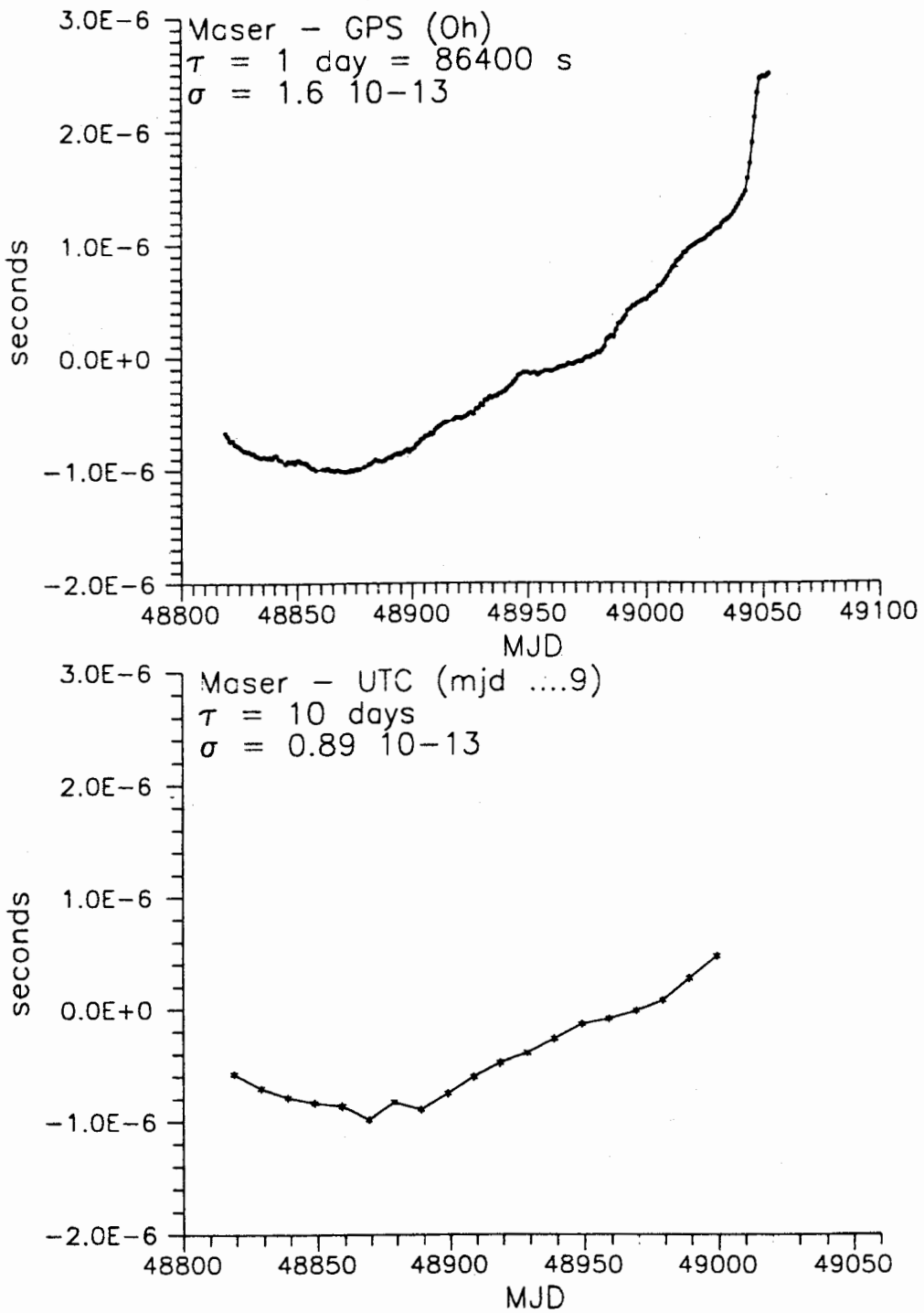


Figure 1:

Top: drift of the Maser CH1-75 with respect to GPS, three periods are to be considered for the drift:

(1) Starting period till MJD = 48865, negative time drift of $-6.6\,10^{-9}$ seconds by day.

(2) Between MJD = 48865 till 49042, positive time drift of $1.4\,10^{-8}$ seconds by day.

(3) On 49042 manual frequency readjustement.

Bottom: Drift of the Maser with respect to UTC.

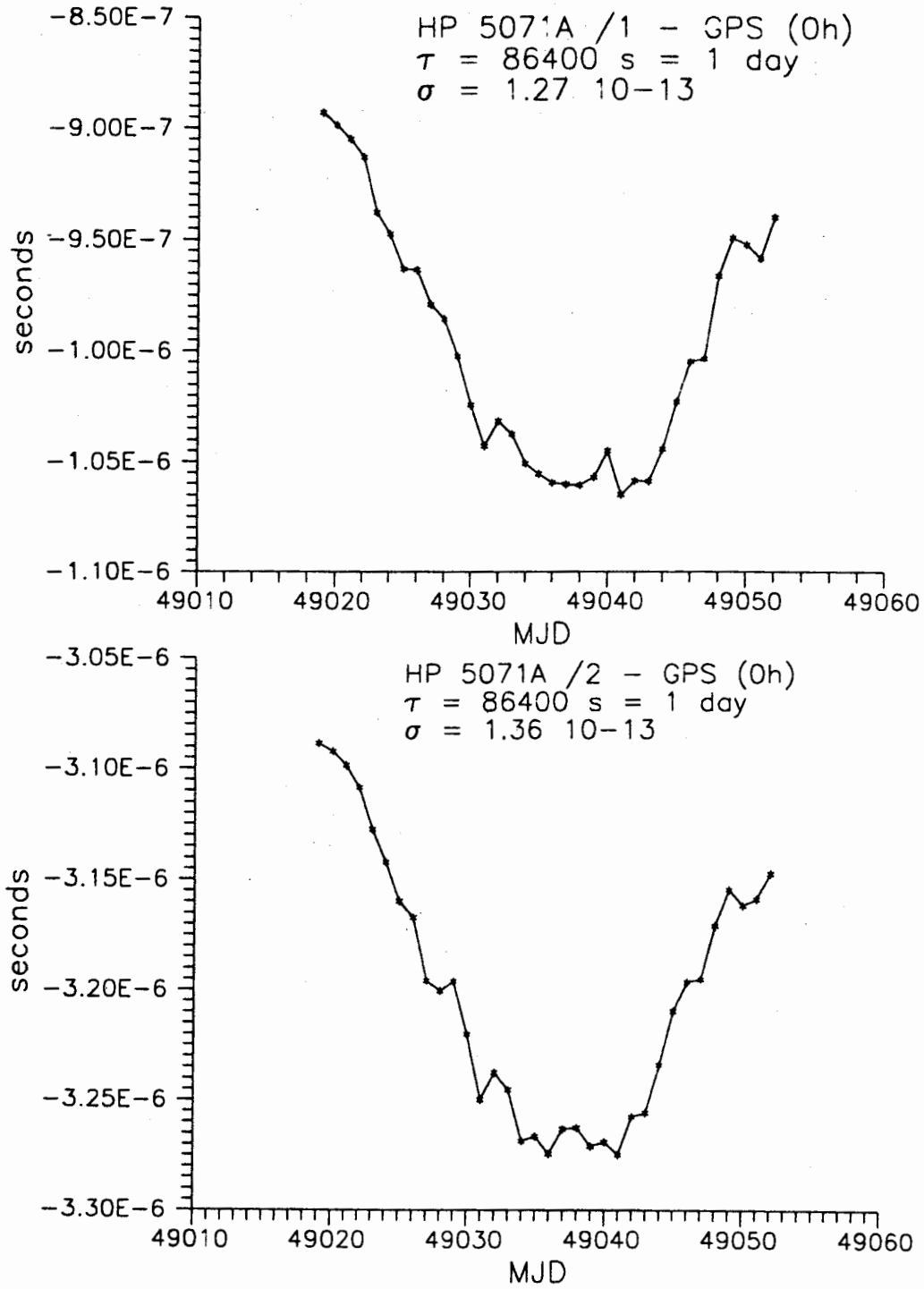


Figure 2: During the first month of operation, drift of two cesium HP 5071A with respect to GPS. σ is the Allan variance and τ is the sampling period.

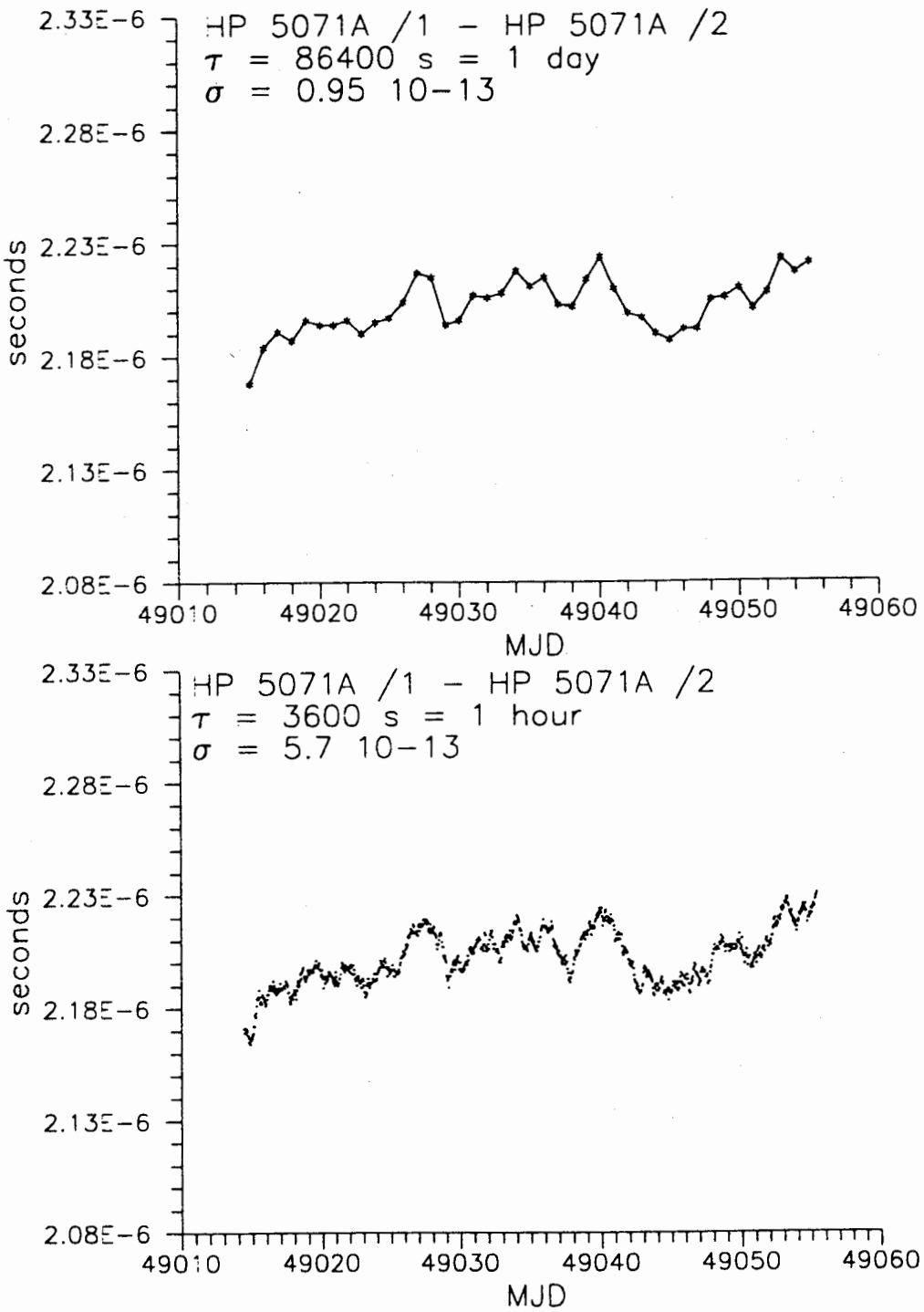


Figure 3: Comparaison of two Cesium HP 5071A with respect to each other.

Top: one measurement every 24 hours.
Bottom : one measurement every hour.
(σ is the Allan variance, τ the sampling period)

TRANSFER OF REFERENCE SIGNALS FOR SPACE VLBI : SHORT AND MEDIUM TERMS

G. Daigne

Observatoire de Bordeaux, F - 33270 Floirac

ABSTRACT

Two space VLBI missions, Muses-B and Radio Astron, are being prepared for launch in 1995-96. High stability of the on-board oscillators will be achieved by the transfer of a reference signal given by H-masers at tracking stations. The coherent two-way transfer, with a transponder in space, will provide both accurate Doppler measurements and correction for atmospheric fluctuations, at least in the medium term. On short time scales, smaller than a few times the one-way transit duration, the correction is no longer possible and we underline the crucial role of the on-board oscillator to ensure the short term phase stability of the receiver.

Two-way transfer, Space VLBI

1. IINTRODUCTION

The frequency and phase stability given by H-Maser clock oscillators has been the key factor for the development of ground-based VLBI, up to 100 GHz or more in frequency. The duration of the coherent integration time is in fact determined by path length fluctuations of the incoming radiation through the atmosphere and/or ionosphere ; it is typically about a few tens of minutes in the centimeter range. The situation is a priori more favorable with one of the radiotelescopes of the array placed into orbit. But for the two missions being prepared for launch in 1995-96, the available technology will not allow to place a H-Maser on board. The frequency stability of the system will then rely on the transfer of a reference signal from ground through a two-way coherent link as shown in Fig. 1. This link will also provide with a very accurate Doppler measurement. It will be implemented in the X-band for RadioAstron (Ref. 1) and in the X- and Ku-band for Muses-B (Ref. 2).

Two-way coherent links are in common use for Doppler measurements and tracking of deep space probes. However they don't seem to have been used for transferring time signals with very high short-term stability, and we develop this kind of analysis in a simple model where the satellite is supposed to have no Doppler. We first recall the stability characteristics required for each element of a non connected interferometer.

2. THE LOSS OF COHERENCE

The phase and frequency fluctuations of each independent oscillators of an interferometer array contribute to a loss of coherence, that is to a loss in sensitivity and to larger uncertainties in the measured fringe amplitude. The coherence function of a heterodyne interferometer has been analysed in detail (Refs. 3, 4).

A simple expression for the loss of coherence, $1 - \rho_c$, is obtained with white phase noise ; the contribution of one arm of the interferometer is, for small values of the loss :

$$1 - \rho_c \simeq \frac{2}{3} \pi^2 \nu_0^2 \alpha \qquad (1)$$

where ν_0 is the observed radio frequency, and α is the Allan variance at 1 second. The upper observed frequency being 22 GHz, a loss of coherence of 1 % is obtained for $\alpha = 2.9\ 10^{-24}$.

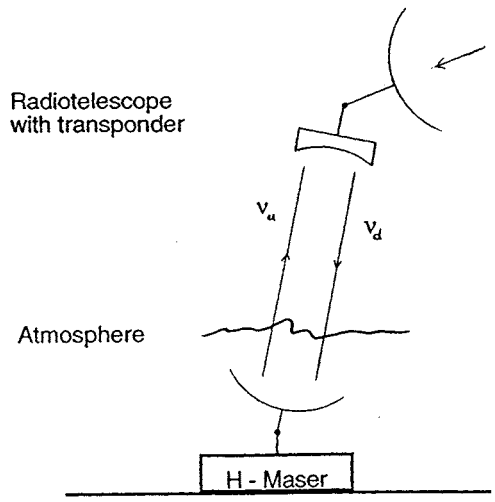


Figure 1. The two-way coherent link for the transfer of a highly stable reference signal to the radiotelescope in space.

A convenient way to visualize the different contributions of an oscillator chain to the coherence losses is to show the Allan standard deviation of the system in a log-log plot as in Fig. 2. The losses will remain smaller than 1 %

as long as the different contributions remain below the straight line with slope – 1 which represents 1 % loss for white phase noise.

3. THE TWO-WAY LINK

3.1 Description and working hypothesis

A two-way link from ground to space, with a coherent transponder on board, is used for precise Doppler measurements. Corrections that have to be made are related to the dispersive effect of the ionosphere and to the fact that the upward and downward frequencies, ν_u and ν_d respectively, are in different frequency bands. The measured Doppler is an average value, on durations greater than 60 s, and its accuracy is expected to be a few 10^{-13} or 0.1 mm/s.

Another purpose of the two-way link is to transfer a reference signal from ground to space, and we shall further analyze the effect of short term path length fluctuations on the stability of the transferred signal.

The velocity or movement of the satellite will be considered to be constant on durations of a few tens of seconds in order to decouple the equations for Doppler measurement and transfer of a time scale. To simplify the mathematical analysis we shall take it equal zero. We suppose that the response of the transponder is instantaneous, without transmission noise. We consider only the atmospheric effects on path fluctuations, neglecting ionospheric effects. The propagation medium being non dispersive, and the phase and group velocities equal, we can consider the propagation of clock ticks as well as propagation of the phase of pure carrier frequencies.

3.2 Spectral analysis of the two-way correction

Let t be the coordinate time of a reference system where the clock on ground is at rest,
t_n , t_n^s and t_n^r times for which clock tick n is respectively transmitted from ground, received at the satellite and received back to ground,
Δt the one way (average) transit duration,
$\tau(t)$ the variation of the transit time mainly due to fluctuations of the wet component of the atmosphere, close to the ground transmitter.
$\tau(t)$ is supposed to have zero mean,
$\gamma(f)$ is the Fourier transform of $\tau(t)$.

The times of arrival of the clock tick n at the satellite and back to ground are :

$$t_n^s = t_n + \Delta t + \tau(t_n) \tag{2}$$

$$t_n^r = t_n + 2\Delta t + \tau(t_n) + \tau(t_n + 2\Delta t) \tag{3}$$

the measured departure to $2\Delta t$ of the transit time is :

$$\delta t_n = \tau(t_n) + \tau(t_n + 2\Delta t) \tag{4}$$

In the two-way correction mode for the transfer of a time scale to the satellite, we say that the time of arrival of clock tick n has to be corrected with

$$\delta T_c(t_n + \Delta t) = 1/2\ \delta t_n \tag{5}$$

Indeed, as long as the process responsible for path length fluctuations is random we cannot formulate any other hypothesis than the equipartition one : half of the two-way measured delay is due to upward propagation and half to downward propagation.

Let $\delta\Gamma_c(f)$ be the Fourier transform of δT_c,

$$\delta\Gamma_c(f) = \gamma(f) \cos(2\pi f\, \Delta t) \tag{6}$$

For short term fluctuations, with $f \geq 1/4\Delta t$, we obtain zeros in the frequency spectrum of the correction to be applied to the transferred signal, whereas the transferred signal should be corrected as long as $\gamma(f)$ is not vanishing.

Another way to imagine this filtering effect of the two-way link for time transfer is to realize that for fluctuations with $f_0 = 1/4\Delta t$ and its odd harmonics, the path delays $\tau(t_n)$ and $\tau(t_n + 2\Delta t)$ have opposite phases and cannot be estimated through a measurement of their sum.

The correction brought with δT_c to the arrival time of the clock tick at the satellite is in error of :

$$\delta T_s(t_n + \Delta t) = \frac{1}{2}[\tau(t_n) - \tau(t_n + 2\Delta t)] \tag{7}$$

whose frequency spectrum is given in terms of the frequency spectrum of the measured correction :

$$\delta\Gamma_s(f) = -j\, tg(2\pi f\, \Delta t)\, \delta\Gamma_c(f) \tag{8}$$

In this simple model, the correction term measures atmospheric fluctuations without transmission noise. With (6) and (8) we can write the relation between the density spectrum of the phase fluctuations at the satellite $S_\phi^s(f)$ and the density spectrum of the atmospheric phase fluctuations $S_\phi^a(f)$ in the two regimes, with or without two-way correction :

$$S_\phi^s(f) = S_\phi^a(f) \sin^2(2\pi f\, \Delta t) \qquad \text{for } f < 1/4\Delta t \tag{9}$$

$$S_\phi^s(f) = S_\phi^a(f) \qquad \text{for } f \geq 1/4\Delta t \tag{10}$$

3.3 A model for the atmospheric fluctuations

With a Kolmogorov spectrum for the turbulence, the structure function of the refractive index at two points with distance R takes the form $C^2R^{2/3}$. The spatial fluctuations are converted into temporal fluctuations above the transmitting antenna through the frozen pattern hypothesis. The wind velocity V, the value of the coefficient C and the scale height h of the wet atmosphere are taken from detailed investigations of atmospheric fluctuations at transmitting antenna sites (Ref. 5). Within this model, the effect of the wind direction relative to the projected direction of the satellite, and the effect of the elevation angle of the satellite has been further studied (Ref. 6). We have plotted on Fig. 2 the Allan standard deviation of the atmospheric fluctuations for a wind velocity perpendicular to the satellite direction, and elevation angles of 10° and 60° respectively, with :

V = 8 m/s
h = 1000 m
$C = 2.4\ 10^{-7}\ m^{-1/3}$

The Allan standard deviation is nearly flat, with slope

−1/6 in the range 1 to 100 seconds. The range for two-way correction is shown on Fig. 2 with a one-way transit duration Δt = 0.25 s. The turning point between no correction and two-way correction is at τ_0 = 1 s. The curves remain below the line showing 1% loss of coherence at 22 GHz, and the benefit of the two-way correction for $\tau > \tau_0$ is clearly seen.

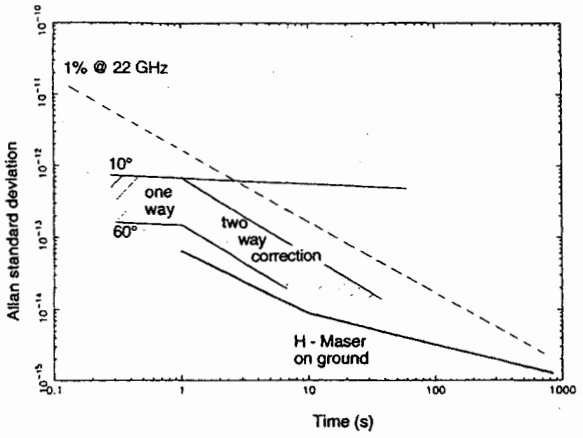


Figure 2. Correction of the effects of atmospheric fluctuations in a coherent two-way link with transponder in space.

3.4 Stability of the on-board oscillator

On time scale smaller than τ_0 , the stability of the on-board oscillator is given either by a high quality quartz oscillator or by the signal transferred from a H-Maser on ground. In the first case, the stability of the quartz oscillator has to be 10^{-11} at 0.1 second, or 10^{-12} at 1 second for a loss of coherence less than 1% at 22 GHz. In the second case, there is no special requirement on the stability of the quartz oscillator, except to match the characteristics of the next element in the oscillator chain.

There are two main differences between the two cases : with an Ultra Stable Oscillator, the signal to noise ratio of the upward link can be reduced accordingly, a second phase lock loop with narrow band controlling the USO. Furthermore, in adverse atmospheric conditions, the loss of coherence could be measured.

4. CONCLUSION

In this simple model we have not considered the ionospheric perturbations nor the effect of the uncertainty in the radial velocity of the satellite. In the frequency bands chosen for the link, the effect of ionospheric fluctuations will certainly be negligible on short time scales, at about 1 second. In the medium-term, where the two way correction would apply, it may contribute and reduce the stability of the transferred signal. On the other hand, an error in the Doppler measurement is a source of uncertainty in the transferred time scale. As long as the radial velocity can be taken as constant in the considered range of fluctuations, the transferred time scale is at most affected to the extent of the inaccuracy in the Doppler measurement. Nevertheless, a complete analysis has to consider both effects, and a detailed knowledge of the procedure for Doppler measurement would be required.

The main result of this contribution is the inadequacy of the two-way link for the transfer of signals which have to be highly stable in the short term. Signals transferred from ground can be disturbed by short time fluctuations with no way for correcting them, and even detecting them. On the other hand any movement of the satellite is observed as a perturbation on the two-way transit duration. This unbalanced sensitivity of the link is due to its asymmetry, the transponder being in space and not on the ground.

5. REFERENCES

1. Kardashev, N.S. & Slysh, V.I. : 1988, "The RadioAstron Project" in The Impact of VLBI on Astrophysics and Geophysics, M.J. Reid and J.M. Moran (eds), IAU Symposium
2. Hirabayashi, H., Inoue, M. & Kobayashi, H. : 1991, Frontiers of VLBI, Universal Academy Press, Inc., Tokyo
3. Rogers, A.E.E. & Moran, J.M. : 1981, *Trans. Instr. and Meas.*, Vol IM-30, **4**, 283
4. Thompson, A.R., Moran, J.M. & Swenson, G.W., eds : 1986, Interferometry and Synthesis in Radioastronomy, John Wiley and Sons, New York.
5. Treuhaft, R.N. & Lanyi, G.E. : 1987, *Radio Sc.* **22**, 251
6. Trama, J.C. : 1992, stage de DEA, Bordeaux Observatory

SAW 1

Chairman: J. Heighway

STATE-OF-THE-ART SAW OSCILLATORS: DESIGN AND PERFORMANCE

G. K. Montress and T. E. Parker

Raytheon Research Division
131 Spring Street
Lexington, Massachusetts 02173
USA

ABSTRACT

State-of-the-art phase noise performance has been demonstrated for both surface acoustic wave (SAW) resonator and delay line stabilized oscillators [Refs. 1 - 5]. The same basic feedback-loop oscillator design philosophy [Ref. 6] was applied in each case in order to achieve these results. This paper details the design, fabrication, component selection, and performance of extremely low-noise SAW resonator and delay line based oscillators. It is shown that a basic 50-Ω modular component, feedback-loop oscillator architecture incorporating a two-port frequency stabilizing device is easy to set-up, and perhaps much more importantly, is easy to trouble-shoot if something should go wrong.

Keywords: *SAW Oscillators, Low-Noise Frequency Sources, Component Residual Phase Noise Measurements, Feedback-Loop Oscillator Design*

1. INTRODUCTION

The residual phase noise properties of an oscillator's electronic components, for example RF amplifiers and electronic phase shifters, are extremely important if reproducibly low white PM and flicker FM oscillator noise levels are to be achieved. For SAW oscillators, it goes without saying that the residual flicker noise of the SAW resonator or delay line must also be minimized if truly state-of-the-art performance is to be achieved. For these reasons, consideration is given to residual phase noise measurement techniques which may be used to screen component noise properties.

To illustrate the feedback-loop oscillator design procedure, several specific examples are discussed. First, prototype SAW resonator oscillators operating in the 350 to 550 MHz range have demonstrated white PM noise floors of -184 dBc/Hz for carrier offset frequencies greater than 100 kHz, while flicker FM noise levels of -88 to -91 dBc/Hz at 10 Hz carrier offset have been achieved. Prototype 400 MHz SAW delay line stabilized oscillators with tuning ranges of ±150 kHz (±375 PPM) have demonstrated white PM noise floors of -170 dBc/Hz, along with flicker FM noise levels of -70 dBc/Hz at 10 Hz carrier offset. In addition, prototype 675 MHz SAW delay line oscillators with tuning ranges of ±600 kHz (±875 PPM) have demonstrated white PM noise floors of -167 dBc/Hz, along with flicker FM noise levels of -60 dBc/Hz at 10 Hz carrier offset. Recently, significant improvements have been realized in the vibration sensitivity performance of All Quartz Package (AQP) SAW hybrid circuit oscillators operating in the 150 MHz to 1 GHz frequency range, with typical vibration sensitivity magnitudes in the range from 1 to $5 \times 10^{-10}/g$. Data is presented to illustrate performance achieved to date. Other aspects of a SAW oscillator's performance, such as long-term stability, temperature stability, AM noise, etc., are discussed. Quite often trade-offs must be made between the various performance criteria that might otherwise be desired. These trade-offs are discussed in the context of the design examples presented.

2. FEEDBACK-LOOP OSCILLATOR DESIGN

Figure 1 shows a simple block diagram for a general purpose SAW device stabilized feedback-loop oscillator design. The feedback-loop oscillator's phase noise performance was first analyzed by Leeson [Ref. 7]. Although individual oscillators will always differ in specifics of their particular implementation, they will all share the following common features: 1) one or more loop amplifiers (G_1, G_2) of sufficient gain to overcome total feedback-loop losses; 2) some means for gain-limiting (compression) within the feedback loop to insure stable oscillation (in many instances this gain-limiting action will actually occur in the second-stage (G_2) or high-level feedback-loop amplifier); 3) provisions for both gain and phase adjustment within the feedback loop in order to establish the proper loop conditions for oscillation to occur, namely approximately 3 dB (2 to 4 dB) of excess small-signal gain (nominally 3 dB of gain compression when equilibrium is reached) and $2\pi N$ radians of net transmission phase shift through the loop, where N is an integer; 4) feedback-loop signal sampling, which may be either capacitive, resistive, etc.; and 5) a buffer amplifier (G_3) to isolate the feedback loop from external load variations. The need for an electronic phase shifter in the feedback loop to provide frequency tuning or modulation and a low-pass filter at the output to suppress undesired harmonic signals may, or may not, be necessary depending upon a specific oscillator's performance requirements.

The basic feedback-loop oscillator circuit of Fig. 1 is particularly well suited for two-port SAW devices and is generally the most commonly used configuration for both resonator- and delay line-based SAW oscillators. It is also particulary convenient to choose components that have been designed to work well in a 50-Ω characteristic impedance environment. By using amplifiers, power splitters or directional couplers, and other electronic components that nominally have been designed with 50-Ω input and output impedances, the proper conditions for oscillation can be

quickly established using a vector network analyzer. By breaking the oscillator's feedback loop open at any convenient location the open-loop gain and transmission phase of the combined oscillator components can then be readily established. The oscillator's equilibrium loop power can be determined by increasing the network analyzer's test signal level until gain compression gives approximately 0 dB of open-loop gain. By observing the gain and phase conditions at frequencies other than the desired oscillation frequency, the potential for spurious oscillations can also be evaluated. Finally, if a CW RF test signal is applied to the input port of the opened-loop at the power level corresponding to 0 dB of open-loop gain, then the oscillator's output power level may be verified. Clearly, other measurements may be performed to establish the oscillator's approximate electronic tuning sensitivity (in the case of a voltage controlled oscillator (VCO)), power supply pushing sensitivity, load pulling sensitivity, etc.

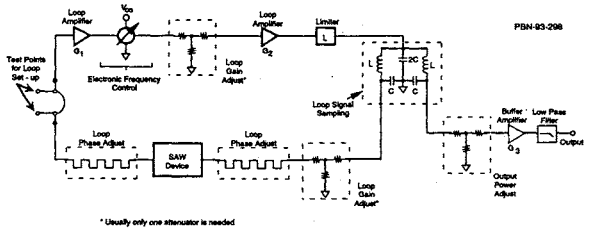


Fig. 1 Circuit Diagram for a General Purpose Feedback-loop Oscillator.

The simplest SAW oscillator may contain only an amplifier, a SAW device, an output coupler, and means for setting the loop's excess gain level and transmission phase shift, e.g., an attenuator and a short length of coaxial cable. Saturation of the loop amplifier's output power provides the necessary gain compression. More capable oscillators may contain a buffer amplifier, an electronic phase shifter, amplitude limiter, or any of the other components shown in Fig. 1.

As indicated in Fig. 1, there is some discretion as to where the loop gain adjust attenuator is placed. For a feedback-loop oscillator with a ***single loop amplifier*** it doesn't matter at all where the attenuator is positioned insofar as its basic contribution to the oscillator's white phase noise floor is concerned, although it may end up affecting the oscillator's overall performance in secondary ways, e.g., if the attenuator were positioned between the output of the high level loop amplifier and the coupling circuitry. However, for a feedback-loop oscillator with ***two loop amplifiers*** there is a distinct performance advantage to be realized if the attenuator is positioned between the output of the first-stage loop amplifier (G_1) and the input of the second-stage loop amplifier (G_2). If the attenuator is not positioned inter-amplifier stage, then the oscillator's phase noise floor will degrade dB for dB according to the value of attenuator used. This will not happen if the attenuator is introduced between the amplifiers, as just noted. In this case the attenuator's contribution to the oscillator's phase noise floor is effectively reduced ("shielded") by the gain of the first-stage loop amplifier. The net effect is to reduce the noise floor degradation factor to only about a tenth of a dB per dB according to the value of the attenuator used. The approximate amount of noise floor degradation, Δ, may be calculated using the following equation

$$\Delta(\text{dB}) = 10\text{x}\log_{10}[(F_1G_1 + F_2L - 1)/(F_1G_1 + F_2 - 1)] \qquad (1)$$

where F_1 (= $10^{[(NF_1)/10]}$) and F_2 (= $10^{[(NF_2)/10]}$) are the ***noise factors*** for the first- and second-stage loop amplifiers, respectively, NF_1 and NF_2 are the ***noise figures*** for the first- and second-stage loop amplifiers, respectively, G_1 is the gain of the first-stage loop amplifier (expressed numerically), and L is the loss of the inter-stage attenuator (expressed numerically). Let's consider a specific example, as follows: $F_1 = F_2 = 3.16$ (i.e., $NF_1 = NF_2 = 5$ dB), $L = 3.16$ (i.e., a 5 dB attenuator), and $G_1 = 25.12$ (i.e., the small-signal gain of the first-stage loop amplifier is 14 dB). If the attenuator in positioned inter-stage, then the oscillator's white phase noise floor degradation will only be about 0.35 dB, whereas if the attenuator is place anywhere else within the feedback loop the oscillator's white phase noise floor will be degraded by the full 5 dB value of the attenuator. This represents a significant noise floor improvement without incurring any performance penalty whatsoever. Clearly, it would be foolish to place the attenuator anywhere except inter-stage when two amplifiers are required in the feedback loop!

Perhaps the greatest advantage which accrues from using the feedback-loop oscillator architecture is the ability to readily isolate problem components, including the ability to pre-screen critical loop components such as the RF amplifiers, SAW device, and electronic phase shifter for their residual phase noise levels before the oscillator is even fully assembled. This ability is particulary useful since these rather basic measurements may be used directly to predict the oscillator's full phase noise spectrum. As has been shown previously, for most SAW resonator oscillators, the following approximate expression may be used to estimate the oscillator's expected phase noise performance for carrier offset frequencies greater than 0.1 Hz [Ref. 8]

$$S_\phi(f_m) = [\alpha_R F_o^4/f_m^3] + [\alpha_E/\{(2\pi\tau_g)^2 f_m^3\}] + [(2\alpha_R Q_L F_o^3)/f_m^2] + [(2GFkT/P_o)/\{(2\pi\tau_g)^2 f_m^2\}] + \alpha_E/f_m + (2GFkT)/P_o \qquad (2)$$

where G is the compressed power gain of the gain-limiting loop amplifier, F is the noise factor of the loop amplifier, or amplifiers, as the case may be, k is Boltzmann's constant, T is the temperature in °K, P_o is the carrier power level (in watts) at the output of the gain-limiting loop amplifier, F_o is the carrier frequency in Hz, f_m is the carrier offset frequency in Hz, τ_g is the SAW resonator's group delay in seconds, Q_L ($=\pi F_o \tau_g$) is the loaded Q of the SAW resonator in the feedback loop, and α_R and α_E are the flicker noise constants for the SAW resonator and loop amplifier(s), respectively. The flicker noise constants may be evaluated from residual noise measurements using the following expressions

$$\alpha_R = [1/(2Q_L F_o)^2][2f_m \text{x} 10^{\{\mathcal{L}'(f_m)/10\}}] \qquad 3(a)$$

and

$$\alpha_E = 2f_m \text{x}[10^{\{\mathcal{L}'(f_m)/10\}}] \qquad 3(b)$$

where Q_L, F_o, and f_m are as previously defined, and $\mathcal{L}'(f_m)$ is taken to represent a measurement of either the resonator's or amplifier's single-sideband residual flicker PM noise. Note that the "prime" symbol is intended to denote a residual rather than oscillator single-sideband noise spectrum.

A similar, although somewhat simpler approximate expression may be used to estimate a SAW delay line oscillator's phase noise spectrum for carrier offset frequencies greater than 1 Hz, namely

$$S_\phi(f_m) = [(\alpha_D + \alpha_E)\{(2\pi\tau_g)^2 f_m^3\}] + [(2GFkT/P_o)/\{(2\pi\tau_g)^2 f_m^2\}] + \alpha_E/f_m + (2GFkT)/P_o \qquad (4)$$

where G, F, k, T, P_o, F_o, f_m, and α_E are as previously defined, τ_g is now the SAW delay line's group delay in seconds, and α_D is the flicker noise constant for the SAW delay line. The amplifier's flicker noise constant in this case is again evaluated using Equation 3(b), while the SAW delay line's flicker noise constant is calculated in a similar manner, i.e., using

$$\alpha_D = 2f_m x[10^{(\mathcal{L}'(f_m)/10)}] \quad (5)$$

where $\mathcal{L}'(f_m)$ now represents the result of a single-sideband residual flicker noise measurement on the SAW delay line device. On occasion, instead of the symbols α_E and α_D, the symbols a_E and a_D have been used to denote an amplifier's and delay line's flicker noise constants, respectively. This was done in order to indicate that they are calculated in a different manner from their respective component's measured residual flicker noise, when compared to the flicker noise constant for a SAW resonator device as defined in Equation (2). As a direct consequence, α_E and α_D have the units of radians^2, while α_R has the units of radians^2/Hz^2. Finally, the magnitude of α_R is markedly different from the magnitudes of either α_E or α_D. A typical magnitude for α_R might be in the range from $1x10^{-41}$ to $1x10^{-39}$, while typical magnitudes for α_E and α_D might range anywhere from $5x10^{-16}$ to $2x10^{-13}$, depending upon the particular amplifier or delay line under consideration.

Since state-of-the-art phase noise performance is the central theme of this paper, Section 3 briefly describes the typical AQP SAW resonator and delay line designs which we routinely use.

3. AQP SAW DEVICE DESIGN

All of our recent SAW oscillator work is based upon the "All Quartz Package" (AQP) [Refs. 9 - 13]. Subsequently, similar AQP devices have been reported by others [Refs. 14 - 17]. The basic AQP structure is shown in Fig. 2, including the ability to laser-trim the device's resonant frequency after AQP sealing [Refs. 12 - 13]. As indicated, two identically oriented pieces of single-crystal quartz are sealed together under high vacuum using a glass frit. The SAW device, which is fabricated on the inside surface of the lower substrate as noted, is thereby protected from any external sources of contamination. AQP SAW devices have demonstrated excellent long-term frequency stability (see Section 6), and the packaging technique is almost ideal for use in hybrid circuit oscillators (see Section 7). Development of the AQP has resulted in a significant reduction in vibration sensitivity at the AQP SAW device level [Ref. 11], as well as more recently for AQP SAW hybrid circuit oscillators [Ref. 18].

This unique packaging concept has also served as the basis for the development of a post-seal frequency-trimming technique [Refs. 12 - 13], which makes it possible to set the frequency of an AQP SAW resonator to within ±1 PPM of a desired value. Although reactive ion etching (RIE) may be used to adjust the frequency of SAW devices prior to sealing [Refs. 19 - 20], it has the disadvantage that unpredictable shifts in frequency can, and indeed do, occur during the high temperature sealing process which is essential to insure good long-term frequency stability [Ref. 6]. Figure 2 serves to illustrate the post-seal frequency-trimming process, which is possible due to the optical transparency of the AQP. During SAW device fabrication a trim pad consisting of aluminum and aluminum oxide is deposited by E-beam evaporation on the inside surface of the cover piece, as indicated in Fig. 2 [Ref. 12 - 13]. The trim pad is situated on the cover in such a way as to be located over the space between the transducers of the SAW device once the AQP is assembled. By focusing high intensity pulsed UV laser light (193 nm) on the $Al_2O_3/Al/Al_2O_3$ trim pad, material can be transported from the cover to the SAW device's acoustically active surface area. The presence of this new material on the acoustically active portion of the SAW substrate slows the acoustic wave and hence lowers the resonant frequency of the SAW device. If the SAW device has been designed so that its resonant frequency after sealing is slightly higher than the ultimately desired value, the laser-trimming process can then be used to bring each device's frequency down to a specified value. Frequency adjustments as large as 100 PPM are possible with an accuracy of better than ±1 PPM. While there is some degradation in long-term frequency stability as a result of the laser-trimming process, we have shown that it is still compatible with long-term fractional frequency oscillator stability requirements of ±10 PPM over ten years [Ref. 4].

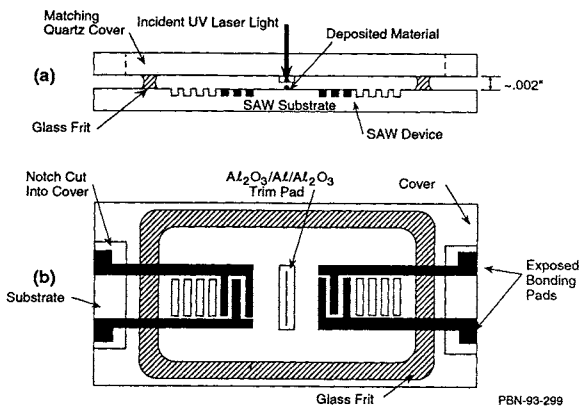


Fig. 2 Side (a) and Top (b) Views of an AQP SAW Resonator, Including Provisions for Laser Frequency-trimming.

AQP SAW resonator or delay line devices are packaged in a similar manner. Typically, our SAW resonator devices are designed as shown in Fig. 3. The active acoustic aperture ranges from 150 to 300 λ, where λ denotes an acoustic wavelength at the device's nominal resonant frequency. An important feature common to all of our SAW resonator designs is the rather large free surface region between the two transducers, typically 200 λ. This results in an effective cavity length [Ref. 21] of approximately 365 λ. As a direct consequence of this large cavity length the resonator will support three distinct cavity modes. This is illustrated in Fig. 4 for a 500 MHz two-port SAW resonator device design with an active acoustic aperture of 300 λ. The three cavity mode resonances are clearly evident, but note that the transmission phase shifts of the two outer resonances are offset by 180° with respect to the transmission phase shift of the central resonance peak. For most filter applications the existence of these satellite resonances would be totally unacceptable, but in an oscillator application the phase reversals exhibited by the satellite peaks totally eliminate them from consideration if the feedback-loop's transmission phase shift has been properly set-up for oscillator operation on the central resonance peak. The large cavity length serves to reduce the fraction of the effective cavity length covered by metal to only about 7 percent, and in many instances eliminates the need to apodize the transducers [Ref. 22], even for the relatively wide apertures (150 to 300 λ) used in our designs. Another advantage of a large cavity size (both length and width) is that it increases the power-handling capability of the resonator [Ref. 23]. SAW resonator devices

incorporating this basic design have typically exhibited 4 to 10 dB of insertion loss, depending upon the operating frequency, number of fingers in each transducer, and the actual acoustic aperture used.

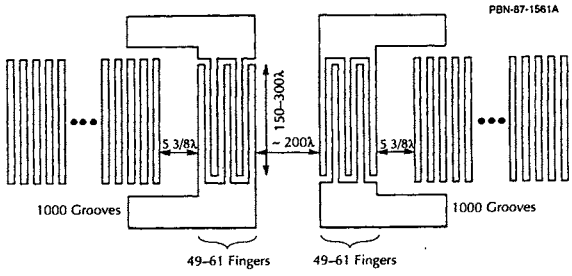


Fig. 3 Basic Two-port SAW Resonator Design.

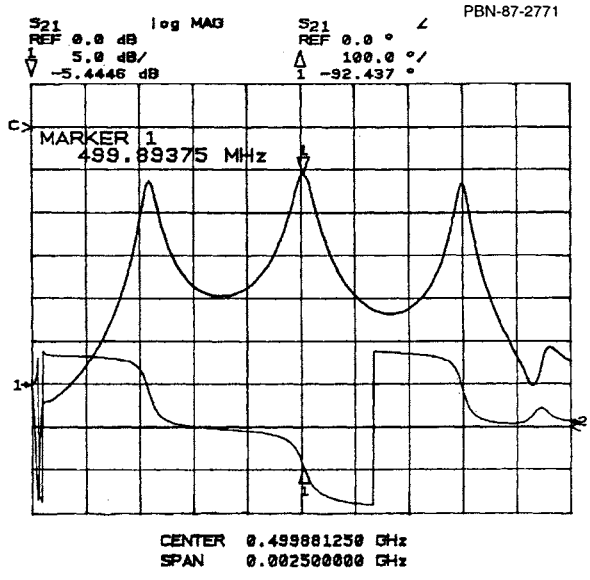


Fig. 4 Insertion Gain and Transmission Phase Shift Responses for a 500 MHz AQP SAW Resonator Device.

The key features of the transducers for a SAW delay line designed for an oscillator application are shown in Fig. 5. For good frequency stability the delay line should have as large a delay time (group delay) as possible. However, the delay line must also provide a frequency filtering function to limit the amplitude response of the delay line to a relatively narrow frequency range. Generally, the value of group delay selected will represent a compromise since most SAW delay line oscillators are intended to operate as frequency agile (voltage tunable) sources. The long "thinned" transducer on the left in Fig. 5 serves this purpose of frequency filtering, while the center-to-center distance between the two transducers determines the group delay, τ_g. To ensure that only one frequency can satisfy the oscillation conditions at any given time, the combined length of the two transducers should be no less than approximately 90 percent of the center-to-center distance between the two transducers. Generally, it has been found desirable to limit the total number of fingers in each transducer to about 120, thus the long transducer is usually thinned into a ladder type structure [Ref. 24], as illustrated in Fig. 5. This results in a device with only about 5 to 6 percent of the active acoustic area covered with aluminum from the transducers. Additional fingers can be used to achieve lower insertion loss, but this increases the undesirable influence of the metalization on turn-over temperature, triple transit reflections, and frequency accuracy. If desired, electrically inactive fingers (so called "dummy" electrodes) may be used in order to reduce ripple in the insertion loss and transmission phase responses. This might be done, for example, in order to achieve a smoother frequency versus tuning voltage transfer characteristic, as well as to achieve a smoother frequency versus temperature characteristic for the oscillator design. Additional information regarding our AQP SAW delay line design approach may be found in Reference 25.

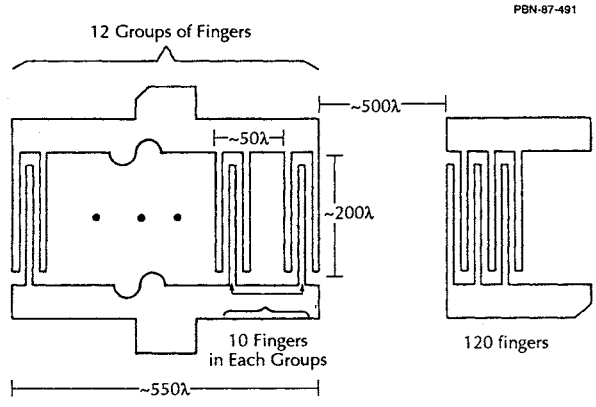


Fig. 5 Typical SAW Delay Line Design.

Additional factors influencing SAW resonator and delay line designs for oscillator applications are discussed in Reference 6.

4. RESIDUAL NOISE MEASUREMENTS

Residual phase noise measurements have been used for some time to characterize devices in the transmission [Refs. 26 - 29], as well as reflection [Ref. 30] modes. The realization of extremely low phase noise VHF, UHF and microwave oscillators is critically dependent upon the use of components, e.g., amplifiers, SAW devices, electronic phase shifters, etc., which possess verifiably low residual phase noise levels. The utility of residual phase noise measurements is due to the fact that there is a one-to-one relationship between an oscillator's closed-loop (oscillator) flicker FM noise ($1/f_m^3$) and open-loop (residual) flicker PM noise ($1/f_m$) level [Ref. 4], namely

$$\mathcal{L}_O(f_m) = \mathcal{L}_R(f_m) + 10x\log_{10}\{1/(2\pi\tau_g)\}^2 + 10x\log_{10}\{1/f_m\}^2 \quad (6)$$

where $\mathcal{L}_O(f_m)$ is the oscillator's single-sideband flicker FM noise spectrum expressed in dBc/Hz, $\mathcal{L}_R(f_m)$ is the residual flicker PM noise spectrum for the oscillator when measured looking through the feedback loop (same test configuration as described previously in Fig. 1 for oscillator set-up)

expressed in dBc/Hz, f_m is the carrier offset frequency in Hz, and τ_g is the SAW device's group delay, in seconds, when the oscillator's feedback loop is closed. For a SAW resonator its group delay may also be found using the relationship $\tau_g = Q_L/(\pi F_o)$, where F_o is the oscillator's carrier frequency in Hz. It is important to realize that $\mathcal{L}_R(f_m)$ represents the algebraic sum of ***all*** of the potential flicker phase noise contributions in the feedback loop due to the loop amplifier(s), SAW device, electronic phase shifter, etc. Therefore, if a given level for the oscillator's flicker FM noise level ($1/f_m^3$) is desired, Equation (6) may be used to set an upper limit on the acceptable residual flicker PM noise for each component in the loop which is found to significantly contribute to the oscillator loop's residual flicker phase noise level. We have found in our own work, for SAW resonator and delay line oscillators such as those described in Section 7, that it is often quite difficult, if not impossible, to actually uniquely identify the particular component which is dominating the oscillator's flicker FM noise level. In reality, one or more components (typically the limiting or gain-compressing loop amplifier and the SAW device) may both be contributing on a more or less equal basis.

The basic residual phase noise measurement set-up is shown in Fig. 6. The choice of a test source depends upon a number of factors. Generally, a frequency synthesizer based test source configuration is most useful when evaluating the residual flicker phase noise level of a SAW resonator or delay line device, since it may readily be adjusted in frequency in order to test a wide variety of relatively narrow bandwidth devices. We have previously shown [Ref. 31] that a state-of-the-art, extremely low-noise (both PM ***and*** AM) SAW resonator oscillator is quite useful when evaluating the residual phase noise properties of components which may exhibit appreciable AM-to-PM conversion sensitivities, e.g., amplifiers, electronic phase shifters, etc.

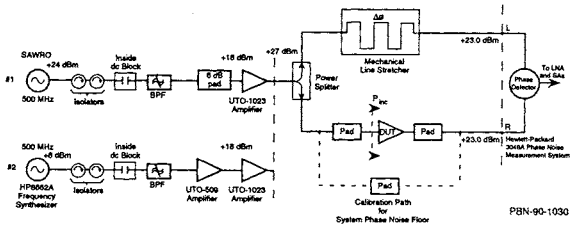


Fig. 6 Residual Phase Noise Measurement Set-ups at 500 MHz.

Figure 7 shows the measured residual flicker phase noise level for a 500 MHz AQP SAW resonator device with an active acoustic aperture of 300 λ and an effective cavity length of approximately 365 λ, with $\tau_g \cong 2.5$ μsec. The solid line ($1/f_m$ slope) shown in Fig. 7 represents the limiting residual phase noise system sensitivity floor for measurements on SAW resonator devices in this frequency range (typical insertion loss less than 7 to 9 dB). A Hewlett-Packard 8662A frequency synthesizer was used as the test source for the measurement. Figure 8 illustrates a typical residual flicker phase noise level measured on a 675 MHz AQP SAW delay line device ($\tau_g \cong 0.45$ μsec). The solid line ($1/f_m$ slope) shown in Fig. 8 represents the limiting residual flicker phase noise system sensitivity floor for measurements on SAW delay line devices in this frequency range (typical insertion loss less than 20 to 22 dB). Once again a Hewlett-Packard 8662A Frequency Synthesizer was used as the test source for the measurement. It should be noted that the measured flicker noise level for both cases just illustrated are each within 3 dB of the limiting system sensitivity noise floor, and therefore the actual residual flicker phase noise levels for the SAW devices are substantially lower (at least 3 dB) than the measurements would otherwise suggest. Generally, residual flicker noise measurements on SAW resonators are limited to carrier offset frequencies less than about 100 Hz, while for SAW delay line devices they are limited to carrier offset frequencies less than approximately 1 kHz, since de-correlation of the test source's phase noise will occur due to the rather substantial unmatched delay introduced in one test leg by the SAW device under test.

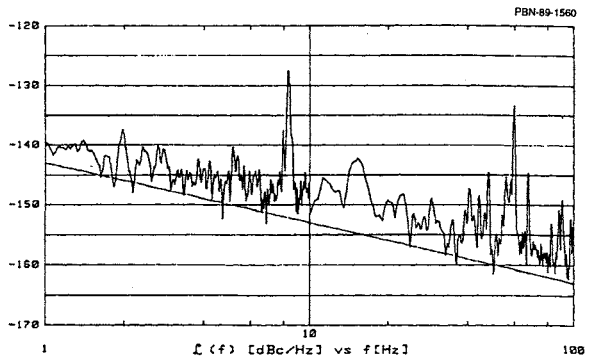


Fig. 7 Measured Residual Flicker Phase Noise Level for a Typical Low-noise 500 MHz AQP SAW Resonator Device After a High Power RF Burn-in.

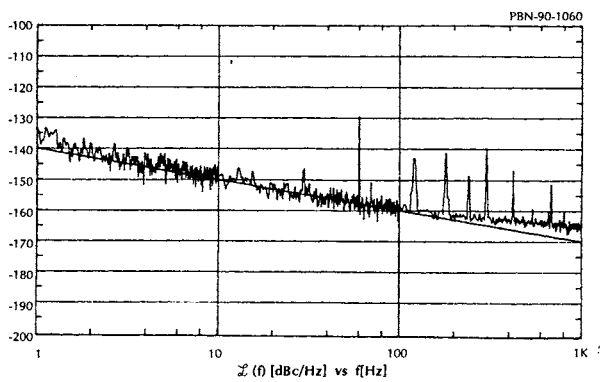


Fig. 8 Measured Residual Flicker Phase Noise Level for a Typical 675 MHz AQP SAW Delay Line Device Without a High Power RF Burn-in.

It is interesting to note that while the AQP SAW resonator device whose residual flicker FM noise was shown in Fig. 7 received a burn-in using a high incident power RF signal level [Ref. 32], the AQP SAW delay line device illustrated in Fig. 8 was not burned-in. In general, we have found that the high power RF burn-in technique may be used to lower the flicker phase noise levels of SAW resonator, as well as SAW delay line, devices. However, many of our devices, both resonators and delay lines, are sufficiently close to the limiting system sensitivity noise floor even prior to burn-in that we are really unable to tell directly whether the burn-in has actually accomplished its objective or not, let alone to what extent the burn-in may have improved the device's flicker PM noise level. In these circumstances, the only way to really tell whether the high power RF burn-in has effected a decrease in a device's flicker noise level is to evaluate the device's performance in a "test bed" oscillator circuit of demonstrated low residual flicker PM noise. Clearly, this approach will ultimately be limited by the residual flicker PM noise level of the "test bed" oscillator circuit itself!

If a suitably low-noise test source is used, then the basic residual phase noise test set-up shown in Fig. 6 is capable of evaluating the residual phase noise properties of electronic components, e.g., amplifiers and electronic phase shifters, for carrier offset frequencies from 1 Hz to 40 MHz with unprecedented sensitivity and reproducibility [Ref. 31]. Figure 9 illustrates the measurement technique for three similar amplifier designs from three different vendors. The specific amplifiers are: 1) Cougar AC-509, 2) Watkins-Johnson PA3-1, and 3) Avantek UTO-509. For all three amplifiers the test conditions were identical, namely an incident RF test power level of +12.0 dBm at the amplifier's input. This incident RF test power level corresponds to approximately 3 dB of gain compression. These three amplifiers are all remarkably similar in their nominal electrical performance parameters, which are summarized in Table I, along with the results of our own electrical measurements on the devices which were tested. The UTO-509's residual phase noise performance shown in Fig. 9 is typical of that observed on more than one hundred devices, while the specific PA3-1 and AC-509 data shown in Fig. 9 was essentially identical for samples of three and four devices, respectively. It is interesting to note that both the PA3-1 and the AC-509 have consistently higher levels of residual flicker noise, when compared to the UTO-509. Also, all four AC-509s which were tested exhibited an anomalous residual phase noise "bump" near 6 kHz carrier offset frequency. Finally, the higher residual white phase noise floor measured for the PA3-1 is consistent with the fact that its noise figure was found to be approximately 2 dB higher than that of the UTO-509. However, the comparatively high white phase noise floor for the AC-509 cannot be explained in a similar manner. Separate residual phase noise measurements were performed on each amplifier using a frequency synthesizer (Hewlett-Packard 8662A) based test source, primarily for its quantitatively higher AM noise level, thereby confirming that the AC-509 does indeed have a considerably higher AM-to-PM conversion factor than either of the other amplifiers [Ref. 31]. These results are shown in Fig. 10. It is evident that even with the extremely low-noise SAW resonator oscillator based test source configuration, it simply isn't possible to accurately measure this particular amplifier's (the AC-509) residual white phase noise floor properties, at least when only a single device is tested at one time.

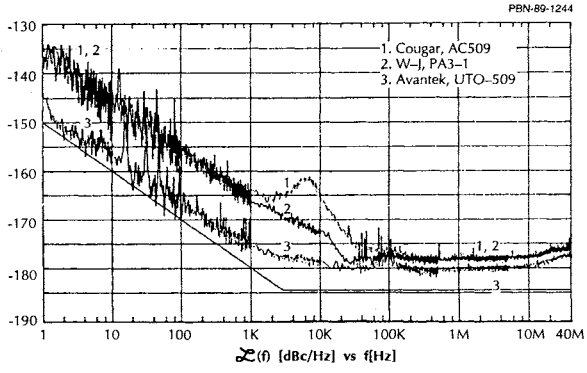


Fig. 9 Residual Phase Noise Measured for Three Similar Amplifiers Using SAWRO Test Source.

While the phase noise of either test source shown in Fig. 6 is really not a limiting factor for residual phase noise measurements on a device under test (DUT), at least for reasonably small mismatches in delay between the two test path legs, the AM noise properties of the test source are actually much more important than thus far indicated. This is illustrated in Fig. 11, which shows the results of three separate amplifier residual phase noise measurements, specifically: 1) the frequency synthesizer based test source used to test a UTO-1023 (Avantek) RF amplifier at approximately 3 dB of gain compression (+17.0 dBm input power), 2) the frequency synthesizer based test source used to test a UTO-509 (Avantek) RF amplifier at approximately 3 dB of gain compression (+12.0 dBm input power), and 3) the SAWRO based test source used to test the same UTO-509 (Avantek) RF amplifier under identical test conditions. The DUTs are clearly converting the frequency synthesizer based test source's AM noise into PM noise, thereby limiting this measurement configuration's ability to characterize the amplifiers' actual residual phase noise levels. In point of fact, the UTO-1023 is decreasing the ***effective*** AM source noise suppression for the measurement to only 2 to 4 dB [Ref. 31]. This observation serves to place a much more stringent requirement upon the test source's AM noise spectrum, if it is to be used for accurate residual phase noise measurements on DUTs with appreciable AM-to-PM conversion factors. As seen in Fig. 11, the SAWRO based test source may be used to characterize the residual phase noise properties of the UTO-509 amplifier for all carrier offset frequencies. However, note that using the frequency synthesizer based source to test a UTO-509 presents a problem not only far-from-the-carrier, but also close-to-the-carrier as well. Reference 31 contains considerable additional information regarding the residual phase noise characterization of a variety of components.

TABLE I

Summary of Typical RF Amplifier Characteristics at T = 25°C

Amplifier	Small Signal Gain	Noise Figure	Pout @ 1 dB Compression	Pout @ 3 dB Compression	I_{DC} @ V_{DC}	Residual Phase Noise[2] ℒ(f = 1 Hz)
Avantek/UTO-509						
Vendor Typical	14.3 dB	4.5 dB	+22.3 dBm	N.A.	90 mA @ +15V	N.A.
Measured	(14.2 dB)[1]	(4.2 dB)[1]	(+22.2 dBm)[1]	(+23.3 dBm)[1]	(100 mA @ +15V)	(~ -148 dBc/Hz)[1]
Watkins-Johnson/PA3-1						
Vendor Typical	14.5 dB	7.0 dB	+22.5 dBm	N.A.	91 mA @+15V	N.A.
Measured	(14.2 dB)[1]	(6.8 dB)[1]	(+22.6 dBm)[1]	(+23.6 dBm)[1]	(92 mA @+15V)	(~ -136 dBc/Hz)[1]
Cougar/AC509						
Vendor Typical	13.5 dB	5.0 dB	+22.0 dBm	N.A.	88 mA @ +15V	N.A.
Measured	(14.0 dB)[1]	(5.3 dB)[1]	(+22.0 dBm)[1]	(+23.1 dBm)[1]	(87 mA @ +15V)	(~ -136 dBc/Hz)[1]

[1] Measured at 500 MHz [2] Measured at 3 dB of Gain Compression N.A. = Not Applicable

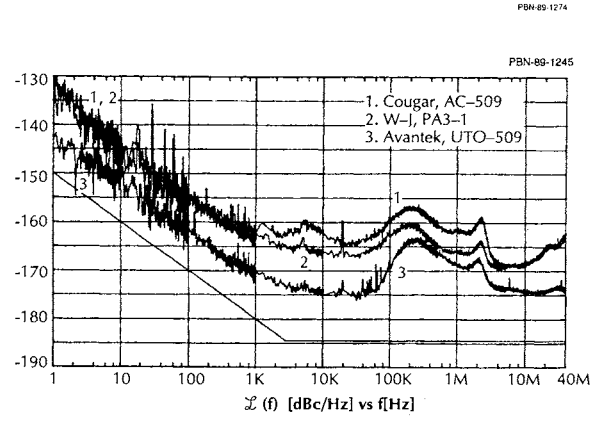


Fig. 10 Residual Phase Noise Measured for Three Similar Amplifiers Using Synthesizer Test Source.

It should perhaps be noted that neither an amplifier's residual flicker noise level, nor the possible presence of "bumps" in its residual phase noise spectrum, are

predictable based upon any other electrical characteristic that we are aware of, including a small-signal noise figure measurement. Furthermore, as has been shown, even very similar amplifier designs from different vendors can differ markedly in their residual phase noise characteristics, even though their other nominal electrical parameters are nearly identical. It has been our experience that every single amplifier must be "screened" before use in an oscillator design where truly state-of-the-art phase noise performance is desired. Otherwise, you will constantly be attempting to diagnose which component, or components, is the culprit when a particular oscillator's phase noise spectrum simply doesn't meet its intended specification.

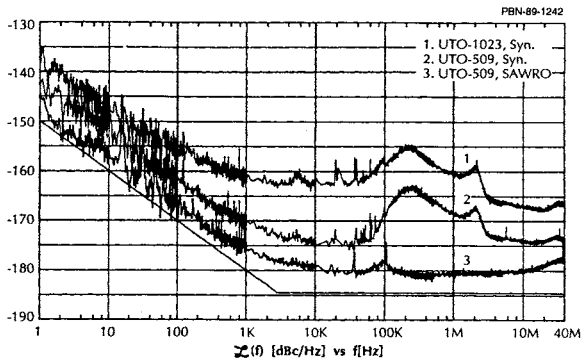


Fig. 11 Residual Phase Noise Measured for Two High Power RF Amplifiers Using SAWRO and Synthesizer Based Test Sources.

Finally, the basic residual phase noise measurement technique may also be applied to other electronic components as well, e.g., electronic phase shifters, mixers, frequency dividers, frequency multipliers, etc., as has been shown in Reference 31. Many of these components are essential parts of low-noise microwave frequency sources based upon lower frequency (VHF or UHF), low-noise SAW oscillators.

It is perhaps worth noting here that the different limiting measurement system sensitivity noise floors shown in Figs. 7, 8, and 9, for residual phase noise measurements on a SAW resonator, a SAW delay line, and an RF amplifier, respectively, are due to the varying loss or gain exhibited by these devices. While an amplifier such as the UTO-509 may be tested as shown explicitly in Fig. 6, one (for a SAW resonator) or more (for a SAW delay line) RF amplifiers must also be used in the test leg when evaluating devices with significant insertion losses (greater than 2 or 3 dB) in order to achieve the best limiting system sensitivity noise floor for the measurement. Thus the amplifiers which must be used to test the residual flicker noise levels of SAW resonators and delay lines are part of the "basic" residual phase noise set-up, and their own flicker noise levels are therefore part of the limiting system sensitivity flicker noise level.

While there are some obvious limitations upon the ability to accurately test devices for their residual flicker noise levels when these levels are less than about -145 dBc/Hz at 1 Hz carrier offset, the residual noise measurement technique just described is still extremely useful as a means of "weeding out" components which are clearly too noisy for the intended application. Finally, as even lower residual noise phase detectors (mixers), RF amplifiers and video (baseband) low-noise amplifiers (LNAs) are developed our ability to characterize component residual noise properties will correspondingly improve. We have now reached the point where a significant improvement in our limiting residual phase noise system floors will really be essential in order to successfully pursue a further order of magnitude reduction (to -95 to -105 dBc/Hz at 10 Hz carrier offset for a 500 MHz AQP SAW resonator based oscillator and to -80 to -90 dBc/Hz at 10 Hz carrier offset for a 400 MHz AQP SAW delay line (τ_g = 1.6 μsec) based oscillator) in a SAW oscillator's flicker FM noise level.

5. TEMPERATURE STABILITY

The dependence of a SAW oscillator's frequency on temperature falls into two general categories, namely: static and dynamic effects. The familiar parabolic fractional frequency versus temperature characteristic for the family of rotated Y-cuts of quartz, of which ST-cut quartz is an example, represents the static temperature effect [Ref. 6]. Very little has changed with regard to the static temperature sensitivity of SAW devices on quartz over the last several years. Although several new cuts of quartz have been identified, they have not found widespread use for oscillator applications because the improvement in temperature sensitivity is only modest, at best, and they all have potentially serious drawbacks such as non-zero power flow angles or else very tight angular fabrication tolerances. Most, if not all, SAW devices for oscillator applications are still fabricated on the ST-cut like family of rotated Y-cuts. Reference 6 provides information regarding the relevant literature.

In addition to the static temperature effect, there is another phenomena observed on SAW oscillators that depends upon the time-rate-of-change of the temperature. This is known as the dynamic temperature effect. Figure 12 shows the response of a SAW oscillator to a 0.2°C temperature step. After the temperature is increased the frequency eventually settles to a slightly higher value. This is the static temperature effect and the change in the equilibrium frequency is small ($\cong 3 \times 10^{-9}$) because the temperature step occurred very close to the oscillator's turn-over temperature point. When the temperature is returned to its original value, the oscillator's frequency also returns to its initial value. However, during the short time periods of rapid temperature change at the leading and trailing edges of the temperature step, there are correspondingly rapid changes in frequency. A transient downward shift in frequency occurs at the upward shift in temperature and the reverse happens at the downward shift in temperature. These transient frequency shifts are a manifestation of the dynamic temperature effect. The magnitude of the frequency change is proportional to the time-rate-of-change of temperature. In this example the dynamic temperature coefficient is about -2.4 PPM/°C/second.

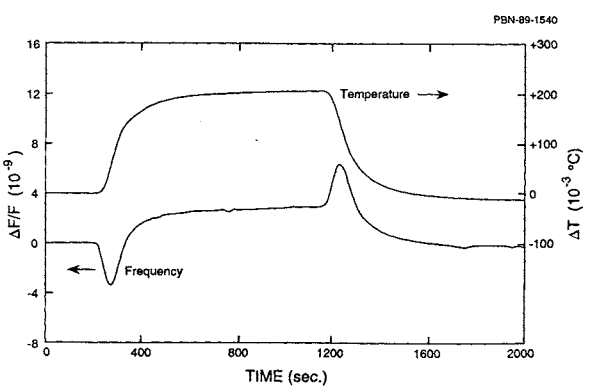


Fig. 12 Fractional Frequency Transient Response of a SAW Oscillator Subjected to a +0.2°C Temperature Step.

The dynamic temperature effect is caused by temperature gradients and is very much dependent upon how the SAW device is mounted [Ref. 34]. Little information on dynamic temperature effects in SAW devices exists in the literature, but dynamic temperature coefficients ranging from 1 to 10 PPM/°C/second (both positive ***and*** negative) have been observed in the authors' laboratory. If there exists a stress compensated cut of quartz for SAW devices similar to the SC-cut of quartz for bulk wave devices, it has not as yet been identified.

It is worth noting that the all quartz packaging technique and the laser frequency-trimming process have produced no observable changes in the basic frequency versus temperature characteristic of SAW devices fabricated on rotated Y-cuts of quartz.

6. LONG-TERM FREQUENCY STABILITY

The long-term frequency stability (aging) of SAW oscillators is an extremely important performance parameter for many system applications. In general, aging rates of less than about ±1 PPM/year (averaged over a period of years) are required for most, but not all, applications of high performance oscillators. For this reason, Raytheon has conducted an extensive aging test program for almost twenty years now. Good long-term frequency stability is critically dependent on the packaging of the SAW device. It is well known that the SAW device must be subjected to a high temperature (greater than 360°C) bake-out for at least one hour prior to sealing in an ultra-clean, hermetic package [Ref. 1]. This point is illustrated in Fig. 13 which shows the beneficial influence of progressively higher bake-out temperatures on long-term stability for five different TO-8 cold-weld sealed SAW resonator based oscillators. Note the difference in vertical scale between the upper and lower portions of Fig. 13. As the bake-out temperature was increased, there is a definite trend toward significantly better long-term frequency stability as a direct result of the higher bake-out temperatures. The cold-weld sealing approach for SAW devices was the first to demonstrate SAW-based oscillator long-term frequency stabilities of ±1 PPM/year, or better. The AQP approach is almost ideal in this respect since the basic sealing procedure involves a high temperature processing step. AQP sealed SAW devices have been evaluated extensively during the last ten years.

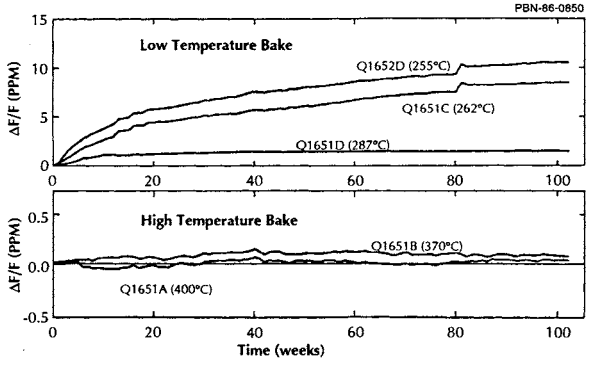


Fig. 13 Long-term Fractional Frequency Stability Measurements on Five SAW Oscillators Illustrating the Beneficial Influence of a High Temperature Bake-out Prior to Sealing for TO-8 Cold-weld Sealed SAW Resonator Devices.

In the past we have published considerable data covering a wide variety of test conditions (dissipated RF power, acoustic stress density, transducer metalization, etc.) for a number of AQP SAW device designs [Refs. 1 - 4, 6, 9, 12 - 13]. Figure 14 shows the measured long-term fractional frequency stability of four AQP SAW resonator oscillators operating in the 300 to 800 MHz frequency range. These are lower power oscillators with approximately +5 dBm incident RF power on the AQP SAW devices. Roughly one-half of the incident RF power is actually dissipated in the resonator. The four curves shown are representative of results obtained on printed circuit board test oscillators where the AQP SAW devices were attached using a four-corner RTV mounting approach. This approach insures a stress-free mounting environment for the AQP SAW device. The AQP SAW oscillator which has been on test for more than nine years incorporates one of the first AQP SAW resonator devices which we fabricated! This particular oscillator's total fractional frequency variation has been less than ±0.2 PPM (peak-to-peak) during the entire test period.

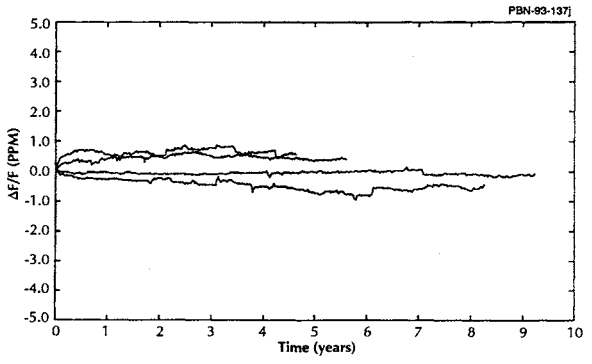


Fig. 14 Long-term Fractional Frequency Stability for Four AQP SAW Resonator Printed Circuit Board Oscillators.

AQP SAW devices are particularly well suited to hybrid circuit applications, and in fact it is in hybrid circuit oscillator designs that their potential for inherently low vibration sensitivity may be leveraged [Ref. 18]. However, in order to achieve good vibration sensitivity performance the AQP SAW device should be mounted up-side-down using a uniform mounting material for support [Refs. 11, 18, 35 - 37]. The mounting material must be sufficiently "soft" so as to not introduce stress and thereby interfere with, or bias, either the device's basic static fractional frequency versus temperature characteristic or else its long-term frequency stability properties. At the same time the mounting material should uniformly support the AQP SAW device, while introducing no mechanical resonances. A number of materials have been examined for this application. At the present time we are using a double-sided silicone pressure sensitive adhesive (PSA) tape with considerable success. While the aging characteristics for hybrid circuit oscillators where the AQP SAW device is mounted using the PSA are not quite as good as those shown in Fig. 14, they are still consistent with long-term fractional frequency stabilities of ±10 to ±15 PPM over a ten year period. However, we have seen instances where the material selected to uniformly support the AQP SAW device can indeed lead to a serious degradation in an oscillator's long-term fractional frequency stability. This point is illustrated in Fig. 15, which shows long-term fractional frequency stability for several 900 MHz AQP SAW resonator hybrid circuit oscillators. For the first four or five months everything appeared to be just fine, and then the oscillators'

frequencies began to drift excessively. The problem was ultimately traced to the acrylic sheet adhesive which was used to uniformly mount the AQP SAW devices into the hybrid packages. While the material appeared to have very good properties for uniformly supporting the AQP SAW device [Ref. 18], it obviously had some built-in limitations related to the material's stability over time. This example demonstrates the care which must be exercised in selecting candidate materials for AQP SAW device mounting, and furthermore shows that there is simply no substitute for oscillator aging studies in order to verify that the candidate material is indeed satisfactory.

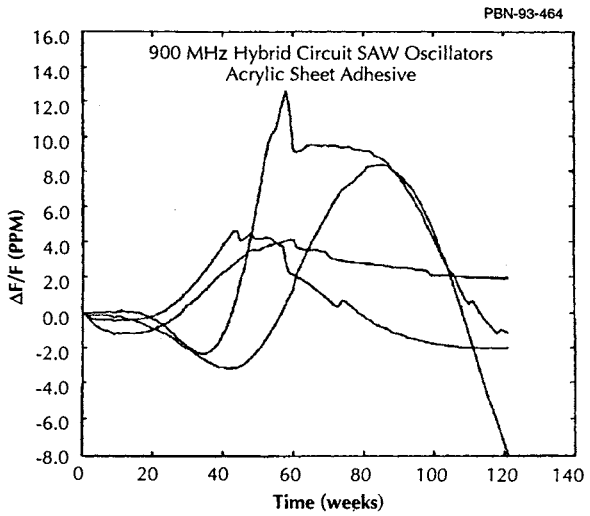


Fig. 15 Anomalous AQP SAW Resonator Hybrid Circuit Oscillator Long-term Fractional Frequency Stability Induced by Mounting Stresses.

7. SAW OSCILLATOR DESIGN EXAMPLES

7.1 *Introduction*

While many factors enter into the design of an oscillator in response to a desired set of performance parameters, we have generally found that the specified phase noise spectrum is given considerable weight when evaluating whether a particular design is truly capable of addressing a specific oscillator requirement. Other performance parameters, e.g., vibration sensitivity, AM noise, harmonic and spurious output signal levels, load pulling sensitivity, dc power supply pushing sensitivity, etc., may take a back seat in comparison to phase noise although an increased emphasis is now evident for low phase noise oscillators with minimal vibration sensitivities ($\equiv 1 \times 10^{-10}/g$ per axis, or lower) for ground-based, as well as for airborne and missile, operating environments. References 6 and 8 present detailed discussions of the various factors which may enter into the design of an oscillator in response to a particular set of requirements. In the examples which follow, the oscillators' phase noise spectra were considered of paramount importance, and every attempt was made to achieve the best overall phase noise performance in each case.

7.2 *Low Phase Noise SAW Resonator PCB VCOs*

We have previously described our efforts to develop extremely low phase noise SAW resonator VCOs [Refs. 1 - 5], based upon the feedback-loop oscillator architecture illustrated in Fig. 1. Typically, for more than twenty engineering prototype oscillators, phase noise performance as follows was routinely achieved: white phase noise floors of approximately -177 dBc/Hz for carrier offset frequencies greater than 100 kHz, and a flicker FM noise level of -75 dBc/Hz at 10 Hz carrier offset. While this is not quite as low as originally described in Reference 1, some compromises were necessary in order that the oscillators could be manufactured in production with acceptable yields. The various electronic components used in this printed circuit board (PCB) based oscillator design were described in Reference 3.

More recently, we have realized noticeable (> 3 dB) improvements in both the close-to-carrier flicker FM phase noise level, as well as the white PM noise level far-from-the-carrier, for laboratory prototype 500 MHz AQP SAW resonator printed circuit board based oscillators. This has been achieved by incorporating a new dual channel resonator design, as shown schematically in Fig. 16, as well as by using an even higher power, yet still suitably low residual flicker noise TO-8 RF amplifier, namely the UTO-1023 (Avantek). The result of these enhancements is shown in Fig. 17 for a laboratory prototype oscillator. The curve labeled "prior art" indicates the level of phase noise performance deemed to be achievable prior to the work described in References 1 and 32. The small "dip" (4 or 5 dB) in measured phase noise near 750 kHz carrier offset is attributable to noise degeneration in the feedback loop due to the two satellite SAW device resonances, since the net transmission phase shift through the loop is almost exactly 180° at this carrier offset frequency! This is the only observable consequence of our basic multi-mode cavity SAW resonator design described earlier. It should be noted that even for two SAW resonators fabricated at the same time in reasonably close proximity on a single quartz substrate, they will quite often not have identical resonant frequencies. This is illustrated by the frequency response characteristic shown in Fig. 18(a). It is readily apparent that the overall frequency response of the composite dual channel device is substantially degraded due to this splitting of the individual device resonant frequencies. As described previously, laser-trimming of the individual resonant frequencies may be used in order to eliminate this problem [Ref. 12 - 13]. The resulting improvement in the device's frequency response is shown in Fig. 18(b), after laser-trimming. The laser-trimming technique was also used to suppress 3rd order transverse modes in each of the individual resonator frequency responses [Ref. 13]. Each resonator device had an active acoustic aperture of 150 λ, thus the dual resonator device was essentially equivalent to the single 300 λ resonator devices which we have described previously [Refs. 1 - 5], except that the unloaded Qs for the dual channel devices were substantially better than those measured on single wide aperture resonators. This is attributable to the significant reduction in finger resistance (and thereby ohmic losses) for the dual channel device design.

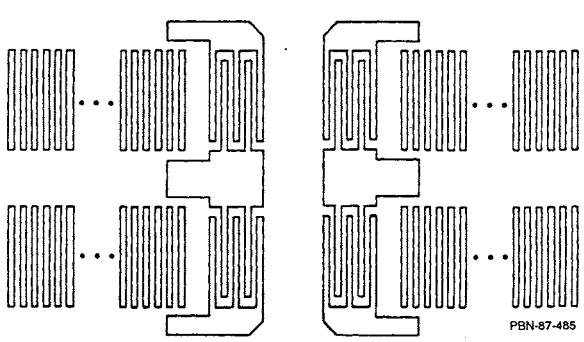


Fig. 16 Basic Dual Channel SAW Resonator Design.

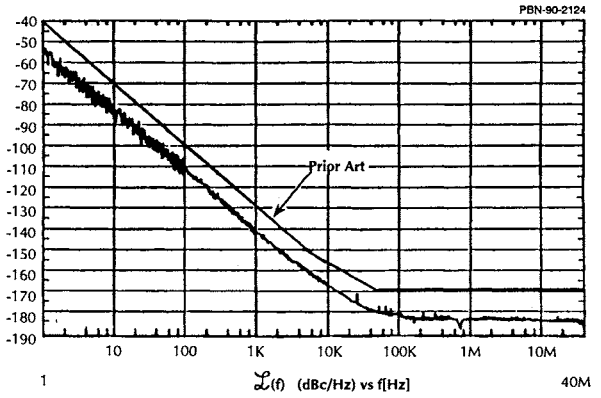
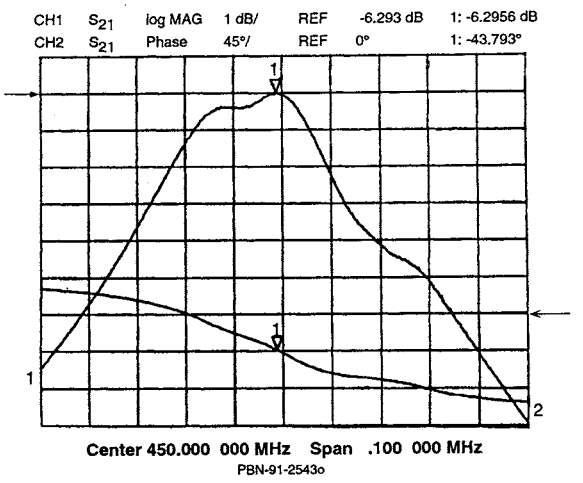
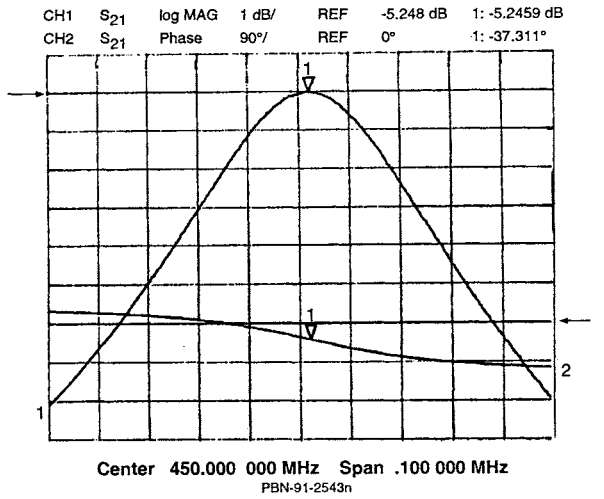


Fig. 17 Phase Noise Spectrum for a Laboratory Prototype 500 MHz AQP SAW Resonator PCB VCO.



(a)



(b)

Fig. 18 Insertion Gain and Transmission Phase Shift Plots for a Dual Channel 450 MHz AQP SAW Resonator Device. (a) Before Laser-trimming. (b) After Laser-trimming.

7.3 *Low Phase Noise SAW Resonator Hybrid Circuit VCOs*

Figure 19 shows a detailed circuit schematic for an AQP SAW resonator hybrid circuit oscillator. The primary goal of this effort was to demonstrate for a hybrid circuit oscillator comparable, or even better, phase noise performance than that described in Section 7.2 for PCB oscillators. Side benefits would be significantly reduced size and improved vibration sensitivity. Figure 20 shows a photograph of a prototype 350 MHz AQP SAW hybrid circuit oscillator in a 1.47" (3.75 cm) x 1.47" (3.75 cm) x 0.275" (0.70 cm) machined Kovar package. The RF amplifiers are substrate carrier versions of the UTO-509 (Avantek). This amplifier is the same as was used for the work described in References 1 through 5, and was selected for its extremely low residual flicker PM noise levels. Figure 21 shows a photograph of the 350 MHz AQP SAW dual channel resonator device used in this oscillator design.

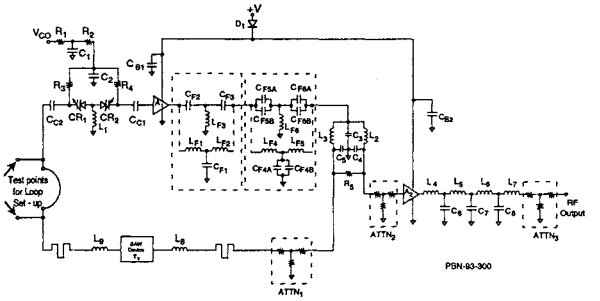


Fig. 19 Circuit Schematic for a SAW Based Hybrid Circuit Oscillator.

The measured single-sideband phase noise spectrum for one 350 MHz AQP SAW resonator hybrid circuit oscillator is shown in Fig. 22. The white PM noise floor is approximately -177 to -178 dBc/Hz, which is comparable to our prior results for PCB based oscillators. The flicker FM noise level of -89 dBc/Hz at 10 Hz carrier offset frequency is noticeably better than previously achieved performance (-83 dBc/Hz for a 500 MHz carrier frequency, in Fig. 17), even after taking into account the difference in carrier frequencies [Ref. 8]. The SAW resonator's loaded Q, Q_L, in the feedback loop was approximately 6500. Several hybrid oscillators demonstrated flicker FM noise levels of -90 to -91 dBc/Hz at 10 Hz carrier offset. The phase noise performance shown in Fig. 22 represents the current state-of-the-art for a SAW resonator hybrid circuit oscillator.

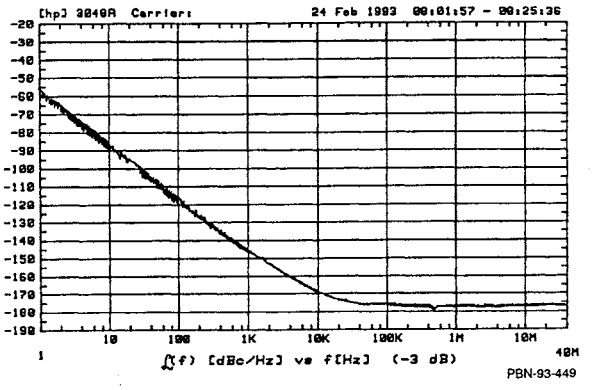


Fig. 22 Single-sideband PM Noise Spectrum for an Extremely Low-noise AQP SAW Resonator Hybrid Circuit Oscillator.

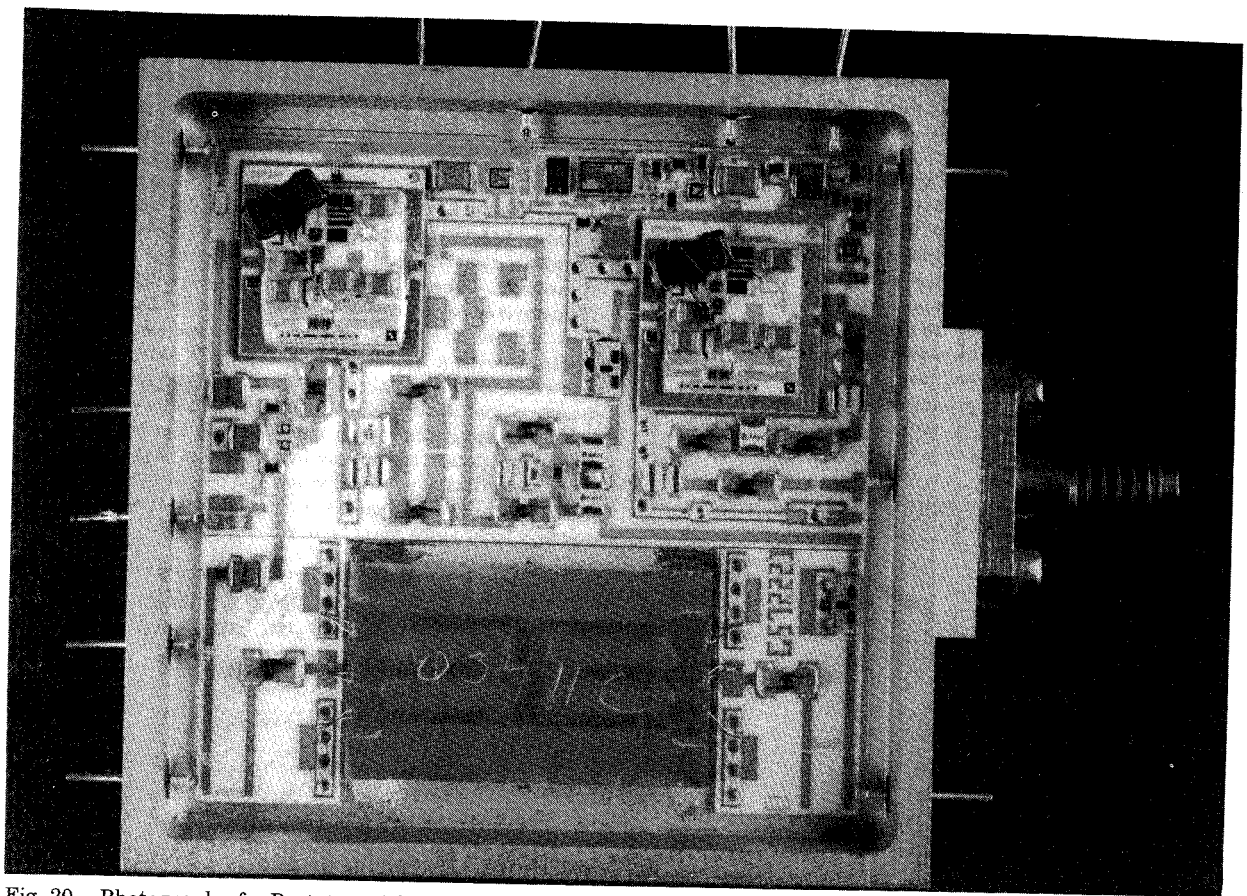


Fig. 20 Photograph of a Prototype AQP SAW Resonator Hybrid Circuit Oscillator. Hybrid Package is 1.47" (3.75 cm) x 1.47" (3.75 cm) x 0.275" (0.70 cm), exclusive of RF connectors and pins.

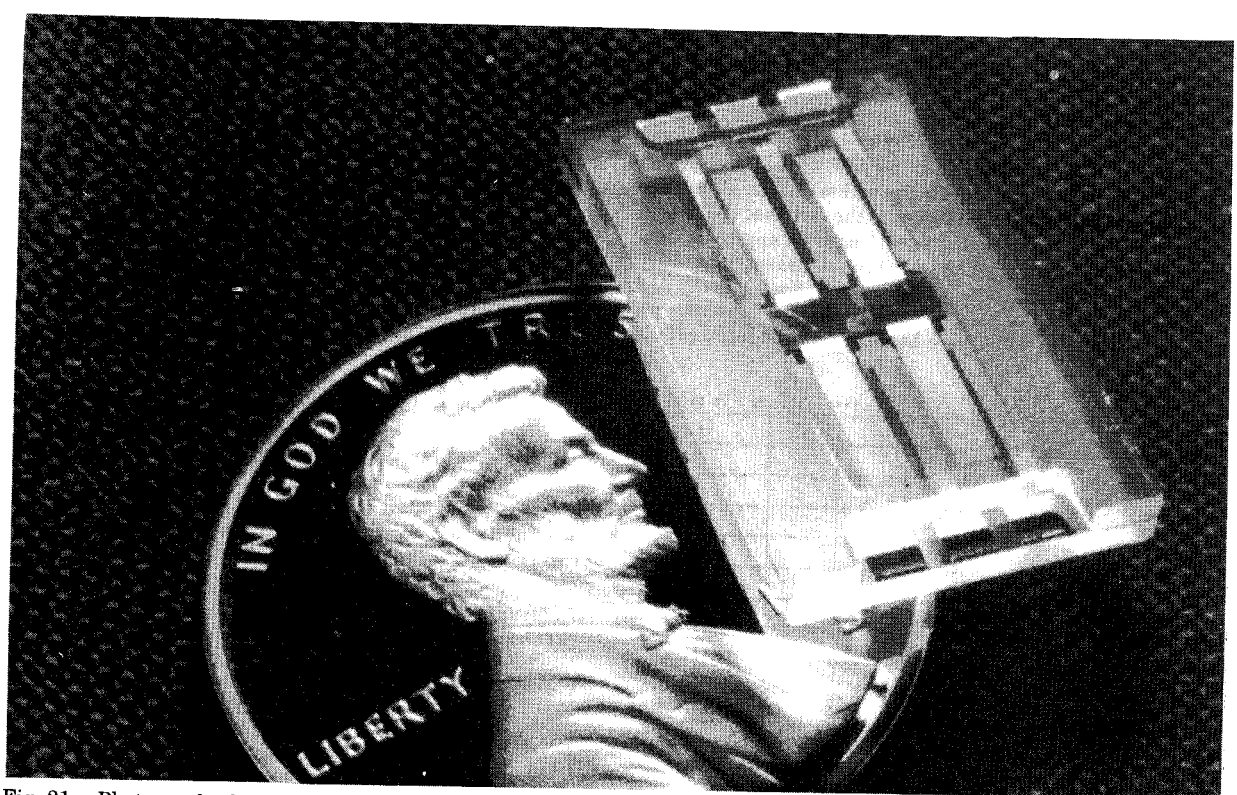


Fig. 21 Photograph of a Dual Channel 350 MHz AQP SAW Resonator Device.

A typical fractional frequency versus temperature characteristic for a 350 MHz AQP SAW hybrid circuit oscillator is shown in Fig. 23. The oscillator's turn-over temperature is approximately 72°C, which is about 4°C to 6°C lower than the AQP SAW resonator device's turn-over temperature. The lower turn-over temperature observed on the oscillator, compared to the SAW device, is attributable to the temperature sensitivity of the oscillator's electronic components. Typically, the variation in turn-over temperature from oscillator-to-oscillator is within a ±3°C to ±5°C window. Note that Fig. 23 also indicates the variation of the oscillator's RF output power versus temperature. As may be readily seen, the oscillator's output power varies by only ±0.2 dB over the temperature range from -55°C to +85°C. This is an extremely good result, and is a tribute to the RF amplifier's stability over temperature.

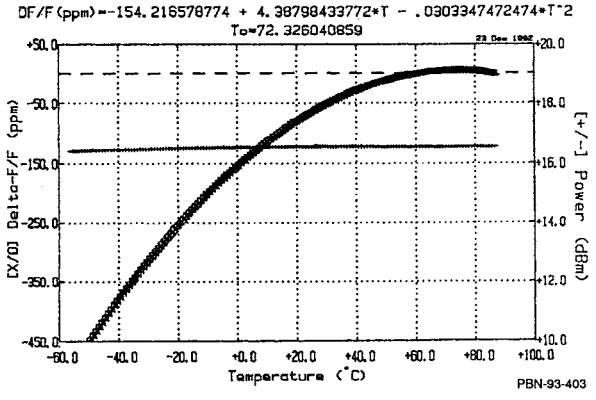


Fig. 23 Fractional Frequency and RF Output Power Stabilities versus Temperature for a 350 MHz AQP SAW Hybrid Circuit Oscillator.

The 350 MHz AQP SAW resonator hybrid circuit oscillator's AM noise spectrum is shown in Fig. 24. This measurement is basically consistent with previous results, although the 3 to 5 dB rise in the AM noise spectrum for carrier offset frequencies from 50 kHz to 2 MHz has not been observed previously. One possible design difference that may account for this behavior is the fact that the hybrid oscillator's buffer amplifier is operating with approximately 3 to 4 dB of gain compression. Additional measurements are planned to investigate this AM noise behavior since the AM noise level would ordinarily be expected to be comparable to, or less than, the PM noise level for all carrier offset frequencies.

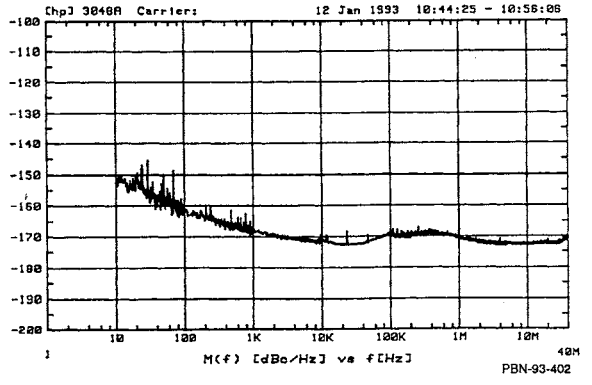


Fig. 24 Single-sideband AM Noise Spectrum for an Extremely Low-noise AQP SAW Resonator Hybrid Circuit Oscillator.

An oscillator's sensitivity to vibration is an important performance criteria. Figure 25 illustrates, for a typical 500 MHz oscillator, the potential phase noise performance degradation which would result from a random vibration spectrum (G = $1x10^{-5}$ g^2/Hz, 20 Hz to 3 kHz) whose RMS g value is only 0.12 gs. The potential phase noise degradation is illustrated for two different vibration sensitivities, namely: $1x10^{-9}/g$ and $1x10^{-10}/g$. This particular random vibration spectrum is typical of ground-based system application, whereas airborne or missile environments are characterized by spectra with RMS g levels of perhaps 5 to 25 gs. Figure 25 shows that the phase noise spectrum for an oscillator with a vibration sensitivity of $1x10^{-9}/g$ will be degraded significantly even for the random vibration spectrum just described. If the vibration sensitivity is an order of magnitude lower, there is still the potential for some degradation to the oscillator's phase noise spectrum. In most applications, even when the oscillator's vibration sensitivity is as low as $1x10^{-10}/g$, additional vibration isolation techniques must be used.

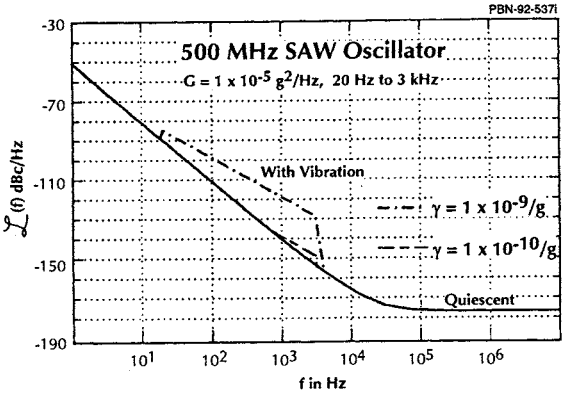


Fig. 25 Calculated Phase Noise Degradation due to a Random Vibration Spectral Density Distribution, for a 500 MHz SAW Oscillator.

The AQP SAW hybrid circuit oscillator must be mounted on a very stiff (non-bending) surface in order to achieve good vibration sensitivity performance [Ref. 18]. Our current approach is shown in Fig. 26, where the base of the hybrid circuit oscillator's package is bonded to a 0.4" (1.0 cm) thick alumina stiffener. Alumina is a good choice of material since it is about 5 times stiffer than aluminum. A thinner (0.040" (0.1 cm)) piece of alumina is bonded to the cover of the oscillator's package in order to eliminate a cover resonance around 3 kHz. This stiffened oscillator assembly is capable of providing exceptionally good vibration sensitivity performance. This is illustrated in Fig. 27 for a typical 350 MHz AQP SAW resonator hybrid circuit oscillator similar to the one shown in Fig. 20. The vibration sensitivity magnitude for this oscillator is only about $2x10^{-10}/g$, dominated by the vibration sensitivity vector's component in a direction normal to the plane of SAW propagation. Using this basic approach, we have consistently been able to achieve vibration sensitivity magnitudes in the 1 to $6x10^{-10}/g$ for AQP SAW resonator hybrid circuit oscillators operating in 300 to 500 MHz frequency range. In fact, one 900 MHz AQP SAW hybrid circuit oscillator exhibited a vibration sensitivity magnitude of $5x10^{-11}/g$. Vibration sensitivity magnitudes this low are approaching theoretically predicted values in the 1 to $2x10^{-11}/g$ range [Ref. 36]. We have not as yet been able to reproducibly achieve oscillators with vibration sensitivities below $1x10^{-10}/g$. Performance ultimately at this level is a major objective of our current work.

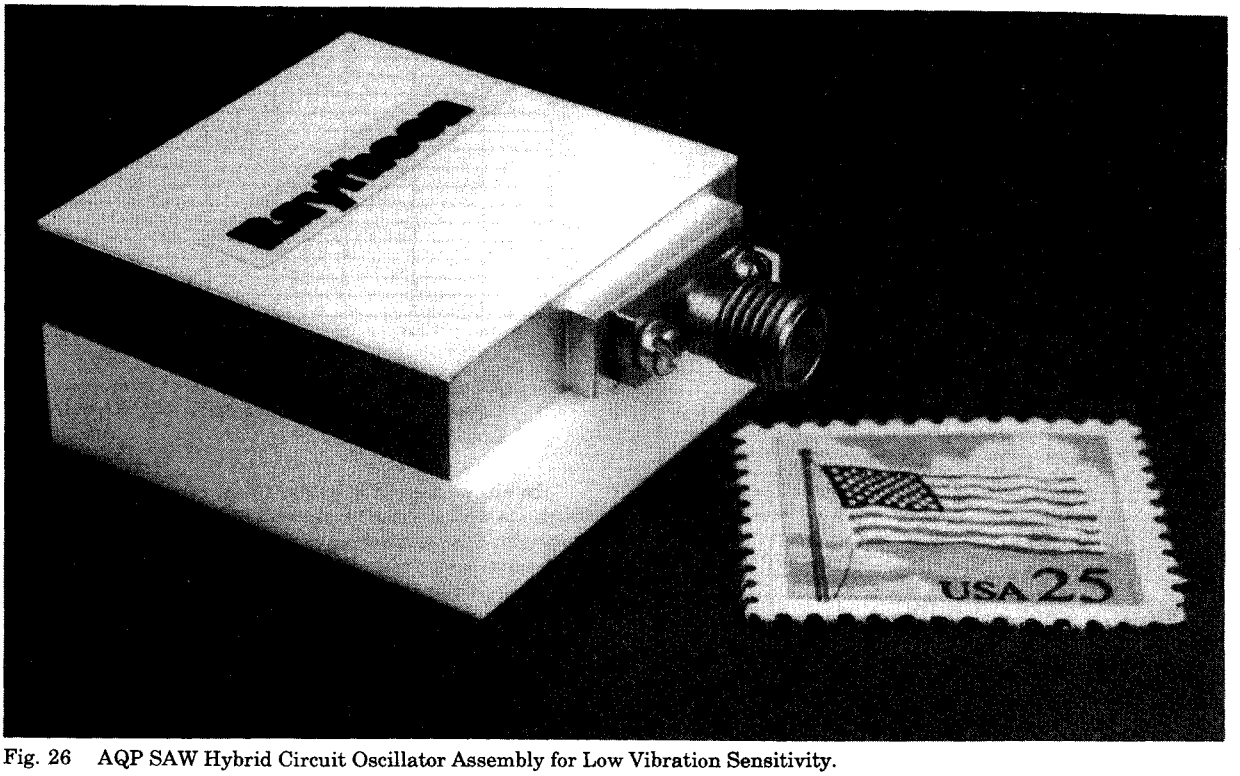


Fig. 26 AQP SAW Hybrid Circuit Oscillator Assembly for Low Vibration Sensitivity.

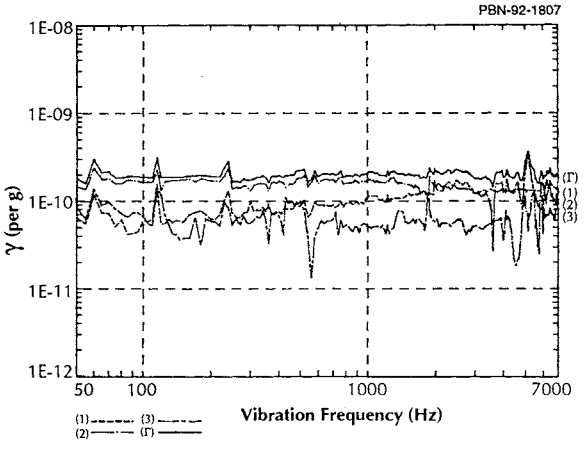


Fig. 27 Vibration Sensitivity Performance for a 350 MHz AQP SAW Resonator Hybrid Circuit Oscillator.

7.4 *Low Phase Noise SAW Delay Line PCB VCOs*

While SAW delay line oscillators have generally been considered too noisy for most low-noise applications, we have recently found that AQP SAW delay line devices can possess flicker phase noise levels which are substantially better than previously expected. This was indicated in Fig. 8 for a 675 MHz AQP SAW delay line device. In general, some of this improvement undoubtedly results from the fact that a high power RF burn-in has also been found to be effective in reducing the flicker phase noise of SAW delay line devices. However, as noted previously, the AQP SAW delay line device whose residual flicker noise level was shown in Fig. 8 ***did not*** receive a high power RF burn-in. Careful design of the AQP structure for SAW delay line devices helps to minimize reflected signals, while the recessed transducer metalization (and "dummy" electrodes if necessary) will help to eliminate spurious triple transit reflections. It has been shown that these spurious signals may contribute to a device's measured flicker phase noise level [Ref. 22]. It goes without saying that the transducer metalization process (Cu-doped aluminum in all of our SAW resonator and delay line devices, deposited using standard E-beam evaporation techniques) should obviously be well controlled and reproducible, since we have seen isolated instances where a device's flicker noise has been significantly degraded due to improper film deposition conditions.

Figure 28 shows the basic circuit schematic for an engineering prototype 400 MHz AQP SAW delay line PCB oscillator. The SAW device's group delay was approximately 1.6 μsec. All three RF amplifiers were Avantek UTO-509s. This particular oscillator has a usable tuning range of ±150 kHz (sufficient to maintain phase lock to a stable reference signal over temperature as well as all other environmentally induced frequency changes), yet it still has a white phase noise floor of approximately -170 dBc/Hz and a flicker FM noise level of -70 dBc/Hz at 10 Hz carrier offset, as illustrated by the measured phase noise spectrum shown in Fig. 29. The broad "dip" in the oscillator's phase noise floor which is evident near 350 kHz carrier offset is due to noise degeneration in the feedback loop at frequencies $\pm 1/(2\tau_g)$ ($\cong \pm 325$ kHz) where the net transmission phase shift through the feedback loop is 180° and the SAW delay line device's insertion loss is not too large. In general, the smaller group delay and higher insertion loss of a SAW delay line in comparison to a SAW resonator will preclude delay line based oscillators from achieving phase noise levels comparable to those demonstrated by SAW resonator oscillators, but in many applications, e.g., very wideband VCOs, the considerably larger tuning range of a SAW delay line oscillator is quite important. The phase noise performance shown in Fig. 29 serves to illustrate that good phase noise performance may still be achieved even for a relatively wideband SAW delay line oscillator.

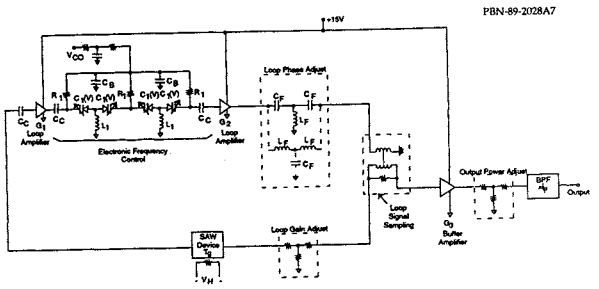


Fig. 28 Circuit Schematic for a Low-noise SAW Delay Line VCO.

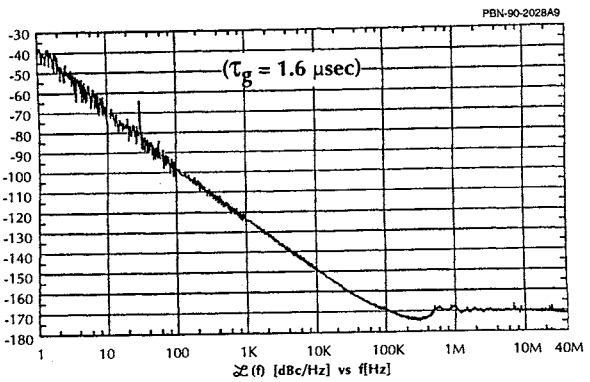


Fig. 29 Measured Phase Noise Spectrum for One Engineering Prototype 400 MHz AQP SAW Delay Line PCB VCO.

More recently, laboratory prototype 675 MHz AQP SAW PCB VCOs were evaluated based upon devices similar to those whose residual flicker noise was shown in Fig. 8. The same basic circuit design as shown in Fig. 28 was once again used, but now the three amplifiers were each UTO-1005s. This particular amplifier is essentially "identical" to the UTO-509, except that reactive broadbanding has been used to achieve comparable electrical performance to 1 GHz, rather than to the 500 MHz upper limit of the UTO-509. The measured phase noise spectrum for one such oscillator is shown in Fig. 30. Since the group delay of this device is only 0.45 µsec, an oscillator properly configured with this delay line device would be capable of electronic frequency tuning over a ±550 kHz range, with negligible degradation in its phase noise spectrum. In spite of this extremely wide tuning range potential, the oscillator still possesses a white phase noise floor of approximately -167 dBc/Hz, and a flicker FM noise level of -60 dBc/Hz at 10 Hz carrier offset. The broad "dip" in the oscillator's phase noise floor near 1 MHz carrier offset is once again attributable to noise degeneration in the feedback loop for frequencies $\pm 1/(2\tau_g)$ ($\cong$ 1.1 MHz) removed from the carrier, where the net loop transmission phase shift is 180°. The approximate 10 dB degradation in flicker FM noise level for this delay line oscillator, in comparison to the result for the 400 MHz SAW delay line oscillator's flicker FM noise level shown in Fig. 29, is due almost entirely to the difference in group delay between the two devices (i.e., $10\text{x}\log_{10}(1.6/0.45)^2$ = 11.0 dB) [Ref. 6]. The difference in white phase noise floor levels between the two delay line oscillators is primarily due to the somewhat inferior noise figure of the UTO-1005 in this frequency range in comparison to the UTO-509 when operating at lower frequencies, as well as to the fact that the output power at 3 dB of gain compression for the UTO-1005 is also somewhat reduced compared to that for the UTO-509.

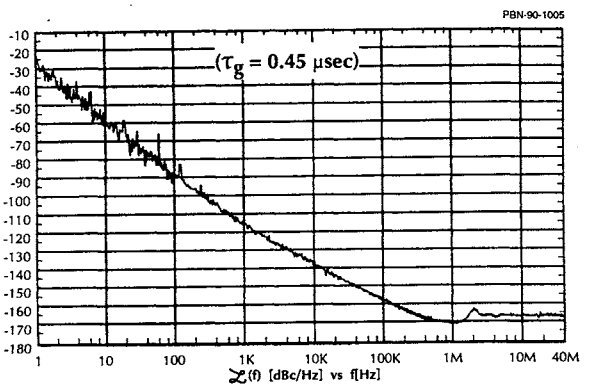


Fig. 30 Measured Phase Noise Spectrum for One Laboratory Prototype 675 MHz AQP SAW Delay Line PCB VCO.

In either case, it is evident that careful SAW device design, as well as close attention to processing parameters, along with close monitoring of electronic component residual phase noise levels can result in SAW delay line VCOs with extremely wide tuning ranges, while simultaneously maintaining very low phase noise levels.

7.5 *Summary*

It is indeed gratifying that the basic feedback loop oscillator design, whose phase noise was first analyzed by Leeson [Ref. 7], may be successfully applied under such a wide variety of circumstances, including the potential use of many different classes of frequency stabilizing element, with equally satisfactory results. In fact, it is perhaps appropriate to conclude with the observation that when properly designed and implemented, the feedback-loop oscillator configuration is well suited to the realization of truly state-of-the-art oscillator performance. Based upon our own experience, this statement applies not only to the exceptionally low phase noise levels which have been achieved, but also to virtually all other measures of an oscillator's performance which might be applied, as previously demonstrated [Refs. 1 - 5].

8. SUMMARY & CONCLUSIONS

We have shown, using several design examples, that the basic feedback-loop oscillator configuration is well suited to high performance oscillator design in the VHF, UHF and microwave frequency ranges, based upon the incorporation of a two-port frequency stabilizing device. State-of-the-art performance was realized for AQP SAW resonator- and delay line-based oscillators.

While the phase noise spectra shown in Figs. 17 and 22 represent the current state-of-the-art for a 500 MHz AQP SAW resonator PCB oscillator and a 350 MHz AQP SAW resonator hybrid circuit oscillator, respectively, by no means do they represent limits to what may ultimately be achieved with further improvements in both SAW resonator design and fabrication techniques, as well as anticipated improvements in RF amplifier residual phase noise levels. For 350 to 500 MHz AQP SAW resonator ***hybrid circuit oscillators*** current goals are to simultaneously demonstrate, while maintaining good long-term frequency stability, the following performance objectives: 1) flicker FM noise levels of -95 to -105 dBc/Hz at 10 Hz carrier offset, 2) -185 dBc/Hz white PM noise floors, and 3) vibration sensitivity ***magnitudes*** reproducibly less than $1\text{x}10^{-10}/g$.

The design and evaluation techniques just discussed have proven to be invaluable in selecting the proper types of components for obtaining truly low-noise oscillator

performance. As shown in Figs. 9 and 11, not all amplifiers are created equal, at least with regard to their residual phase noise levels. Residual phase noise measurements are also very valuable in screening out obviously defective components that would otherwise contribute unacceptably high levels of residual phase noise. The proper evaluation of amplifier characteristics has allowed us to predict an oscillator's white phase noise floor to within about ±1 to ±2 dB. However, the ability to predict flicker FM noise levels on individual oscillators (based upon component residual noise measurements) has been somewhat limited due to the fact that in many cases the residual flicker PM noise levels of the components are at or below measurable levels. Experience with more than one-hundred fully characterized printed circuit board and hybrid circuit oscillators (SAW resonator and delay line based) has given us considerable confidence in the feedback-loop oscillator based design procedures which have been described.

REFERENCES

[1] G. K. Montress, T. E. Parker, and M. J. Loboda, "Extremely Low Phase Noise SAW Resonator Oscillator Design and Performance", in *Proceedings of the IEEE Ultrasonics Symposium*, 1987, Vol. 1, pp. 47-52.

[2] G. K. Montress, T. E. Parker, M. J. Loboda, and J. A. Greer, "Extremely Low-Phase-Noise SAW Resonators and Oscillators: Design and Performance", *IEEE Transactions on Ultrasonics, Ferroelectrics, & Frequency Control*, Vol. 35, No. 6, pp. 657-667, November 1988.

[3] T. E. Parker and G. K. Montress, "Low Noise SAW Resonator Oscillators", in *Proceedings of the 43rd Annual Symposium on Frequency Control*, 1989, pp. 588-595.

[4] T. E. Parker and G. K. Montress, "Frequency Stability of High Performance SAW Oscillators", in *Proceedings of the IEEE Ultrasonics Symposium*, 1989, Vol. 1, pp. 37-45.

[5] G. K. Montress and T. E. Parker, "Design Techniques for Achieving State-of-the-Art Oscillator Performance", in *Proceedings of the 44th Annual Symposium on Frequency Control*, 1990, pp. 522-535.

[6] T. E. Parker and G. K. Montress, "Precision Surface Acoustic Wave (SAW) Oscillators", *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, Vol. 35, No. 3, pp. 342-364, May 1988.

[7] D. B. Leeson, "A Simple Model of Feedback Oscillator Noise Spectrum", *Proceedings of the IEEE*, Vol. 54, No. 2, pp. 329-330, February 1966.

[8] T. E. Parker, "Characteristics and Sources of Phase Noise in Stable Oscillators", in *Proceedings of the 41st Annual Symposium on Frequency Control*, 1987, pp. 99-110.

[9] T. E. Parker, J. Callerame, and G. K. Montress, "A New All Quartz Package for SAW Devices", in *Proceedings of the 39th Annual Symposium on Frequency Control*, 1985, pp. 519-525.

[10] G. K. Montress, T. E. Parker, and J. Callerame, "A Miniature Hybrid Circuit SAW Oscillator Using an All Quartz Packaged Resonator", in *Proceedings of the IEEE Ultrasonics Symposium*, 1985, Vol. 1, pp. 277-282.

[11] J. A. Greer and T. E. Parker, "Improved Vibration Sensitivity of the All Quartz Package Surface Acoustic Wave Resonator", in *Proceedings of the 42nd Annual Symposium on Frequency Control*, 1988, pp. 239-251.

[12] J. A. Greer, T. E. Parker, M. Rothschild, and D. J. Erlich, "A Novel Technique for Trimming the Frequency of a Sealed Surface Acoustic Wave Resonator", in *Proceedings of the 41st Annual Symposium on Frequency Control*, 1987, pp. 351-359.

[13] J. A. Greer, G. K. Montress, and T. E. Parker, "Applications of Laser-Trimming for All Quartz Package, Surface Acoustic Wave Devices", in *Proceedings of the IEEE Ultrasonics Symposium*, 1989, Vol. 1, pp. 179-184.

[14] A. D'Almeida, P. Leblois, D. Hauden, S. Calisti, J. Y. Duquesnoy, and L. Penavaire, "Aging in SAW Oscillators", in *Proceedings of the 2nd European Frequency and Time Forum*, 1988, pp. 739-751.

[15] S. Calisti, P. Defranould, L. Penavaire, D. Hauden, and A. D'Almeida, "All Quartz Packaged S.A.W. Resonator", in *Proceedings of the 3rd European Frequency and Time Forum*, 1989, pp. 165-168.

[16] L. Penavaire, S. Calisti, M. Solal, and P. Dufilie, "State of the Art SAW Resonators in France", in *Proceedings of the IEEE Ultrasonics Symposium*, 1989, Vol. 1, pp. 185-190.

[17] P. Defranould, L. Penavaire, S. Calisti, and F. Raymond, "State-of-the-Art of SAW Resonators", in *Proceedings of the 4th European Frequency and Time Forum*, 1990, pp. 395-404.

[18] T. E. Parker, J. A. Greer, and G. K. Montress, "SAW Oscillators with Low Vibration Sensitivity", in *Proceedings of the 45th Annual Symposium on Frequency Control*, 1991, pp. 321-329.

[19] W. J. Tanski, "Surface Acoustic Wave Frequency Trimming of Resonant and Travelling-Wave Devices on Quartz", *Applied Physics Letters*, Vol. 39, No. 1, pp. 40-42, July 1, 1981.

[20] P. S. Cross and W. R. Shreve, "Frequency Trimming of SAW Resonators", *IEEE Transactions on Sonics and Ultrasonics*, Vol. SU-29, No. 4, pp. 231-234, July 1982.

[21] W. J. Tanski and H. van de Vaart, "The Design of SAW Resonators on Quartz with Emphasis on Two Ports", in *Proceedings of the IEEE Ultrasonics Symposium*, 1976, pp. 260-265.

[22] W. R. Shreve and P. S. Cross, "Surface Acoustic Wave Resonators", in *Precision Frequency Control: Acoustic Resonators and Filters*, Vol. 1, E. A. Gerber and A. Ballato, Eds. New York: Academic Press, 1985, pp. 118-145.

[23] W. R. Shreve, R. C. Bray, S. Elliott, and Y. C. Chu, "Power Dependence of Aging in SAW Resonators", in *Proceedings of the IEEE Ultrasonics Symposium*, Vol. 1, 1981, pp. 94-99.

[24] J. Crabb, M. F. Lewis, and J. D. Maines, "Surface-Acoustic-Wave Oscillators: Mode Selection and Frequency Modulation", *Electronics Letters*, Vol. 9, No. 10, pp. 197-199, May 17, 1973.

[25] G. K. Montress, T. E. Parker, D. W. Kress, and J. A. Kosinski, "Design and Performance of a Low Noise, Wide Tuning Range AQP SAW Delay Line VCO", in *Proceedings of the IEEE Frequency Control Symposium*, 1992, pp. 356-370.

[26] F. L. Walls and A. E. Wainwright, "Measurement of the Short-Term Stability of Quartz Crystal Resonators and the Implications for Crystal Oscillator Design and Applications", *IEEE Transactions on Instrumentation and Measurement*, Vol. IM-24, No. 1, pp. 15-20, March 1975.

[27] F. L. Walls and C. M. Felton, "Low Noise Frequency Synthesis", in *Proceedings of the 41st Annual Symposium on Frequency Control*, 1987, pp. 512-518.

[28] T. R. Faulkner and R. E. Temple, "Residual Phase Noise and AM Noise Measurements and Techniques", Hewlett-Packard Application Note, H-P Part Number 03048-90011.

[29] G. S. Curtis, "The Relationship Between Resonator and Oscillator Noise, and Resonator Noise Measurement Techniques", in *Proceedings of the 41st Annual Symposium on Frequency Control*, 1987, pp. 420-428.

[30] T. F. O'Shea, M. R. Lewis, and B. R. Horine, "Low-Noise Voltage Controlled SAW Oscillator for Phase Lock Loop Application", in *Proceedings of the IEEE Ultrasonics Symposium*, 1987, Vol. 1, pp. 53-59.

[31] G. K. Montress, T. E. Parker, and M. J. Loboda, "Residual Phase Noise Measurements of VHF, UHF, and Microwave Components", in *Proceedings of the 43rd Annual Symposium on Frequency Control*, 1989, pp. 349-359.

[32] 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", in *Proceedings of the IEEE Ultrasonics Symposium*, 1987, Vol. 1, pp. 43-46.

[33] R. L. Baer, "Phase Noise in Surface-Acoustic-Wave Filters and Resonators", *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, Vol. 35, No. 3, pp. 421-425, May 1988.

[34] B. K. Sinha, "On the Interaction of Temperature Gradients with Surface Wave Propagation in Anisotropic Solids", *IEEE Transactions on Sonics and Ultrasonics*, Vol. SU-32, No. 4, pp. 583-591, July 1985.

[35] H. F. Tiersten, D. S. Stevens, and R. D. Weglein, "Acceleration Sensitivity of Surface Acoustic Wave Resonators", *Applied Physics Letters*, Vol. 48, No. 2, pp. 97-99, January 13, 1986.

[36] H. F. Tiersten and D. V. Shick, "An Analysis of Normal Acceleration Sensitivity of ST-cut Quartz Surface Acoustic Wave Resonators with Substrate Extending Beyond the Supports", in *Proceedings of the 2nd European Frequency and Time Forum*, 1988, pp. 753-765.

[37] H. F. Tiersten and D. V. Shick, "An Analysis of the In-Plane Acceleration Sensitivity of ST-cut Quartz Surface Acoustic Wave Resonators with Interior Rectangular Support", in *Proceedings of the 43rd Annual Symposium on Frequency Control*, 1989, pp. 396-404.

MODELING OF A GENERAL STRUCTURE OF SAW ACOUSTICALLY AND ELECTRICALLY COUPLED RESONATOR FILTERS

D.P. Chen, J. Heighway and C.S. Hartmann*

Advanced SAW Products SA, CH-2022 Bevaix, Switzerland
*Hartmann Research Inc., Dallas, TX, USA

ABSTRACT

To accurately model SAW filters which are formed by acoustically and/or electrically coupled single mode resonators is of great practical interest. Since 1980s several successful kinds of coupled resonator filters have been reported. The structures of these devices seem different, but because of common features shared by these structures it is possible to give them a unified model. In this paper the new modeling approach we used for SAW transversely coupled resonator filters is extended to model a general 4 electric port structure. Based on this model, both transverse and longitudinal acoustic coupling effects can be well treated. This model has been used to analyze and design transversely coupled resonator filters(including two lowest modes plus spurious high order transverse modes) with narrow band on Quartz . Also it has been used for multiple longitudinal mode coupled resonator filters on 36°-$LiTaO_3$ with wider bandwidth. The comparisons of the theoretical predictions to the experiments show a excellent agreement for the devices on Quartz and the reasonably good agreement for the devices on 36°-$LiTaO_3$, which are working with a leaky surface wave.

1. INTRODUCTION

A great challenge for the use of SAW techniques comes from the increasing requirement for miniaturized, low loss and moderate bandwidth filters for use in the front end of compact, mobile communication systems. Coupled resonator filters (CRF) are widely recognized as a potential means to meet these needs rather than conventional SAW transverse filters. In fact, since 1980s several kinds of successful CRF devices have been reported[1-3]. CRF are defined as a filter formed with at least two single mode resonators, which may correspond to different working modes (for example, different orders of modes in longitudinal dimension and/or modes in transverse dimension), through various coupling mechanism (acoustical and/or electrical). In principle, they are different from conventional transverse bandpass filter because they are based on resonators. Also they are different from single mode resonators due to the extra coupling mechanisms which usually are crucial issues for getting the required performance. Therefore, accurate modeling of these CRF devices has become a new challenge in SAW techniques. Up to now, as we know, different approaches have been chosen to model these devices. For example, the equivalent circuit model is used for two coupled longitudinal mode filters[2], the combined coupling-of-mode formalism and the transmission-matrix approach for longitudinally coupled resonator filters on Quartz [4] and coupling-of-modes in time and in space approach for two port two pole waveguide-coupled SAW resonator filters[5]. In our previous paper[6] an innovative modeling approach was introduced for acoustic-transversely coupled resonator filters based on the closed form expressions of mixed P matrices of the grating and IDT which are derived by coupling-of-mode theory and an excellent agreement between theory and experiments for devices on ST-Quartz was found over 80 dB dynamic range. However, the basic method of this model is also available to other coupled resonator filters. In this paper this model will be extended to the structures with both transverse modes and longitudinal modes and to the materials which possess a high piezoelectric coupling coefficient and can support non-Rayleigh surface modes.

Any modeling must be related to some specific structure. The basic structure described by our model is one track 4 electric port structure composed of 9 gratings and 4 IDT's with some spacers between them. The next section discusses why this structure is chosen and how the often used coupled resonator filters can be represented by it. In section 3 the model for this structure is described by a 4x4 Y matrix which is derived based on cascading mixed P matrices of gratings and IDT's. The parasitic electrostatic elements are able to be properly added. The model was at first applied to a test structure which is designed for experimental determination of coupling-of-modes parameters used in transversely coupled resonator filters on ST-Quartz[7]. The theory and the experiment are in a excellent agreement with each other over 80 dB dynamic range. Finally, an application of this model is made to an analysis of a transversely coupled resonator filter (TCF) on ST-Quartz, which works with Rayleigh wave and over a narrow pass bandwidth of less than 0.1%, and where the spurious high order transverse modes are included. The theoretical results show a good agreement with experimental results. Further, this model is applied to one track acoustically coupled resonator filter in 36°-$LiTaO_3$ which works with leaky surface wave and a pass bandwidth of more than 2%. Even though the closed form P matrix used for the gratings and the IDTs doesn't include all the specific features possessed by the leaky surface waves the theoretical prediction is in reasonably good agreement with experiment. This means that the model based on the closed form P matrix for the grating and IDT is still valid for high piezoelectric materials,

non-Rayleigh waves, and thick metal strips over a moderate bandwidth to the first order approximation.

2. THE MODELED GENERAL STRUCTURE

Fig. 1 shows a one track structure composed of 5 gratings, 4 IDT's and some spacers between them. It is a longitudinally resonating type of structure. The surface acoustic waves propagating underneath can be represented by longitudinal cavity modes which are coupled to each other. We choose it as the basic structure to model because various structures of interest could be represented by it.

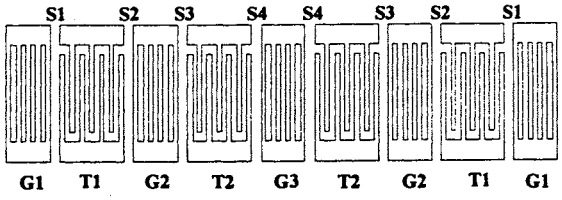


Fig. 1 A basic 4 electric port structure to be modeled

For example, as far as the transverse acoustic coupling effect is concerned, the structure to be treated should be the two track structure shown in Fig. 2. In terms of transverse eigenmode presentation it can be reduced into one track structure as shown in Fig. 1. This structure is described with a 8x8 Y matrix,

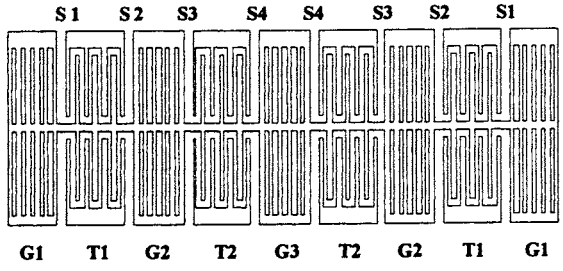


Fig. 2 A two track structure where transversely acoustic coupling effect is involved

$$[I_i]=[Y_{ij}]\times[V_j] \qquad i,j=1,2,.....,7,8 \qquad (1)$$

Let the superscripts S and a denote the symmetric and the antisymmetric transverse eigenmodes respectively. The following conditions are satisfied,

$$I_i^S = I_i + I_{i+4} \quad \text{and} \quad V_i^s = V_i = V_{i+4}, \qquad (2)$$

for the symmetric modes and

$$I_i^a = I_i + I_{i+4} \quad \text{and} \quad V_i^a = V_i = -V_{i+4}, \qquad (3)$$

for the anti-symmetric modes, where $i=1,2,3,4$. Therefore the Y matrices for both modes are as follows,

$$Y_{ij}^s = 2(Y_{ij} + Y_{i(j+4)}) \qquad (4)$$

and

$$Y_{ij}^a = 2(Y_{ij} - Y_{i(j+4)}) \qquad (5)$$

where $i,j=1,2,3,4$ and Y_{ij}^s and Y_{ij}^a should be thought as a summation of Y matrices for all of the possible symmetric and anti-symmetric modes respectively. Inversely, the Y matrix of 8-port structure can be expressed by the Y matrices of these eigenmodes,

$$Y_{ij} = \frac{1}{4}(Y_{ij}^s + Y_{ij}^a) \quad \text{and} \quad Y_{i(j+4)} = \frac{1}{4}(Y_{ij}^s - Y_{ij}^a) \qquad (6)$$

and by reciprocity other elements can be found. This means that the 8-port structure could be deduced by only considering a 4-port structure provided that the Y matrices for the eigenmodes in this 4-port structure are known. Now the mode analysis problem with both transverse and longitudinal distributions involved can be dealt with in the one track 4 electrical port structure.

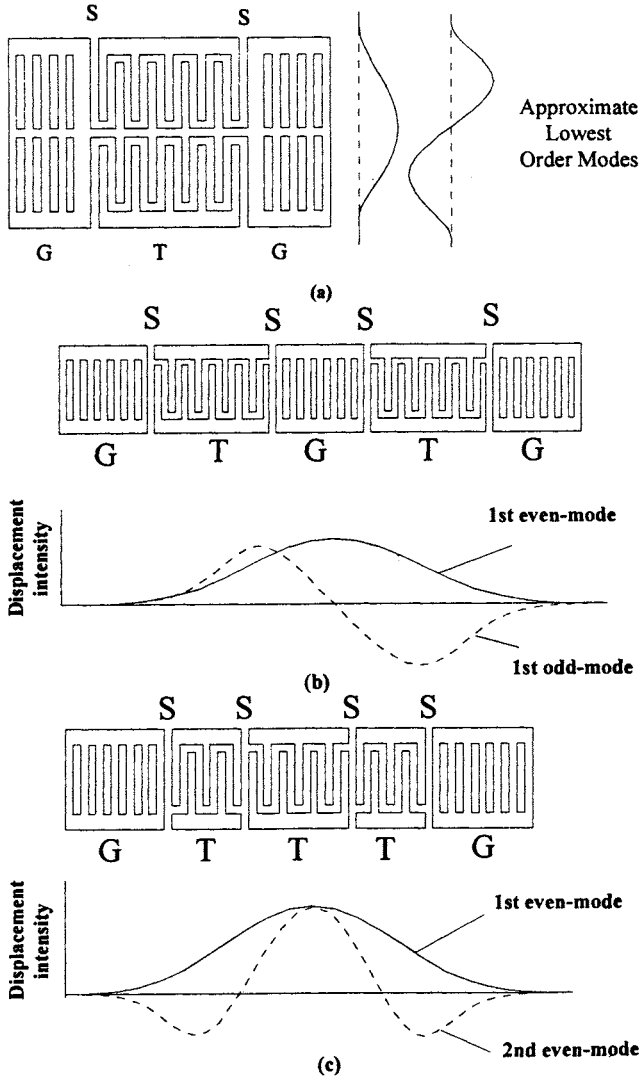


Fig. 3 Three coupled resonator filter structures of interest: (a) two track transversely coupled resonator filter, (b) one track two port longitudinal coupled resonator filter working with both even and odd order modes and (c) one track two port longitudinal coupled resonator filter with even order modes only.

Fig. 3 (a), (b) and (c) show three basic structures of coupled resonator filters which have been widely investigated. Here (a) is two track transversely coupled resonator filter and it is a simplified version of Fig. 2. So it can be represented by the basic structure model. Both (b) and (c) are one track two port coupled resonator filters, but (b) is working with both even and odd longitudinal modes and (c) only with even modes due to the structure symmetry. They can be deduced from the basic structure.

In the next section, the Y matrix of the 4 electric port structure shown in Fig. 1 is deduced.

3. Y Matrix Description for The Basic Structure

The SAW grating and IDT can be described by 2-port and 3-port P matrices respectively, defined in Fig. 4 (a) and (b). The elements of the P matrices are given in previous paper[6] with closed forms.

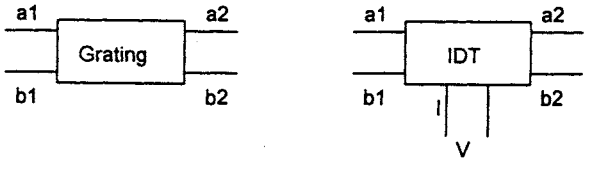


Fig. 4 The diagrams for defining the mixed P matrices of grating and IDT : (a) for grating, (b) for IDT.

$$\begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} P_{11}^G P_{12}^G \\ P_{21}^G P_{22}^G \end{bmatrix} \bullet \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} \qquad \begin{bmatrix} b_1 \\ b_2 \\ I \end{bmatrix} = \begin{bmatrix} P_{11}^T P_{12}^T P_{13}^T \\ P_{21}^T P_{22}^T P_{23}^T \\ P_{31}^T P_{32}^T P_{33}^T \end{bmatrix} \bullet \begin{bmatrix} a_1 \\ a_2 \\ V \end{bmatrix}$$

The basic 4 electric port structure can be described with a 6-port P matrix obtained by cascading the P matrices of all the 9 gratings, 4 IDT's and the spacers between according to the cascade conditions,

$$a_2^i = b_1^{i+1} \qquad \text{and} \qquad b_2^i = a_1^{i+1} \tag{7}$$

where the superscript i and $i+1$ indicate the variables corresponding to ith and (i+1)th sections respectively.

$$\begin{bmatrix} a_1 \\ a_2 \\ I_1 \\ I_2 \\ I_3 \\ I_4 \end{bmatrix} = \begin{bmatrix} P_{11}P_{12}P_{13}P_{14}P_{15}P_{16} \\ P_{21}P_{22}P_{23}P_{24}P_{25}P_{26} \\ P_{31}P_{32}P_{33}P_{34}P_{35}P_{36} \\ P_{41}P_{42}P_{43}P_{44}P_{45}P_{36} \\ P_{51}P_{52}P_{53}P_{54}P_{55}P_{56} \\ P_{61}P_{62}P_{63}P_{64}P_{65}P_{66} \end{bmatrix} \bullet \begin{bmatrix} b_1 \\ b_2 \\ V_1 \\ V_2 \\ V_3 \\ V_4 \end{bmatrix} \tag{8}$$

By definition, the Y matrix for the 4 electric port structure is as follows,

$$[Y_{ij}] = [P_{(i+2)(j+2)}] \tag{9}$$

Note that expression (9) is obtained by the cascading of P matrixes in the longitudinal dimension and only longitudinal modes are involved. An assumption that all the involved longitudinal modes possess the same transverse behavior is implied. When transverse acoustic coupling is considered expression (9) needs to be repeatedly used for each transverse mode and the final Y matrix is given by summation of every mode contribution according to equation (6). Then all the parasitic electrostatic elements, such as parallel capacitances, ground resistances and inductances, series resistances and inductances and bridge capacitances between different ports shown in Fig. 5, can be added to get final expressions for the Y matrix elements. For specific application the required 2-port Y matrix description can be easily deduced according to the connection requirement. For example, for the structure shown in Fig. 3 (c) the Y matrix will be a 3-port matrix and then the final 2-port Y matrix for the device can be obtained by connecting the first and the third ports in parallel.

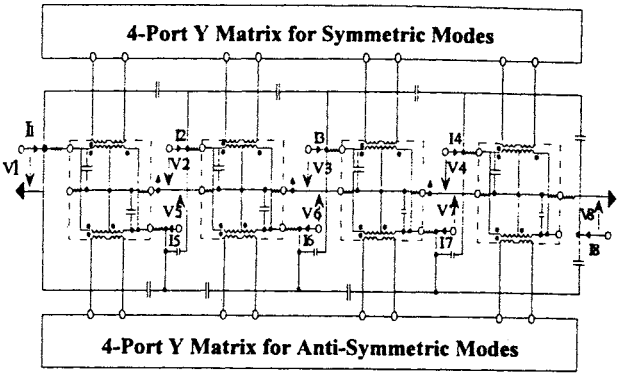


Fig. 5 The schematic diagram including possible parasitic electrostatic elements

4. EXPERIMENTAL VERIFICATIONS

At first this model was used to design a special two track test structure on ST-Quartz. Each track has 3 electric ports and two tracks are transversely coupled to each other. The test structure is designed for experimental determination of COM parameters for transversely coupled resonator filters[7]. Fig. 6 shows a typical result for the fundamental anti-symmetric mode obtained from this structure. An excellent agreement between the theory and the experiment is obtained over 80 dB dynamic range.

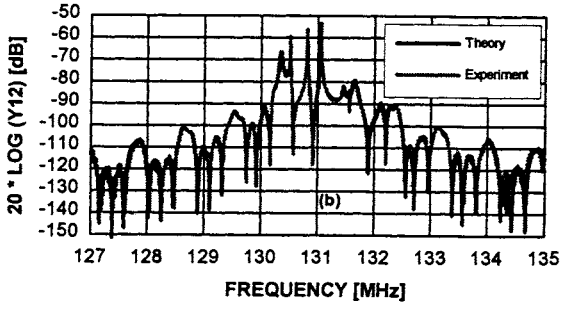


Fig. 6 The theoretical and the experimental results of longitudinal mode distribution which are the anti-symmetric in transverse dimension within the test structure

To verify the suitability of this model to various coupled resonator filters, it has been used to analyze two different kinds of coupled resonator filters. One is the two pole transversely coupled resonator filter on ST-Quartz shown in Fig. 3 (a), which has a fractional bandwidth of less than 0.1% and employs Rayleigh waves. Another is the one track two port longitudinally coupled resonator filter on 36°-$LiTaO_3$ shown in Fig. 3 (b), which has a fractional bandwidth about 2.5% and employs leaky surface waves. The experimental results are shown in Fig. 7 (a) and (b) separately along with the theoretical analyses. For the former an excellent agreement between the experiment and the theory can be seen even for the spurious high order transverse modes located at high frequency side. The latter show a reasonable good agreement between the theoretical prediction and the measured result where a higher loss parameter is used over the high frequency stopband region than over the passband and the low frequency stopband regions, because of, as is well known [8], the loss effect caused by mode conversion into bulk waves. This result is very encouraging. It indicates that the model based on the

closed form solutions of the conventional COM theory may still be valid for a high piezoelectric coupling material, non-Rayleigh waves, and wide bandwidth at least to the first order approximation. Of course, the COM parameters for different materials, different type of waves and transverse modes are different. The accuracy of the theoretical prediction is strongly dependent on the determined COM parameters for the various materials and the device geometric parameters. This issue is, however, beyond the scope on this paper.

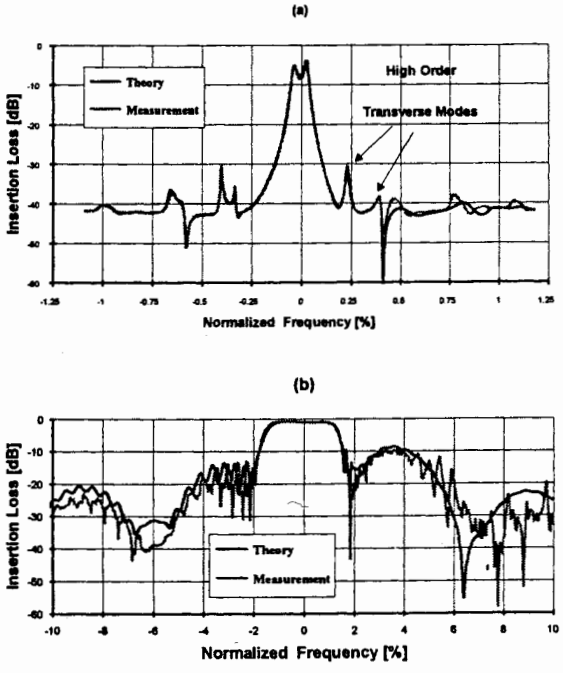


Fig. 7 The comparisons between the theory and the experiment for two kinds of coupled resonator filters: (a) two pole transversely coupled resonator filter on ST-Quartz, (b) one track longitudinally coupled resonator filter on 36°-$LiTaO_3$.

5. CONCLUSION

A modeling for a general structure, where both transverse and longitudinal mode coupling effects are present, has been done based on the cascading of mixed P matrix closed form expressions of the SAW grating and IDT obtained by COM theory. Applications of this model to various cases prove that it is excellent for Rayleigh wave narrow band devices and reasonably good for leaky surface wave devices with moderately wide band on 36°-$LiTaO_3$. Great simplicity and flexibility are the advantage of this model based on COM theory. The crucial issue for usage of this model is to accurately determine the relations of the COM parameters to the various materials and to the device geometric parameters.

6. ACKNOWLEDGMENT

Authors are very grateful to V. Plessky T. Thorvaldsson and P. Durrant for many helpful discussions and gratefully acknowledge the technical support from M. Schwab and M. Sharif.

7. REFERENCES

[1] M. Tanaka, T. Morita, K. Ono and Y. Nakazawa, "Narrow Bandpass Filter Using Double-Mode SAW Resonators on Quartz", 38th Annual Frequency Control Symposium, 1984, pp. 286-293.

[2] M. Hikita, T.Tabuchi, H. Kojima, A. Sumioka, A. Nakagoshi and Y. Kinoshita, "High Performance SAW Filters With Several New Technologies For Cellular Radio", IEEE Ultrasonics Symp. Proc., 1984, pp. 82-92.

[3] K. Anemogiannis, F. Muller, W. Ruile and G. Riha, "High Performance Low-Loss SAW Filters For Mobile Radio With Improved Stoband Rejection", IEEE Ultrasonics Symp. Proc., 1988, pp. 77-82.

[4] G. Scholl, A. Christ, H.P. Grassl, W. Ruile, P. Russer and R. Weigel, "Efficient Design Tool For SAW-Resonator Filters", IEEE Ultrasonics Symp. Proc., 1989, pp. 135-140.

[5] C.K. Campbell, P.M. Smith and P.J. Edmonson, "Aspects of Modeling The Frequency Response Of A Two-Port Waveguide-Coupled SAW Resonator-Filter", IEEE Transaction on UFFC, Vol. 39, No. 6, Nov. 1992, pp. 768-773.

[6] C.S. Hartmann, D.P. Chen and J. Heighway, "Modeling Of SAW Transversely Coupled Resonator Filters Using Coupling-Of-Modes Modeling Technique", 1992 IEEE Ultrasonics Symp. Proc.

[7] C.S. Hartmann, D.P. Chen and J. Heighway, "Experimental Determination Of COM Model Parameters For SAW Transversely Coupled Resonator Filters", 1992 IEEE Ultrasonics Symp. Proc.

[8] V. Plessky and T. Thorvaldsson, "Rayleigh Waves And Leaky SAW In Periodic Systems Of Electrodes : Periodic Green's Function Analysis", 1992 IEEE Ultrasonics Symp. Proc

DESIGN AND PERFORMANCE OF GHz RANGE, LOW-COST OSCILLATORS USING SHEAR HORIZONTAL SURFACE ACOUSTIC WAVES

Ivan D. Avramov

Institute of Solid State Physics, 72 Tzarigradsko Chaussee, 1784 Sofia, Bulgaria

ABSTRACT

A simple oscillator circuit using a two-stage RC bipolar transistor amplifier as an active element and a two-port shear horizontal (SH) surface acoustic wave (SAW) resonator as a frequency stabilizing element is presented, analyzed and experimented. The model, used to analyze the circuit, is based on Barkhausen's feedback theory and allows the oscillator eigenfrequency and the gain compression to be calculated. The circuit allows an optimization of the matching conditions between the SH-wave device and the active element and a variation of the phase and amplitude condition of oscillation only by varying the values of the series and parallel feedback components and without using additional matching circuits other than the bond leads of the acoustic wave device. It was used to build stable SH-wave oscillators in the 750÷2000 MHz range. Data on phase noise, short term- and temperature stability of the oscillators will be presented.

1. INTRODUCTION

Resonator stabilized SAW oscillators, using bipolar modular amplifiers for operation in a 50 Ω environment, have demonstrated outstanding performance with respect to phase noise, aging and vibration sensitivity (Ref. 1). These modular amplifiers are relatively expensive, have fixed gain and need external matching circuits if desired matching conditions to the SAW device are required. For several applications, however, in which the oscillator specifications are not very tight, much simpler circuits may perform satisfactorily. This paper will present analysis and experimental data from a simple SH-wave resonator stabilized oscillator using a two-stage bipolar transistor RC amplifier which allows the gain and phase, the input and output admittances and the output power to be varied in wide limits by proper selection of the series and parallel feedback elements. Thus the circuit offers a great flexibility in the design of low-cost SAW resonator oscillators, is easily analyzed and optimized on a personal computer and can be used for frequencies up to 2 GHz and higher if proper high frequency transistors are used. The circuit is broad band, very stable with temperature and is readily realized in surface mount device (SMD) technology. Data on phase noise, short-term and temperature stability of microwave oscillators, stabilized with two-port STW resonators and based on this circuit will be presented and discussed.

2. ANALYSIS OF SAW TWO-PORT RESONATOR OSCILLATORS

The simplest SAW resonator feedback oscillator consists of an amplifier as an active element and a passive frequency stabilizing element which in our case is a two-port SAW or SH-wave resonator. The basic oscillator schematic is shown in Fig. 1a) while Fig. 1b) shows the oscillator equivalent circuit (Ref. 2).

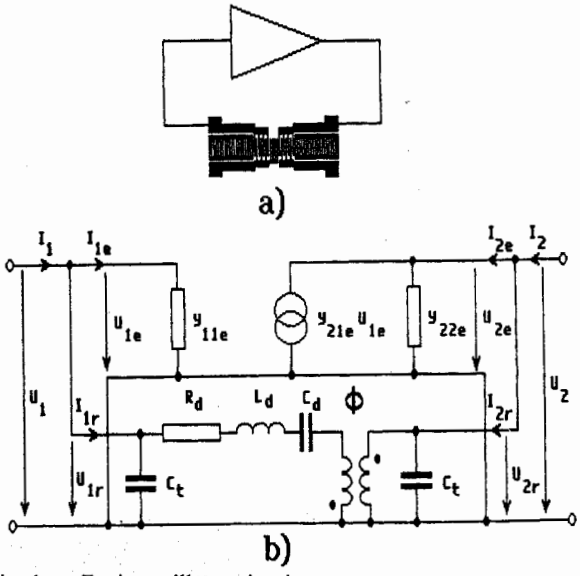


Fig. 1. Basic oscillator circuit:
a) oscillator schematic
b) equivalent electrical circuit.

In this circuit the active element is represented by its Y-matrix $[Y_e]$ as follows:

$$[Y_e]=\begin{bmatrix} Y_{11e} & Y_{12e} \\ Y_{21e} & Y_{22e} \end{bmatrix}=\begin{bmatrix} g_{11}+jb_{11} & 0 \\ Y_{21}\exp(j\phi) & g_{22}+jb_{22} \end{bmatrix} \quad (1)$$

where $g_{11}+jb_{11}$ and $g_{22}+jb_{22}$ are the input and output short circuit admittances of the active element and $Y_{21}\exp(j\phi)$ is its forward transfer coefficient. The backward transfer coefficient Y_{12e} here has been neglected for the sake of simplicity.

The two-port acoustic resonator is described by its equivalent Y-matrix:

$$[Y_r]=\begin{bmatrix} Y_{11r} & Y_{12r} \\ Y_{21r} & Y_{22r} \end{bmatrix} \tag{2}$$

the components of which are calculated as:

$$Y_{11r} = Y_{22r} = j\omega C_T + 1/[R_d + j(\omega L_d - \frac{1}{\omega C_d})] \tag{3}$$

and

$$Y_{12r} = Y_{21r} = -\exp(j\Phi)/[R_d + j(\omega L_d - \frac{1}{\omega C_d})] \tag{4}$$

Here R_d, L_d and C_d are the dynamic resonant resistance, inductance and capacitance respectively, C_T is the static interdigital transducer (IDT) capacitance and Φ is the additional phase shift of 180 deg which can be added or subtracted by reversing the resonator bond leads. The dynamic parameters R_d, L_d and C_d for a given resonator can easily be calculated from the measured values of the resonant frequency f_o, the device insertion loss and its 3 dB bandwidth in a 50Ω system.

Since both 4-poles in Fig 1 are connected in parallel, for the currents and voltages at their input and output terminals, it follows:

$$U_1 = U_{1r} = U_{1e} \qquad I_1 = I_{1r} + I_{1e} \tag{5}$$

$$U_2 = U_{2r} = U_{2e} \qquad I_2 = I_{2r} + I_{2e} \tag{6}$$

From (5) and (6) it follows that the Y-matrix of the equivalent quadripole $[Y]$ is equal to the sum of both matrices $[Y_e]$ and $[Y_r]$:

$$[Y]=\begin{bmatrix} Y_{11} & Y_{12} \\ Y_{21} & Y_{22} \end{bmatrix}=\begin{bmatrix} Y_{11e}+Y_{11r} & Y_{12r} \\ Y_{21e}+Y_{21r} & Y_{22e}+Y_{22r} \end{bmatrix} \tag{7}$$

As I_1 and I_2 are zero, from the definition of Y-parameters it follows that the equivalent 4-pole in Fig. 1 is described by the following system of linear equations:

$$I_1 = Y_{11}U_1 + Y_{12}U_2 = 0$$

$$I_2 = Y_{21}U_1 + Y_{22}U_2 = 0 \tag{8}$$

The system (8) has one or more non-zero solutions for the voltages U_1 and U_2 (oscillation case) if:

$$Det[Y] = 0 \tag{9}$$

Substituting (1), (2), (3) and (4) in (7) and solving (8) with respect to the modulus and phase of Y_{21} we obtain the critical value of the forward transmission coefficient of the system $|Y_{21ck}|$, necessary to maintain oscillation at the k-th eigenmode (k is a positive integer), as well as the eigenfrequency f_k at which oscillation occurs. The amplitude condition of oscillation is considered satisfied if:

$$|Y_{21e}| \geq |Y_{21ck}| \tag{10}$$

which means that the actual value of the amplifier forward transmission coefficient should exceed the calculated critical value according to (9). The oscillator eigenfrequency f_k is calculated from the phase condition of oscillation which will be satisfied if the total phase shift around the loop, according to (9) is a multiple of 2π, namely:

$$\varphi(f_k) = 2k\pi \tag{11}$$

Fig. 2 is a plot of the function $|Y_{21c}(f)|^{-1}$ over the resonance bandwidth. Obviously oscillation will occur at an eigenfrequency f_k for such a value of the integer k which is

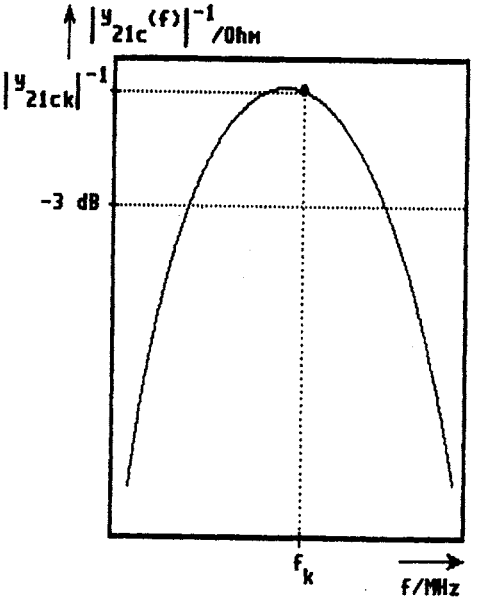


Fig. 2. Graphical visualization of the amplitude and phase condition of oscillation

closest to the resonant frequency where the critical value of the system's forward transmission coefficient achieves minimum. Generally f_k does not correspond to the resonant frequency f_o but can be adjusted to be very close to it in the 1 dB resonant bandwidth by adjusting the phase ϕ. Usually stable oscillation is maintained if sufficient gain compression σ_k at f_k is guaranteed. According to Ref. 2 the system can be considered linear if the gain compression is chosen as:

$$\sigma_k = |Y_{21e}|/|Y_{21ck}| = 2...3 \tag{12}$$

The analysis of the system takes only a few seconds on a personal computer and the final result, shown in Fig. 2, can be optimized for a given SAW or STW two-port resonator as a function of the Y-parameters of the active element so that:

1. a minimum value of $|Y_{21ck}|$ for a desired loaded Q value of the resonant system (3 dB bandwidth of the curve in Fig.2) is achieved,

2. condition (12) for the gain compression is satisfied and

3. the frequency of oscillation is well within the 1 dB bandwidth of the resonant system where the loaded

group delay value is maximum and, therefore, best phase noise performance is achieved (Ref. 1).

Obviously this optimization can be achieved only with an active element which allows a variation of its input and output admittances and of the modulus and phase of its forward transfer coefficient, according to (1), in wide limits. An amplifier which makes such an optimization possible is described in the next section.

3. AN UNIVERSAL BROAD-BAND AMPLIFIER FOR SAW AND STW TWO-PORT RESONATOR OSCILLATORS

Fig. 3a) is a schematic of the broadband RC-amplifier used in this study. It consists of two transistor stages using the transistors T1 and T2. Each transistor has an independent series and parallel feedback, realized with the components R_{e1}, C_{e1}, R_{kb1} and R_{e2}, C_{e2}, R_{kb2} respectively. The

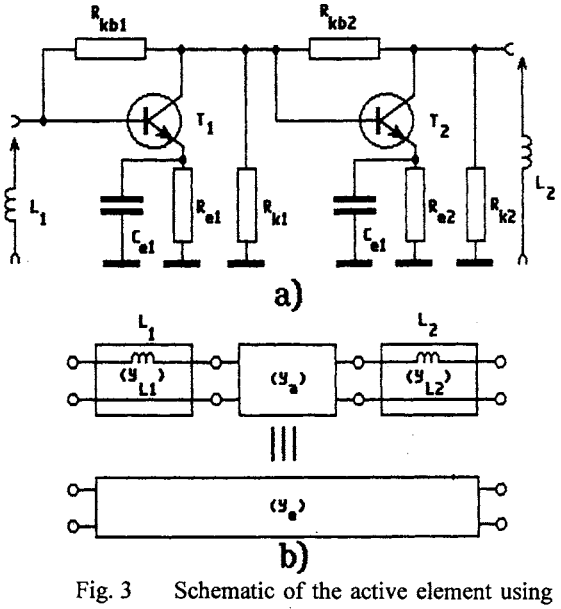


Fig. 3 Schematic of the active element using the RC-amplifier.
a) amplifier circuit
b) building the Y-matrix of the active element including the bond lead inductors L1 and L2.

collector resistors R_{k1} and R_{k2} are also included in the analysis by adding their addmittances to the active component of the output admittances of each transistor stage. The series inductors L1 and L2 represent the bond lead inductances of the wires connecting the two-port acoustic resonator to the amplifier. The Y-parameters of the entire active element according to (1) is obtained by cascading the Y-matrix of the RC-amplifier with the Y-matrices of the inductors L1 and L2 as shown in Fig.2b).

As shown in Ref. 4 the Y-matrix $[Y^S]$ of a transistor stage with a series and parallel feedback, represented by the feedback admittances Y_1 and Y_2 in Fig.4, are calculated from the transistor Y-matrix:

$$[y^T]=\begin{bmatrix} y_{11}^T & y_{12}^T \\ y_{21}^T & y_{22}^T \end{bmatrix} \tag{13}$$

as follows:

$$Y_{11}^S = y_{11}^T + Y_2 - \frac{(y_{11}^T + y_{12}^T)(y_{11}^T + y_{21}^T)}{Y_1 + y_{11}^T + y_{12}^T + y_{21}^T + y_{22}^T} \tag{14}$$

$$Y_{12}^S = y_{12}^T - Y_2 - \frac{(y_{11}^T + y_{12}^T)(y_{12}^T + y_{22}^T)}{Y_1 + y_{11}^T + y_{12}^T + y_{21}^T + y_{22}^T} \tag{15}$$

$$Y_{21}^S = y_{21}^T - Y_2 - \frac{(y_{21}^T + y_{22}^T)(y_{11}^T + y_{21}^T)}{Y_1 + y_{11}^T + y_{12}^T + y_{21}^T + y_{22}^T} \tag{16}$$

$$Y_{22}^S = y_{22}^T + Y_2 - \frac{(y_{21}^T + y_{22}^T)(y_{12}^T + y_{22}^T)}{Y_1 + y_{11}^T + y_{12}^T + y_{21}^T + y_{22}^T} \tag{17}$$

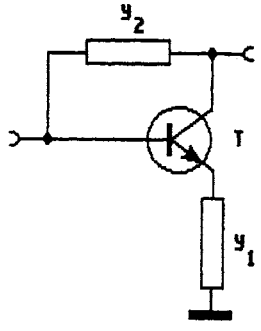


Fig. 4. Schematic of a transistor stage with series and parallel feedback.

In our case the feedback admittancies Y_1 and Y_2 are calculated as:

$$Y_1 = j\omega_o C_e + 1/R_e \text{ and } Y_2 = 1/R_{kb} \tag{18}$$

Then the Y-matrix of the RC-amplifier $[Y^A]$ is obtained by cascading the Y-matrices of both transistor stages $[Y^{S1}]$ and $[Y^{S2}]$ as follows (Ref. 4.):

$$Y_{11}^A = Y_{11}^{S1} - \frac{Y_{12}^{S1} Y_{21}^{S1}}{Y_{11}^{S2} + Y_{22}^{S1}} \tag{19}$$

$$Y_{12}^A = -\frac{Y_{12}^{S1} Y_{12}^{S2}}{Y_{11}^{S2} + Y_{22}^{S1}} \tag{20}$$

$$Y_{21}^A = -\frac{Y_{21}^{S1} Y_{21}^{S2}}{Y_{11}^{S2} + Y_{22}^{S1}} \tag{21}$$

$$Y_{22}^A = Y_{22}^{S1} - \frac{Y_{12}^{S2} Y_{21}^{S2}}{Y_{11}^{S2} + Y_{22}^{S1}} \tag{22}$$

Finally, for obtaining the entire Y-matrix of the active element $[Y_e]$, represented by (1), the Y-matrices of the series indictors, represented by a matrix of the type:

$$[Y^L]=\begin{bmatrix} 1/j\omega_o L & -1/j\omega_o L \\ -1/j\omega_o L & 1/j\omega_o L \end{bmatrix} \tag{23}$$

are necessary. This cascading calculation is performed using equations (19)...(22).

4. PROPERTIES OF THE RC-AMPLIFIER

The RC-amplifier, described in the previous section, has following features which make an easy and quick optimization of the oscillator circuit on a personal computer possible:

1. Each transistor stage has an independent series and parallel feedback which allows the amplifier input and output admittances to be controlled in wide limits by varying the emitter feedback components R_{e1}, C_{e2}, R_{e2} and C_{e2}. Thus the input and output admittances can be made capacitive to achieve a trade-off between loaded resonant Q and insertion loss in the system. If a minimum loaded Q and low-loss operation are desired, e.g. to achieve a maximum tuning range in a voltage controlled oscillator (VCO), then inductive input and output admittances will be necessary. Fig. 5 shows plots of the Y11 and Y22 parameters of an amplifier optimized for this case. The inductive components of 18.73 nH and 17.45 nH of the input and output admittances have been optimized in such manner that both static transducer capacitances of 1.4 pF at each IDT are completely tuned out at the frequency of 1 GHz.

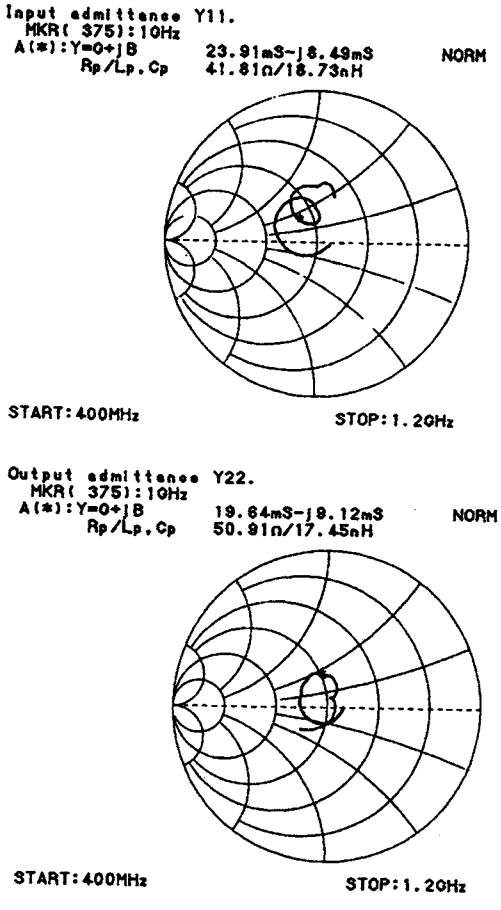


Fig. 5. Input and output admittances of an amplifier for inductive loading.

2. The amplifier gain and the phase in the oscillator loop can be controlled in wide limits by altering the values of the collector/basis feedback resistors R_{kb1} and R_{kb2} without seriously affecting the matching conditions, established with the emitter feedback components. Thus the gain compression can be adjusted according to condition (12) while the oscillator eigenfrequency can be placed within the 1 dB bandwidth of the resonance.

3. The circuit is broad band, has a very low phase slope and, because of that, is stable over a wide temperature range (see below). Fig. 6 is the frequency and phase response of the amplifier over a frequency span of 2 GHz. Here transistors with a transition frequency of 7.5 GHz were used.

4. The d.c. bias currents of the transistors can be adjusted so that the first transistor stage operates at a minimum noise figure while the second one, at the desired output power level. Thus optimum white phase noise performance can be achieved.

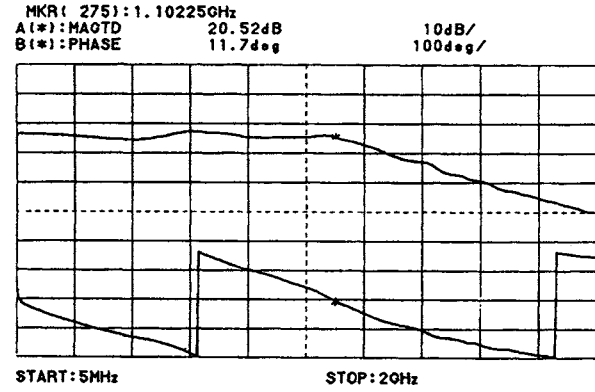


Fig. 6. Frequency response (upper curve) and phase response (lower curve) of the RC-amplfier.

5. The circuit is readily realized in SMD-, thick-film or thin-film technology and occupies an area of less than 2 cm^2.

The next section will present some experimental data obtained from microwave oscillators stabilized with STW two-port resonators and using the RC-amplifier circuit.

5. EXPERIMENTAL DATA

4.1. Phase noise and short-term stability

Computer optimized RC-amplifier circuits were used to build low-cost STW oscillator circuits operating in the frequency range of 0.75-2 GHz. Fig. 7 is the frequency and phase response of a 776 MHz two-port STW resonator using a SAW design (Ref. 5). This device was built on 36 deg rotated Y-cut quartz, occupies a substrate area of 3 mm^2 and has a typical unloaded Q of about 5000. The oscillator circuit, using this device, was designed for operation at a loaded Q value of 3000 and a gain compression of 4 dB. The oscillator output power was 13 dBm. Fig. 8 shows the phase noise performance of the oscillator. The single sideband phase noise at Fourier frequency offsets of 1 Hz and 1 KHz away from the carrier frequency is suppressed by 20 dBc/Hz and 104 dBc/Hz respectively. The curve is dominated by a slope of 30

dB/decade which is due to 1/f noise in the STW-device and the transistors.

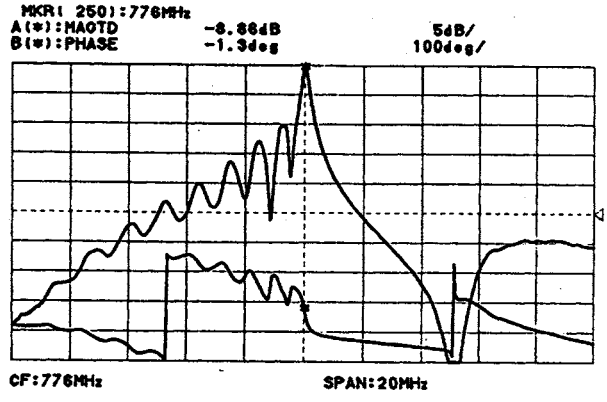


Fig. 7. Frequency and phase responses (upper and lower curve respectively) of a 776 MHz STW resonator using a SAW design.

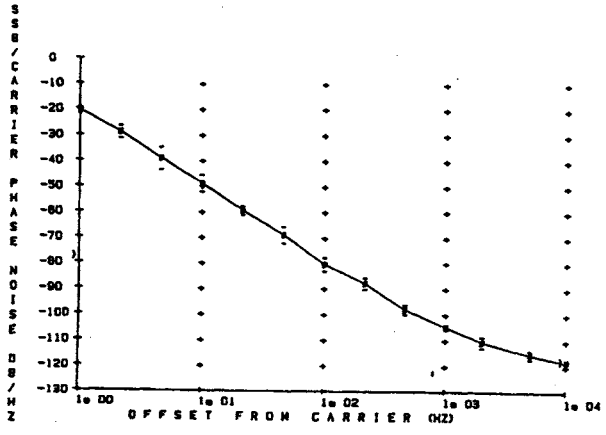


Fig. 8. Phase noise performance of a 776 MHz STW low-cost oscillator.

Fig. 9. is the Allan-Variance of the same oscillator for measurement times from 100 µs to 1s. The short-term stability is constant from 1 ms to 1 s and does not exceed the $2.10^{-10}/s$ value.

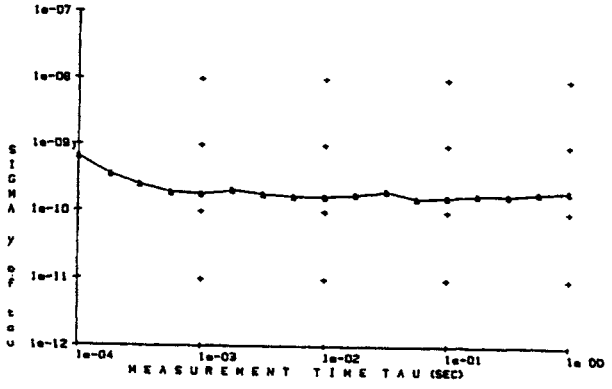


Fig. 9. Allan-Variance of the 776 MHz oscillator.

As shown in Ref. 3, metalization in SH-wave based resonant devices offers an additional degree of freedom in the control over the resonant Q compared to their SAW counterparts. For example, devices of the same design can have quite different loaded Q-values but almost equal insertion loss, dependent on what metalization was used. This feature offers great flexibility in the design of a VCO in which a trade-off between tuning range and phase noise performance has to be achieved. Fig. 10 shows the close-to-carrier phase noise performance of a 1 GHz VCO using a low-Q surface skimming bulk wave (SSBW) resonator with a 3 dB band-

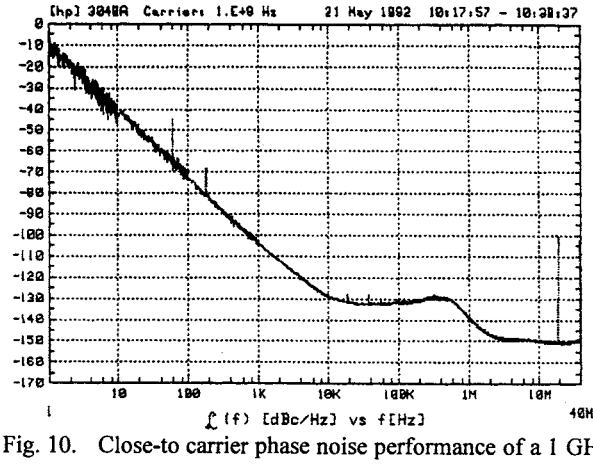


Fig. 10. Close-to carrier phase noise performance of a 1 GHz SSBW VCO with a tuning range of 300 KHz.

width of about 500 KHz. The oscillator uses the RC-amplifier described above, only here a varactor tuned variable phase shifter, inserted between the first and second transistor stages, was used to tune the VCO over a frequency range of 300 KHz. As shown in Fig. 10 this VCO still has a phase noise suppression of 10dBc/Hz at an offset of 1 Hz away from carrier. The measurement was performed with a high spectrum purity frequency synthesizer as a reference source.

The operating frequency range of SH-wave based resonant devices can readily be extended to frequencies above 2 GHz using low-cost photolitographic techniques. Fig. 11 shows the frequency and phase responses of a 1.957 GHz extended cavity single mode STW two-port resonator (Ref. 3).

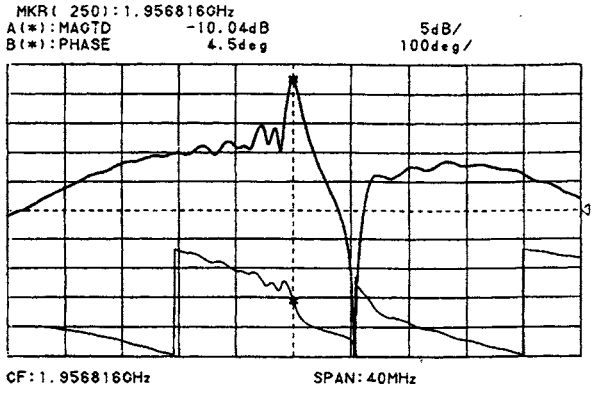


Fig. 11. Frequency response (upper curve) and phase response (lower curve) of a 1.957 GHz extended cavity single mode STW resonator.

The device has a loaded Q of 2600 and an unloaded Q of 3800 which is about 71% of the material limit for SAW while the insertion loss is only 10 dB. This value can easily be compensated for by a two-stage RC-amplifier as the one discussed above. Fig. 12 is the narrow band output spectrum of such a low-cost 2 GHz STW oscillator using this circuit.

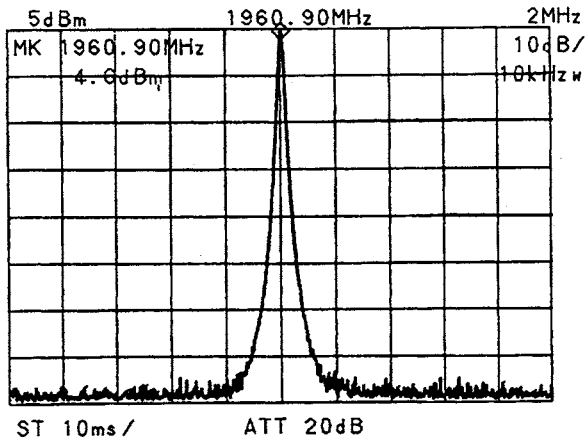


Fig. 12. Narrow band output spectrum of a 2 Hz low-cost GHz oscillator.

The transition frequency of the bipolar transistors, utilized here, is 12 GHz. The oscillator output power is 5 dBm.

4.2 Temperature stability

The RC-amplifier circuit was found to perform very well in oscillators operating over a wide temperature range. The plots in Fig. 13 are the experimental temperature frequency characteristics of 766 MHz STW oscillators in which the acoustic devices were fabricated on different quartz

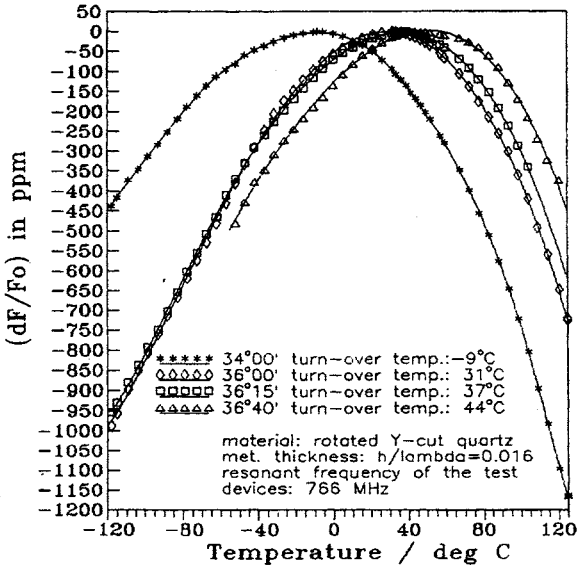


Fig. 13 Temperature instability of STW oscillators for four different quartz cuts.

cuts. All oscillators were found to perform without failure in the temperature range (-120...+120) deg C. The characteristics are close to parabolic and similar to those obtained from SAW devices. The only difference is that the curves are slightly asymmetric and their turnover temperature is somewhat more sensitive to cut orientation tolerances and metalization.

6. SUMMARY AND CONCLUSIONS

This paper has described the analysis, design and performance of low-cost STW resonator oscillators using a simple two-stage RC-amplifier circuit. Despite the circuit simplicity the oscillators were found to have a respectable phase noise performance, high short-term stability and a temperature stability comparable to that of SAW oscillators. Also the oscillators were found to perform well over a very wide temperature range of (-120...+120) deg C. The oscillator circuit allows a quick optimization on a personal computer and can be used for oscillators operating at frequencies as high as 2 GHz. It is the author's belief that the circuit would be competitive for a variety of low-cost applications.

Acknowledgments

The author wishes to gratefully acknowledge Drs. J. Heighway and V. Plessky from Advanced SAW Products SA in Bevaix, Switzerland for making the presentation of this paper possible by organizing and supporting the author's trip to Switzerland. He would like to thank also Drs. I. Yakovkin, S. Kozlov and A. Aseev from the Institute of Semiconductor Physics in Novossibirsk, Russia for the opportunity to have the STW devices fabricated in those laboratories. Special thanks are expressed to Drs. T. Parker and G. Montress from Raytheon Research Division in Lexington, Massachusets for their assistance in measuring the phase noise performance of the 1 GHz STW VCO.

This work was supported by the National Research Fund of Bulgaria.

References

1. T. E. Parker and G. K. Montress, "Precision SAW Oscillators", IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control", Vol. (UFFC-35), No. 3, pp. 342-364.

2. Dvornikov A, Ogurtsov V. and Utkin G., "Stabilnie Generatori s Filtrami na Poverhnostnih Acusticheskih Volnah", Moscow, Radio i Swyaz, 1983.

3. I. D. Avramov, "Microwave Oscillators Stabilized with Surface Transverse Wave Resonant Devices", Proc. 46-th IEEE Frequency Control Symposium, Hershey, PA, USA, 27-29May, 1992, pp. 391-408.

4. H. Lingenfelder, "Berechnung der Y-Kennwerte Linearer Verstaerkerschaltungen.", Radio, Fernsehen, Elektronik, Vol. 29, No. 10, 1980, Germany, pp. 655-700.

5. I. D. Avramov, "High-Q Metal Strip STW Resonators Using a SAW Design" IEEE Control, IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control, Vol. (UFFC-37), No. 6, November, 1989, pp. 530-534.

A NEW TECHNIQUE FOR SAWR FREQUENCY TRIMMING

G. Avitabile, L. Roselli, G. Manes

Dipartimento di Ingegneria Elettronica, Florence, Italy

C. Caliendo, E. Verona

Inst. of Acoustics "O. Corbino" - C.N.R., Rome, Italy

Abstract

The paper deals with a new, low-cost technique for an easy trimming of SAW resonator center frequency. A standard synchronous resonator mask is first designed, then, aligning the mask on the substrate by a controlled axis rotation, a simple but accurate control of the resonant frequency is attained. Negligible insertion loss increase and turn over temperature shift are observed.

SAW devices, SAW Resonators, frequency control.

1. Introduction

Surface-acoustic-wave resonators (SAWR) represent a key component for the implementation of stable frequency sources and narrow-band filters in the UHF-L frequency range.

Exact resonant frequency is a critical parameter to be achieved, as it is affected by uncontrolled factors.

Post fabrication trimming is in general requested for tolerance compensation, resulting in a high cost and time-consuming process.

A new trimming technique is introduced and experimented, allowing frequency tuning to be achieved by controlling one of the technological steps of the process with a minimum increase in the cost and process complexity.

Experimental results showing a 4500 ppm frequency trimming capability are demonstrated, with marginal performance degradation.

2. Basic principle

Exact determination of SAWR resonant frequency is affected by a number of factors depending on the approximations in defining circuit model parameters as well as on tolerances in the technological process such as metallization thickness and electrode over-etching.

As a result, a certain degree of indetermination, owing to systematic errors, should be accepted in defining the exact value of the resonant frequency. Usually, post-fabrication trimming techniques are required to implement fine frequency adjustments. Techniques such as chemical etching provide good results at the expenses of an increase in process costs.

An alternative approach to the fine frequency tuning is described, based on the possibility of precisely controlling the exact relative angle between photo mask and crystallographic axis during the fabrication process. This allows to control the propagation velocity, and consequently the resonant frequency while avoiding any specific step in the fabrication process.

As well known, the dependence of SAW velocity on the propagation direction can be expressed by the following formula introduced by Bell (Ref. 9)

$$\frac{\Delta V}{V} = \left[343\delta\phi + 8\delta\phi^2 + 99\theta^2 + 58\psi^2\right] * 10^{-6} \quad (1)$$

where the numerical coefficients are empirically defined; $\delta\phi$ and $\delta\theta$, represent the relative angles between the propagation

direction and X-axis in the propagation plane; $\delta\psi$ represents the rotation angle with respect to the reference cut (fig. 1).

For SAW resonator applications $\delta\phi$ is in general chosen to fulfill specific turn-over temperature requirements, while ψ and θ are set to zero in order to minimize the effects of the related parabolic terms in (1), which may result from misalignments and tolerances. Once the cut angles have been chosen and substrate orientation defined, the resulting propagation velocity could be evaluated by (1).

Tolerances on $\delta\phi$ e $\delta\theta$ are determined by the cutting process and, in fact, affect the resonant frequency, while ψ is related to the mask alignment precision which could be controlled at the fabrication stage.

The possibility of precisely controlling the mask orientation provides a tool for adjusting the resonant frequency. A rotation offset can be intentionally introduced, between X-axis and propagation direction for the purpose of properly altering the propagation velocity, resulting in corresponding variation of the SAWR resonant frequency.

Based on this principle the tuning process requires the fabrication of a first set of resonators with nominal zero rotation offset and subsequent SAWR resonant frequency measurement.

As a second step, a proper rotation offset is introduced, in order to modify the propagation velocity and tune the resonant frequency, accordingly.

As the tuning process only results in increasing the resonant frequency, the original SAWR design should necessarily be set to a lower value than requested.

3. Experiments

In order to investigate the effectiveness of this technique several series of SAW synchronous (Ref. 4) resonators on ST quartz with different rotation offset have been realized and tested.

The design of these devices is characterized by the following parameters:

number of reflector electrodes = 380

number of transducer electrodes = 141

electrode width = 5 μm

maximum active electrode length = 50 wavelength

nominal resonant frequency = 157.5 MHz

Apodized transducers are used to minimize the excitation of transverse modes. An array of dummy fingers is placed inside the cavity and a slanted shape is applied to the bonding pads to reduce feed-through between transducers.

4. Results

In order to demonstrate the viability of this new technique for SAWR trimming some experiments have been made using the previously described photolithografic mask rotated over the substrate with angles varying from 0 to 9 degrees (0,2.5,5,9).

In figure 4 is shown the tuning range obtained by varying the rotation angle from 0 to 9 degrees. According to the previous formula, a parabolic behavior of the frequency variation versus rotation angle is obtained.

SAWR thermal behavior, as previously mentioned, is a key performance of the device. In figure 5 is reported the frequency variation versus temperature for all the considered rotation angles. It is worth to note that no substantial change from 0 degree rotation is detected, except for a little turn-over temperature (TOT) shift.

Angle	f_0 MHz	$\Delta f/f_0$ (ppm)
0°	157.65	0
2.5°	157.7	270
5°	157.79	860
9°	158.37	4500

Table 1: data of Fig 4

5. Conclusion

A trimming technique has been introduced and experimented for precision tuning of SAW resonators.

The new technique does not require any additional technological step, as requested by others previously reported procedures, being capable to provide the needed frequency trimming by simply rotating the photo mask axis with respect to the crystallographic axis by a proper angle.

Experimental results, showing the effectiveness of the approach have been given for a 157 MHz resonator prototype. TOT shift has been also observed, showing negligible variations within a rotation angle less than ten degrees.

6. References

[1] C. S. Hartman and R. C. Rosenfeld: "Acoustic surface wave resonator devices",United States Patent n° 471,616 May 27 1975.

[2] W. J. Tanski: "Surface acoustic wave resonators on Quartz", IEEE Transactions on Sonics and Ultrasonics SU-29 Vol. 2 March 1979.

[3] P.S. Cross, W. H. Haydl and R. S. Smith: "Design and applications of two-port SAW Resonators on YZ-Lithium Niobate", Proceedings of the IEEE, Vol. 64, n° 5, May 1976.

[4] P. S. Cross, W. R. Shreve and Tun Sein Tan: "Synchronous IDT SAW Resonators with Q above 10000", Ultrasonic Symposium 1979.

[5] W. H. Haydl P. S. Cross: "Fine tuning of surface-acoustic-wave resonator filters with metallization thickness", Electronics letters May 1975.

[6] G. Avitabile, L. Roselli, C. Atzeni, G. Manes: "Surface acoustic wave resonators", ETT Vol. 2 n° 5 Sept.-Oct. 1991.

[7] W. H. Haydl, P. Hiesinger, R. S. Smith, B. Dischler and K. Heber: "Design of Quartz and lithium niobate saw resonator using aluminum metallization", Proc. of the 30th Freq. Cont. Symp., 1976 pp. 346 - 357.

[8] P. S. Cross, W. R. Shreve: "Frequency Trimming of SAW Resonators", IEEE Transactions on Sonics and Ultrasonics SU-29 pp. 231-234 July 1982.

[9] D. T. Bell, Jr.: "Phase error in long surface wave devices", in J. de Klerk, Ed., Proc. 1972 IEEE Ultrasonics Symposium, IEEE, New York,1972,pp.420-423.

[10] D. P. Morgan: "Surface-wave devices for signal processing", Elsewier Publisher, 1985.

[11] H. Matthews: "Surface wave filters design, construction, and use", John Wiley & Sons, Inc. Toronto 1977.

7. Figure captions

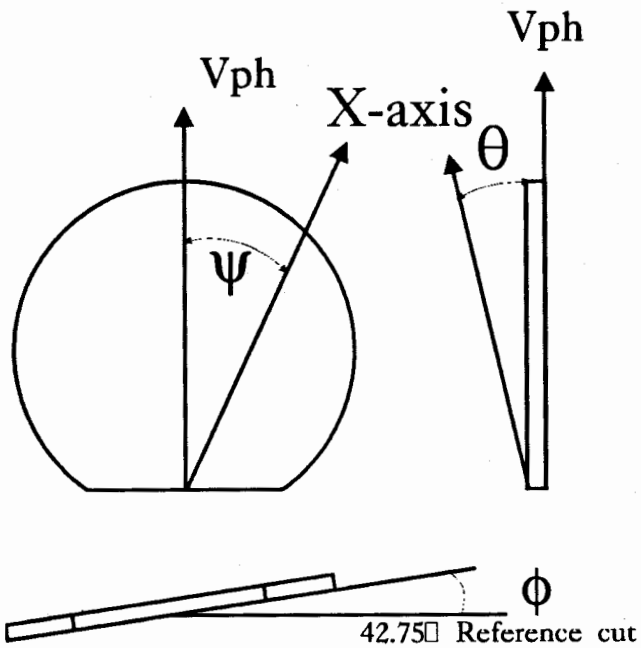


Fig. 1 Characteristic cut angles for ST quartz.

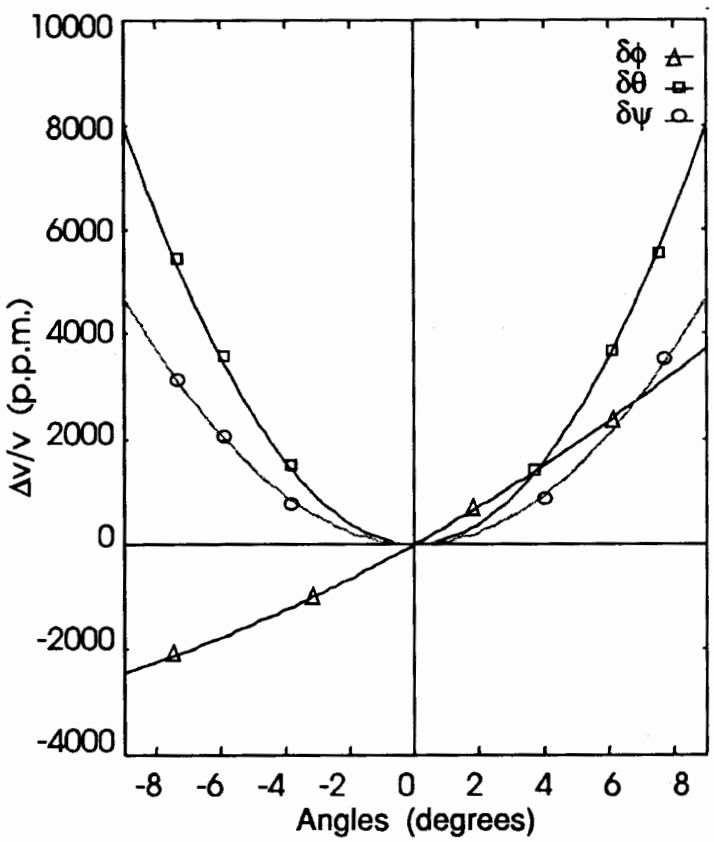


Fig. 2 Theoretical velocity variation versus rotation angles

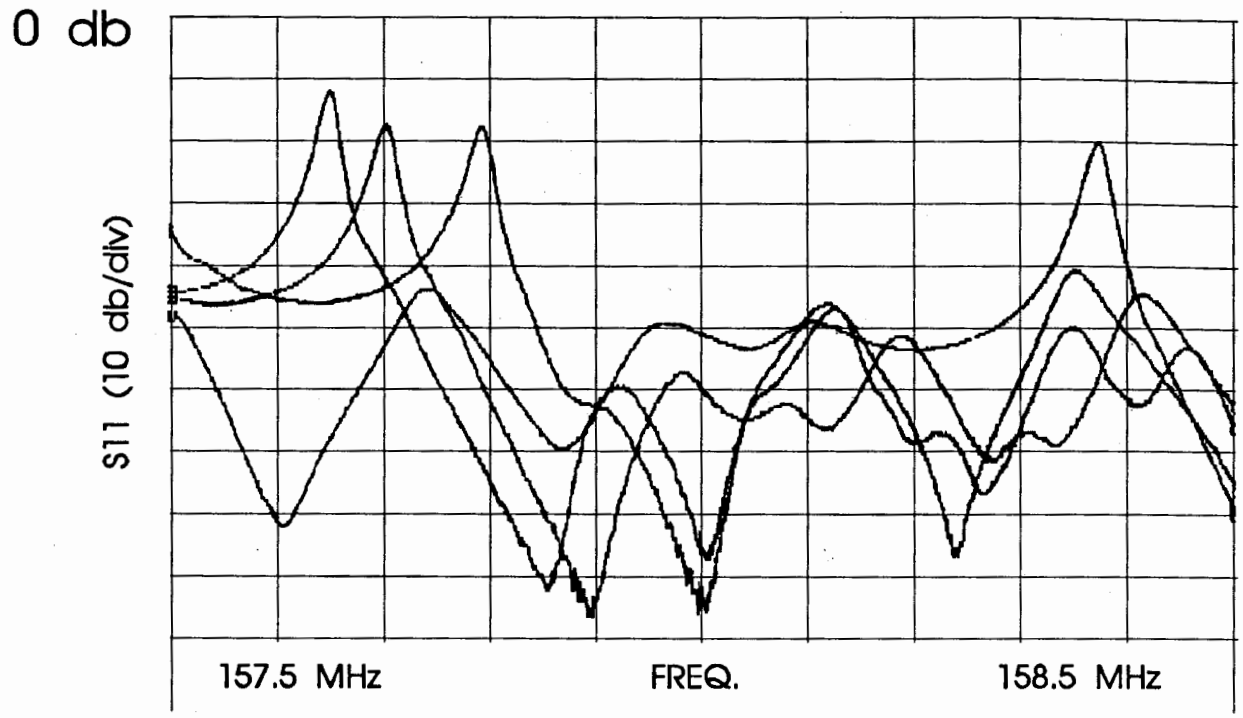


Fig. 3 Frequency response of SAWRs with different rotation angles: 0,2.5,5,9

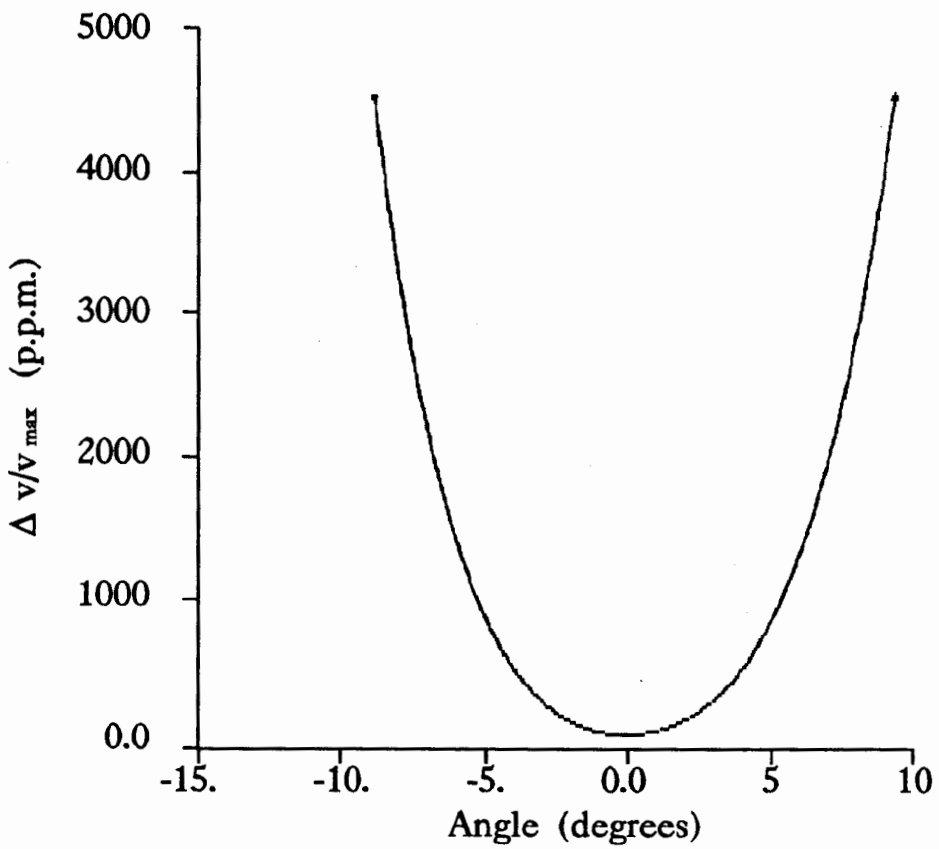


Fig. 4 Resonant frequency variation versus rotation angle

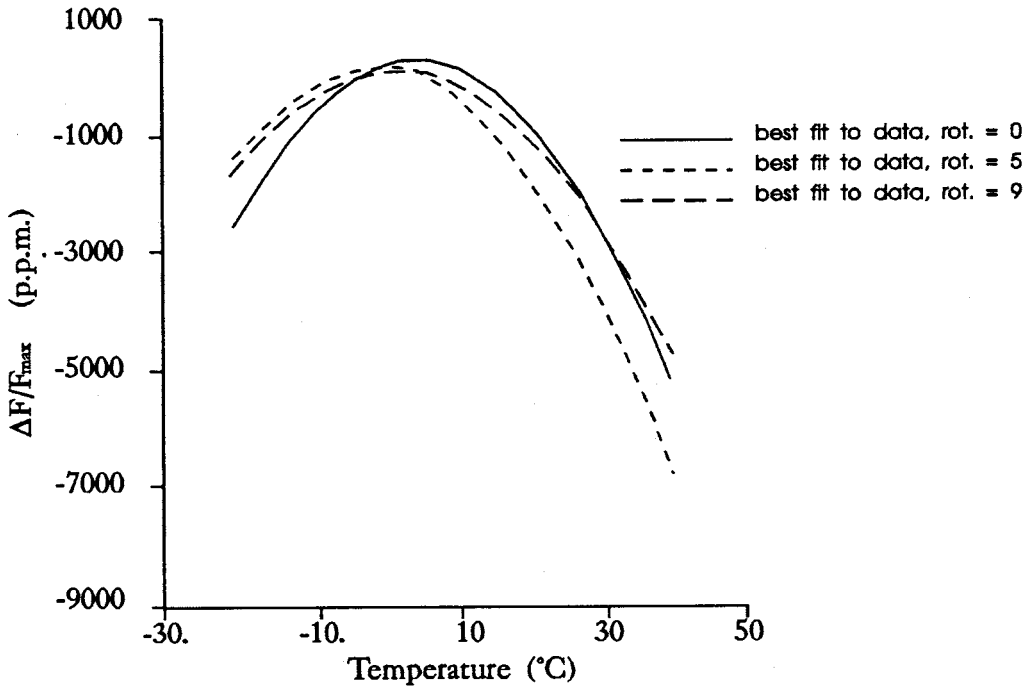


Fig. 5 Temperature behavior for specified rotation angle

CESIUM 1

Chairman: L. Cutler

OPTICAL PUMPING IN PTB's FREQUENCY STANDARD CSX

S. Weyers, R. Schröder, T. Heindorff and A. Bauch

Physikalisch-Technische Bundesanstalt, Braunschweig, Germany

ABSTRACT

At PTB a conventional cesium atomic beam apparatus was provided with an optical system detecting the clock transition. By this it was possible directly to compare the frequency results obtained with an optical detector and with an hot-wire detector. These data agree within $1 \cdot 10^{-13}$.

Keywords: atomic clock, optical pumping

1. INTRODUCTION

Several years ago optical pumping was suggested as a method to reduce both, instability and uncertainty of cesium atomic clocks (Ref. 1). The question was, whether the clock transition frequency realized in a conventional clock with magnetic beam selection might be systematically shifted compared to the value realized in a clock with optical pumping. Our aim is to contribute to the solution of this problem.
For this purpose we reconstructed the PTB's conventional atomic beam apparatus CSX. The intention was to preserve the possibility of an easy interchange of the optical and the conventional detection mode. Thus frequency comparisons between the CSX's classical mode of operation and the optically pumped version can be carried out.
Under the prevailing experimental situation we have to put up with some restrictions concerning the polarization of the laser light, the fraction of irradiated atoms, the collection efficiency and difficulties in the optical alignment. Nevertheless promising results were obtained.

2. OPTICAL PUMPING OF ^{133}Cs

The relevant energy levels of the cesium atom are shown in Fig. 1. To excite optically the atoms a laser diode with a wavelength of λ = 852.124 nm is employed.
Due to the selection rules $\Delta F = 0, \pm 1$ there are two transitions, $F_g=3 \Rightarrow F_e=2$ and $F_g=4 \Rightarrow F_e=5$, called "cycling transitions". The use of one of these transitions provides a particularly high amount of fluorescence light as optical pumping into the other respective ground level is avoided. To obtain a high signal-to-noise ratio these transitions seem most promising for detection, if the laser noise is not important.
For comparison, also pumping transitions have been used for the detection in our measurements ($F_g=3 \Rightarrow F_e=3$ and $F_g=4 \Rightarrow F_e=4$).

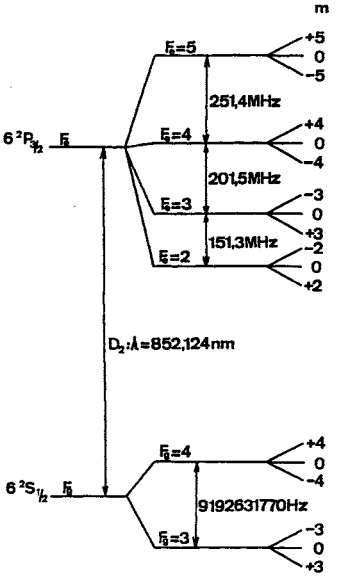


Fig. 1. Energy level diagramm of ^{133}Cs of relevance for to discuss the excitation of the D_2 line.

3. THE EXPERIMENTAL SET-UP

3.1 The apparatus CSX

In Fig. 2 a vertical section of the CSX is shown. The apparatus is cylindrically symmetric. The conventional atomic beam system consists of the oven and the Langmuir-Taylor detector. The two magnets P and A serve for state selection. Together with a diaphragm in front of the second irradiation section (not seen in the figure) their

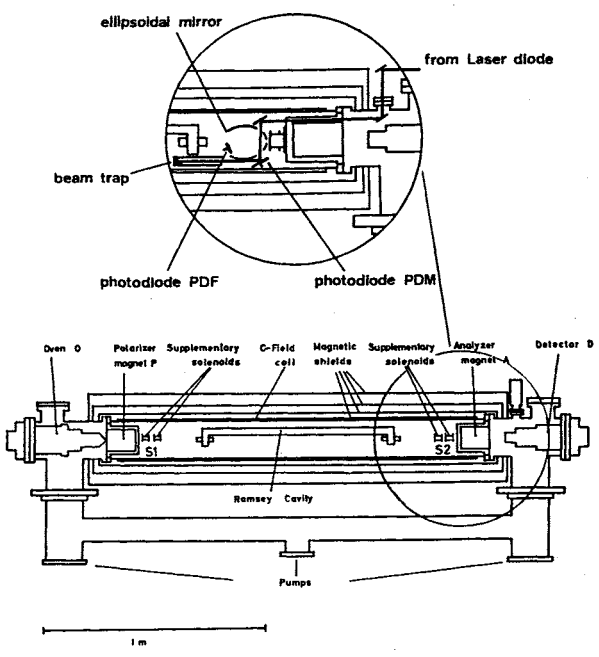


Fig. 2. Vertical section of the PTB's experimental frequency standard CSX. The enlarged section shows the optical detection system.

apertures limit the atomic beam diameter to 3 mm.
In the conventional mode the clock transition is detected by the flop-out method. The signal is used to lock the VCXO in the microwave frequency synthesis chain to the atomic resonance frequency. A cylindrical coil, which extends over the magnetic systems P and A, generates the longitudinal C-field. The supplementary solenoids serve for providing a smooth transition of the magnetic field from the C-field region to the magnets. Four coaxial Mumetal cylinders shield the outer magnetic field. The mean velocity of the atoms contributing to the clock transition is 412 m/s resulting in a linewidth of 259 Hz.

3.2 The optical detection system

The pumping light is fed into the vacuum envelope from a place outside the four-fold magnetic shielding and is guided to the interaction zone by a tube with mirrors at each end and diaphragms (see the enlarged section of Fig. 2).
The amount of the fluorescence signal shows that about 90% of the atoms are irradiated by the laser beam, 3 mm in diameter. The interaction zone with the atomic beam lies in the focal point of an ellipsoidal mirror made of nickel coated with a highly reflective rhodium film. Due to the ferromagnetic properties of nickel the direction of the magnetic field in the interaction zone may differ from the C-field direction. The following results are obtained with the laser light polarization transversal to the C-field direction.

The ellipsoidal mirror collects the fluorescence light and focuses it to the photodiode PDF (active area 5,6 mm^2) in the second focal point. The photodiode PDM monitors the incident light intensity and deflects the laser light into a beam trap. In the $F_g=3 \Rightarrow F_e=2$ excitation scheme the stray light detected by PDF is about 20% of the useful fluorescence signal.

3.3 The laser diode system

In an exchange with the LPTF, Paris, we got a laser diode system consisting of a Hitachi HLP1400 laser diode with external Fabry-Perot resonator (Refs. 2,3). The linewidth of the output beam is reduced to about 100 kHz.

4. FLUORESCENCE SIGNAL OF THE CSX

Behind the polarizing magnet P the atomic beam consists only of atoms in the states $F_g=(4,m)$ with m = -3,-2,...,+4. If part of the atoms have made the clock transition (4,0)-(3,0), the atomic beam in the optical interaction zone consists of about 7/8 atoms in the state (4,m) with m = -3,-2, -1,+1,+2,+3,+4 and about 1/8 in the state (3,0). In this case the fluorescence signal shown in Fig. 3 is observed. The largest signal is obtained for the two cycling transitions. The $F_g=3 \Rightarrow F_e=2$ transition is more suitable because only atoms in the state $F_g=(3,0)$ contribute to the fluorescence signal. So the shot noise $N_S = (2eI)^{1/2}$ (e: charge of the electron, I: photo current) is lower than for the other cycling transition by a factor of about 4.

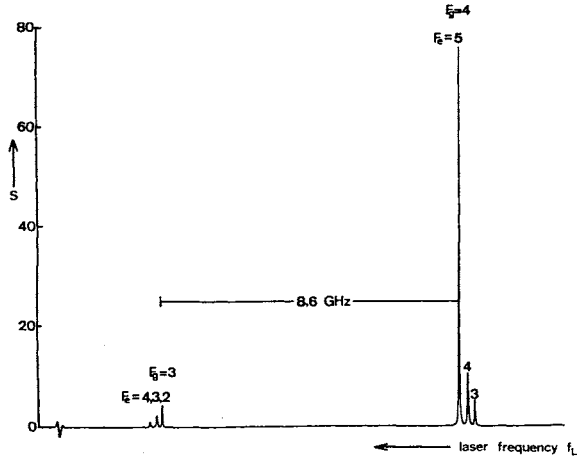


Fig. 3. Fluorescence light S in arbitrary units in dependence upon the laser frequency f_L. In the atomic beam the energy levels (3,0) and (4,m), m = -3,-2,-1,+1,+2,+3,+4, are populated.

5. FREQUENCY MEASUREMENTS IN THE OPTICAL MODE

The measurement of the clock frequency is affected by a number of frequency shifts. The most of the shifts will remain unchanged in either mode of detection and are well known from the conventional mode of operation. The use of the optical detection may result in a light shift but also in a change of the shifts related with the atomic velocity. In the following it is tried to separate the individual effects.

5.1 Light shift

When we used the optical detector for the first time, the beam trap shown in Fig. 2 had been omitted. The result was that the light reflected by the photodiode PDM propagated inside the CSX vessel. For an incident light power of 0,5 mW we measured a relative frequency shift of $8 \cdot 10^{-11}$, when the $F_g=3 \Rightarrow F_e=2$ transition was used for detection. This gross effect had to be suppressed by the installation of the beam trap.
Further possible causes of light shift are the stray light of the apertures of the ellipsoidal mirror and the fluorescence light emitted into the direction of the atomic beam in the cavity.
At present, i.e. with a beam trap mounted, a residual light shift is surely smaller than $1 \cdot 10^{-13}$, as the frequency results were found independent of the incident laser power at a level of less than $5 \cdot 10^{-14}$.

5.2 Frequency related with the velocity of the atoms

Both, the second-order Doppler effect and the cavity phase difference shift, are velocity-dependent. Their contributions to the frequency shifts are:

$$\frac{\Delta f_D}{f_0} \cong -\frac{\langle v \rangle^2}{2c^2} \quad (1) \qquad \text{and} \qquad \frac{\Delta f_{Ph}}{f_0} = \frac{\Delta\phi \ \langle v \rangle}{2\pi L} \frac{1}{f_0}, \quad (2)$$

respectively (Ref. 4). f_0 is the clock transition frequency, $\langle v \rangle$ the mean atomic velocity, c the speed of light in vacuum and L the separation of the two microwave interaction zones. The cavity phase difference $\Delta\phi$ of the CSX cavity cannot be measured by the usual method of a beam reversal. Instead, it is assumed that the mean frequency difference between the CSX and the PTB's CS2 is attributable to the $\Delta\phi$ in the CSX. Hence $\Delta\phi = 220\ \mu$rad is obtained.
Fig. 4 shows the central fringes of the Ramsey resonances of the clock transition line obtained with the optical detector ($F_g=3 \Rightarrow F_e=2$, cycling transition). Curve a) was obtained with the microwave power adjusted to give maximum resonance signal at the conventional Langmuir-Taylor detector. The pattern is distorted compared with curve b), where the microwave power was reduced by 1,5 dB.

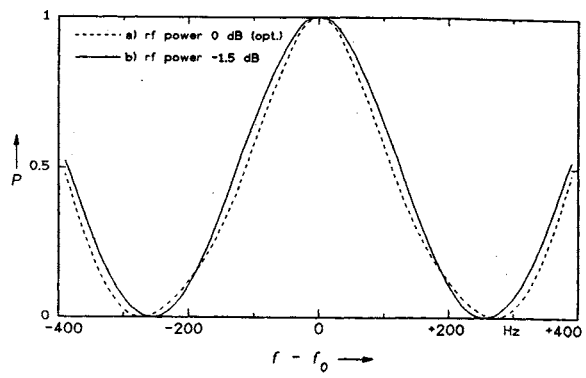


Fig. 4. Central Ramsey fringe of the (4,0)-(3,0) clock transition detected with the transition $F_g=3 \Rightarrow F_e=2$ at different rf powers (see text). P is the transition probability. f_0 = 9 192 631 770 Hz.

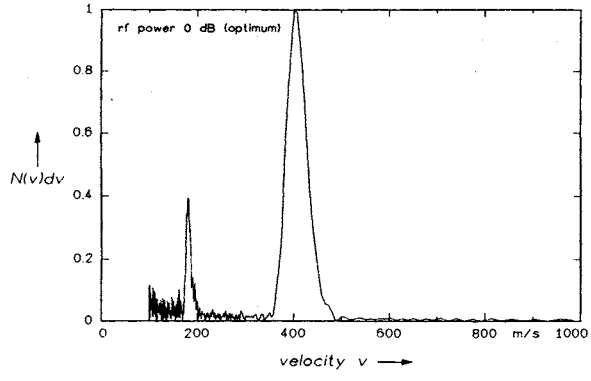


Fig. 5. Normalized velocity distribution N(v)dv calculated from Fig. 4, curve a) (full Ramsey pattern).

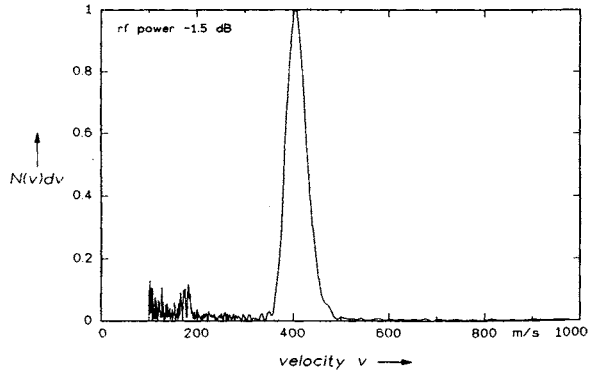


Fig. 6. Normalized velocity distribution N(v)dv calculated from Fig. 4, curve b) (full Ramsey pattern).

The Ramsey patterns recorded between $f-f_0 = \pm 10$ kHz were used to obtain the velocity distribution of the atoms by a Fourier transformation. The results are shown in Figs. 5 and 6. In case of Fig. 5 there is a distinct portion of atoms with a mean velocity of about 170 m/s.
The deflection properties of the polarizing magnet P entail the contribution of these slow atoms. They occur with high probability in the thermal atomic beam ($\sqrt{2kT/m}$ = 240 m/s) and make a 3/4-period of oscillation around the magnet's symmetry axis, while the fast atoms make a 1/4-period only.
In the optical mode the signal of these atoms is emphasized if a cycling transition for detection is used, because then the signal is approximately inversely proportional to the velocity of the atoms. In case of a pumping transition operated in saturation this effect plays a minor role.
In the conventional mode of operation of the CSX the analyzer magnet defocuses these atoms, which thus cannot contribute to the clock transition signal.
To have a clear view of the distortion see Fig. 7. In this simulation the Ramsey resonance R is a superposition of the two cosine functions C1 and C2 with different periods corresponding to the velocities 412 m/s and 170 m/s. The shift between the two central Ramsey fringes of C1 and C2 is exaggerated to show clearly the distortion of R. Fig. 7 shows that the value of the measured centre frequency will depend on the modulation width, if the probing of the atomic line centre is done by a square wave frequency modulation technique.
Usually, the correction (1) is determined using

$$\frac{\Delta f_D}{f_0} = - \frac{(2\ L^* \cdot FWHM)^2}{2c^2}, \qquad (3)$$

where $L^* = L + 1{,}20 \cdot l$ (Ref. 5) with the length of the irradiation zone l. The FWHM is experimentally determined as the frequency separation between the resonance maximum and the first minimum (see Fig. 4)

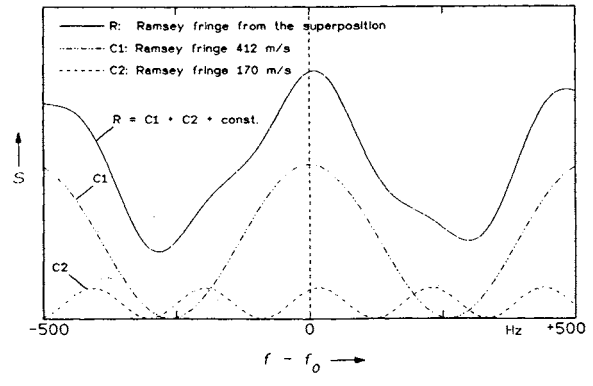


Fig. 7. Simulation of the adding of Ramsey fringes to a distorted pattern (see text for explanation).

using again square wave frequency modulation.
In case of the distorted Ramsey pattern this method to determine Δf_D is no longer applicable. Instead the corrections (1) and also (2) would have to be calculated as an average over the actual velocity distributions (Figs. 5 and 6) taking into account the modulation scheme to probe the line centre and the phase difference $\Delta\phi$ = 220 μrad. These calculations are on the way.

5.3 Agreement of frequency measurements

The contribution of the slow atoms to the resonance pattern can be minimized by a proper reduction (-1,5 dB) of the rf power so that the distortion of the Ramsey fringes disappears almost completely. In this case the corrections (1) and (2) can be used to determine the frequency for the optical detector in a good approximation, too. The results are compared with the long-term average CSX frequency obtained in the conventional detection mode.
We obtained the quantity

$$F_\varphi(CSX) = \frac{f_0(CS2) - f_\varphi(CSX)}{f_0(CS2)}, \qquad (4)$$

where $f_0(CS2)$ is the clock transition frequency realized with the primary clock CS2, and $f_\varphi(CSX)$ is the clock transition frequency realized with the CSX but not corrected for the phase difference. $F_\varphi(CSX)$ is obtained in the optical and the conventional mode of operation, which should ideally be equal.

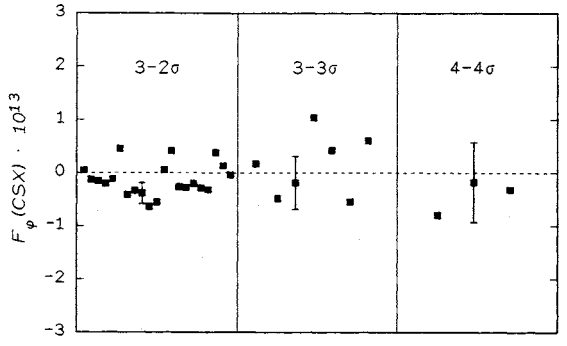


Fig. 8. Relative frequency deviation $F_\varphi(CSX)$ obtained with the optical detection system using three different excitation schemes. $F_\varphi(CSX) = 0$ corresponds to the value obtained with the conventional detector, which is known with a statistical uncertainty of $2 \cdot 10^{-14}$. The data points have been collected in a random sequence. The measurements lasted between 8 to 18 hours. Typical error bars are indicated.

The results of the frequency measurements obtained using three different optical excitation schemes are shown in Fig. 8, where $F_\varphi(CSX) = 0$ corresponds to the value obtained with the conventional detector, which is known with a statistical uncertainty of $2 \cdot 10^{-14}$. The small apparent systematic deviation from zero might still be caused by the contribution of the slow atoms. A detailed analysis of the corrections (1) and (2) as outlined in chapter 5.2 is on the way.

6. CONCLUSIONS

It is shown that an atomic clock with an optical detection system may be operated in good agreement with the results obtained by the conventional mode of detection.
For the correct evaluation of the systematic frequency shifts due to the second-order Doppler effect it is necessary to know the velocity distribution of the atoms contributing to the clock transition. In CSX this distribution is determined by the polarizing magnet, which selects only a small portion of atoms from the Maxwellian velocity distribution.
However, there is a little contribution of slow atoms leading to a distortion of the Ramsey fringes, particularly if a cycling transition is used.
In case of an optically pumped atomic clock without velocity selective magnets the broad velocity distribution has to be well-known to carry out the correct evaluation of the second-order Doppler effect.

This work is supported by the Deutsche Forschungsgemeinschaft and the European Community.

7. REFERENCES

[1] A. Kastler, Le Journal de Physique et le Radium **11**, 255 - 265 (1955)

[2] B. Dahmani, L. Hollberg, and R. Drullinger, Optics Letters **12**, No.11, 876 - 878 (1987)

[3] Ph. Laurent, A. Clairon, Ch. Bréant, IEEE J. Quantum Electron. **QE-25**, No.6, 1131 - 1142 (1989)

[4] J. Vanier and C. Audoin, The Quantum Physics of Atomic Frequency Standards, Bristol and Philadelphia, A. Hilger, (1989)

[5] P. Thomann, Proc. of the 22nd Annual PTTI Applied and Planning Meeting, 331 - 335 (1991)

MICROWAVE SPECTRUM ASYMMETRY IN AN OPTICALLY PUMPED CESIUM BEAM CLOCK

V. Giordano, N. Dimarcq, P. Cérez, G. Théobald, C. 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

In this paper, we present an experimental study of the microwave spectrum asymmetry in an optically pumped cesium beam resonator. This machine operates with sigma polarized laser light tuned to the $F = 3 \rightarrow F' = 3$ optical transition at $\lambda = 852$ nm . The asymmetry of the microwave spectrum appears as soon as exists a detuning between the laser frequency and the atomic frequency of the $F = 3 \rightarrow F' = 3$ transition. The critical parameters which influence the asymmetry of the spectrum are determined. Various assumptions concerning the origin of this asymmetry are listed. In standard operating conditions, an estimate of the Rabi pulling frequency shift caused by a small laser detuning is finally made.

Keywords : Optically pumped cesium beam frequency standards ; laser detuning ; microwave spectrum asymmetry ; Rabi pulling frequency shift.

1. INTRODUCTION

Significant causes of frequency shift in cesium beam frequency standards are related to an asymmetrical distribution of the atomic populations among the hyperfine ground state Zeeman sublevels which is due to non perfect atomic beam preparation. Beam tubes operating with magnetic state selectors always exhibit an asymmetrical microwave spectrum that comprises seven transitions $\Delta F = \pm 1$; $\Delta m_F = 0$. Consequently, the central line of this spectrum (familiar clock transition) is distorted leading to Rabi and Ramsey pulling [1].

Rabi pulling occurs because the aisles of the pedestal of the adjacent $(3,1) \leftrightarrow (4,1)$ and $(3,-1) \leftrightarrow (4,-1)$ transitions overlap the line of the $(3,0) \leftrightarrow (4,0)$ transition. The shape of the central line is distorted by an added slanted base line which depends on the microwave power and on the value of the static magnetic induction.

Ramsey pulling is a more complicated effect due to a coupling between the clock transition and unwanted $\Delta m_F = \pm 1$ transitions which are induced when the static and the magnetic microwave inductions have mutually orthogonal components [2]. This always occurs to some degree as a consequence of the curvature of the magnetic lines of force in the Ramsey cavity. It can be shown that this effect adds a term to the transition probability and the central line has a Ramsey shape with an offset depending on the asymmetry of the microwave spectrum.

The scientific community involved in research and developments on optically pumped cesium beam and clocks has emphasized a major advantage offered by optical method of state preparation : the microwave spectrum of the cesium atom can be made highly symmetrical around the central $(3,0) \leftrightarrow (4,0)$ line.

In principle, optical pumping achieved with linearly polarized laser light (delivered for example by commercially available laser diodes) should not introduce an unbalance between the atomic populations of the Zeeman sublevels with opposite m_F values. However, experiments have shown that the microwave spectra delivered by optically pumped cesium beam resonators (OP Cs BR) exhibit a slight asymmetry which depends on the laser frequency detuning with respect to the optical transition frequency choosen to induce optical pumping [3].

In order to study more precisely this effect, we have operated our OP Cs BR named Cs 3 with two independent lasers. Let us recall that this machine regularly works with a single laser diode for optical state preparation and detection [4]. Its frequency is tuned to the $F = 3 \rightarrow F' = 3$ optical transition at $\lambda = 852$ nm (D_2 line). The fine tuning of the laser is accomplished by looking at fluorescence light of the cesium atoms in the pumping region : modulation of the laser frequency and synchronous detection of the fluorescence flux provide an error signal for diode laser locking purposes.

The role of the two independent lasers is the following : the first one (pumping laser) can be frequency locked with a variable and controlled detuning on the $F = 3 \rightarrow F' = 3$ transition. The second laser (detection laser) remains always tuned to the center of the considered $F = 3 \rightarrow F' = 3$ transition. Owing to this versatile arrangement, the influence of the laser frequency detuning, laser intensity and static magnetic field on the different ground state Zeeman populations can be studied more easily.

In this paper, we first describe the experimental set-up. Then we present results showing that asymmetry varies mainly with the laser detuning and the magnitude of the static magnetic field in the pumping region. Moreover, it is possible with a proper adjustement of optical pumping parameters to achieve a perfect symmetry of the microwave spectrum.

Finally, an attempt to explain the experimental observation will be presented as well as a calculation of the Rabi pulling frequency shift due to a given amount of asymmetry in the microwave spectrum delivered by our optically pumped cesium beam clock.

2. EXPERIMENTAL SET-UP

The experimental set-up is shown in Fig. 1. Cs 3 is the OP Cs BR. More details concerning this resonator can be found in a previous publication [4]. The optical frequencies of LD_1 (pumping light) and LD_2 (detection light) are stabilized independently with the help of two auxiliary cesium beams.

Owing to the servo system and the acousto optic modulator, LD_1 operates at a stable frequency which can be continuously varied around the resonance frequency f_{33} of the $F = 3 \rightarrow F' = 3$ optical transition. Indeed we have :

$$f_{LD1} = f_{32} + 2\Delta$$

f_{32} is the frequency of the F = 3 → F' = 2 optical transition given by Cs_A whereas 2 Δ is the frequency shift given by the AOM operating in the double pass configuration. With Δ ≃ 75 MHz, $f_{LD1} \simeq f_{32} + 150$ MHz ≃ f_{33}.
Let us also mention that, thanks to the double pass method, the direction of the deflected beam is not modified by any change of the frequency Δ. LD_2 is directly prestabilized on the F = 3 → F' = 3 optical transition given by Cs_B.

For each component of the microwave spectrum, we first adjust the microwave power at its optimum level and then we measure the corresponding amplitude at resonance. The relative asymmetry defined by :

$$a_{|m|} = \frac{A_{(3,m \leftrightarrow 4,m)} - A_{(3,-m \leftrightarrow 4,-m)}}{A_{(3,0 \leftrightarrow 4,0)}}$$

|m| = 1, 2, 3

is finally determined as a function of the detuning Δ of the pumping laser.

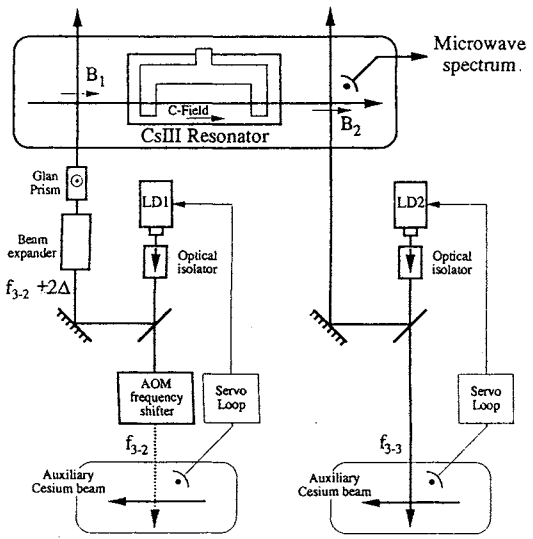


Figure 1: The experimental setup. The auxiliary cesium beam are CsA (left) and CsB (right).

3. EXPERIMENTAL RESULTS

3.1 Microwave spectra

Fig. 2a,b,c shows typical microwave spectra recorded when $f_{LD1} - f_{33}$ = - 20 Mhz ; 0 ; + 20 MHz respectively. In this experiment, the microwave level is set around the value which maximizes the clock transition. The C field is B_C = 63 mG ; the magnetic static inductions B_1 and B_2 in the optical interaction regions are large enough to avoid the detrimental Hanle effect that otherwise affects the pumping and detection efficiencies. We have $B_1 = B_2$ = 1 G.

We can see in Fig. 2a,c that large asymmetries occur when the laser detuning is significant. Moreover, the asymmetry changes its sign with ($f_{LD1} - f_{33}$).

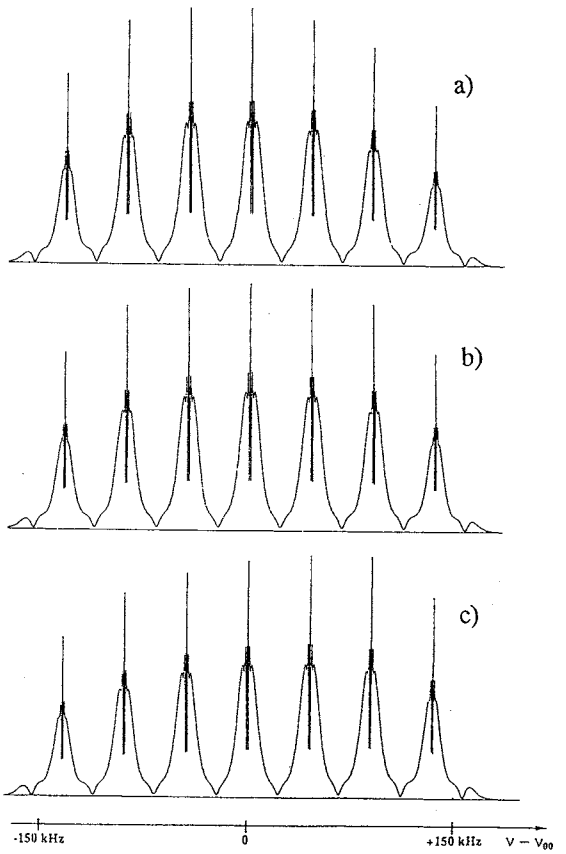


Figure 2: Typical records of the microwave spectra when:
a) $f_{LD1} = f_{3-3}$ -20MHz
b) $f_{LD1} = f_{3-3}$
c) $f_{LD1} = f_{3-3}$ +20MHz

3.2 Shape of the asymmetry curves

Fig. 3 shows the evolution of $a_{1,2,3}$ as a function of 2Δ. In this experiment, we have fixed B_C = 63 mG ; $B_1 = B_2$ = 0.5 G. The pumping intensity P/S is set to its usual value 20 mW x cm^{-2}. To our knowledge, this ensures a complete pumping process at resonance with a mean interaction time $T_i \simeq$ 20 µs. We estimate that less than 1 % of the atomic population remains unpumped at resonance in such conditions. The experimental curves of Fig. 3 are "dispersion shaped" and fractional asymmetries as large as ≃ 20 % are observed for |m| = 3 and $f_{LD1} - f_{33} \simeq$ 24 MHz.

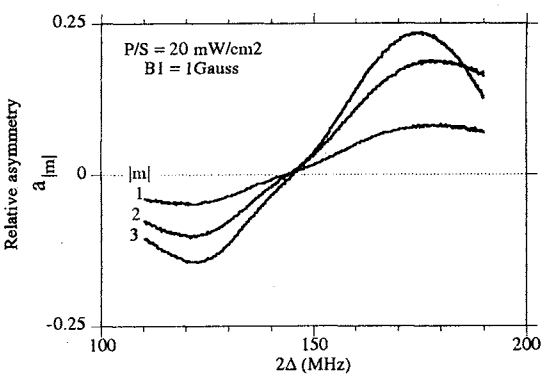


Figure 3: Evolution of the asymmetry a|m| as a function of 2Δ

3.3 Influence of the pumping laser intensity

Fig. 4 and 5 show the evolution of $a_{|1|}$ (2Δ) and $a_{|3|}$ (2Δ) respectively for 3 values of the pumping intensity parameter P/S. At small values of the detuning, we observe that large variations of P/S give rise to slight variations of the asymmetry.

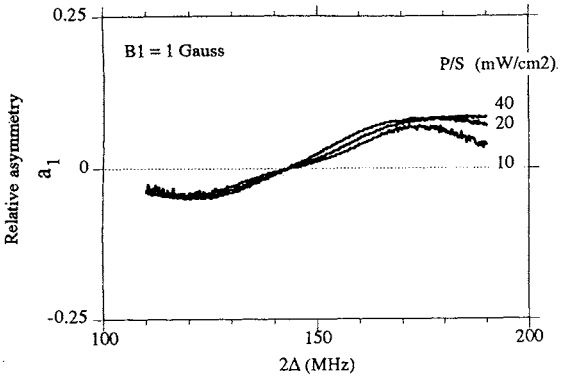


Figure 4: Evolution of the asymmetry a1 versus 2Δ.
The parameter is the pumping intensity P/S.

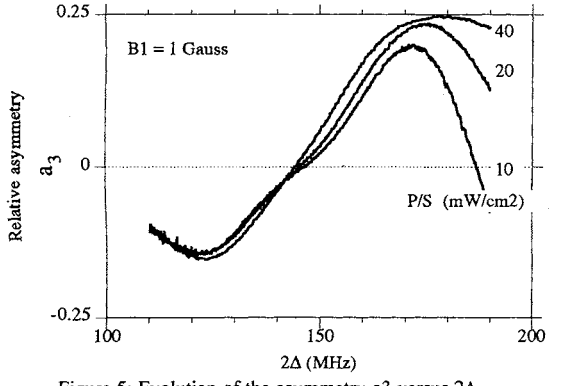


Figure 5: Evolution of the asymmetry a3 versus 2Δ.
The parameter is the pumping intensity P/S

3.4 Influence of the static magnetic field in the pumping region

Fig. 6 shows that $a_{|1|}$ (2Δ) varies strongly with the value of the static magnetic field B_1. It becomes independent of B_1 when $B_1 \geq 0.5$ G.

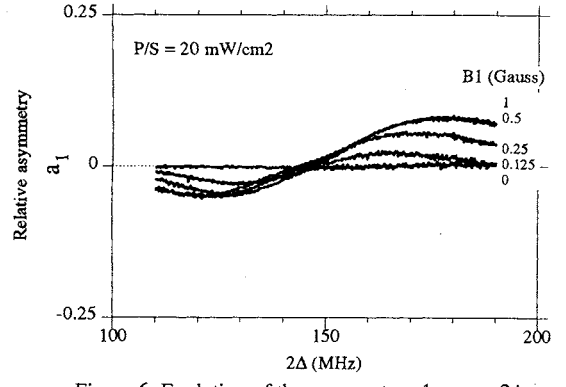


Figure 6: Evolution of the asymmetry a1 versus 2Δ.
The parameter is the amplitude of the static magnetic field B1 in the pumping region.

3.5 Influence of the interaction time T_i

Further experiments have established that T_i is not a critical parameter governing the asymmetry evolution. This can be shown in Fig. 7.

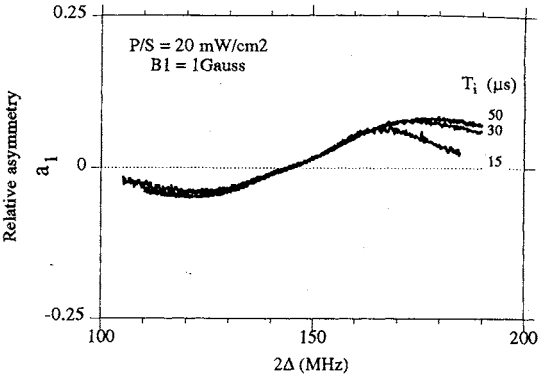


Figure 7: Evolution of the asymmetry a1 versus 2Δ.
The parameter is the mean interaction time Ti.
$T_i = L_i/(1.22\alpha)$ with $\alpha = 216$m/s

4. COMMENTS CONCERNING THE RESULTS

It is clear that in our optically pumped cesium beam machine operating with σ polarized laser light tuned to the $F = 3 \rightarrow F' = 3$ transition, the microwave spectrum may be asymmetrical if the laser is detuned. For a small detuning (a few MHz), the critical parameter is the value of the magnetic static field B_1 in the pumping region. When $B_1 \geq 0.5$ G, the dependence of the asymmetry versus B_1 becomes smaller. Currently, we have no explanation of such a behaviour taking into account the fact that, the number of unpumped atoms is negligible (< 1 %). In standard operating conditions, we think that the pumping efficiency is independent of the velocity distribution of the atoms.

An attempt has been made to calculate the population asymmetry taking into account a laser detuning. The values predicted by the model do not agree with the experimental ones. We note a ratio of 10 in favour of the experiment. Such a discrepancy has already been mentioned [5].

There is obviously a need for further investigations. The problem of population leakages due to the overlap of the $3 \rightarrow 3$ and $3 \rightarrow 2$ fluorescence lines which are closely spaced could be responsible for an imbalanced distribution of the Zeeman sublevels populations in the ground state.

We have finally made an estimation of the Rabi pulling frequency shift [6] caused by a laser detuning Δ. A typical asymmetry $a_{|1|} = 0.55 \times 10^{-3}$/MHz has been measured with the Cs3 resonator operating in the standard conditions. Moreover, we have assumed that :

1°) the velocity distribution of all the Zeeman sublevels in the ground state has the same Maxwellian profile,

2°) the effect of frequency modulation of the microwave interrogation signal has been neglected.

This calculation gives $\Delta f/f = 5\ 10^{-14}$ for a laser detuning of 1 MHz and a static magnetic field $B_c = 60$ mG.

5. CONCLUSION

In this paper, we have pointed out the link between the asymmetry of the microwave spectrum delivered by an OP Cs BR operated with the familiar $3 \rightarrow 3\ \sigma$ transition and a possible laser detuning. To minimize this detuning, it seems interesting to lock the frequency of the laser diode using the third harmonic technique [7]. The frequency offset due to the background added to the fluorescence profile should be reduced. It is also obvious that theoretical efforts have to be undertaken to determine the origin of the asymmetry.

Let us also mention that when the D_1 line is considered at $\lambda = 894$ nm, the fluorescence resonant lines are well separated. Consequently, the laser frequency is accurately locked to the center of the considered transition and there are no leakage ; this leads to highly symmetrical sublevels populations [3].

Acknowledgements

We gratelully acknowledge the support of STEI and PTF.

REFERENCES

[1] de Marchi A., Rovera G.D. and Premoli A. Metrologia, 1984, 20, 37-47.

[2] Cutler L.S., Flory C.A., Giffard R.P. and de Marchi A. J. of Applied Physics, 1991, 69, n° 5, 2780-2792.

[3] Dimarcq N., Giordano V., Théobald G., Cérez P. J. of Applied Physics, 1991, 69, n° 3, 1158-1162.

[4] Giordano V., Hamel A., Petit P., Théobald G., Dimarcq N., Cérez P., Audoin C. IEEE Trans. on Ultrasonics, Ferroelectrics and Frequency Control, 1991, 38, 350-357.

[5] Nafcha Y. Albeck D., Rosenbluh M. Phys. Rev. Lett., 1991, 67, n° 17, 2279-2282.

[6] Vanier J. and Audoin C. *The Quantum Physics of Atomic Frequency Standards*. Adam Hilger, Bristol, U.K., 1989, p. 834, Vol. 2.

[7] Cérez P., Brillet A. Metrologia, 1977, 13, 29-33.

THE EFFECT OF MAJORANA TRANSITIONS ON THE FREQUENCY OF A CESIUM ATOMIC CLOCK

Roland Schröder, Andreas Bauch

Physikalisch-Technische Bundesanstalt, Braunschweig, Germany

ABSTRACT

A theoretical model is summarized which describes the effect of Majorana transitions on the frequency of a cesium atomic clock. For the first time the phase correlation between successive excitations in inhomogeneous magnetic fields and in the microwave resonator are taken into account. It follows that π transitions with the selection rule $\Delta F=1$, $\Delta m=\pm 1$ together with Majorana transitions are of particular importance. The theory is supported by a number of experimental findings which are outlined and interpreted in terms of the predictions of the theory.

Keywords:
Cesium atomic clocks, Majorana transitions, frequency shifts

1. INTRODUCTION

Observations were made which indicated frequency shifts in cesium atomic clocks caused by Majorana transitions in the atomic beam /1-3/. These transitions obey the selection rules $\Delta F=0$, $|\Delta m| = 0,1,2..2\cdot F$ and they occur when the atomic beam passes through weak inhomogeneous magnetic fields, as Majorana demonstrated as early as 1932 /4/. At PTB, from 1983 to 1987, investigations into this subject were carried out on the CSX experimental apparatus /5/. Frequency shifts were observed which were surprisingly large: they reached the 10^{-11} level. Recently the authors developped a theoretical model which permits to explain these findings /6/. It is the subject of this paper at first to summarize this theory. Then the essential predictions are listed. A looking into old laboratory notebooks revealed that additional experimental data had been collected which can now be used to confirm the validity of the theoretical model. This material is presented together with a short description of the CSX apparatus.
Equations numbered (p.q) here can be found as #q in chapter p in /6/.

2. SUMMARY OF THE THEORY

The interaction region of a cesium atomic clock is illustrated in Fig.1. Consider at first the absence of Majorana transitions. Atoms in a pure state enter the resonator. The microwave irradiation takes place at R1 and R2 in two successive phase-coherent pulses of the duration $\tau=l/v$ separated by the interval $T=L/v$, v being the velocity of the atom. In R1 or R2 it is now possible for an atom to get from the initial to the final state. The interference between the two possibilities of transition leads to the well-known Ramsey fringes as the line shape of the allowed microwave transitions. Due to the finite time of interaction, the excitation has a spectral distribution by which also neighbouring microwave transitions are induced. Furthermore, the magnetic field lines of the microwave field - as shown in Fig.1 - are curved and, hence, not parallel to the direction of the C-field over the whole length l. As a result, it is basically possible to excite transitions which allow changes $|\Delta F| = 0,1$ and $|\Delta m|=0,1,2$, both in R1 and R2. They will in general lead to shifts of the clock transition frequency which were named "Rabi pulling" /7/ or "Ramsey pulling" /8/, depending on the mechanism involved.

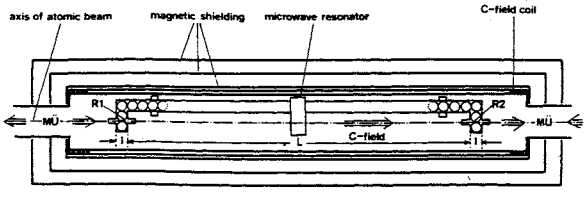


Fig.1: Interaction region of an atomic clock. R1, R2: sections of the interaction with the microwave field of the length l; L: distance between the end parts of the resonator. In the resonator the course of the magnetic microwave field component is indicated. MÜ are the places where Majorana transitions can be induced in the atomic beam.

Consider now Majorana transitions occuring at the places MÜ indicated in Fig.1. An atom leaving the polarizer magnet in a pure state (4,m) can now reach a final state (3,n) along several more different ways which exist as a consequence of the subsequent action of inhomogeneous magnetic fields and the microwave fields. The probability amplitude for the occupation of each final level is a sum of terms which describe the individual paths. By forming the probability density (square of the sum of amplitudes) interference terms appear which are of particular weight if they contain a product of the transition amplitude of a parasitically induced microwave transition (e.g. a π transition) times the transition amplitude of the very probable clock transition . The calculation could therefore be restricted to these terms. To simplify matters, in /6/ the approach that Majorana transitions occur either in front of the resonator (case 1) or behind it (case 2) was made. Case 1 was treated in detail. Here six relevant individual paths all ending in the (3,0) level can be identified which are listed in table 1.

Table 1: Potential paths from the initial states (4,0) and (4,±1) to the final state (3,0). The notation is the following: $\bar{1}$:= -1, Y $\hat{=}$ yes, N $\hat{=}$ no, superscripts: place of the transition (R1 or R2), subscripts: initial level of the transition. The transition amplitudes are given in chapter 4 in /6/.

Path	Transition in R1	amplitude	in R2	amplitude
1	(4,0)⇒(3,0)	$Y_0^{(1)}$	(3,0)⇒(3,0)	$N_0^{(2)}$
2	(4,0)⇒(4,0)	$N_0^{(1)}$	(4,0)⇒(3,0)	$Y_0^{(2)}$
3	(4,1)⇒(3,0)	$Y_1^{(1)}$	(3,0)⇒(3,0)	$N_0^{(2)}$
4	(4,$\bar{1}$)⇒(3,0)	$Y_{\bar{1}}^{(1)}$	(3,0)⇒(3,0)	$N_0^{(2)}$
5	(4,1)⇒(4,1)	$N_1^{(1)}$	(4,1)⇒(3,0)	$Y_1^{(2)}$
6	(4,$\bar{1}$)⇒(4,$\bar{1}$)	$N_1^{(1)}$	(4,$\bar{1}$)⇒(3,0)	$Y_1^{(2)}$

An atom initially in the state (4,m), will take one of these ways if a Majorana transition to one of the three states (4,0) or (4,±1) had been stimulated in the region MÜ preceding R1. It is explicitely distinguished in table 1 whether the microwave transitions are made in R1 or R2. The atom will finally be found in the (3,0) state with the probability (density) P given by

$$P := P(m,v,\varepsilon) = P_0 + P_s \quad (2.2),(2.3) \qquad (1)$$

with $\varepsilon = f_p - f_0$

where f_p is the probing microwave frequency and f_0 the unperturbed clock transition frequency. P_0 (4.10) results from the paths 1 and 2 to

$$P_0 = \tilde{T}_{0m}^2 \cdot \sin^2(2\, b_\sigma \tau) \cdot (1+\cos(2\pi\varepsilon T))/2 \qquad (2)$$

which is easily recognized as the transition probability of the clock transition in the near-resonant approximation (e.g. eq. 5.2.46 in /9/) multiplied by the probability $\tilde{T}_{0m}^2$ that the atom reaches the resonator in the (4,0) state due to a Majorana transition (4,m)⇒(4,0).

P_s describes the "perturbations" originating from the additional paths 3 - 6 and is in detail explained in (2.2), (4.11) and (4.12) /6/. It is sensitive to the phase relation between the states involved: (4,±1) and (4,0) (see table 1). The coherence between these neighbouring states evolves in time with the Larmor frequency

$$\omega_L = g\, \mu_B \,|\, B \,|\, / \,\hbar \qquad (3)$$

as the characteristic frequency (3.7). The g factor of cesium is 1/4, μ_B is the Bohr magneton, $\hbar$ the Planck constant divided by (2π) and B the magnetic flux density (named magnetic field furtheron) at the location of the atom.
As a characteristic quantity the parameter K was identified which is the number of 2π phase evolutions (with angular velocity ω_L) between the place where for the first time a coherence between Zeeman states is induced and R1 or R2, respectively (5.17).

The frequency shift $\delta\varepsilon_m$ caused by atoms starting at the initial level (4,m) was calculated to be

$$\delta\varepsilon_m = \delta\hat{\varepsilon}_m \cdot \langle \sin(\psi_0 - 2\pi\, K\, \langle v\rangle_m / v\,) \rangle . \qquad (4)$$

Infact, there are two terms of the kind (4) coming from the interactions in R1 and R2.

Here < > denotes the averaging over the velocity distribution $n_m(v)$. To obtain (4) it was presupposed that the width Δv of the velocity distributions is narrow and the value of K is inserted for an atom with mean velocity $\langle v\rangle_m$. Besides, it is supposed that the centre ot the resonance line is detected using square-wave frequency modulation with the modulation width equal to the line width. The angle ψ_0 (3.13) is determined by the interaction with the inhomogeneous fields and, in case of narrow velocity distributions, it is approximately constant.

For the maximum amplitude $\delta\hat{\varepsilon}_m$ one finds

$$\delta\hat{\varepsilon}_m \sim \tilde{T}_{0m}(\, \tilde{T}_{1m} + \tilde{T}_{\bar{1}m}\,) \cdot \Pi . \qquad (5)$$

Here $\tilde{T}_{fi}$ is, apart from a pure phase factor, the complex transition amplitude (4,i) ⇒ (4,f) due to Majorana transitions (3.12) and Π is the transition amplitude for the π transitions (5.16).

Fig.2 illustrates the dependence of the frequency shift $\delta\varepsilon_m$ upon K for a velocity distribution having a width $\Delta v = 0{,}15 \cdot \langle v \rangle_m$. The beat characteristic and the general decrease of the envelope with $1/|K|$ follow directly from the averaging process stated in (4). The total frequency shift is the sum over all terms $\delta\varepsilon_m$, m = -3,..4, a superposition of several curves like that in Fig.2, but with different amplitudes and oscillation periods. It turns out that in practical cases a few terms with specific values of m dominate. Thus the total frequency shift will still exhibit an oscillatory character in K /6/.

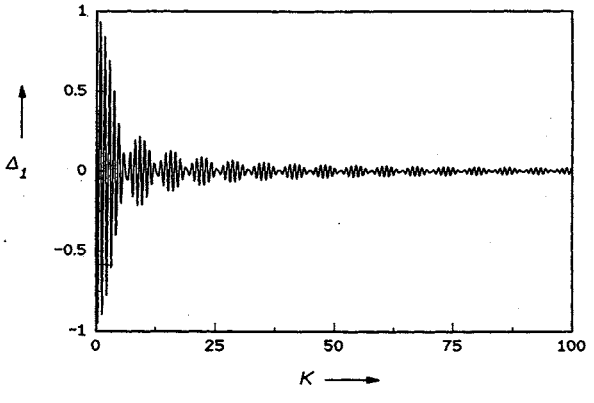


Fig.2: Normalized frequency shift $\Delta_1 := \delta\varepsilon_m/\delta\hat{\varepsilon}_m$ calculated with the aid of (4) and a width of the velocity distribution $\Delta v_m/\langle v \rangle_m = 0{,}15$ as a function of the number of 2π phase evolutions K.

3. GENERAL REMARKS ON THE FREQUENCY SHIFTS

As $\tilde{T}_{\bar{f}\bar{i}} = (-1)^{f+i} \cdot \tilde{T}_{fi}$ is valid, it follows from (5) that $\delta\hat{\varepsilon}_0 = 0$ and $\delta\hat{\varepsilon}_{\bar{m}} = -\delta\hat{\varepsilon}_m$. Therefore atoms entering the interaction region in the state (4,0) do not contribute to the frequency shift $\delta\varepsilon$. Furthermore, a shift $\delta\varepsilon_m$ is compensated by a shift $\delta\varepsilon_{\bar{m}}$ if the velocity distributions $n_m(v)$ and $n_{\bar{m}}(v)$ are identical with regard to the total number of atoms and their velocity. A perfect compensation, however, cannot be expected if magnetic state selection is employed as the state selection is based on the force - different in amount and sign - acting on the atoms in the different quantum states m. Certainly, the state selection by optical pumping could be advantageous here.

Fig.2 illustrates that large frequency shifts occur for small values of K only. A small separation in space of the perturbing fields will favour large effects: If the Majorana transitions occur in front of the resonator, the π transitions in R1 are harmful, if the Majorana transitions occur behind the resonator, those in R2. As K is proportional to the magnetic flux density along the beam path (via ω_L), $\delta\varepsilon$ should decrease with increasing C-field B_C at otherwise unchanged conditions in the clock. The amplitude of the excitation of the π transitions, Π, behaves similarly (see (4.3)). The maximum observable frequency shift was indeed measured to be roughly proportional to $1/B_C^2$ as it is shown in Fig.3.

In detail, however, Π is a periodic function of the magnetic field, and in particular for

$$\omega_L \tau /2 = (2j + 1)\, \pi/2 \quad , \; j = 1,2,.. \qquad (6)$$

the π transitions become ineffective and the frequency shift is zero (5.18). The averaging over the atomic velocities will prevent a perfect zero frequency shift at the predicted settings of B_C (3), (6).

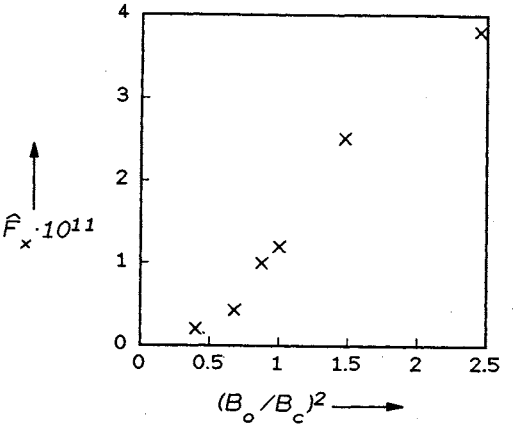


Fig.3: Maximum values $\hat{F}_x$ of the CSX frequency $F_x := (f(CSX)-f(CS1))/f(CS1)$ as a function of the C-field B_C from data sets like the one in Fig.7. The value $B_0 = 10{,}4\ \mu T$ ($\omega_L = 2\pi \cdot 36{,}5 \cdot 10^3\ s^{-1}$) used for the normalization is the routine value of B_c (see table 2). f(i) are the resonance frequencies of the clock transition realized in the CSX and in the primary clock CS1.

Because of the proportionality to Π, δε will strongly vary with the strength of the microwave excitation field. Any setting away from the optimum excitation condition will reduce the unshifted signal which is proportional to P_0 (2). $|\delta\varepsilon|$ will in particular rapidly increase at microwave power levels above the optimum value but will only slightly decrease at levels below. This behaviour can be found in the experimental data shown in Fig.4. For comparison, the frequency versus power dependence measured when Majorana transitions were carefully avoided is given, too.

The stimulation of Majorana and π transitions is also a function of the trajectory of the atom, more precisely, of the distance y from the symmetry line y=0 at the place MÜ and in the irradiation sections R1 and R2 (see Fig.1). As Π is proportional to y, it follows that the frequency shift will disappear if the distribution of the trajectories around y=0 is perfectly symmetrical (4.3, 4.5) and the stimulation of Majorana transitions has the same property of axial symmetry. Some reasons were given that this perfect symmetry was never achieved during the experiments with the CSX, explaining the large frequency shifts observed /6/. Further experimental evidence will be given subsequently.

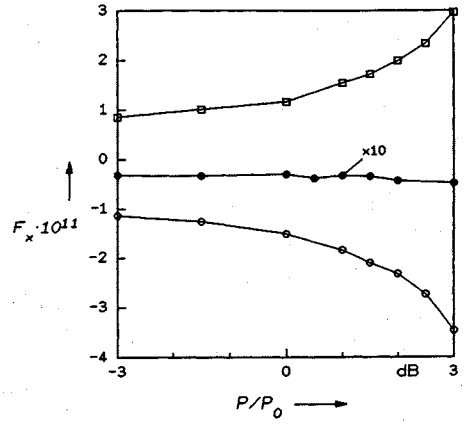


Fig.4: Relative CSX frequency F_x as a function of the power P of the microwave probing signal, measured while Majorana transitions were induced, "□ ", "○", and in the unperturbed mode of operation "●" (F_x multiplied by 10). The mean frequency offset of $\simeq -4 \cdot 10^{-13}$ in the data "●" is caused by the cavity phase difference. P_0 is the "optimum" microwave power in the unperturbed mode of operation leading to a maximum resonance signal.

4. SURVEY OF THE EXPERIMENTAL WORK

- The CSX atomic beam apparatus

The experiments described here were carried out using the CSX apparatus, the PTB's experimental atomic beam frequency standard. In Fig.5 it is schematically represented, characteristic data are collected in table 2. A point beam source and a point detector in an axially symmetrical arrangement are used. The properties of the atomic beam are governed by the polarizer and analyzer magnets. They consist of quadrupole and hexapole magnets. The C-field is generated by a solenoid surrounded by a magnetic shield.

For the investigations into Majorana transitions, in the places S1 and S2 (see Fig.5) pairs of coils were mounted which allowed additional static magnetic fields to be generated.

Furthermore, the microwave cavity can be slightly rotated, the centre of rotation lying in the mid of R1. This precaution was originally taken to measure the spatial variation of the phase of the microwave field in R2 /10/. It allowed, in the present context, to detect the influence of the π transitions in connection with Majorana transitions.

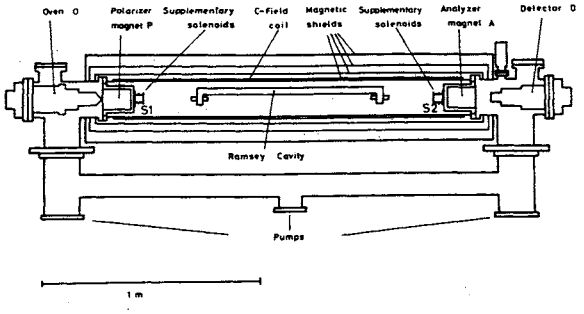


Fig.5: Schematic diagram of the Cs atomic beam frequency standard CSX (vertical section).

Table 2: Parameters of the frequency standard CSX during the experimental investigations into Majorana transitions in the years 1984 to 1987

Diameter of the atomic beam contributing to the signal	3 mm
Mean velocity $\langle v \rangle_0$ of atoms in the initial state (4,0)	410 m/s
Length L	0,7677 m
Length l	23,2 mm
Distance S1 - R1, R2 - S2 (typical value)	0,27 m
Line width of the clock transition	262 Hz
C-field flux density (typ. value)	10,4 μT

• Results and their interpretation

In this paper only the double zero experiments shall be recalled. The coils in S1 and also in S2 were used to generate a field with two zeroes in the z-component at a small distance from one another in z_{01} and z_{02}, as it is illustrated in Fig.6. Under these conditions, the clock frequency was found shifted by up to $\pm 5 \cdot 10^{-11}$. As an example, Fig.7 shows the results obtained during one measurement run. A careful inspection of the prevailing experimental situation permitted to explain the data reasonably well /6/. Here only the key to the understanding shall be repeated in a simplified version.

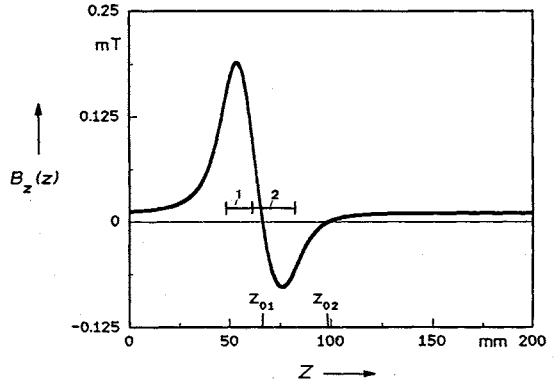


Fig.6: Field component $B_z(z)$ calculated as a superposition of the constant C-field of 10,4 μT and the field generated with the coils 1 and 2 in S1, whose position is outlined; for comparison with Fig.7: $I_1(S1) = 250$ mA, $I_2(S1) = 84$ mA.

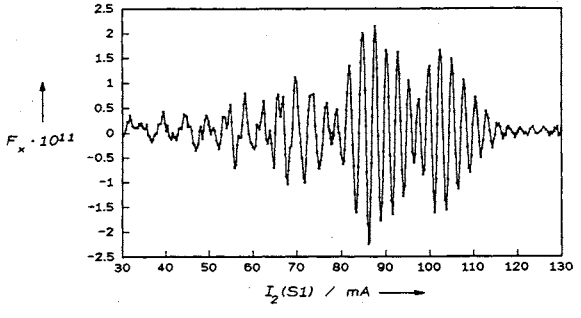


Fig.7: Relative CSX frequency F_X as a function of the current I_2 through coil 2 in S1 in the case of a field configuration as in Fig.6. Time of measurement per value is 30 min, statistical uncertainty of the frequency measurement $\sigma_y(30\ \text{min}) = 3 \cdot 10^{-13}$.

Having passed the first zero of $B_z(z)$ in z_{01} the atom is in a coherent superposition of Zeeman eigenstates. The coherence evolves with ω_L given by the local magnetic field. At $z = z_{02}$, a value of K has accumulated depending on ω_L and the atom's time of flight. The second zero in the CSX had an extremely non-adiabatic character. Its effect on the atomic state vector is in a good approximation to exchange the coefficients v_m by $v_{\bar{m}}$, the v_i being the coefficients of a series expansion of eigenfunctions of the atomic state. This in turn entails, that the accumulated phase between z_{20} and R1 enters in (4) with a different sign compared with those from the preceding section from z_{01} to z_{02}. In consequence, the final value of K in (4) can have positive and negative values and, in particular, can become zero. By manipulating the currents through the coils and thereby changing the magnetic fields in between the two zeroes, the value of K can be swept around zero. Indeed, the maximum frequency shifts shown in Fig.7 were obtained when K was around zero. The period of the frequency oscillations also coincides with a change of K by ±1. The same explanation holds for the data shown in Figs. 3 and 4.

Frequency shifts similar to those shown in Fig.7 were also observed when there was a double zero field distribution behind R2 and Majorana transitions were carefully avoided before R1. In this case, the π transitions induced in R2 are particularly

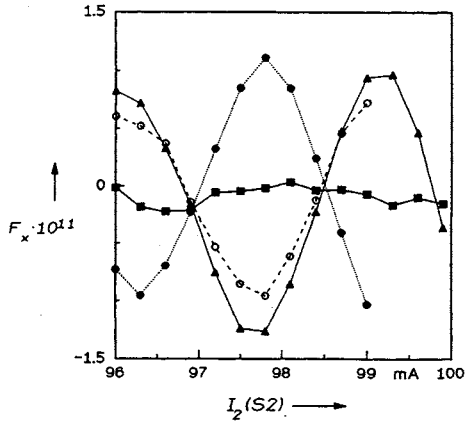


Fig.8: Relative CSX frequency F_x as a function of the current I_2 through coil 2 in S2 for four positions of the second irradiation section of the resonator; "■" $\Delta y = 0$ (cavity axis coincides with the atomic beam axis), "▲" $\Delta y = -1{,}5$ mm, "0" $\Delta y = -1{,}2$ mm, "●" $\Delta y = +1{,}2$ mm. Additional explanations given for Fig. 7 apply here, too.

harmful, as discussed in chapter 3. This could be tested by slightly rotating of the CSX microwave cavity. In Fig.8 frequency measurement results are shown which were taken at different positions of the microwave cavity with respect to the atomic beam axis. Note, that at a fixed setting of the current $I_2(S_2)$ the frequency shift even changes its sign when the atomic beam is admitted to pass the cavity opening above or below the symmetry line. In a different view, the frequency dependence on Δy is plotted in Fig.9 for two settings of the magnetic fields. For comparison, the original "distributed phase difference" data are given /10/.

Frequency shifts provoked by inhomogeneous fields in front of R1, however, were almost insensitive to a slight rotation of the cavity which mainly affects the interaction in R2. Both, the positive and the negative results are considered as further evidence for the validity of the theoretical treatment.

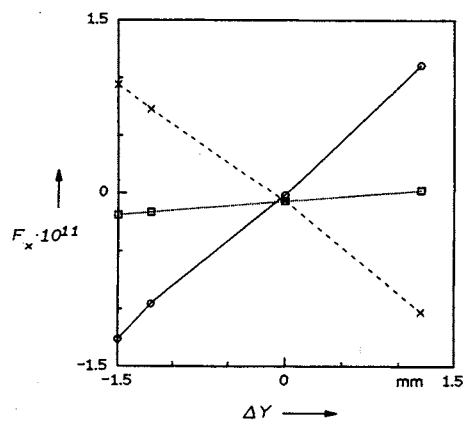


Fig.9 Relative CSX frequency F_x as a function of the cavity position Δy for two settings of the current $I_2(S2)$, "X", 97,7 mA, and "O", 99,0 mA, (see Fig 8) and in the unperturbed mode of operation " □ " /10/.

5. CONCLUSION

A recently developed theoretical model was used to explain frequency shifts which were observed in the CSX atomic clock during experiments with Majorana transitions. It was found that the frequency shifts could in general be suppressed by

- employing an atomic beam with a highly symmetrical population of the Zeeman sublevels,

- avoiding any asymmetry in the beam forming elements and in the magnetic field and interaction geometry,

- employing a large C-field such that the phase coherences dephase in the average due to the large Larmor frequency and due to the spread of atomic velocities.

Last not least, it would be the best solution to avoid the occurrence of Majorana transitions by carefully trimming the magnetic fields along the beam path between the state selective element and the state analyzer. Rules for to prove the success of such measures have been given in /5,6,11,12/. They apply likewise to cesium atomic clocks with magnetic state selection and with optical pumping.

6. REFERENCES

[1] D.W Allan et al., Proc. 31st Ann. Symp. on Frequ. Contr. (1977) 555

[2] S.Urabe et al., IEEE Trans. Instr. Meas. **IM-29** (1980) 304

[3] G. Becker, IEEE Trans. Instr. Meas. **IM-27** (1978) 319

[4] E. Majorana, Nuovo Cimento **9** (1932) 34

[5] A. Bauch, Dissertation Universität Mainz, PTB - Bericht **PTB-Opt-20**, PTB, Braunschweig 1986.

[6] A. Bauch, R. Schröder, subm. to Annalen der Physik (Leipzig)

[7] A. De Marchi, G.D. Rovera, A. Premoli, Metrologia **20** (1984) 37

[8] L.S. Cutler, C.A.Flory, R.P.Giffard, A. De Marchi, J. Appl. Phys **69** (1991) 2780

[9] J.Vanier, C.Audoin: The Quantum Physics of Atomic Frequency Standards. Adam Hilger, Bristol and Philadelphia 1989

[10] A. Bauch, T. Heindorff, R. Schröder, IEEE Trans. **IM-34** (1985) 136

[11] A. Bauch, T.Heindorff, Proc. 5th EFTF (Besancon 1991), 243

[12] A. Bauch, R. Schröder, Proc. 5th EFTF (Besancon 1991), 11

ACCURACY ANALYSIS OF THE ATOM-MICROWAVE INTERACTION IN A TE_{017} CYLINDRICAL RABI CAVITY FOR A HIGH C-FIELD CESIUM BEAM RESONATOR

Andrea De Marchi*, Elio Bava**, Patrizia Tavella!

* Dip. di Elettronica, Politecnico di Torino, Torino, Italy.
** Dip. di Elettronica e Informazione, Politecnico di Milano, Milano, Italy
! Istituto Elettrotecnico Nazionale Galileo Ferraris, Torino, Italy.

Abstract

The possibility of using a TE_{017} cylindrical cavity for the excitation of microwave transitions in a Cesium atomic beam is analyzed in this paper from the point of view of the biases that it may produce on the observed resonance frequency. Doppler related effects, including distributed phase shift, and pulling by neighboring transitions are the sources of bias which must be studied. Relevant considerations concern travelling wave symmetry, transversal phase variations across the beam for the used longitudinal field component, and magnitude of transversal field components. The proposed cavity appears particularly suited for the high C-field configuration, in which the beam can be very wide because transversal field components are irrelevant. The TE_{01} mode, in fact, accepts with little disturbance wide open holes in the terminal caps. In this way the wide central area of quasi-stationary phase, which it offers for the longitudinal field component, can be exploited thoroughly. High beam intensity should then be achievable without negative effects.

Evaluation of the overall Doppler effect by numerical integration of the Schrödinger equation confirms a rough estimate based on phase variations, derived analytically from the expression of the field within the cavity, and suggests feasibility of accuracies in the 10^{-14} range.

1 Introduction

An interesting feature of the high C-field concept for Cesium beam standards is that neighbouring transitions are so far away in frequency from the clock transition that no fear is left of exciting $|\Delta m| = 1$ transitions and suffering from Ramsey pulling, even if the rf magnetic field component normal to C-field is not at all negligible. It is then possible to open up the beam passing holes, in a suitable microwave cavity, and greatly increase in this way the beam flux without incurring in an accuracy reduction. The TE_{01} mode in circular waveguide appears ideal in this sense, because it guarantees in its center a wide area of quasi-stationary phase [1] for the relevant longitudinal magnetic field component, and accepts with little disturbance wide open holes in the terminal caps. A Ramsey structure with TE_{011} cavities at the ends would then appear to be the best solution. With the elimination by high field operation of Rabi and Ramsey pulling, cavity phase shift would remain the greatest cavity induced bias, at the level of almost 10^{-13}. The possibility suggested in [2, 3] that multilambda Rabi cavities may be relatively free from Doppler related biases, makes then attractive the adoption of such a solution when the accuracy goal is 10^{-14}.

In this paper a report is given on the design and realization of a TE_{017} cylindrical cavity which can be used in the prototype high field beam machine which is under construction at the Politecnico. An accuracy analysis is also offered for a beam resonator employing this cavity. The first order Doppler effect is studied as a function of relevant parameters, which turn out to be asymmetry in conductivity of the end caps, and microwave power level.

An estimate is also given of the expected level for other biases in this configuration.

This research is partially funded by:
- Consiglio Nazionale delle Ricerche
- Istituto Nazionale di Fisica della Materia
- Agenzia Spaziale Italiana

2 Cavity design and realization

A cylindrical cavity was designed, which resonates on different TE_{01p} modes at different Cesium transition frequencies at the field of 82 mT.

The choice was made to work close to cutoff for the considered waveguide mode. The advantages of this solution were seen to be:

1. The reduced transversal cavity dimension, which then allows the use of a small C-field magnet bore, with gain in power consumption and magnet weight.
2. The minimization, as will be shown below, of the first order Doppler effect resulting from asymmetries in the absorption of the two end caps.
3. The maximization of cavity length, for interaction time purposes, with a minimal number of wavelengths in the cavity. This is interesting because it simplifies the theoretical analysis of the interaction.
4. A slight reduction of cavity Q in this very low loss mode, with consequent increased external electronic tuning range.

The main drawback is that the internal diameter of the cavity becomes the most important dimension in determining its resonance frequencies. Mechanical construction tolerances are then quite tight. This is all the more true if specific resonance frequencies are desired for two different cavity modes, because then both length and diameter are fixed by the two imposed conditions.

For the design of the considered cavity the $7\lambda_g/2$ mode was tuned to the minimized ν_{-1} frequency of the clock transition, and the $11\lambda_g/2$ was tuned to ν_0 for field measurements. Pedices attached to frequencies indicate the m_F quantum number of the considered Cesium hyperfine transition. These frequencies are $\nu_{-1} \approx 8.9007$ GHz and $\nu_0 \approx 9.475$ GHz at 82 mT. As a result, the $3\lambda_g/2$ mode will not be too far from ν_{-2}, and the $9\lambda_g/2$ mode will be close to the ν_0 at low field. This feature can be used to operate the machine at low fields for comparison purposes. Cavity dimensions are fixed by the following two equations, where $p_{-1} = 7$ and $p_0 = 11$:

$$\left(\frac{2\pi a}{c\chi_{01}}\right)^2 = \frac{p_0^2 - p_{-1}^2}{(p_0\nu_{-1})^2 - (p_{-1}\nu_0)^2} \quad (1)$$

$$\left(\frac{c}{2L}\right)^2 = \frac{\nu_0^2 - \nu_{-1}^2}{p_0^2 - p_{-1}^2} \quad (2)$$

which yield a diameter $2a = 43.082$ mm and a length $L = 391.34$ mm. Relative resonance variations with cavity dimensions are given by:

$$\left.\frac{\delta\nu}{\nu}\right|_a = -\left(\frac{c\chi_{01}}{2\pi a\nu}\right)^2 \frac{\delta a}{a} \quad (3)$$

$$\left.\frac{\delta\nu}{\nu}\right|_L = -\left(\frac{p}{\nu}\right)^2 \left(\frac{c}{2L}\right)^2 \frac{\delta L}{L} \quad (4)$$

from which it can be seen that the ratio of frequency sensitivities to a and L is $(\chi_{01}L/p\pi a)^2$, that is about 10 for $p = 7$. This imposes mechanical tolerances that are quite loose on the length, particularly by taking into account the flexibility allowed by mode filters, but are very tight on the diameter. Fine stone finishing on Aluminum may be able to guarantee a tolerance of 5 μm on an internal tube diameter; alternatively, electroforming of Copper on a precisely turned Aluminum mandrel should be capable of 1 μm accuracy. The latter would amount to a frequency uncertainty of the order of $5{\cdot}10^{-5}$, which is barely within the resonance width of the used high Q mode. The former would produce a cavity which may need electronic tuning to guarantee on-resonance operation. However, this shouldn't be seen as a drawback because servo control on the cavity frequency is almost compulsory on such a high Q mode, for the reduction of cavity pulling, if only because temperature variations move the resonance by the linear thermal expansion coefficient $\alpha = 1.6{\cdot}10^{-5}/°C$. Only a metalized glass cavity might allow elimination of a servo system.

3 First order Doppler effect

The frequency ν', in the atom's reference frame, of a microwave signal of frequency ν, can be calculated from the time derivative of the total phase of the signal as seen by the atom. It is then

$$\nu' = \frac{1}{2\pi}\left(2\pi\nu + v\frac{d\varphi}{dx}\right) \quad (5)$$

where $d\varphi/dx$ is the phase gradient along the atomic trajectory and v is the velocity of the atom. In the case of a pure standing wave the phase gradient vanishes and $\nu' = \nu$ for any velocity. At the other extreme, the maximum possible gradient is the one generated by a pure travelling wave, which is given by $k_0 = 2\pi/\lambda$. This means that the phase changes by 2π every wavelength. In this case, for an atom running directly against the wave, the Doppler shift $\delta\nu_{Da} = \nu' - \nu$ as seen by the atom is given by v/λ, or $\nu{\cdot}v/c$. In intermediate cases the shift is proportionally smaller, and can be calculated with the formula:

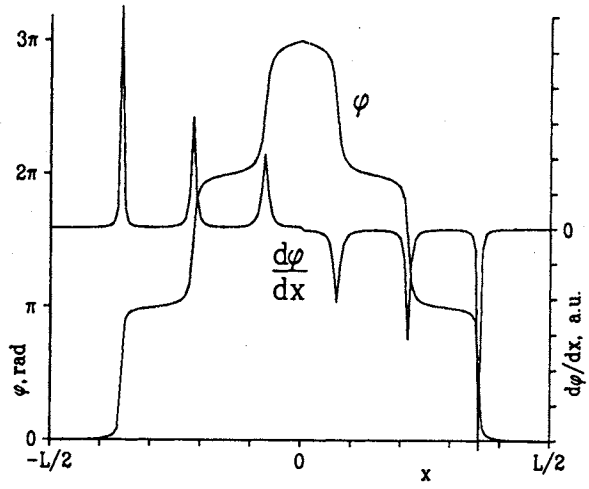


Figure 1: Qualitative plot of phase and phase gradient on the axis of a TE_{017} cylindrical cavity.

$$\left.\frac{\delta\nu}{\nu}\right|_{Da} = \frac{v}{c}\frac{1}{k_0}\frac{\mathrm{d}\varphi}{\mathrm{d}x} \qquad (6)$$

if the phase gradient is known.

In the case at hand, where the atom passes through a $7\lambda_g/2$ microwave cavity, the situation is more complicated because both phase and amplitude of the relevant H_x magnetic field component change with position. As an atom goes through the cavity, at each location it will experience a different Doppler shifted frequency, but the field will not drive the atom's phase with the same strength everywhere, because its amplitude varies. Since the relevant quantity in determining how much the atom is driven is always the product of field amplitude and exposure time (Ramsey's $b\tau$), it can be expected that the average Doppler shift at the end of the interaction may be not too different from:

$$\left.\frac{\delta\nu}{\nu}\right|_{Da} = \frac{v}{c}\frac{\beta}{k_0}\frac{1}{L}\int M(x)\frac{\mathrm{d}\varphi}{\mathrm{d}(\beta x)}\,\mathrm{d}x \qquad (7)$$

with $\beta = 2\pi/\lambda_g$ and $M_{<,>}(x) = |H_{<,>}(x)|/H_0$. This can be evaluated approximately in a rather straightforward way from the expression of $H(x)$, given in [3] and given below, referred to the maximum H_0 of the exciting field. It is obvious that the maximum of $H(x)$ is $2H_0$ at the antinodes, where counter- running waves sum in phase.

$$M_{<,>}(x) \approx 2\sinh\left[\gamma\left(x \pm \frac{L}{2}\right) \pm \frac{\xi \mp \chi}{2}\right] \qquad (8)$$

The same symbols adopted in [3] are used here: the waveguide mode propagation constant is $\gamma = \alpha + j\beta$, and the reflection coefficient at the end caps is given by $\Gamma_{<,>} = -e^{-(\xi \mp \chi)}$, with $\xi = \xi' + j\xi''$ and $\chi = \chi' + j\chi''$. With these definitions the second cap dissipates more than the first if $\chi' > 0$, in which case the residual running wave from the central feed to the caps is stronger in the second half of the cavity than in the first, and the total Doppler shift seen by the atoms is expected to be negative. In (8) the origin of the x axis is in the center of the cavity, and no other asymmetry is considered besides the absorption in the end caps. This is justified by the calculations in [3], which show how the latter is the most important contribution. From (8) it is not hard to write the full analytical expression of phase and phase gradient as a function of x. The shape of both is shown in fig. 1. The phase gradient peaks come from the phase steps of π occurring at each half wavelength, where the amplitude of the field is minimum because the standing wave vanishes and the small progressive wave which carries the power to be dissipated in the walls prevails. The phase gradient can be written as:

$$2\frac{\mathrm{d}\varphi_{<,>}}{\mathrm{d}(\beta x)} \approx \frac{\sinh A_{<,>} - (\alpha/\beta)\sin B_{<,>}}{M^2_{<,>}(x)} \qquad (9)$$

with $A_{<,>} = \alpha(2x \pm L) \pm (\xi' \mp \chi')$ and $B_{<,>} = \beta(2x \pm L) \pm (\xi' \mp \chi'')$, and the integral in (7) can be rewritten, after some algebra, as an integral in $\mathrm{d}|x|$, from 0 to $L/2$, of the sum $\mathrm{d}\varphi_</\mathrm{d}(\beta x) + \mathrm{d}\varphi_>/\mathrm{d}(\beta x)$ divided by $M(x)$. Two regimes can be recognized for the integral, which suggest easy approximations. One is where $M \approx 2$, that is around $|x| = n\lambda_g/2$, the other is at the phase steps, where $|x| = n\lambda_g/2 + \lambda_g/4$. In the first case it can be readily seen that the most important term in the sum of derivatives with respect to βx is $\sinh A_< + \sinh A_>$, which is approximately equal to $-\chi'$. In the second case, the same result is found after considering that the width of the gradient peaks is given by the local (minimum) value of M. The Doppler shift as calculated from (7) is then evaluated to be:

$$\left.\frac{\delta\nu}{\nu}\right|_{Da} \approx -\chi'\frac{\beta}{k_0}\frac{v}{c} \qquad (10)$$

and since $\beta/k_0 \approx 1/4$ in our case, by assuming $\chi' < 10^{-2}\xi' \approx 3\cdot10^{-7}$, it is found to be smaller than $3\cdot10^{-14}$ at 100 m/s.

The result (10) is a rough estimate of the frequency shift experienced in the atom's reference frame. In order to excite the atom with its resonance frequency it is then necessary to drive the cavity with a signal shifted in frequency of this same amount, but on the opposite side. The Doppler effect observed in the atomic resonator is then given by (10), with a change in sign.

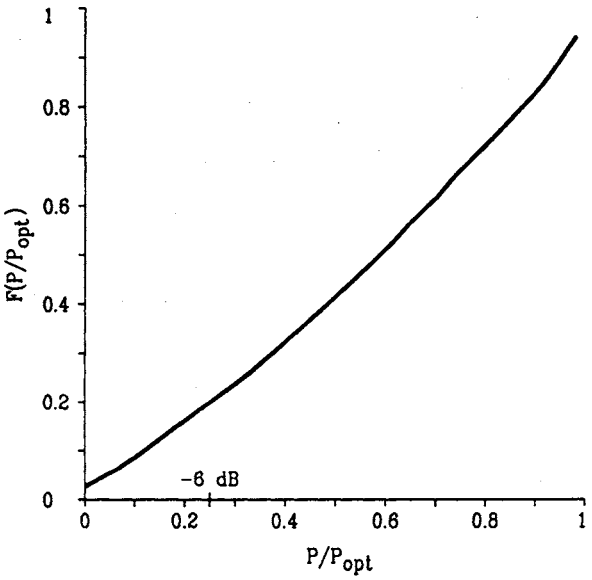


Figure 2: Calculated power dependent factor to be introduced in formula (11).

$$\left.\frac{\delta\nu}{\nu}\right|_D \approx \chi' \frac{\beta}{k_0}\frac{v}{c} \qquad (11)$$

By using the Magnus expansion [4], and an appropriate partition of the integration field in order to avoid convergence problems, an approximate analytical solution of the Schrödinger equation can be obtained. Maximization of the latter vith respect to frequency locates the shifted peak of the transition probability and allows one to write an expression for the Doppler effect. Its form substantially confirms (11), except for an extra power dependent proportionality factor which appears in the more refined calculation. A step by step numerical integration of the Schrödinger equation [3] also confirms this behaviour and gives the slope of power dependence. In fig. 2 the shape of this power dependent factor is shown. It is interesting to note that this function $F(P/P_{opt})$ does not seem to vary with the number p of half wavelengths in the cavity, for $p \geq 3$. At the optimum power P_{opt}, we find $F \approx 1$, which means that the estimate given above is quite accurate for that power level. At a power level 6 dB lower than P_{opt}, which is the level suggested in [3] for optimized operation, $F \approx .2$. In these conditions the first order Doppler effect is expected to be smaller than $6{\cdot}10^{-15}$ for atoms at 100 m/s, with the assumptions indicated above on the end cap symmetry.

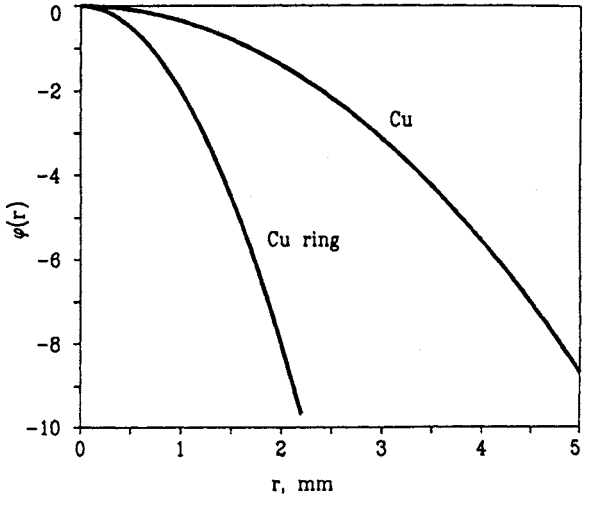


Figure 3: Transversal variations of the phase as a function of the distance r from center in the TE_{017} cavity. The phase curve of a ring cavity is reported for comparison.

4 Other systematic effects

A quick review of the traditional items usually appearing in accuracy budgets as related to cavity problems is given here for completeness.

Distributed Phase shift

As mentioned in the introduction, one of the reasons to select this cavity mode is the very wide area of stationary phase that it offers in the center of its cross section. In fig. 3 the phase curve as a function of radial distance from cavity center is shown, as calculated for the considered cavity using the analysis spelled out in [1]. The phase distribution curve of the ring cavity [5] adopted in NIST-7 is also shown for comparison. It can be seen that the phase experienced by an atom in an 8 mm beam cannot vary by more than 6 μrad under any circumstances, with beam holes of 8 mm in diameter. The average phase variation from end to end can then be guessed to be most likely much smaller than 1 μrad. This would set a very conservative limit of 10^{-14} to this effect.

Cavity pulling

More critical than phase shift, can be expected to be cavity pulling, as the cavity mode under consideration is a very high Q mode. Operation close to cutoff in fact does in general make for a low cavity Q, but the Q of this mode is so high that even so it remains higher than customary for Cesium beam standards. In a test cavity a Q of more than $2{\cdot}10^4$

was measured. On the other hand, since in the considered arrangement, unlike in Ramsey cavities, the beam will pass through the whole cavity and not just at the two ends, a pure mode must be guaranteed everywhere within the cavity, even near the feed. The fact that operation is planned with a waveguide mode which is not the fundamental one does not simplify the task. Discrimination against lower modes is essential. Mode filters at the end caps give flexibility in performing it, but in any case the job is made easier by a high Q. This fact, and the perturbations which would be introduced by overcoupling or other de-Q-ing practices, suggest as the best solution to resort to continuous cavity tuning with electronic control. A phase modulation "Pound" scheme with, for example, a 500 kHz modulation (more than twice the cavity resonance width) seems perfectly adequate to keep the cavity tuned as well as desired at all times, and is too far from the atomic resonance to introduce stark shift with the fraction of power which can enter the cavity.

Pulling by neighboring transitions

This item is here only as a reminder of the existence of this problem in other configurations.

5 Conclusions

In this paper an estimate was spelled out of the bias effects which may arise in a Cesium beam resonator from the use of a TE_{017} cylindrical microwave cavity. The results of this analysis appear promising on all accounts, even in the projected magnitude of the first order Doppler effect, which would at first sight seem to impose prohibitive symmetry requirements, in a cavity in which the atoms travel along the microwave power. Particularly interesting is the fact that no phase effect greater than 10^{-14} is expected, which might require a cavity (or beam) reversal for its evaluation. This is certainly an advantage with respect to Ramsey cavities at the target level of accuracy. The only partial drawback of this cavity seems to be its high Q, which forces the adoption of a servo control to keep the cavity tuned on the atomic resonance at all times.

Acknowledgments: the authors wish to acknowledge the appreciation of C. Salomon, A. Clairon and colleagues, who are planning to incorporate in their microgravity clock [6] the cavity model described here and in [2, 3].

References

[1] G. Vecchi, A. DeMarchi: "Spatial phase variations in a TE_{011} microwave cavity for use in a Cesium fountain primary frequency standard," CPEM-92, Paris. Also to be published on *IEEE Trans. Instrum. Meas.* (March 1993).

[2] A. De Marchi, R. E. Drullinger, J. H. Shirley: "Interference fringes from single-cavity excitation of an atomic beam," *Proc. 44th FCS* p. 34–38 (1990).

[3] E. Bava, P. Tavella, A. DeMarchi: "Doppler effects in a multi-wavelength Rabi cavity for the excitation of atomic beams in microwave frequency standards," *Proc. 46th FCS* p. 81–85 (1992).

[4] E. Bava, A. G. Shahian: "Interaction between atom and radiation with sinusoidally shaped amplitude: application to frequency standards," *Appl. Phys. B*, **B54**, 500–505, (1992).

[5] A. De Marchi, J. Shirley, D. J. Glaze, R. Drullinger, "A new cavity configuration for Cesium beam primary frequency standards," *IEEE Trans. Instrum. Meas.*, IM-37, 2 (1988).

[6] A. Clairon, Ph. Laurent, A. Nadir, M. Drewsen, D. Grison, B. Lounis, C. Salomon: "A simple and compact source of cold atoms for Cesium fountains and Microgravity clocks," *Proc. 6th EFTF*, p. 27, Noordwijk, March 1992.

SAW 2

Chairman: Ph. Defranould

A COMPARISON OF LOW LOSS SAW FILTER TECHNIQUES

Clinton S. Hartmann*, Victor Plessky and Dong Pei Chen

ASP Advanced SAW Products, CH 2022 Bevaix, Switzerland
***Hartmann Research, Inc. Dallas, Texas 75243**

Over the past few years, a variety of new approaches for achieving Low Loss SAW filters have been successfully demonstrated. These include various practical implementations of Single Phase Unidirectional Transducer (SPUDT) Filters, multi-strip-coupled ring filters, Interdigitated Interdigital Transducer (IIDT) Filters, Transversely Coupled Resonator Filters (TCF), Longitudinally Coupled SAW Resonator Filters (CRF), Electrically Coupled SAW Resonator Filters (ERF), and a variety of Lattice filter structures in which several one-port SAW resonators and/or separate SAW transducers are used as two-terminal building blocks in the Lattice filter network. Many of the devices reported in the literature can properly be described as a combination of several of the above techniques.

In addition to these new design approaches, the range of low loss SAW filter technology has been expanded by use of surface acoustic propagation modes which differ significantly from the Rayleigh type SAW modes which traditionally formed the basis for SAW technology. There are a variety of these modes including STW (surface transverse waves), "leaky" surface waves (in which the leakage rate is very small), SSBW (surface skimming bulk waves), and Bluestein-Gulyaev-Shimizu waves. All of these modes are characterized by the fact that the acoustic particle displacement is largely (and often totally) parallel to the substrate surface; therefore we will refer to this class of modes as surface shear acoustic waves. The well developed device modeling techniques used for Rayleigh type SAW devices are less accurate when used for surface shear acoustic wave devices because these modes are always closely coupled to a bulk shear mode. In spite of the theoretical design difficulties, these modes are of great interest for UHF low-loss SAW filters because they can provide both larger piezoelectric coupling which reduces loss and a larger wave velocity which eases lithography constraints.

This paper will survey these various filter types and compare their performance characteristics including insertion loss, amplitude flatness, delay flatness and out-of-band rejection.

A NEW COMPACT SAW LOW LOSS FILTER FOR MOBILE RADIO

Marc SOLAL, Jean-Michel HODE

Thomson Sintra DTAS, 399 Route des Crêtes, 06903 Sophia-Antipolis, France

ABSTRACT

The key objective of low loss filters for mobile radio applications is small size. For some years, SAW manufacturers have been working on SPUDT based filters which combine the advantages of low insertion loss, simple tuning, and single level processing, nevertheless the size is still too large for pocket telephone applications.

We present a new SPUDT based low loss filter utilizing an improved architecture which gives a larger impulse response for a given size by means of folded acoustic paths. This technique was used to design a 71 MHz SAW filter on quartz. This filter fits in a SIP 5 package while the classical SPUDT filter requires for equivalent characteristics a DIP 14 package, corresponding to a 50% size reduction.

Keywords : SAW, Filter, SPUDT, DART,Low Loss.

1. INTRODUCTION

The Pan European digital cellular radio system (GSM/DCS-1800) requires that the receivers incorporate highly selective filters to pass the processed channel and to protect the system from co-channel interference. Typically, the filtering function is now split between the intermediary frequency and the baseband processing section. The aim of the IF filter is to protect the receiver against interferences while the baseband insures the channel selection. Typical IFs are between 45 MHz and 220 MHz, insertion loss has to be kept within the 10 dB range for 200 kHz bandwidths. The required selectivity depends on receiver designs with a mean specification of +/-300 kHz for the-10 dB bandwidth and a 40 dB rejection. One main issue, important for pocket telephone applications, is to achieve very small sizes. As an example, the smallest GSM telephone available today is 147 X 67 mm^2.

For this kind of IF filter, SAW is now recognized as the technology offering the best compromise between performance, size, and cost.

For some years, SAW manufacturers have been working on Single Phase Unidirectional Transducer (SPUDT) (Refs 1-8) techniques. By combining both reflectors and sources inside transducers with adequate phase shifts for unidirectionality, the SPUDT combines the advantages of single step manfacturing, low insertion losses and simpler tuning than multiphase transducers, but filters sizes are not small enough for pocket telephone applications. An alternate technique, resonator filters can achieve a very small size. However, the required fractional bandwidth is too large to use a temperature insensitive material such as quartz. Depending of the filter structure, this relatively wide bandwidth results in a poor rejection or may make the design impossible to attain.

We present here a new SPUDT based lowloss filter architecture which gives a larger impulse response for a given size than a classical SPUDT and so allows a more efficient filtering. This architecture was used to design a 71 MHz IF filter for GSM allowing a reduction by a factor 2 of the substrate area when compared to the classical SPUDT.

2. ARCHITECTURE DESCRIPTION

Figure 1 shows the architecture of the filter which is made of two acoustical tracks each comprisedof two SPUDTs separated by a reflection grating. From one track to the other, the offsets between the transducers and the center grating are different by a quarter wavelength.

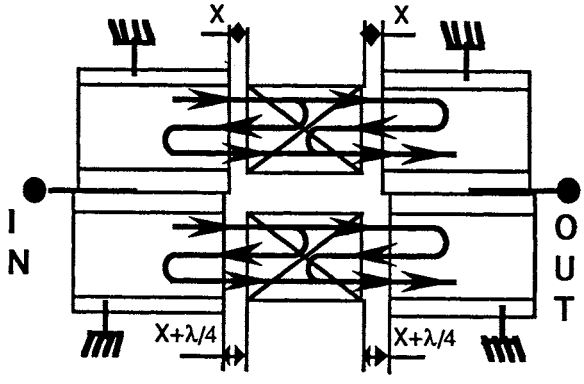


FIGURE 1 : STRUCTURE OF THE NEW FILTER

By construction, the phase difference between the two tracks is 180° for the direct path. So, there is no current generation in the load impedance and the SPUDT behaviour is similar to its short circuit condition where only the acoustical reflection is seen.

After reflection on the center grating, the waves on the 2 tracks become in phase again and give a useful electric signal after transduction through the SPUDT. Similarly, fig 1 shows there is a dual path consisting of a reflection on the center grating, a reflection on the SPUDT and a transmission through the center grating. As for the first path, the waves have opposite phase on the first SPUDT and are in phase again on the 2nd SPUDT.

These two paths add each other to generate the filtered signal. For design simplicity, we prefer to make the wave amplitudes and phases equal for the 2 paths. That leads to the following constraints :

- the center grating has to be designed symmetric
- the input and output SPUDT must have the same short circuit reflection.

With those conditions, the filtering function is given by the cascade of :

- the input and output SPUDT transductions
- the center grating reflection
- a SPUDT short circuit reflection
- transmission through the center grating

Then, the filter transfer function F(f) is at first order :

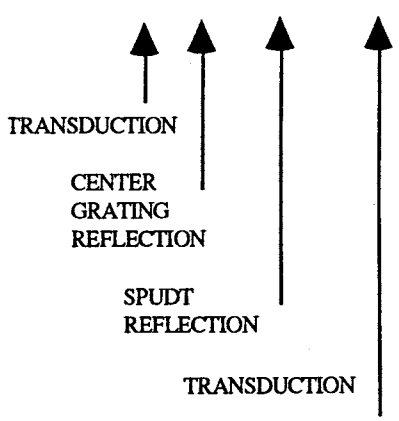
$$F(f) \cong 2\, Tr(f)^2\, T(f)\, R(f)\, R_{SPUDT}(f)$$

where :

- Tr(f) is one SPUDT transduction
- T(f) is the transmission through the center grating.
- R(f) is the reflection on the center grating
- $R_{SPUDT}(f)$ is the acoustical reflection on a SPUDT

If T_{SPUDT} is the duration corresponding to the SPUDT length and T_G is the duration corresponding to the center grating length, the whole impulse response duration is

$$T = T_{SPUDT} + 2T_G + 2T_{SPUDT} + T_{SPUDT}$$



or : $T = 4\, T_{SPUDT} + 2\, T_G$

And the duration corresponding to the substrate length is only

$$T_S = 2\, T_{SPUDT} + T_G$$

For the same substrate length, the new filter has an impulse response duration which is 2 times the impulse response of a classical SAW filter. So, using the new structure *the substrate size can potentially be reduced by a factor 2.*

Center grating reflection coefficient

If we assume the losses in the center grating are neglectible, we can define the parameter Δ by :

$$\begin{cases} |R(f)| = \sin(\Delta) \\ |T(f)| = \cos(\Delta) \end{cases}$$

To first order, the filter transfer function is then :

$$|F(f)| \cong 2|Tr(f)|^2 \cos(\Delta)\sin(\Delta)|R_{SPUDT}(f)|$$

$$|F(f)| \cong |Tr(f)|^2 |R_{SPUDT}(f)| \sin(2\Delta)$$

The filter insertion loss will be minimum when :

$$\sin(2\Delta) = 1 \text{ i.e. } \Delta = \frac{\pi}{4}$$

$$\text{or : } |R(f)| = |T(f)| = \frac{\sqrt{2}}{2}$$

The filter insertion loss is minimum when the center grating reflection (and transmission) is -3 dB at center frequency. Under this condition, the insertion loss will be slightly higher than the insertion loss for a classical SPUDT design due to the acoustical reflection on the transducers. With appropriate design, this increase in loss will be no more than 1 or 2 dB.

Suppression of Bulk modes

On quartz substrates, it is well known that bulk mode spurii are generated around 1.05 f_0, 1.6 f_0 and 1.8 f_0. To attenuate these spurii, the only known way is to machine the back side of the substrate, which is cost consumming and is only efficient for the 1.05 f_0 spurious.

The bulk modes are not very dependant on surfac conditions and the reflection on the grating at the surface is neglectible. So, for the new structure, only the direct path between transducers is involved for bulk modes. By construction, there is a λ/2 difference in separation for the tracks (the wavelength at the bulk mode frequency is the same than the SAW wavelength at f_0) and the bulk spurii on the 2 tracks cancel each other.

The only remaining spurii are due to cross talk between the 2 tracks due to diffraction or to the reflected waves on the substrate edges which do not cancel.

Balanced drive:

Integrated circuits for GSM/DCS 1800 receivers often need differential input and output drive. The new structure is well suited for balanced drive. That can be accomplished by connecting the 2 tracks in series as shown on fig 2.

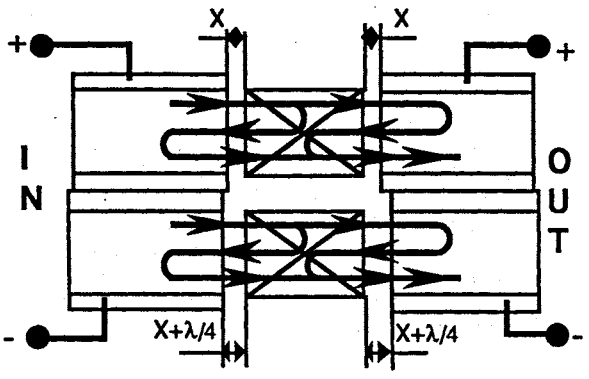


FIGURE 2 : BALANCED DRIVE STRUCTURE

3. DESIGN PRINCIPES

3.1 Design of the center grating

The center grating has to be designed for -3 dB reflection. We developed a reflector design procedure similar to that already designed for SPUDT (ref. 5). This procedure performs the minimization of the error (in the Tchebicheff sense) between the grating reflection transfer function and an ideal transfer function. A large size reduction was shown for a high reflection grating design. For gratings with reflection levels as low as -3 dB, it has been found that multiple reflections are neglectible and results are close to those of the classical Remez exchange algorithm.

The result of the first step is a function r(t) or r(x) which gives the reflection distribution in the grating. As usual for other kinds of SAW filters, the second step consists of designing the grating to obtain the desired reflection distribution. It is important to keep in mind that, due to the multiple reflections effect, the grating reflection is not linearly dependant on the reflection coefficient, that is to say that it is the absolute value of r(t) which has to be designed.

Two different methods were investigated to weight the grating. The first one is to use a position weighting technique similar to the technique used to weight metallic RACs (ref. 9). Since the influence of the center grating is neglectible far away from the center frequency, it was found preferable to use a classical withdrawal weighting technique. To maintain a constant metallization ratio along the grating, the regions without sources are filled with λ/4 period electrodes which have 0 reflection.

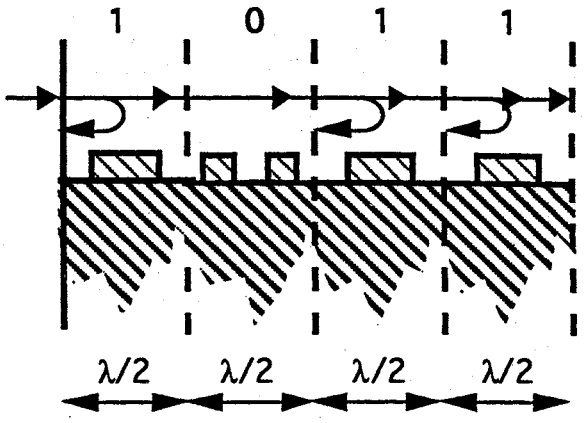


FIGURE 2 : WITHDRAWAL WEIGHTING FOR A GRATING

An important advantage of that kind of withdrawal weighting is the suppression of the 2nd harmonic. That can be explained easily by noting that the null cells in the grating have their stop-band around 2 f_0. As the number of 0s in the withdrawal sequence is generally large for good close in rejection, the grating transmission is very small at 2 f_0 and the second harmonic is rejected. Figures 3 show an example of grating transmission fo a filter a 83 MHz.

Scales : vertical 10 dB/div, horizontal 20 MHz/div

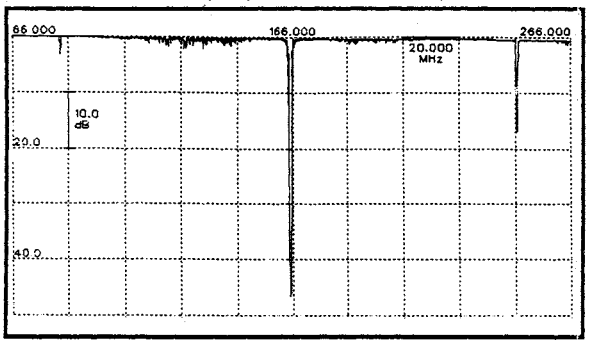


FIGURE 3 : TRANSMISSION OF A WITHDRAWAL WEIGHTED GRATING (f_0=83 MHz)

3.2 SPUDT design

One of the main criteria when designing the SPUDT is the filter insertion loss. Two parameters have a large influence.

- The SPUDT short circuit reflection has to be as large as possible because this term is a factor on the filter transfer function. In addition, the transducer directivity and by consequence the transduction efficiency are directly related to it (ref. 4). Furthermore, the SPUDT are generally electrically driven near the acoustical matching condition to obtain low triple transit. Consequently, a high short circuit reflection reduces the electrical mismatch.
- The SPUDT electrical Q (i.e. the ratio of the susceptance to the conductance) has to be as low as possible so as to reduce the insertion loss due to matching inductors as to reduce the sensitivity of the filter characteristics to the tolerances of the tuning elements. For this reason, the number of active electrodes has to be as large as possible.

Another reason is that too high source and charge impedances are often unacceptable for the filter user and it is always preferred to keep small acoustical apertures to reduce the size.

The close in transfer function shape is given mainly by the shapes of the SPUDT and center grating reflection transfer functions. The transduction affects far out frequencies. Due to the structure symmetry, the two SPUDTs must have the same short circuit reflection. Then, if possible, it is always preferred for small sizes to perform the larger part of filtering in the center grating.

This shows that a trade off must be made between size and insertion loss.

For low insertion loss and easy tuning, SPUDTs have to be as large as possible.

For small sizes, SPUDTs have to be as small as possible.

To decouple these two effects it is important to choose an internal SPUDT structure which combines low electrical Q an high reflectivity. Various structures were proposed in the past (refs. 1,2,3,4,8). The EWC structure first proposed by HARTMANN (ref. 4) allows the design of transducers with both one $\lambda/4$ width reflective electrode and one active electrode per wavelength. The only problem is that the phase shift between transduction and reflection centers of the generic structure is not optimum for directivity. This can be corrected by minor modifications.

4. EXPERIMENTAL RESULTS

Amplitude reference : -8 dB
Scales : vertical 10 dB/div, horizontal 0.4 MHz/div

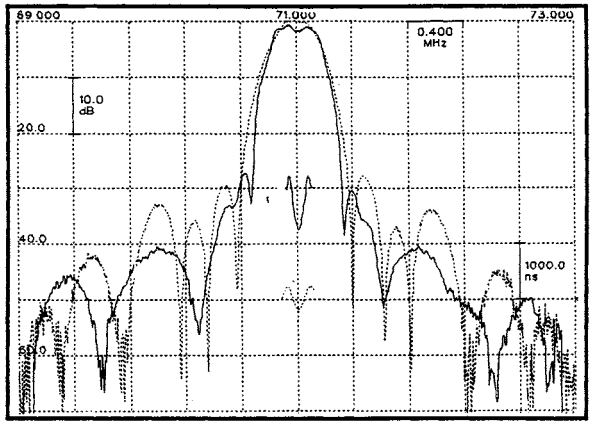


FIGURE 4 : COMPARISON OF THE NEW FILTER (FULL LINE) AND THE CLASSICAL SPUDT (DASHED LINE) AT 71 MHZ

We used our new structure to design a 71 MHz filter for the GSM mobile radio application. The -3 dB bandwidth was specified to be around +/-100 kHz, the -10 dB less than +/-300 kHz the rejection 40 dB and the insertion loss less than 10 dB. These characteristics are equivalent to those of an existing classical SPUDT filter. This filter is housed in a DIP 14 package (internal size 20 X 5 mm^2). The aim was to reduce the size of this filter down to a SIP 5 package (internal size 13 X 3 mm^2) with equivalent performances. Figure 4 shows a comparison of the previous filter and the new one. It appears that althrough it is smaller in size,the new filter has a better shape factor that the old one. As expected, a moderate insertion loss increase (about 1 dB) is obtained. The only difficult point is the group delay ripple (due to the larger delay of the new filter) but it is still acceptable for GSM application and was predicted by simulations.

Figure 5 shows plots for a 83 MHz filter which was designed using the same technique but with the balanced drive structure. This filter has a similar -3 dB bandwidth to the 71 MHz one but is narrower (+/- 300 kHz at -20 dB) and has a greater rejection. Its insertion loss is about 8 dB. It is housed in a DIP 14 package while DIP 16 was necessary for a classical SPUDT design. The wideband plot (fig 7) exhibits a 90 dB ultimate rejection with all spurii below -70 dB.

Amplitude reference :- 8dB
Scales : vertical 10 dB/div, horizontal 0.8 MHz/div

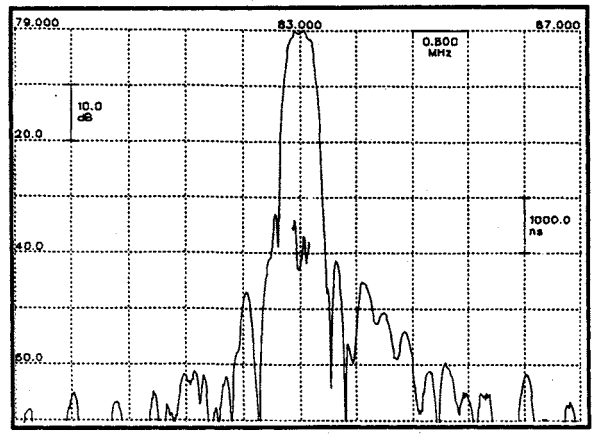


FIGURE 5: PLOT OF A 83 MHz BALANCED DRIVE FILTER

Scales : vertical 10 dB/div, horizontal 20 MHz/div

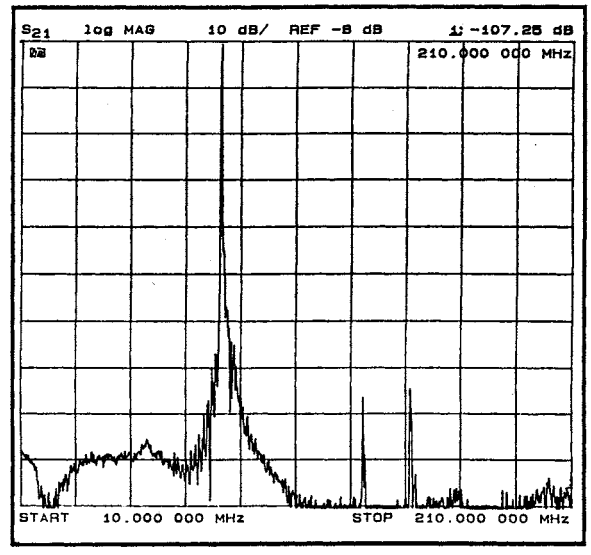


FIGURE 6 : WIDE BAND PLOT OF THE 83 MHz FILTER

5.-CONCLUSION

We developed an original SPUDT based structure to solve the mobil radio low loss filters size reduction problem. This technique uses 2 nearly identical acoustical tracks with

proper separations in order to fold the acoustical paths and leads to an impulse response duration twice that of a classical SPUDT. Furthermore, this technique allows natural suppression of bulk modes and 2nd harmonic. Applying it to a 71 MHz filter design for mobile GSM applications led to a size reduction by a factor of 2 for equivalent and even better performances. As a second example of our design techniques possibilities, a 83 MHz filter with all spurii below -70 dB was presented.

REFERENCES

[1] K. HANNA, B. HUNSINGER, "A Triple Transit Suppression Technique", 1976 IEEE U.S proc., p328

[2] M. LEWIS, "Low loss SAW Devices employing Single Stage Fabrication", 1983 IEEE U.S. proc., p104

[3] T. KODAMA et al. , "Design of Low Loss SAW Filters employing Distributed Acoustic Reflection Transducers", 1986 IEEE U.S. proc., p59

[4] C.S. HARTMANN and B.P. Abbott, "Overview of Design Challenges for Single Phase Unidirectionnal SAW Filters", 1989 IEEE U.S. proc., p 79

[5] P. VENTURA, M. SOLAL, P. DUFILIE, J. DESBOIS, M. DOISY, JM. HODE, "Synthesis of SPUDT Filters with Simultaneous Reflection and Transduction Optimization", 1992 IEEE U.S. proc.

[6] R.B. BROWN, S. GAPANI, "Apodized Single Phase Unidirectional Transducer SAW Devices", 1991 IEEE U.S. proc., p 231

[7] C. KAPPACHER et al., "Design and Analysis of Single Phase Unidirectional Transducers", 1991 IEEE U.S. proc., p 1

[8] P.V. WRIGH, " A Uniformly Sampled Single Phase Unidirectional Transducer with Sub -Nyquist Spatial Sampling", 1992 IEEE U.S. proc.

[9] G. RIHA et al., "RAC Filters with Position Weighted Metallic Strips", 1982 IEEE U.S. proc.,pp 83-87

[10] G. SCHOLL et al., "New Resonator Filter with High Sidelobe Suppression", 1992 IEEE U.S. proc.

THE MODELING OF HEAVY ELECTRODE GRATINGS FOR NARROW BAND SAW DEVICES.

VICTOR P. PLESSKY, THOR THORVALDSSON

Advanced SAW Products SA, CH-2022 Bevaix, Switzerland

ABSTRACT

Dispersion curves and attenuation coefficients for Rayleigh and leaky surface acoustic waves (SAW) propagating in a periodic system of thin electrodes on a piezoelectric crystal surface are calculated. The case of finite thicknesses electrodes is also considered for leaky waves on 36°YX-$LiTaO_3$ substrates. Such basic parameters as stopband width, stopband center frequency and attenuation coefficients of the wave for different metal film thicknesses are found. The analysis includes the bulk wave radiation and scattering. A simple closed form formula for the dispersion curve is proposed. Numerical examples are given for Rayleigh waves on YZ-LiNbO3 and leaky waves on 36°YX-$LiTaO_3$.

Keywords: *SAW, Leaky waves, mechanical perturbation, dispersion curves.*

1. INTRODUCTION

We consider here the propagation of Bleustein-Gulyaev (BG) type surface acoustic waves in periodic structures of metal electrodes of finite thickness. The main idea of this paper is to include mechanical perturbations in an earlier developed approach [1],[2]. The presence of the electrodes on the surface of a piezoelectric substrate influence the propagation of SAWs both through the perturbation of the electric fields and additional mechanical stresses on the surface. Physically there is an essential difference between these perturbations. The perturbations of electric fields are always strong in a sense that the structure of the fields is changed radically by the presence of metal electrodes, the charge density distribution being determined mainly by electrostatic interactions and does not follow basical harmonic dependence of the mechanical components of the wave. However since the electromechanical interaction is usually small this radical reconstruction of the electric fields near electrodes results in relatively small reflection coefficients. On the contrary, mechanical perturbations are always small if the relative thickness of the electrodes is small $(h/\lambda << 1)$. That means that mechanical displacements in the wave remain approximately the same as on free surface and the reflection coefficients are determined by the perturbation parameter $(h/\lambda << 1)$. From a mathematical point of view, these differences mean that while analyzing wave propagation we must retain a large number of harmonics of the electric field but only a few harmonics of the mechanical components are sufficient. This approach is close to the one recently developed by Danicki [3]- [5].

Since surface transverse waves were discovered simultaneously and independently by two groups [6,7], efforts were made to develop a theory of STW propagation and reflection in a periodic structure at frequency ranges corresponding to the Bragg stopband [8]-[10]. These theories can also include the case of Bleustein-Gulyaev waves (BGW), since the spatial structure of the waves is basically the same. The theory, based on numeric solutions of the equations of coupled harmonics was reported some time ago [8]. For these waves the localization depth is strongly frequency dependent inside the stopband, being minimal at the lower edge of the stopband and increasing at the upper edge. In the case of a strong periodic perturbation the stopband frequency region can merge with the frequency range in which bulk wave scattering takes place. Similar behavior is characteristic of leaky waves in 36°YX-$LiTaO_3$ [1],[2]. The theories based on numeric solutions of the equations coupling Fourier harmonics of the electric and acoustic fields of the waves propagating in periodic systems of electrodes, including the theory described here, are computationally intensive and can not be used directly in design software. For a workable design software, simple closed form solutions, like COM, are desirable. In this paper it is shown that by introducing non critical simplifications, the same as those usually made in Coupling of Mode (COM) approaches, we can obtain a system of equations for coupled space harmonics of propagating waves which have an explicit analytic solution.

2. FORMULATION OF THE PROBLEM

We consider a piezoelectric half space occupying the region $z<0$ and the periodic system of metal electrodes is placed on its surface parallel to y-axis. The electrodes are ideally conducting, have finite thickness h and period of the structure is p, $(h/p << 1)$. The waves of Bleustein-Gulyaev (BG) type are propagating along x-axis and have the only component of the mechanical displacement $\vec{u} = (0,u,0)$ along y-axis. The wavelength λ is supposed to be close to $2p$, that is we consider the frequencies equal or close to the first Bragg stopband frequency. Our aim is to find the main characteristics of the propagation: dispersion curves, the

stopband frequencies, reflection (coupling) coefficients for opposite propagating waves .

Omitting mechanical harmonics with numbers other than "0" or "-1" we ignore the mechanical part of the so called "storage energy effect", which is proportional to $(h/p)^2$ and is small in comparison with the first order uniform mass loading effect (and $\Delta V/V$ effect) included in the analysis. Numerically obtained dispersion curves are approximated by simple closed form dispersion equation, accounting both for Bragg reflection of the waves in grating and bulk wave scattering at frequencies higher than the stopband.

3. THE MAIN EQUATIONS

3.1. GREEN'S MATRIX

All wave components at the surface $z=0$ are represented in the form of Bloch waves in accordance with the Floquet theorem: $a(x)=a_0\exp(jqx)F(x)$ where $F(x)$ is a periodic function of x with the period p. For charge density distribution, we write this equation in the form:

$$\sigma(x)=\sigma_0\exp(j\beta x)\sum_{n=-\infty}^{+\infty}f(x-np) \qquad (1)$$

where the function $f(x)$ is non zero only within one period of the structure and σ_0 is an amplitude factor. That allows us to introduce the unknown values of charge density only on one period [1],[2]. Now using the equivalence

$$\sum_{n=-\infty}^{+\infty}f(x-np)=\sum_{m=-\infty}^{+\infty}\Pi_m\exp(jmQp) \qquad (2)$$

we can rewrite the equation (1) in the form:

$$\sigma(x,\beta)=\exp(j\beta x)\sum_{m=-\infty}^{+\infty}\Pi_m\exp(jmQx)= \\ =\sum_{m=-\infty}^{+\infty}\Pi_m\exp(j\beta_m x) \qquad (3)$$

where β is the wavenumber of the wave composed of the set of harmonics with the wavenumbers $\beta_m=\beta+mQ$,

$$\Pi_m=\frac{1}{p}\int_{-l/2}^{+l/2}f(x)\exp(-jmQx)dx$$

for the mechanical stresses arising due to perturbations we use only two main harmonics :

$$T(x)=T_0\exp(j\beta x)+T_{-1}\exp(j(\beta-Q)x) \qquad (4)$$

corresponding to the two opposite propagating mutually reflecting waves. In the above formulas $Q=2\pi/p$ is the "wavenumber" of the grating of electrodes.

The charges and the stresses are the sources of the electric fields and the mechanical displacements in a substrate.

For a piezoelectric halfspace the corresponding relations are given by the Green's matrix [3],[11],[12], also called the dyadic Green's function [13] :

$$\varphi(q,\omega)=G^{E,E}(q,\omega)\sigma(q,\omega)+G^{E,M}(q,\omega)T(q,\omega) \qquad (5a)$$

$$u(q,\omega)=G^{M,E}(q,\omega)\sigma(q,\omega)+G^{M,M}(q,\omega)T(q,\omega) \qquad (5b)$$

valid due to the linearity of the media for each separate Fourier component. For BG-waves, closed form expressions for Green's matrix components are known [11],[12]. In some other important cases, like for leaky waves in $36^0-LiTaO_3$, the formulas (5) look to be good approximation. For this cut we have used numerically calculated curves of the Green's matrix components.

The charge distribution on each electrode is described by the function:

$$f(x)=\frac{\sum_{n=0}^{N}A_nCh_n(2x/l)}{\sqrt{1-(2x/l)^2}} \qquad (6)$$

where $Ch_n(\varsigma)$ are Chebyshev polynomials and A_n corresponding amplitudes . The function (6) reflects correctly the peculiarities of the charge distribution at the edges of the electrodes. It is known that only a few polynomials are sufficient ($N=3\div5$) for an accurate description [14]. In the case when the charges are introduced by subdivision of the electrode into substrips essentially more unknown values need to be used to obtain the same accuracy. The amplitudes of the Fourier harmonics Π_m is given in this case by the following expression:

$$\Pi_m=\frac{l}{2p}\sum_{n=0}^{N}A_nI_{nm} \qquad (7)$$

where

$$I_{nm}=(-j)^n\pi J_n(\pi\frac{l}{p}m) \qquad (8)$$

and $J_n(\varsigma)$ is Bessel function of rank n of first kind.

If we could ignore as previously [1], [2] the mechanical perturbations , ($T_n=0$), than substituting (3) and (7) into (5a) we have

$$\varphi(x)=\exp(j\beta x)\frac{l}{2p}\sum_{n=0}^{N}A_nC_n(x) \qquad (9)$$

with

$$C_n=\sum_{m=-\infty}^{+\infty}G^{E,E}(\beta_m)I_{nm}\exp(jmQx). \qquad (10)$$

The equation (9) gives the value of the potential in any point x. Supplied with corresponding boundary conditions (see section 3.3) for potential in $n+1$ points on the electrode it gives the closed system of equations for determining the coefficients A_n and, hence , all other characteristics of propagation of the wave.

3.2. MECHANICAL PERTURBATIONS

Adding two harmonics of mechanical stress instead of (9) we get:

$$\varphi(x) = \exp(j\beta x)\frac{l}{2p}\sum_{n=0}^{N} A_n C_n(x) + G^{EM}(\beta)\exp(j\beta x)T_0 + \\ + G^{EM}(\beta - Q)\exp(j(\beta - Q)x)T_{-1}. \tag{11}$$

Now we must take into account also mechanical displacements u_0 and u_{-1}, coupled by the equations (5) to the corresponding harmonics of charge *and stress* , which are in turn created by the perturbations of the mechanical boundary conditions. Using equation (5b) we have

$$u_0 = G^{ME}(\beta)\Pi_0(q,\omega) + G^{MM}(\beta)T_0(q,\omega) \tag{12a}$$

$$u_{-1} = G^{ME}(\beta - Q)\Pi_{-1}(q,\omega) + G^{MM}(\beta - Q)T_{-1}(q,\omega) \tag{12b}$$

Of course, if we have no mechanical perturbations ($T_n = 0$), the displacements u_n can be found in the same way, but previously [1] ,[2] we had no need to determine it, since the boundary conditions included in that case were only electric values.

3.3. Boundary conditions

3.3.1. Mechanical boundary conditions

We write down the mechanical boundary conditions in a form close to that used by Danicki [3], [5] retaining terms responsible for uniform mechanical loading:

$$T_0 = g_0 u_0 + g_{-1} u_{-1} \tag{13a}$$

$$T_{-1} = g_{-1}^{*} u_0 + g_0 u_{-1} \tag{13b}$$

here coefficients $g_{0,-1}$ are proportional to a small parameter of mechanical perturbation h/p

$$g_0 = -\frac{h}{2}(\rho\omega^2 - \mu\beta^2) \tag{14a}$$

$$g_{-1} = -\frac{h}{\pi}(\rho\omega^2 - \mu\beta(\beta - Q)) \tag{14b}$$

for $l/p = 0.5$. For arbitrary l/p ratio the following formulas can be used

$$g_0 = -h\frac{l}{p}(\rho\omega^2 - \mu\beta^2) \tag{15a}$$

$$g_{-1} = -h\frac{\sin(\pi l/p)}{\pi}(\rho\omega^2 - \mu\beta(\beta - Q)). \tag{15b}$$

3.3.2. Electrical boundary conditions

In the case of *shorted electrodes* the potentials at all points on the electrodes must be put equal to zero . One can take $N+1$ points uniformly distributed over the electrode of the width l with coordinates

$$x_k = l\frac{2k - N - 2}{2(N+1)} \tag{16}$$

then the boundary conditions will be written in the form

$$0 = \exp(j\beta x_k)\frac{l}{2p}\sum_{n=0}^{N} A_n C_n(x_k) + \\ + G^{EM}(\beta)\exp(j\beta x_k)T_0 + \\ + G^{EM}(\beta - Q)\exp(j(\beta - Q)x_k)T_{-1}. \tag{17}$$

which generates the system of $k \in 1,\ldots,N+1$ equations (the factor $\exp(j\beta x_k)$ can be omitted here). Together with the equations (12) and (14) we get a complete set of $N+5$ equations for determining $N+1$ amplitudes A_n and 4 mechanical amplitudes $u_{0,-1}$, $T_{0,-1}$.

For *open electrodes* we must put some constant non zero value φ_0 in left hand side of equation (17). This value is one more unknown amplitude. So we need one more equation which is given by the condition of electroneutrality:

$$0 = \sum_{n=0}^{N} A_n (j)^n J_n(\beta l/2) \tag{18}$$

Let us note that the potential remains constant on each of the open electrodes, but is of course shifted by the factor $\exp(j\beta p)$ from one electrode to its neighbor in accordance with the Floquet theorem.

4. DISCUSSION

The procedure of numerical solution used here was discussed previously [1,2]. For Green's matrix components we used numerically obtained data which were extrapolated also for complex-valued wavenumbers of the wave to account for Bragg reflection and attenuation of the waves propagating in the structure. From a mathematical point of view, in all cases a system of linear homogeneous equations is obtained. To get a nontrivial solution (to find the eigenmodes of the system) the determinant must be set to zero. Thus an equation for the wavenumber β is obtained which is generally complex. In numerical calculations β is normalized as follows $\xi = \frac{\beta}{Q/2} = p\beta/\pi$, with $\xi = 1$ for the first Bragg stopband. The normalized frequency is $\varpi = \omega/\omega_B = p\omega/(\pi v_0)$, where $\omega_B = v_0(Q/2)$ is the "Bragg frequency" corresponding to negligible periodic perturbation of the wave with unchanged velocity v_0. In coordinates (ϖ, ξ) the line $\xi = \varpi$ corresponds to a wave with velocity v_0 point (1,1) being the center of the Bragg stopband (edge of the first Brillouin zone) for extremely weak periodic perturbation of the surface.

Fig. 1 illustrates the behavior of the dispersion curves for Rayleigh waves in shorted (*sh*) and open (*op*) periodic systems of thin metal electrodes with width l of electrodes equal to half of the period $l = p/2$. In accordance with previous results, the stopband widths obtained were almost equal:

$(2\Delta\omega/\omega)_{shorted} = 1.04\%$ and
$(2\Delta\omega/\omega)_{open} = 0.99\%$.

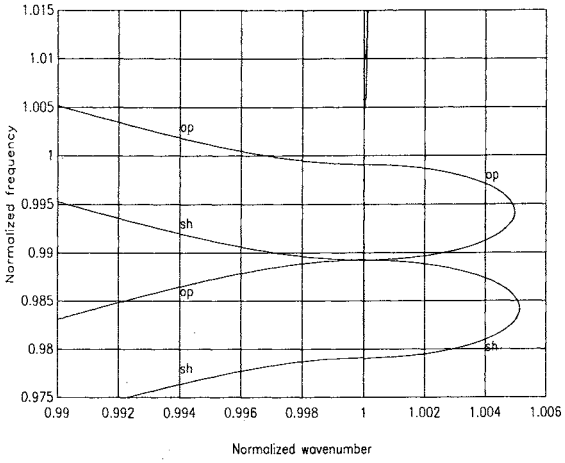


Fig. 1 Dispersion curves for YZ-$LiNbO_3$ (*op*=open electrodes, *sh*=shorted electrodes).

This corresponds to a coupling per wavelength, in the COM model , of

$$\kappa\lambda = \pi(\frac{2\Delta\omega}{\omega})\frac{v_0}{v_{stb}} = 3.32\% \text{ (sh) or } 3.13\text{(op)}$$

with v_{stb} as the SAW velocity in the center of the stopband. The upper edge of the stopband for shorted electrodes coincides with the lower edge of the stopband for the case of open electrodes. The figures also show that the attenuation due to bulk wave scattering starts at different frequencies for shorted and open electrodes, which is explained by different slowing of the SAW's in these cases.

Figure 2 shows analogous curves for leaky waves on 36°YX-$LiTaO_3$ in the case of thin electrodes. The leaky SAW velocity dependence on frequency on 36°YX-$LiTaO_3$ is shown in Fig.3. Although the curves on Fig.2 look similar to the ones on the Fig. 1 a few essential differences should be stressed.

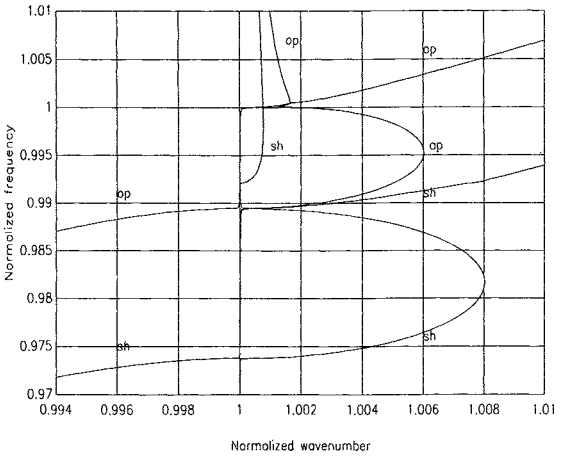


Fig. 2 Dispersion curves for 36-YX $LiTaO_3$ (*op*=open electrodes, *sh*=shorted electrodes)

First of all, the width of stopbands are different for shorted $(2\Delta\omega/\omega)_{shorted}$=1.56% and open (1.11%) electrodes. These numbers are a little higher than published earlier [15].

Secondly, the bulk wave scattering starts at frequencies only slightly higher than the upper edge of the stopband for shorted electrodes, and immediately near the edge of the stopband for open electrodes. This effect for BG-waves was discussed some time ago [8,9]. This attenuation is larger than in the case of YZ-$LiNbO_3$, reaching $0.1\ dB/\lambda$ in the right hand side of the stopband for open electrodes.

Fig.4 illustrates the influence of mechanical perturbations on the width of the stopband and its center frequency. With 3% relative thickness of Al electrodes the effectiveness of the coupling of opposite propagating waves on the grating is increased almost two times.

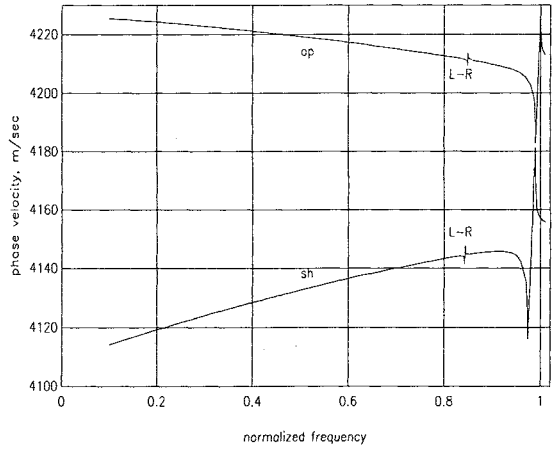


Fig. 3 Phase velocity of the leaky waves in periodic structure of electrodes; 36-YX-$LiTaO_3$. (*op*=open electrodes, *sh*=shorted electrodes. L-R stands for leaky-Rayleigh stopband).

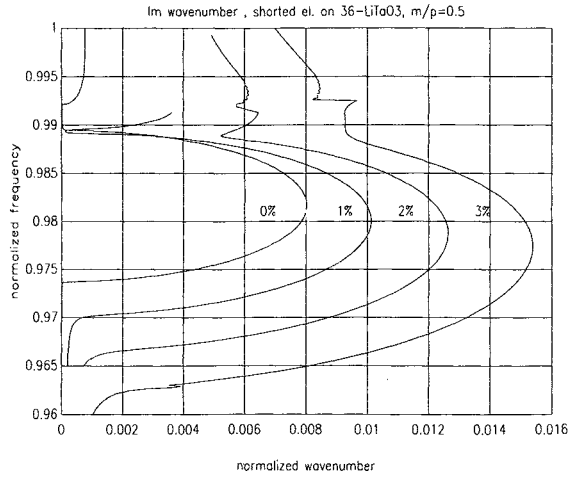


Fig. 4 Dispersion curves for 36°YX $LiTaO_3$

4.1. COM approach

In the framwork of the COM approach only two main harmonics are retained. It is supposed that the waves can be described by the sum of harmonics satisfying the Floquet's theorem:

$$u = u_+ \exp(\kappa_+ y + j(Q/2+\delta)x) + u_- \exp(\kappa_- y + j(-Q/2+\delta)x) + \ldots \quad (19)$$

In formula (19) u_+ and u_- stand respectively for the amplitudes of the waves propagating along the x-axis and in the opposite direction. The wavenumber of the wave $\beta = Q/2 + \delta$, $|\delta| << Q/2$, (where $Q = 2\pi/p$) is supposed to be almost equal to the Bragg wavenumber $Q/2 = \pi/p$, the deviation δ can be complex-valued, and the frequency of the wave is close to the Bragg frequency $\omega = \omega_b + \Delta\omega$, $\Delta\omega << \omega_b$ $\omega_b = V(Q/2)$, where V is the velocity of the wave in the limit of very weak perturbation. The decay constants $\kappa_\pm$ describe the localization depth of the waves near to the surface, ($\mathrm{Re}\,\kappa_\pm > 0$). It is assumed that, as for BG-waves in hexagonal crystals,

$$\kappa_\pm = \sqrt{(Q/2 \pm \delta)^2 - k_t^2} \quad (20)$$

where $k_t = \omega / V$ is the wavenumber of the bulk shear wave.

For this problem the corresponding Coupling-of-Modes equations including electric and mechanical perturbations of the boundary were derived earlier [8] in the following form:

$$\left(\kappa_+ - \eta_0(Q/2+\delta) - 2\varepsilon_0 \frac{k_t^2}{Q}\right) u_+ = \varepsilon_1 \frac{2k_t^2}{Q} u_- \quad (21a)$$

$$\left(\kappa_- - \eta_0(Q/2-\delta) - 2\varepsilon_0 \frac{k_t^2}{Q}\right) u_- = \varepsilon_1^* \frac{2k_t^2}{Q} u_+ \quad (21b)$$

where η_0 is the piezoelectric coupling constant, ε_0 is a parameter of the uniform mass loading and ε_1 is a parameter of a periodic boundary perturbation. For a narrow frequency range including the Bragg stopband, the frequency and wavenumber dependence in the terms proportional to the small parameters η, $\varepsilon_{0,1}$ can be neglected. This allows the equations (21) to be rewritten in a simplified form (see also [10]):

$$(\kappa_+ - \eta(Q/2))u_+ = \varepsilon(Q/2)u_- \quad (22a)$$

$$(\kappa_- - \eta(Q/2))u_- = \varepsilon^*(Q/2)u_+ \quad (22b)$$

or, by introducing a normalized wavenumber deviation q and a normalized frequency shift Δ,

$$\xi = (Q/2+\delta)/(Q/2) = 1+q$$

$$\Delta = (\Delta\omega / \omega_b)$$

the following is obtained:

$$(\sqrt{2(q-\Delta)} - \eta)u_+ = \varepsilon u_- \quad (23a)$$

$$(\sqrt{2(-q-\Delta)} - \eta)u_- = \varepsilon^* u_+. \quad (23b)$$

here, as well as in (21) and in (22), ε^* is complex conjugate of ε. As a further simplification second order terms, proportional to Δ^2, q^2 have been omitted. The first parameter η in (23) is responsible for the spatial structure (i.e. the depth of localization) of the wave and its velocity. The second parameter (ε) describes the coupling strength of the opposite propagating waves.

From the system (23) the dispersion equation is obtained:

$$\left(\sqrt{2(q-\Delta)} - \eta\right)\left(\sqrt{2(-q-\Delta)} - \eta\right) = |\varepsilon|^2. \quad (24)$$

This equation can be solved analytically. But first, the particular frequency points at which the wavenumber is equal to the Bragg wavenumber $Q/2$ and its imaginary part is equal to zero can be found. From equation (24), putting $q = 0$, it follows

$$\sqrt{-2\Delta} = \eta \pm |\varepsilon| \quad (25)$$

These frequency points look like the edges of the stopband [8], [16]:

$$\Delta_- = -\frac{1}{2}(\eta + |\varepsilon|)^2, \qquad \Delta_+ = -\frac{1}{2}(\eta - |\varepsilon|)^2 \quad (26)$$

where Δ_- stands for the lower stopband edge and Δ_+ for the upper one. It can be shown that for Δ_+ the statement is valid only for $|\varepsilon| < 0.5\eta$. From (25) it can be seen that if $|\varepsilon| > \eta$ the second edge of the stopband cannot be separated from the frequencies at which the bulk-wave scattering takes place. More detailed analysis shows that it happens already if $|\varepsilon| > 0.5\eta$. For $|\varepsilon| < 0.5\eta$ the center frequency of the stopband

$$\omega_c = \frac{1}{2}(\Delta_- + \Delta_+) = -\frac{\eta^2 + |\varepsilon|^2}{2} \quad (27)$$

is shifted down both due to the uniform loading (parameter η) and as a result of the interaction of the waves (parameter ε). The stopband width is determined by both the localization depth of the waves ($\propto \eta^{-1}$) and by the coupling coefficient ε and is given (for $|\varepsilon| < 0.5\eta$) by:

$$\Omega = (\Delta_+ - \Delta_-) = 2\eta|\varepsilon| \quad (28)$$

The exact analytical solution of equation (24) can be written in the form:

$$q = \sqrt{(\Delta)^2 - \frac{1}{4}\left(|\varepsilon|^2 \pm \eta\sqrt{2|\varepsilon|^2 - \eta^2 - 4\Delta}\right)^2} \quad (29)$$

Analysis of the solutions (29) shows that in most cases physical solutions correspond to the sign "+". The stopband range of frequencies begins at the frequency: $\Delta = \Delta_-$ and has the upper edge either at the frequency $\Delta = \Delta_+ < \Delta_b$, if $|\varepsilon| < 0.5\eta$ (where $\Delta_b = -\frac{\eta^2 - 2\varepsilon^2}{4}$) or at the frequency Δ_b if $|\varepsilon| > 0.5\eta$. In the latter case the stopband is already not separated from the frequency range which corresponds to the scattering of the bulk waves.

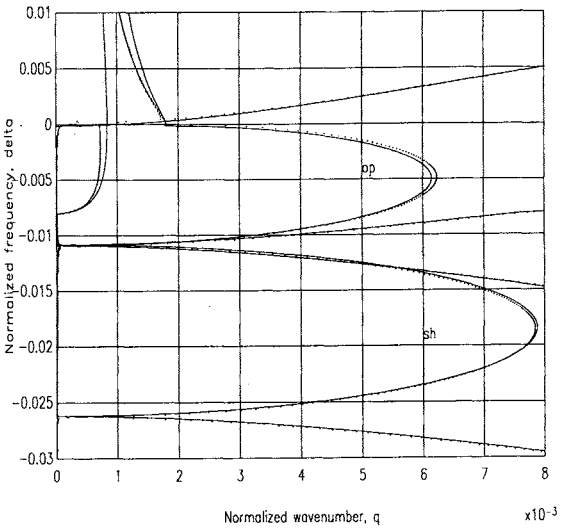


Fig. 5 Dispersion curves for 36-YX-$LiTaO_3$ (dotted:numerical, solid: formula (29); *op*=open electrodes, *sh*=shorted electrodes).

Figure 5 illustrates the dispersion curves calculated above numerically for shorted and open thin metal electrode structures on 36°YX-$LiTaO_3$ and the fitted analytic solutions (29). It is seen that the agreement is quite satisfactory. The analytic formula gives a slightly smaller magnitude of attenuation which results from bulk wave scattering.

5. CONCLUSIONS

An effective technique for the determination of the propagation characteristics of leaky SAW in strong piezoelectrics having a periodic system of electrodes on the surface has been developed. The material is fully characterized by its Green's matrix components which are calculated only once and then are used as known functions, extrapolated into the complex plane near the real wavenumber axis. Dispersion curves and attenuation frequency dependence have been studied in detail for Rayleigh waves in YZ-$LiNbO_3$ and leaky waves in 36°YX-$LiTaO_3$, especially near the stopband regions. For the leaky wave case the bulk wave scattering and strong attenuation of the wave takes place at frequencies only slightly higher than the upper edge of the stopband for shorted electrodes and immediately near the edge of the stopband for open electrodes. A relatively weak coupling between leaky and Rayleigh SAW (L-R) results in another narrow (0.13%) stopband at a frequency 16% lower than the main stopband for leaky waves. Also an extremely narrow (0.01%) stopband for Rayleigh waves (R-R) exists at a frequency $\varpi \approx 0.74$ due to Raleigh waves reflection on the grating. The mechanical load introduced by thick electrodes can essentially increase Bragg coupling of opposite propagating leaky waves. A simple analytic dispersion equation for STW, BG-waves and quasi shear leaky waves is reported for the first time, accounting both for reflection and scattering of the waves on the gratings,

6. ACKNOWLEDGMENTS

This work was supported by Swiss PTT, which is gratefully acknowledged. The authors are grateful to Dr. J. Heighway for his interest, help and permanent support in this work. We also wish to thank C. S. Hartmann, Prof. D. P. Chen, Dr. D. P. Morgan and P. Durrant for fruitful discussions and Dr. B. P. Abbott for software for Green's matrix calculations.

7. REFERENCES.

[1] V. P. Plessky, T. Thorvaldsson, "Rayleigh and leaky SAW's in periodic systems of electrodes: periodic Green's functions analysis", *Electronics Letters*, 1992, Vol. 28, No. 14, pp. 1317-1319.

[2] V. P. Plessky, T. Thorvaldsson, "Rayleigh waves and leaky SAW in periodic systems of electrodes: periodic Green's function analysis", *Proceedings of IEEE Ultrasonics Symposium*, Tucson, 1992

[3] Eugene J. Danicki , "Propagation of transverse surface acoustic waves in rotated Y-cut quartz substrates under heavy periodic metal electrodes", *IEEE Trans. on Sonics and Ultrasonics*, Vol., SU-30, No. 5, 1983, pp. 304-312.

[4] E. Danicki, "Spectral theory of natural unidirectionality of SPUDT for B-G wave", *Proceedings of IEEE Ultrasonics Symposium*, 1990, pp. 113-116.

[5] E. Danicki, "Spectral theory of NSPUDT - Rayleigh wave case", *Proceedings of IEEE Ultrasonics Symposium*, 1991, pp. 145-148.

[6] B. A. Auld, J. J. Gagnepain, and M. Tan, "Horizontal shear waves on corrugated surfaces", *Electron. Lett.*, **12** (24), pp. 650-651, (25 Nov. 1976).

[7] Yu. V. Gulyaev and V. P. Plesskii, "Slow" acoustic surface waves in solids, *Sov. Tech. Phys. Lett.*, Vol.**3**, No. 3, pp. 87-88, March 1977, (submitted Nov. 25,1976).

[8] Yu. V. Gulyaev, T. N. Kurach, V. P. Plesski, and V. P. Tereshkov, *Akust. Zh.* **30**, 47, 1984 (in Russian), *Sov. Phys. Acoust.* **30**, 27 1984 (English translation) .

[9] Yu. V. Gulyaev and V. P. Plesskii, "Propagation of acoustic surface waves in periodic structures", *Sov. Phys. Usp.* **32**(1), pp. 51-74, Jan. 1989.

[10] E. Bigler, B. A. Auld, E. Ritz, and E. Sang, " An analysis of the influence of design parameters on the resonant frequency and Q-factor of surface transverse wave (STW) resonators", *Proc. 45th Ann. Frequency Control Symp.*, Los Angeles, pp. 222-229 ,June 1991, (see here also Ref. at earlier papers of B. A. Auld and co-workers).

[11] M. K. Balakirev, I. A. Gilinskii, *Waves in piezoelectrics*, Nauka, Novosibirsk, 1982, p.239, (in Russian).

[12] Wang Chenghao and Chen Dongpei, "Generalized Green's functions of surface excitation of elastic wave fields in a piezoelectric half-space", *Chinese Journal of Acoustics*, 1985, Vol. 4, No. 4, pp. 297-313.

[13] A. R. Baghai-Wadji, H. Reichinger, H. Zidek and Ch. Mecklenbräuker, "Green's function applications in SAW devices, invited", *Proceedings IEEE Ultrasonics Symposium*, 1991, pp. 11-20.

[14] Ken-ya Hashimoto, Masatsune Yamaguchi, "Derivation of coupling-of-modes parameters for SAW device analysis by means of boundary element method", *Proceedings of IEEE Ultrasonics Symposium*, 1991, pp. 21-26.

[15] K.Y. Hashimoto, M. Yamagnchi, H. Kogo, "Interaction of high-coupling leaky SAW with bulk waves under metallic grating structure on 36°YX $LiTaO_3$", *Proceedings of IEEE Ultrasonics Symposium* 1985., pp.16-21.

[16] B. I. Boyanov, and V. L. Strashilov, "Transmission and Reflection Coefficients for Surface Transverse Waves Propagating under Shorted Metal Strip gratings", *IEEE Trans. Ultrason. , Ferroelec. Freq. Contr.*, Jan. 1992, Vol. 39, No. 1, pp. 119-121.

THE DESIGN AND DEVELOPMENT OF A 950 MHz LOW LOSS SAW COUPLED RESONATOR FILTER FOR A MOBILE COMMUNICATION RECEIVER APPLICATION

M.-A.Schwab and P.A.Durrant

Advanced SAW Products SA, CH-2022 Bevaix, Switzerland

ABSTRACT

The increasing demand for miniature low loss filters for use in the front end of mobile communication systems has led to significant development in the area of SAW coupled resonator filters which are capable of providing the combined advantages of excellent selectivity and very low loss (< 4 dB) in an extremely small package outline.

This paper describes the design procedures and development of such a filter for use in the RF circuit of an NMT (Norsk Mobile Telephone) mobile communication receiver.

The filter was designed using the coupling of modes approach [1] combined with rigorous analysis of the grating reflection coefficient using the periodic Green's Function model including mechanical perturbations [2] for leaky waves substrates.

Comparison of experimental results with theoretical simulations show excellent agreement.

1. INTRODUCTION

The device under development was intended for use as an RF interstage filter to enhance the rejection of the transmitter signal and image frequency spurious.

The use of SAW coupled resonator techniques for RF filtering applications has become more widespread with recent developments in leaky surface wave materials [3]. The high acoustic coupling coefficient of these materials (k^2>10%) has enabled the development of wideband filters (up to 4%) with the low insertion loss (<4 dB) required in the RF circuits in modern mobile communications systems.

An extension of new developments of the Coupling of Modes model (COM) [1] combined with accurate measurements of material parameters was used to synthesize filters using compensatory transducer designs to extend the filter bandwidth.

The purpose of this approach is to allow full compliance of the filter with the specification allowing for both temperature and realistic manufacturing tolerances on a leaky surface wave cut of Lithium Tantalate rather than the more commonly used cuts of Lithium Niobate. The choice was based on an acceptable temperature coefficient, a workable acoustic coupling efficiency and a lower process sensitivity. The advantage of Lithium Tantalate for this application is that filters on this material are less sensitive to processing variations and have a lower frequency temperature coefficient. This results in the realization of a compliant filter using a two track design which occupies less substrate area than the three track approach required on a Lithium Niobate substrate, in addition the design imposes a less stringent specification for processing parameters, both important factors in volume manufacturing

2. SPECIFICATION

The filter specification is detailed in Table 1.

Parameter		Specification		
Centre Frequency		947.5	MHz	
Bandwidth		25	MHz	
Insertion Loss		4	dB	max.
Passband Ripple		1	dB	max.
Inband VSWR.		2.5	dB	max.
Rejections				
DC-890	MHz	35	dB	max.
890 -915	MHz	20	dB	max.
980 - 1025	MHz	15	dB	max.
1025 - 1500	MHz	35	dB	max.
1500 - 3000	MHz	15	dB	max.
Operating Temperature		- 30 , + 70 °C		
Terminating Impedance		50	Ohms	
Package See Fig 1		SMD 5 x 5 mm.		

Table 1 Electrical specification.

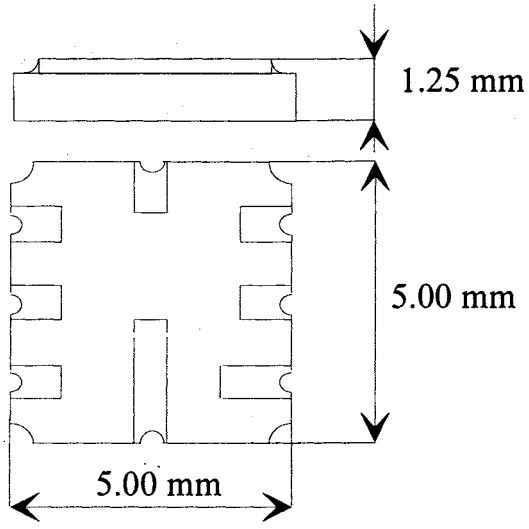


Fig. 1 Outline drawing of a 5 mm x 5 mm SMD package

3. PARAMETER EXTRACTION

The Coupling of Modes approach to filter design requires detailed knowledge of a number of parameters related to substrate material constants and process variables.

Initial extraction of parameters for 36 degree Lithium Tantalate was carried out using a series of test structures developed specifically for high coupling materials [4].

Results from these measurements were then used to design a series of devices of similar fractional bandwidth to the final device but at a lower frequency to reduce the effect of process variations. The COMs input parameters were then fine-tuned to match measured device performance prior to application to the final design.

In reality very little modification was required as can be seen in Figs. 5 and 6 where theoretical simulation and measured results are overlayed without modification

Figures 2 and 3 show sensitivity of two key parameters, leaky wave velocity and electrode reflectivity to metal film thickness (h/λ) and electrode mark and period ratio (m/p) variation.

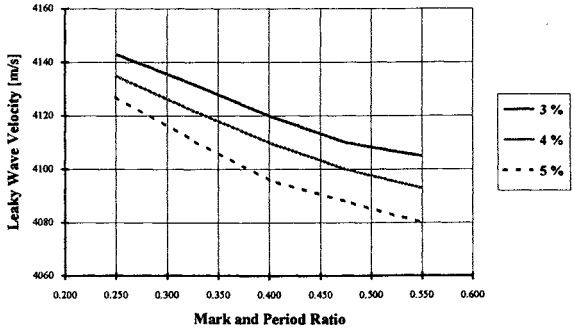


Fig. 2 Plot of the leaky wave velocity versus mark and period ratio for three different metal thicknesses. 3%, 4% and 5% of the wavelength (h/λ)

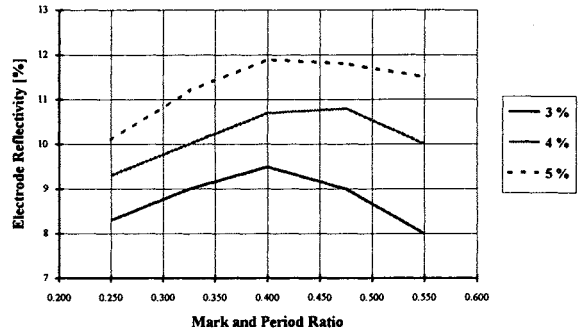


Fig. 3 Plot of the electrode reflectivity versus mark and period ratio for three different metal thicknesses. 3%, 4% and 5% of the wavelength (h/λ)

4. PROCESS SENSITIVITY

In order to ensure that the filter is capable of manufacture with acceptable yield in high volume, it is essential to take account of the sensitivity of design parameters to process tolerance in the design specification for the device..

With the knowledge of the sensitivity of these parameters to process variables obtained during the parameter extraction exercise and the specification of process limits it is possible to calculate the anticipated frequency deviation expected during manufacture.

Table 2 details the velocity shift associated with processing variations for 36 degree Tantalate and 64 degree Lithium Niobate for comparison.

	$LiTaO_3$	$LiNbO_3$
h/λ variation +/-2%	**+/- 1 m/s**	**+/- 2 m/s**
m/p variation +/- 0.5%	**+/- 0.85 m/s**	**+/- 1.35 m/s**

Table 2 Velocity shift associated with processing variation for 36 degree Tantalate and 64 degree Lithium Niobate

In addition it was discovered that the sensitivity of coupling coefficient to these process variations was two times greater for $LiNbO_3$ than for $LiTa0_3$. This is highly significant for production yield since variation in coupling coefficient results in increased passband ripple.

5. FILTER BANDWIDTH CALCULATION

The 25 MHz bandwidth requirement relates to the system signal bandwidth, the filter itself requires this to be significantly increased to allow for temperature, aging and manufacturing setting tolerance. Table 3 shows the allowances calculated for each, again comparing those for $LiTaO_3$ and $LiNbO_3$ assuming current process specifications.

	$LiTaO_3$	$LiNbO_3$
Temperature (-30- +70°C)	3.51 MHz	6.63 MHz
Aging (5 years estimated)	0.047 MHz	0.047 MHz
Manufacturing	1.8 MHz	3 0 MHz
Total	5.36 MHz	9.68 MHz
Required Bandwidth	30.36 MHz 3.20 %	34.68 MHz 3.66 %

Table 3 Comparison of the bandwidth required for 36 degree Lithium Tantalate and 64 degree Lithium Niobate

6. MATERIAL SELECTION

Initial simulations showed that whilst the wider bandwidth required for the use of Lithium Niobate as a substrate material was readily achievable owing to its higher coupling coefficient, the rejection specification becomes difficult to achieve without the addition of an extra coupled resonator filter track.

Lithium Tantalate was therefore selected as the preferred substrate material.

7. FILTER DEVELOPMENT

Initial trials were concentrated on a straightforward 2 transducer track structure to verify feasibility and COMs design parameters. See Fig. 4.

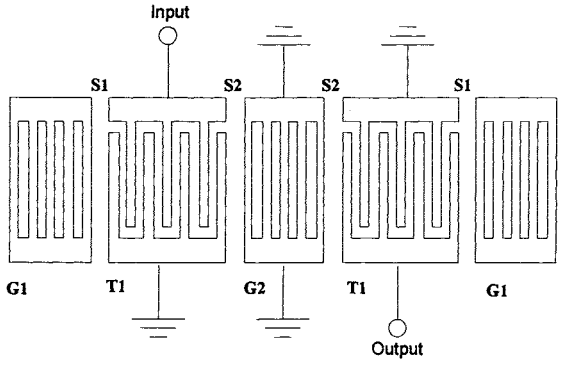


Fig. 4 Schematic drawing of a 2 Transducer Track structure.

The results were encouraging with extremely good correlation between theory and measurement using input parameters exactly as derived from the initial parameter extraction exercise. See Fig. 5.

The measured 1 dB band width for the filter corresponded to a fractional bandwidth of 2.7 % with an insertion loss of < 1 dB.

The out of band rejections of this filter as predicted, were not sufficient to satisfy the specification.

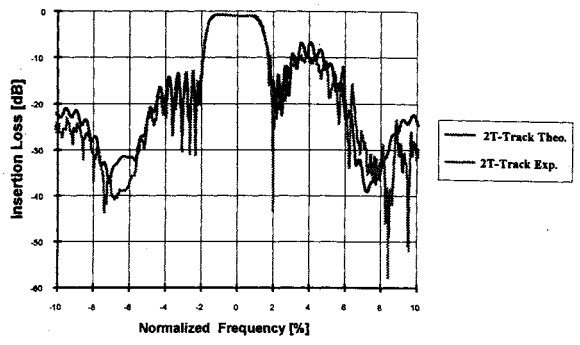


Fig. 5 Comparison of experimental response with theoretical simulation of a 2 Transducer Track structure.

To improve the rejections, a cascaded two track version of the same filter was realized. See Fig. 6.

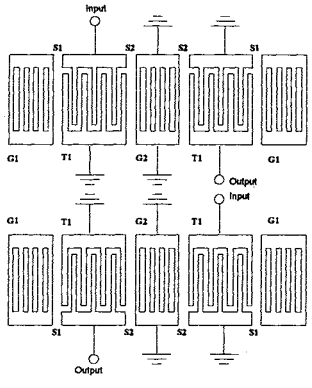


Fig. 6 Schematic drawing of a cascaded 2 Transducer Tracks structure.

This resulted in a similar bandwidth and an insertion loss of < 2 dB but with greatly improved rejections. Again excellent correlation was observed between the original simulation and measurement. See Fig. 7.

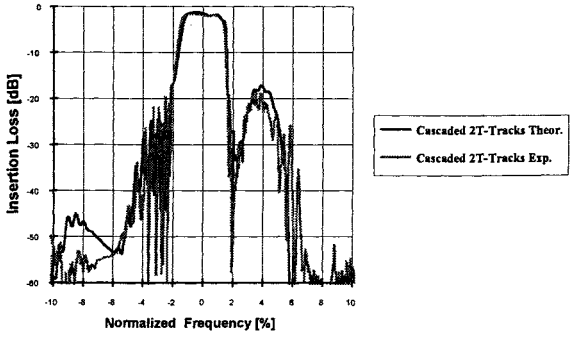


Fig. 7 Comparison of experimental response with theoretical simulation of a cascaded 2 Transducer Tracks structure.

This device had sufficient fractional bandwidth and rejections to satisfy the specification over the temperature range but did not allow sufficient margin for manufacturing variations . See Fig. 8.

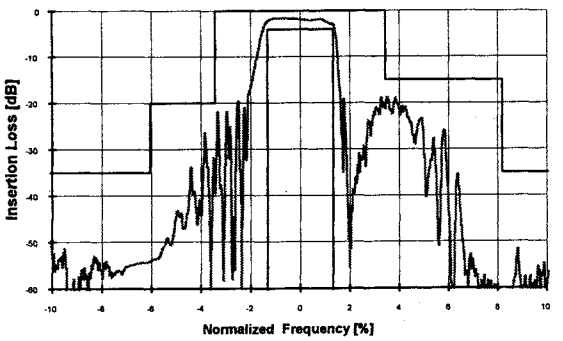


Fig. 8 Comparison of experimental response of a cascaded 2 Transducer Tracks structure with the specification.

To increase the bandwidth, a cascaded two track filter was simulated which utilizes three transducers (see Fig. 9) to cancel alternate longitudinal modes increasing, the mode separation and hence the bandwidth.

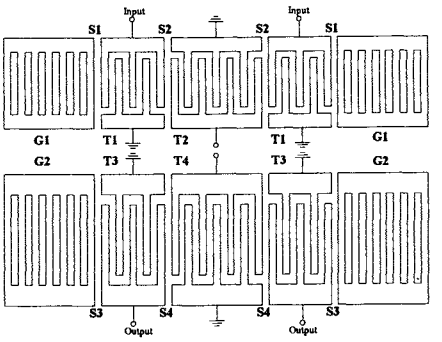


Fig. 9 Schematic drawing of a cascaded 3 Transducer Tracks structure.

To obtain maximum bandwidth, an iterative optimization procedure was used to independently modify the designs of the individual tracks, allowing increased ripple in each track .The tracks were designed to be complimentary to achieve the ripple specification for the overall cascaded two track filter. See Fig. 10.

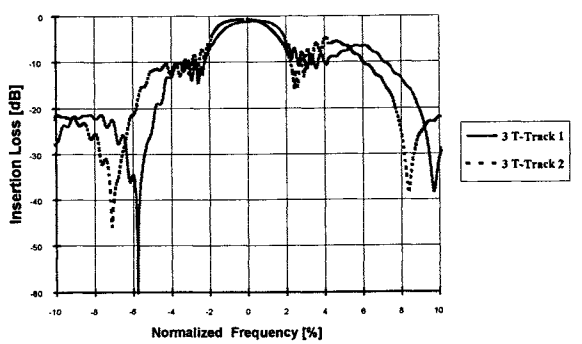


Fig. 10 Theoretical response of the two 3 Transducer Track structure used for the cascading.

In addition, the transition band ripples in the filter caused by sidelobes from the reflector stopband were partially canceled by offsetting the frequency of the two tracks, enhancing close in rejections, as the increased bandwidth design has exhibited higher sidelobes, violating the specification. See Fig. 11.

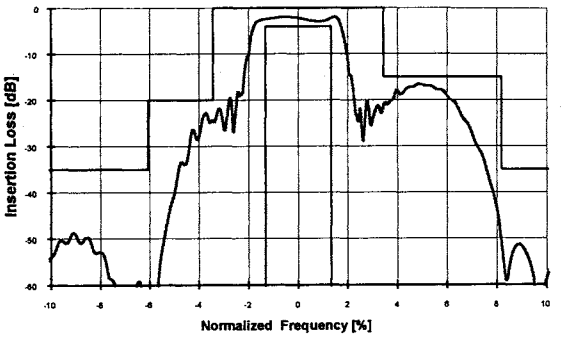


Fig. 11 Comparison of theoretical response of a cascaded 3 Transducer Tracks structure with the specification.

This design satisfies the requirement in all electrical respects .Measured results for this device were not available at the time of writing, but are expected to be consistent with the simulation in view of the excellent correlation previously seen during the development.

The package was custom designed for the application and includes special pattern recognition features for automatic die-bonding and wire bonding procedures see fig. 12

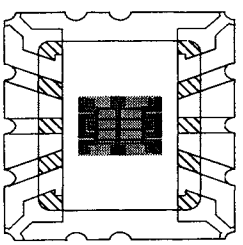


Fig. 12 Drawing of a chip in a SMD package.

8. CONCLUSIONS

The design of a 950 MHz filter for a mobile communications receiver has been demonstrated taking advantage of the low process and temperature sensitivity of 36 degree Lithium Tantalate.

A novel compensatory two track approach has been utilized to extend filter bandwidth on that material.

The filter design using this approach satisfies the requirements of a filter specification for a mobile communications receiver application. Full account has been taken of temperature and processing sensitivities so that the device will be manufacturable with high yield in volume.

The device is housed in a 5 mm x 5 mm miniature SMD package

Excellent agreement has been shown between theoretical simulations and measurements based on comprehensive measurements of Coupling of Modes parameters for the substrate material.

ACKNOWLEDGMENTS

The authors would like to express gratitude to D.P.Chen, V.Plessky for many helpful discussions and gratefully acknowledge the processing support from S.Mückenhirn and D.Pujol. The original development work for the modeling of structures on leaky surface wave materials was funded by the Swiss PTT and our thanks go to them for their support.

REFERENCES

[1] C.S.Hartmann, D.P.Chen and J.Heighway : " SAW transversely coupled resonator filters using Coupling of Modes modeling technique ", 1992 Ultrasonoics Symposium Proceedings.

[2] V.P.Plessky, T.Thorvaldsson : " Rayleigh waves and leaky SAW in periodic systems of electrodes : periodic Green's functions analysis ", Electronics Letters, Vol. 28, N0. 14, pp. 1317-1319.

[3] Takao Morita, Yoshitaka Watanabe and Yuzo Nakazawa "SAW Resonator Filters for Mobile Communications", Proc.Int. Symposium. SAW devices for Mobile Communications 186-193, 1992

[4] C.S.Hartmann and B.P.Abott : " Experimentally determining the transduction magnitude and phase and the Reflection Magnitude and phase of SAW SPUDT Structures", 1992 Ultrasonoics Symposium Proceedings 37-42.

Effects of Surface Electrical Boundary Condition on Excitation and Propagation of Highly Piezoelectric Leaky Surface Acoustic Waves

Ken-ya HASHIMOTO and Masatsune YAMAGUCHI

Dept. of Elec. and Electron. Engrg., Chiba University, Inage-ku, Chiba-shi 263 Japan

Abstract

The paper describes how the surface electrical boundary conditions affect the excitation and propagation of leaky surface acoustic wave (LSAW) on highly piezoelectric substrates, i.e., 36°YX-$LiTaO_3$, 41°YX-$LiNbO_3$ and 64°YX-$LiNbO_3$. It is shown that LSAW behaviours are somewhat different from those for conventional Rayleigh-type SAW and can be characterised to a great extent by means of the effective permittivity. LSAW excitation by surface-sources and its propagation characteristics under metallic-gratings are discussed. Furthermore, IDT properties for LSAW excitation are discussed, and the coupling-of-modes analysis for LSAW-IDTs is carried out. Various parameters characterising LSAW behaviours are also derived.

1 Introduction

Leaky surface acoustic waves (LSAWS) on 41°YX-$LiNbO_3$ (41-LN) and 64°YX-$LiNbO_3$ (64-LN)[1], which possess an extraordinarily large electromechanical coupling factor K^2, have been studied for wide bandwidth, low-loss filter applications during the past two decades[2][3]. LSAW on 36°YX-$LiTaO_3$ (36-LT)[4] has also been expected to be promising for filter and oscillator applications with moderate temperature stability[2][5].

Although LSAW radiates with propagation bulk acoustic waves (BAWs) into a substrate, it has conventionally been considered to be one of the SAW families provided that the loss with propagation (BAW radiation) is small.

LSAW propagating on some highly piezoelectric substrates, however, has two distinct features, both of which originate from the fact that LSAW velocity is in close proximity to surface skimming bulk wave (SSBW) velocity[6]. One is the effect of surface metallisation which significantly changes LSAW propagation parameters such as velocity, attenuation, penetration depth, etc. The other is the interaction between LSAW and SSBW. Since these two waves are easy to satisfy the phase-match condition each other, LSAW propagation becomes considerably complicated under periodic grating structures. These two features make it difficult, for example, to estimate K^2 of LSAW by the fractional velocity change due to surface metallisation.

From this point of view, this paper is aimed at discussing the effect of surface electrical boundary conditions upon LSAW excitation and propagation characteristics on highly piezoelectric substrates of 36-LT, 41-LN and 64-LN. Various behaviours of LSAW on these substrates are compared with those of SAW on 128°YX-$LiNbO_3$ (128-LN).

LSAW excitation by surface-sources and propagation characteristics under metallic-gratings are discussed, and it has been made clear how the LSAW behaviours are different from those for SAW. IDT properties for LSAW excitation are also discussed employing the coupling-of-modes (COM) analysis[7]. Various parameters characterising the LSAW behaviours are derived through the discussion, and it was shown that most of LSAW behaviours can be characterised by means of effective permittivity[8][9].

2 Effective permittivity

Assume that waves propagate toward X direction, and the plane Z=0, where an IDT is placed, coincides with an interface between a piezoelectric substrate for Z<0 and a region of free space for Z>0.

Displacement components u_i (i=1,2,3), respectively, in the X, Y, Z directions and electric potential u_4 in the substrate are governed by the Newton and Maxwell equations. By imposing appropriate boundary conditions at Z=0, u_i can be expressed as a sum of partial waves of the form of Fourier integral[8][9],

$$u_i(X,Z) = \int_{-\infty}^{+\infty} \frac{Q(k)}{|k|} \sum_{n=1}^{4} A_{in} exp(k\Omega_n Z - jkX)dk, \quad (1)$$

where k and $Im(k\Omega_n)$ are the wave numbers in the X and -Z direction, and A_{in} are the coefficients determined by the boundary conditions. $Q(k)$ is the Fourier transform of the surface charge density $q(X)$ and is given by,

$$Q(k) = (2\pi)^{-1} \int_{-\infty}^{+\infty} q(X) exp(+jkX)dX. \quad (2)$$

For each k, the Newton and Maxwell equations give eight partial wave solutions with different Ω_n. Four of them correspond to the waves coming out of the inside of the substrate, which do not contribute to the total field in the system considered here. The other four proper partial waves can be identified by Ω_n: they must be either evanescent toward -Z direction, i.e., $Re(\Omega_n) > 0$, or have power flow toward -Z direction, i.e., $Re(\Omega_n)=0$ and $\partial\{Im(k\Omega_n)\}/\partial k < 0$[10].

From eq.(1), the surface electric potential $\phi(X) = u_4(X,0)$ is written as,

$$\phi(X) = \int_{-\infty}^{+\infty} \frac{Q(k)}{|k|\epsilon(k/\omega)} exp(-jkX)dk, \quad (3)$$

where $\epsilon(V^{-1})$ is the effective permittivity[8][9] given by,

$$\epsilon(V^{-1}) = \left[\sum_{n=1}^{4} A_{4n}\right]^{-1} = -\frac{D_f(V^{-1})}{D_m(V^{-1})}. \quad (4)$$

In eq.(4) $V=\omega/k$, and D_f and D_m are the determinants of the matrices describing the boundary conditions on free and metallised surfaces[11]. Thus $\epsilon(V^{-1})$ can readily be calculated by an SAW velocity calculation program.

Note that the static permittivity ϵ_∞ which determines IDT capacitances is,

$$\epsilon_\infty = \epsilon_0 + \sqrt{\epsilon_{11}^T \epsilon_{33}^T - \epsilon_{13}^{T\,2}} = \epsilon(\infty), \quad (5)$$

where ϵ_0 is the permittivity in vacuum.

The total power P radiated from $q(X)$ is given by $\epsilon(k/\omega)$ as,

$$P = \frac{1}{2} Re\left[\int_{-\infty}^{+\infty} \phi(X) \frac{dq(X)^*}{dt} dX\right] \quad (6)$$

$$= \pi\omega Im\left[\int_{-\infty}^{+\infty} \frac{|Q(k)|^2}{|k|\epsilon(k/\omega)} dk\right]. \quad (7)$$

Fig.1 shows $\epsilon_\infty/\epsilon(V^{-1})$ for 128-LN as a function of V. The pole located at V=3,994m/s represents the SAW velocity V_{Sf} on a free surface. Because all of the partial waves are evanescent toward -Z direction for V <4,079m/s, $\epsilon(V^{-1})^{-1}$ is pure real. For V >4,079m/s, $\epsilon(V^{-1})^{-1}$ becomes a complex number suggesting that some of partial waves radiate BAWs into the bulk of a substrate.

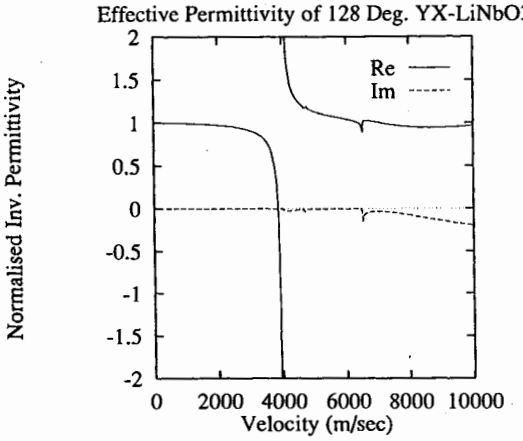


Figure 1: $\epsilon_\infty/\epsilon(V^{-1})$ for 128-LN. (ϵ_∞=56.68ϵ_0)

Applying the residue theorem to eqs.(3) and (7), one may obtain the electric field $\phi_{Sf}^{\pm}(X)$ and power $P_{Sf}^{\pm}$ associated with SAW propagating toward ±X directions as[9],

$$\phi_{Sf}^{\pm}(X) = \pi j K_{Sf}^2 \epsilon_\infty^{-1} Q(\pm k_{Sf}) exp(\mp j k_{Sf} X), \quad (8)$$

and

$$P_{Sf}^{\pm} = 2^{-1}\pi^2 \omega K_{Sf}^2 \epsilon_\infty^{-1} |Q(\pm k_{Sf})|^2, \quad (9)$$

where $k_{Sf} = \omega/V_{Sf}$. K_{Sf}^2 is the effective electromechanical coupling factor on a free surface[9] and is given by,

$$K_{Sf}^2 = 2V_{Sf}\epsilon_\infty \left[\frac{\partial\{\epsilon(V^{-1})\}}{\partial V}|_{V=V_{Sf}}\right]^{-1}. \quad (10)$$

Note that eq.(8) is not valid under an IDT.

When the propagation surface is metallised, it is convenient to employ the electric field $e(X) = -\partial\phi(X)/\partial X$ instead of $q(X)$. If $h(X) = \int q(X)dX$ is defined as a field quantity, then it is given by,

$$h(X) = \int_{-\infty}^{+\infty} \frac{\epsilon(k/\omega)E(k)}{|k|} \exp(-jkX)dk. \quad (11)$$

In the similar way, P is obtained as,

$$P = -\pi\omega Im\left[\int_{-\infty}^{+\infty} \frac{\epsilon(k/\omega)|E(k)|^2}{|k|} dk\right], \quad (12)$$

where

$$E(k) = (2\pi)^{-1} \int_{-\infty}^{+\infty} e(X) \exp(+jkX)dX. \quad (13)$$

Fig.2 shows $\epsilon(V^{-1})/\epsilon_\infty$ for 128-LN as a function of V. There appears a pole at V=3,886m/s which represents the SAW velocity V_{Sm} on a metallised surface. Compared with Fig.1, it is seen that $\epsilon(V^{-1})$ and $\epsilon(V^{-1})^{-1}$ are very similar to each other for 128-LN: the sign of the imaginary parts and the locations of the SAW poles are only different. This suggests that the acoustic wave excitation and propagation on 128-LN are quite similar on both free and metallised surfaces.

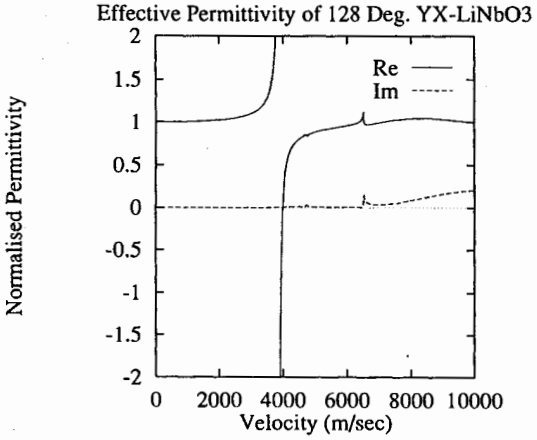


Figure 2: $\epsilon(V^{-1})/\epsilon_\infty$ for 128-LN.

By applying the residue theorem to eqs.(11) and (12), $h_{Sm}^{\pm}(X)$ and $P_{Sm}^{\pm}$ associated with SAW propagating toward ±X directions are also obtained as,

$$h_{Sm}^{\pm}(X) = -\pi j K_{Sm}^2 \epsilon_\infty E(\pm k_{Sm}) exp(\mp j k_{Sm} X), \quad (14)$$

and

$$P_{Sm}^{\pm} = 2^{-1}\pi^2 \omega K_{Sm}^2 \epsilon_\infty |E(\pm k_{Sf})|^2, \quad (15)$$

where $k_{Sm}=\omega/V_{Sm}$. K^2_{Sm} is the effective electromechanical coupling factor on a metallised surface and is given by,

$$K^2_{Sm} = -2V_{Sm}\epsilon_\infty^{-1}[\frac{\partial\{\epsilon(V^{-1})^{-1}\}}{\partial V}|_{V=V_{Sm}}]^{-1}. \quad (16)$$

Note that eq.(14) is not valid under an IDT.

For 128-LN shown in Figs.1 and 2, K^2_{Sf} and K^2_{Sm} are estimated to be 5.75% and 5.70%, respectively, which nearly coincide with K^2_V of 5.62% defined by the fractional velocity change due to surface metallisation.

From eqs.(7) and (12), the radiated BAW power P_{Bf} and P_{Bm} for the surface being free and metallised, respectively, are given by,

$$P_{Bf} = \pi\omega \int_{-\infty}^{+\infty} \frac{|Q(k)|^2}{|k|} Im[\epsilon(k/\omega)^{-1}]dk, \quad (17)$$

and

$$P_{Bm} = \pi\omega \int_{-\infty}^{+\infty} \frac{|E(k)|^2}{|k|} Im[\epsilon(k/\omega)^*]dk. \quad (18)$$

The BAW radiation conductance of an IDT is roughly proportional to $Im[\epsilon(1/fp)/\epsilon_\infty]$ and $Im[\epsilon_\infty/\epsilon(1/fp)]$[12][13] ($p$ is the IDT finger periodicity), respectively, for free and metallised surfaces.

Under the electrostatic approximation and/or for $\omega \cong 0$, $E(k)$ is expressed by $jsgn(k)\epsilon_\infty^{-1}Q(k)$. Thus, if $Im[\epsilon_\infty/\epsilon(V^{-1})] \cong Im[\epsilon(V^{-1})^*/\epsilon_\infty]$, this means that BAWs having $k=\omega/V$ are radiated deep into the bulk of the substrate and scarcely affected by the surface electrical boundary conditions. In the other words, when the behaviours of $Im[\epsilon_\infty/\epsilon(V^{-1})]$ and $Im[\epsilon(V^{-1})^*/\epsilon_\infty]$ are much different, BAWs are radiated very close to the surface and susceptible to the surface electrical boundary conditions.

3 Excitation and propagation

Fig.3 shows $\epsilon(V^{-1})/\epsilon_\infty$ for 36-LT. In addition to the SAW pole at V=3,132m/s, another pole appears at V=4,112m/s, which is accompanied by a narrow and symmetrical imaginary part. The guided wave represented by a non-pure-real pole is so-called LSAW. Since $Im[\epsilon(V^{-1})]$ and $Im[\epsilon(V^{-1})^{-1}]$ represent the energy dissipation (see eqs.(7) and (12)), the non-pure-real pole suggests that LSAW radiates (leaks) with propagation its energy into the bulk of a substrate.

Fig.4 shows $\epsilon_\infty/\epsilon(V^{-1})$ for 36-LT. Both SAW and LSAW poles exist at V=3,133m/s and V=4,212m/s, respectively. As can be seen from the figure, a very large imaginary part appears for V >4,212m/s, just above the LSAW pole. Since its shape is steep and asymmetrical with V, it may indicate the radiation of (fast-shear) SSBW: asymmetry reflects the cut-off nature of SSBW radiation[6][12][13]. SSBW is radiated parallel to the surface at the cut-off velocity V_B.

When V is complex, four proper Ω_n in eq.(1) should carefully be chosen out of eight Ω_n: Ω_n must be chosen so that the corresponding partial wave is proper even-when $Im(V)$ goes to zero. Since $Re(V_{Lf}) \gg |Im(V_{Lf})|$ and $Re(V_{Lm}) \gg |Im(V_{Lm})|$ in most cases, the selection could be done with the aid of slowness curves[10].

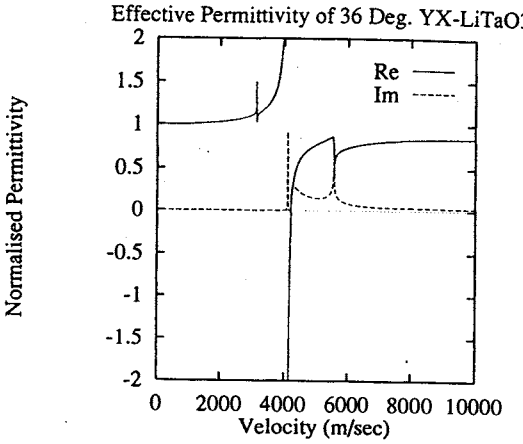


Figure 3: $\epsilon(V^{-1})/\epsilon_\infty$ for 36-LT. (ϵ_∞=51.15ϵ_0)

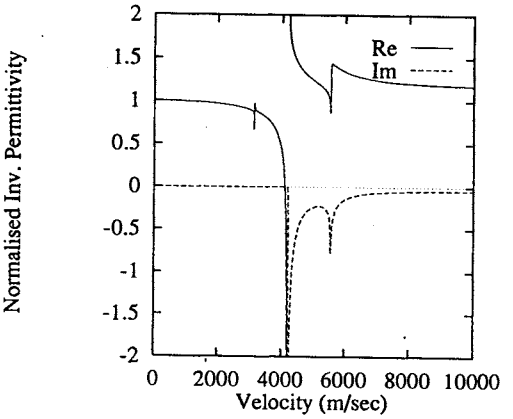


Figure 4: $\epsilon_\infty/\epsilon(V^{-1})$ for 36-LT.

If V_{Lf} and V_{Lm} are properly determined, the attenuation coefficients α_{Lf} and α_{Lm} can be obtained from the imaginary parts of V_{Lf} and V_{Lm}. The electromechanical coupling factors K^2_{Lf} and K^2_{Lm} of LSAW for free and metallised surfaces, respectively, can also be calculated from eqs.(10) and (16).

Table 1 shows thus estimated LSAW parameters for 36-LT, 41-LN and 64-LN with K^2_V for comparison. It is

Table 1: LSAW parameters for 36-LT, 41-LN and 64-LN

substrate	36-LT	41-LN	64-LN
$V_{Lf}(m/s)$	4,211.5	4,792.3	4,742.3
$V_{Lm}(m/s)$	4,113.0	4,379.7	4,474.8
$V_B(m/s)$	4,211.6	4,793.1	4,793.1
$\alpha_{Lf}(dB/\lambda)$	2.5×10^{-3}	4×10^{-6}	3.59×10^{-2}
$\alpha_{Lm}(dB/\lambda)$	7.65×10^{-4}	4.38×10^{-2}	2×10^{-6}
$K^2_{Lf}(\%)$	0.43	3.97	9.63
$K^2_{Lm}(\%)$	6.72	20.2	12.3
$K^2_V(\%)$	4.56	16.0	11.0

seen from the table that compared with 64-LN, K^2_{Lf} is much smaller than K^2_{Lm} for 36-LT and 41-LN. This is caused by the fact that V_{Lf} of 36-LT and 41-LN are in close proximity to V_B: LSAW energy is not effectively trapped close to a free surface, which makes LSAW most insusceptible to the surface electrical perturbation and, accordingly, weakly radiated by the surface charge.

As is expected from the above discussion, it has experimentally been confirmed that LSAW is very weakly

radiated on a free surface of 36-LT and 41-LN, where SSBW is the dominant propagation mode[18]. In other cases, LSAW is the dominant propagation mode.

When the propagation characteristics are most dependent upon the surface electrical boundary conditions, one may be unable to make a rough estimation of the electromechanical coupling factor by the fractional velocity change due to surface metallisation. Similar situation occurs for higher-order modes of SAW near their cut-off frequencies, SH-SAW[15], etc.

Note that when P_{Bf} and P_{Bm} are estimated by eqs.(17) and (18), they include LSAW power as well. This is because the radiated LSAW is finally converted into BAWs.

4 Propagation under metallic-gratings

Refs.[16] and [17] showed that the dispersion relation of SAW propagating under open-circuited (OC) metallic-gratings with periodicity p is given by,

$$1 + F_s R_s + F_{1-s} R_{1-s} = 0, \qquad (19)$$

where $s=p/\lambda$, λ is the wavelength, $F_x = P_x(0)/P_{1-x}(0)$, $P_v(w)$ is the vth-order Legendre function, and

$$R_s = \{1 - \epsilon_\infty/\epsilon(V_s^{-1})\}/\{1 + \epsilon_\infty/\epsilon(V_s^{-1})\}. \qquad (20)$$

In eq.(19), V_s=$V(s^{-1}$-1) and the strip width to space ratio is assumed to be unity for simplicity. The dispersion relation of SAW under short-circuited (SC) metallic-gratings is simply given by replacing $\epsilon_\infty/\epsilon(V_s^{-1})$ in eq.(20) by $\epsilon(V_s^{-1})/\epsilon_\infty$.

As a typical example, the velocity dispersion of SAW propagating under metallic-gratings on 128-LN was calculated, and is shown in Fig.5. There appears a stopband at p/λ=0.5, and its fractional widths $\delta V_{Go}/V_o$ and $\delta V_{Gs}/V_s$ for OC and SC gratings, respectively, are 1.30% and 1.31%. Out of the stopband, the absolute value of the gradient of the velocity dispersion for OC gratings is almost identical with that for SC gratings.

Although eq.(19) takes only the first term into consideration, it gives SAW dispersion relations with sufficient accuracy.

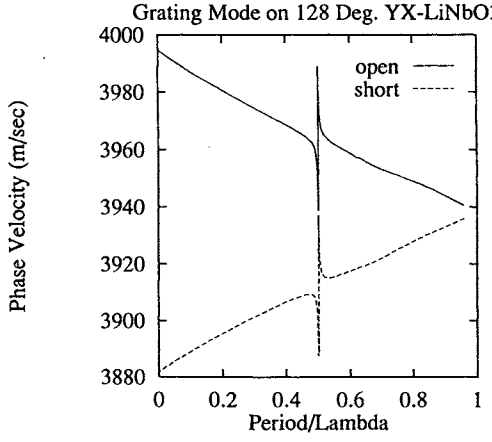


Figure 5: Velocities dispersion of SAW grating-mode on 128-LN

Fig.6 shows the velocity dispersion of LSAW propagating under metallic-gratings on 36-LN. Unlike the SAW grating-mode on 128-LN, $\delta V_{Go}/V_o$ and $\delta V_{Gs}/V_s$ are quite different from each other, and estimated to be 0.85% and 1.4%, respectively. The dependence of velocity change upon p/λ for OC gratings is relatively flat compared with that for SC gratings. This is because the LSAW velocity under OC gratings is in close proximity to V_B.

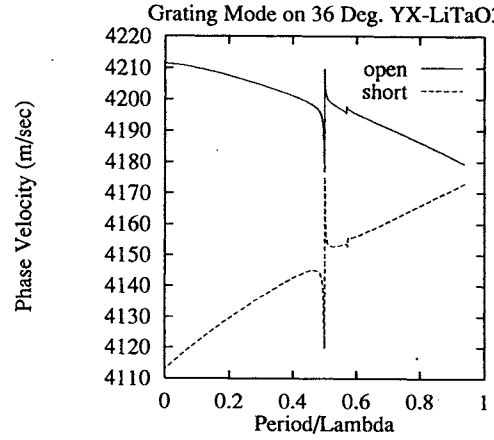


Figure 6: Velocity dispersion of LSAW grating-mode on 36-LT

Since the strip width to space ratio is assumed to be unity, the LSAW phase velocities V_{Go} and V_{Gc}, respectively, at the upper and lower stopband edges for OC gratings are simply determined as,

$$\epsilon(V_{Go}^{-1})/\epsilon_\infty = -0.04499, \qquad (21)$$

$$\epsilon(V_{Gc}^{-1})/\epsilon_\infty = -1. \qquad (22)$$

For SC gratings, the LSAW phase velocity at the upper stopband edge is identical with V_{Gc}, and the velocity V_{Gs} at the lower edge is,

$$\epsilon_\infty/\epsilon(V_{Gs}^{-1}) = -0.04499. \qquad (23)$$

Thus, these velocities can be estimated directly from $\epsilon(V^{-1})$ and $\epsilon(V^{-1})^{-1}$ curves. Table 2 shows V_{Go}, V_{Gc} and V_{Gs} estimated for 36-LT, 41-LN and 64-LN.

Table 2: Phase velocities of LSAW grating-mode at the first stopband for 36-LT, 41-LN and 64-LN

substrate	36-LT	41-LN	64-LN
$V_{Go}(m/s)$	4,210.7	4,785.1	4,732.3
$V_{Gc}(m/s)$	4,176.0	4,606.0	4,613.5
$V_{Gs}(m/s)$	4,118.9	4,398.2	4,486.9

Fig.7 shows the attenuation of LSAW under metallic-gratings on 36-LT. For both OC and SC gratings, the attenuation is negligible for p/λ <0.5, whereas it significantly increases when p/λ slightly exceeds 0.5. Since LSAW reflection by the gratings only causes attenuations at stopbands, the attenuation shown in the figure is due to BAW radiation through the interaction of LSAW with BAWs: LSAW is backward-scattered as BAWs propagating into the bulk of a substrate[18]. Two large attenuation peaks are observed at $p/\lambda \cong$0.5 and 0.57, which correspond to the phase matching of LSAW under

grating structures with backward-propagating fast-shear and longitudinal SSBWs[18]. Comparison of Fig.6 with Figs.3 and 4 suggests that the attenuation due to the LSAW backward-scattering is given by $Im[\epsilon_\infty/\epsilon(V_{1-s}^{-1})]$ and $Im[\epsilon(V_{1-s}^{-1})/\epsilon_\infty]$, respectively, for OC and SC gratings.

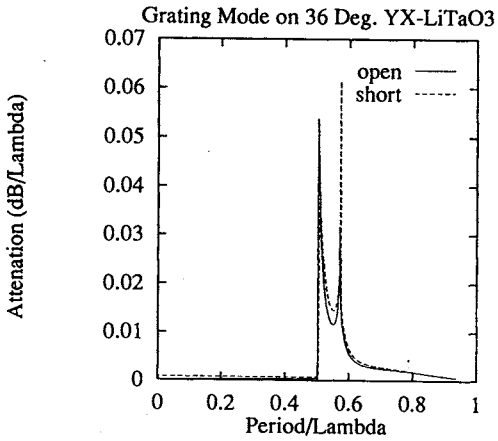


Figure 7: Attenuation of LSAW grating-mode on 36-LT

5 IDT characteristics

Fig.8 shows the experimentally obtained radiation conductance of an IDT fabricated on 36-LT: the finger periodicity is $57.8\mu m$ and the number of finger pairs L is 10[19].

The mainlobe appears at 71MHz, which corresponds to the LSAW radiation. At frequencies higher than the mainlobe, a relatively large conductance of 0.5-0.8mS is observed. Compared with $Im[\epsilon(V^{-1})]$ in Fig.3, it is suggested that this is due to the fast-shear and longitudinal BAWs radiated into the bulk of a substrate. In the figure, the theoretical conductance calculated by the Green function method[8] is also shown, which is in good agreement with the experiment. However, the calculation needs long computational time, and is not practical for IDTs with large L.

It may be of interest that although the insertion loss between two IDTs is considerably reduced by metallising the propagation surface, the radiation conductance is scarcely affected by the electrical condition of the propagation surface (either free or metallised).

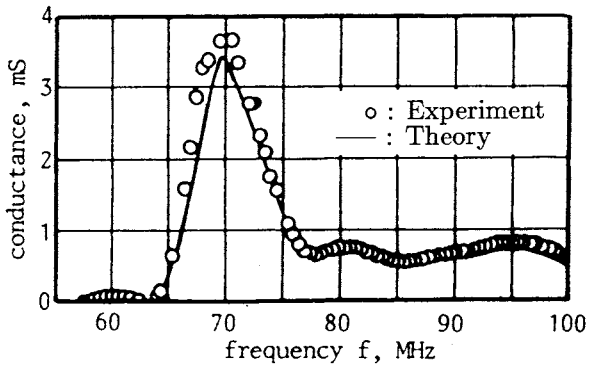


Figure 8: Radiation conductance of solid-IDT on 36-LT

IDTs with small L usually exhibit very complex behaviour[20]. However, if L is sufficiently large, the IDT behaviour for LSAW and BAWs could separately be analysed. This is because the conductance for BAW radiation is proportional to L to $L^{1.5}$[12], whereas the conductance for LSAW radiation to L^2. In addition, the IDT bandwidth is inversely proportional to L. Hence, for an increased L, there is a situation where the BAW radiation scarcely affects the LSAW radiation, in particular, near the LSAW resonance.

The COM theory[7] is often used for SAW device analysis. The IDT behaviour can be predicted most accurately if four COM parameters are properly determined[21].

So we have also attempted to make a further analysis of LSAW-IDTs by the COM theory.

Let $\phi^{\pm}(X)$ be the mode amplitude for SAW propagating toward $\pm$X direction and $I(X)$ be the current at the bus bar. Then, the COM equation for SAW device analysis is given by,

$$\frac{d\phi^+(X)}{dX} = -j\theta\phi^+(X) - j\kappa_{12}\phi^-(X) + j\zeta V, \quad (24)$$

$$\frac{d\phi^-(X)}{dX} = +j\kappa_{12}^*\phi^+(X) + j\theta\phi^+(X) - j\zeta^* V, \quad (25)$$

$$\frac{dI(X)}{dX} = -4j\zeta^*\phi^+(X) + 4j\zeta\phi^-(X) + j\omega CV, \quad (26)$$

where $\theta = \kappa_{11} + \omega/V_r - 2\pi/p$, V is the applied voltage, C is the capacitance per unit length of an IDT finger pair, V_r is the approximate SAW velocity under an IDT, κ_{11} and κ_{12} are the self- and mutual-coupling coefficients, respectively, and ζ is the transduction coefficients.

The COM parameters (κ_{11}, κ_{12}, ζ and C) are conventionally related to the phase velocities at the stopband of metallic-gratings equivalent to IDTs. For solid-IDTs, the parameters are given by,

$$\kappa_{11}p = 2\pi - \pi(V_{Gs} + V_{Gc})/V_r, \quad (27)$$

$$\kappa_{12}p = \pi(V_{Gc} - V_{Gs})/V_r, \quad (28)$$

$$\zeta p = -j2^{-1}\sqrt{\{\pi\omega Cp(V_{Go} - V_{Gs})/V_r\}}, \quad (29)$$

where $Cp=\epsilon_\infty$.

Since the stopband behaviour of OC and SC gratings is almost identical for conventional SAW, it is possible to determine the COM parameters for SAW-IDTs from eqs.(27)-(29).

However, eqs.(27)-(29) may not directly be applicable to LSAW-IDTs on 36-LT and 41-LN. This is because an LSAW-IDT resonates at $\theta \cong 0$, i.e., $\omega p \cong \pi(V_{Gs} + V_{Gc})$, and its performance is substantially characterised by the SC grating behaviour. That is, in order to derive the COM parameters, it does not always seem adequate to employ the result for OC gratings.

Since κ_{11} and κ_{12} are determined by the sift of the centre frequency and stopband width for the SC gratings, eqs.(27) and (28) could reasonably be accepted even for an LSAW-IDT. Since ζ determines the current flowing in at the center frequency of SC gratings, eq.(29) may be replaced by,

$$\zeta p = -j2^{-1}\sqrt{\{2\pi\omega Cp(V_{Gc} - V_{Gs})/V_r\}}, \quad (30)$$

The COM parameters were calculated by substituting the values listed in Table 2 into eqs.(27), (28) and

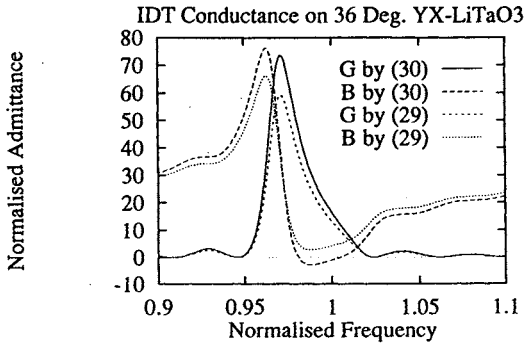


Figure 9: Calculated input admittance of solid LSAW-IDT on 36-LT

(30), and the input admittance of a solid LSAW-IDT was calculated. The number of finger pairs L is chosen as 26, which is suitable for the image connection configuration[5].

Fig.9 shows the calculated result. The horizontal axis is the normalised frequency ω/ω_B, and the vertical axis is the normalised conductance $G/\omega_B\epsilon_\infty$ and susceptance $B/\omega_B\epsilon_\infty$, where $\omega_B=2\pi V_B/p$. It is seen that $|B|$ is very small over a wide range of frequencies higher than the resonance. This is most desirable performance for the image connection configuration. For comparison, the admittance calculated using eq.(29) is also shown. As can be seen, the frequency range, where B should become negative, does not appear. This is because ζ is estimated to be too small by eq.(29).

It is also possible to determine the COM parameters from the theoretical response obtained by the Green function method[22]. The method has a merit to applicable to IDTs having arbitrary electrode configuration. Yamanouchi, et al.[23] have successfully applied the method for designing single-phase unidirectional transducers on 41-LN.

6 Conclusions

This paper is aimed at discussing the effect of surface electrical boundary conditions upon LSAW excitation and propagation characteristics on highly piezoelectric substrates of 36-LT, 41-LN and 64-LN.

It was shown that most of LSAW behaviours can be characterised by means of effective permittivity. The excitation of LSAW by surface-sources and LSAW propagation under metallic-gratings were discussed, and it was made clear how the LSAW behaviours are different from those for SAW. Furthermore, IDT properties for LSAW excitation were discussed, and the COM analysis for an LSAW-IDTs was carried out. Various parameters characterising the LSAW behaviours were also derived.

References

[1] K.Yamanouchi and K.Shibayama: J. Appl. Phys., Vol.43 (1972) 856-862.

[2] K.Yamanouchi and M.Takeuchi: Proc. IEEE Ultrason. Symp. (1990) 11-18.

[3] T.Morita, K.Watanabe, M.Tanaka and Y.Nakazawa: Proc. Int. Symp. SAW Devices for Mobile Comm. (1992) 186-193.

[4] K.Nakamura, M.Kazumi and H.Shimizu: Proc. IEEE Ultrason. Symp. (1977) 819-822.

[5] M.Hikita, H.Kojima, T.Tabuchi and Y.Kinoshita: IEEE Trans. Microwave Theory and Tech., Vol.MTT-33 (1985) 510-518.

[6] M.Lewis: Proc. IEEE Ultrason. Symp. (1977) 744-752.

[7] D.P.Chen and H.A.Haus: IEEE Trans. Sonics and Ultrason., Vol.SU-32 (1985) 395-408.

[8] R.F.Milsom, N.H.C.Reilly and M.Redwood: IEEE Trans. Sonics and Ultrason., Vol.SU-24 (1977) 147-166.

[9] R.F.Milsom, N.H.C.Reilly and M.Redwood:'The interdigital transducers', in Surface wave filters edited by H.Mathews (Wiley-Interscience, 1977) 55-108.

[10] B.A.Auld:'Acoustic waves and fields in solids', Vol.I (John Wiley and Sons, 1973).

[11] J.J.Campbell and W.R.Jones: IEEE Trans. Sonics and Ultrason., Vol.SU-15 (1968) 209-217.

[12] K.Hashimoto and M.Yamaguchi: Trans. IEICE (C), Vol.J71-C (1988) 1297-1302 (in Japanese).

[13] K.Hashimoto and M.Yamaguchi: J. Fac. Engrg., Chiba Univ., Vol.41 (1990) 9-17.

[14] K.Hashimoto, M.Yamaguchi and H.Kogo: Proc. IEEE Ultrason. Symp. (1983) 345-349.

[15] K.Hashimoto and M.Yamaguchi: Proc. IEEE Ultrason. Symp. (1988) 97-101.

[16] K.Blϕtekjær, K.A.Ingebrigtsen and H.Skeie: IEEE Trans. Electron. Devices, Vol.ED-20 (1973) 1133-1138.

[17] K.Blϕtekjær, K.A.Ingebrigtsen and H.Skeie: IEEE Trans. Electron. Devices, Vol.ED-20 (1973) 1139-1146.

[18] K.Hashimoto, M.Yamaguchi and H.Kogo: Proc. IEEE Ultrason. Symp. (1985) 16-21.

[19] K.Hashimoto and M.Yamaguchi: Jpn. J. Appl. Phys., Vol.26, Suppl.26-1 (1987) 159-161.

[20] K.Hashimoto and M.Yamaguchi: Trans. IEICE (C), Vol.J71-C (1988) 479-485.

[21] C.S.Hartmann and B.P.Abbott: Proc. IEEE Ultrason. Symp. (1990) 37-42.

[22] K.Hashimoto and M.Yamaguchi: Proc. IEEE Ultrason. Symp. (1991) 21-26.

[23] K.Yamanouchi, M.Takeuchi, T.Meguro, K.Doi and K.Murata: Jpn. J. Appl. Phys., Vol.30, Suppl.30-1 (1991) 173-175.

CESIUM 2

Chairman: P. Thomann

REALISATION OF A FREQUENCY STANDARD USING AN ATOMIC FOUNTAIN

A. Michaud, M. Chowdhury, K.P. Zetie, C.J. Cooper
G. Hillenbrand, V. Lorent, A. Steane, and C.J. Foot

University of Oxford, Clarendon Laboratory,
Oxford, OX1 3PU, UK

ABSTRACT

We report the realisation and preliminary study of a frequency standard using a fountain of laser cooled caesium atoms. Our apparatus uses a magneto-optical trap as a source of cold atoms and optical pumping to prepare the atoms in the correct state before they enter the microwave cavity.

KEYWORDS: Atomic clock, atomic fountain, laser cooling.

1. INTRODUCTION

Caesium atoms at a temperature of only 3μK can now be produced by laser cooling. These atoms have a velocity of only about 10mms^{-1} – a few times the recoil velocity from a single photon. Such ultracold atoms allow much longer measurement times than have previously been possible, together with smaller systematic shifts. These features have been demonstrated for RF and microwave transitions by Kasevich *et al.* (Ref. 1) who made the first working atomic fountain using sodium, and Clairon *et al.* (Ref. 2,3,4) who made a caesium atomic fountain. The advantages of this technique were first considered by Zacharias who attempted to use velocity selection to implement the Ramsey separated oscillatory field method in a fountain by allowing the atoms to interact twice with the same microwave cavity, once on the way up and again on the way down (Ref. 5). This symmetric configuration reduces the cavity phase shift which is one of the most troublesome problems in present primary frequency standards. In this paper, we report on our work to make a frequency standard following this approach. Firstly we give an overall view of the operation of the fountain and then, in the following sections, a more detailed description of those aspects which are specific to our system.

2. OVERVIEW OF EXPERIMENT

The physical arrangement of the apparatus is shown in figure 1. The magneto-optical trap (MOT), used as a source of cold atoms, comprises three orthogonal pairs of laser beams of the requisite polarisations ($\sigma^+\sigma^-$) which intersect in a region where a gradient of magnetic field is produced by a pair of anti-Helmholtz coils (Ref. 6). Typically there are 10^7 atoms in the trap after a 1s loading time. These atoms are launched upwards to form a fountain of cold atoms using the following sequence: The magnetic field of the trap is switched off, to leave the atoms in optical molasses. The atoms are then launched upwards by introducing a frequency difference between the vertical beams so that the rest frame of the molasses is moving upwards at the required velocity ("moving molasses technique", Ref. 4) and the cooling gives a narrow spread in velocities (the laser detuning below resonance is changed smoothly from a -10MHz to -65MHz during this cooling period). The final stage of the launch sequence is to optically pump all the atoms into the lower hyperfine level (we

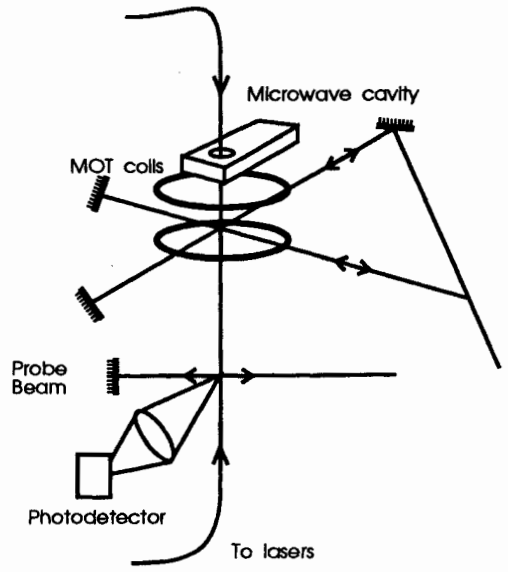


Figure 1: Experimental arrangement of the atomic fountain

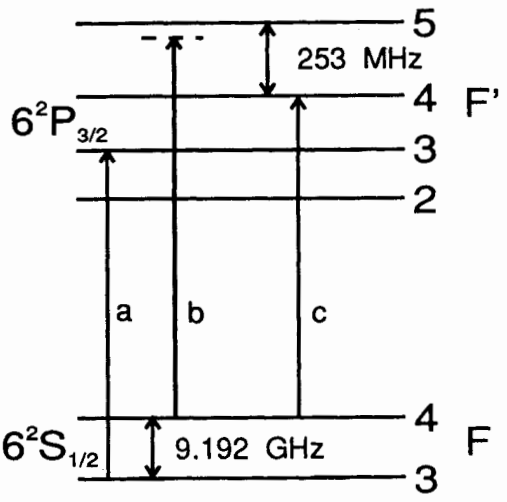


Figure 2: Energy level scheme D_2 line in caesium. The transitions used are: (a) repumper (b) trapping and cooling (c) pumping

discuss state selection in section 3). After launching, the atoms pass upwards through a 12mm hole in a microwave cavity (a shorted section of X-band waveguide located 40mm above the trap centre) and after a period of 0.5 s (for our highest launches) they fall back under gravity passing through the cavity a second time. Both the atoms which have undergone a transition and the total number of atoms are measured by monitoring the fluorescence from a 6mm diameter probe beam (the photodiode and lens have a combined efficiency of 5%).

3. OPTICAL PUMPING AND STATE SELECTION

During the loading of the trap and launching by molasses, two laser beams are present. The main beam is tuned slightly below the F=4 → F'=5 cycling transition (figure 2) and provides the cooling and trapping forces. Off-resonant excitation of the F=4 → F'=4 transition causes some of the atoms to leak into the lower hyperfine level (F=3), where they are lost. Therefore, a second laser beam is necessary, tuned to the F=3 → F'=3 or F'=4 transitions, to repump those atoms.

Caesium atomic clocks are based on the magnetic dipole transition $|3,0\rangle$ to $|4,0\rangle$ between the ground state hyperfine levels since neither of these states has any first order Zeeman shift (Ref. 7). Thus to observe the clock transition there must be a state selection process to create a population imbalance between these two states. The current working primary time standards use magnetic state selection but optical pumping is the natural choice for laser cooled atoms since no additional lasers are required. The advantages of optical pumping have long been recognized and there are a considerable number of published papers in which various techniques have been analyzed theoretically and experimentally (Ref. 8). Optical pumping methods for clocks can be divided into two categories, namely one-laser and two-laser schemes; we have used both types.

In a simple one-laser scheme all the atoms are transferred into one of the hyperfine levels. This hyperfine-pumping was carried out in our initial experiments (and those of Clairon *et al.*) by using a laser resonant with the F=4 → F'=4 transition which emp-

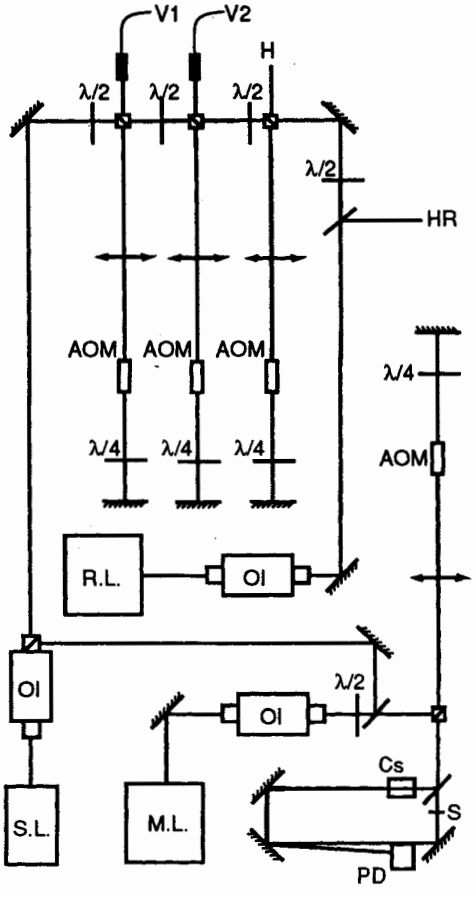


Figure 3: Laser System: AOM: Acousto-Optic Modulator, Cs: Caesium Cell, ML: Master Laser, P: Photodetector, RL: Repumper Laser, S: Shutter, SL: Slave Laser, OI: Optical Isolator, $\lambda/2$, $\lambda/4$: Retardation plates

tied the upper level into F=3 with each atom scattering an average of 2.4 photons. We have achieved effective pumping by this method by simply cutting the repumping beam after the cooling period in molasses (when the main laser is detuned 65MHz below F=4 $\rightarrow$ F'=5) and deliberately allowing the atoms to leak out of the cooling cycle. Experimentally we have found that this alternative method does not always give complete pumping and the slower rate of pumping will lead to a spread in the vertical velocity distribution (this is possibly an advantage since it reduces the atomic density without atoms being lost by transverse spreading). However this simplified method can be used as part of a two-laser scheme.

Two-laser schemes exploit the fact that the electric dipole matrix element between two $M_F = 0$ states of the same total angular momentum is zero e.g. $\langle F = 3, M_F = 0|\mathbf{d}.\mathbf{E}|F' = 3, M_F = 0\rangle = 0$ for the scheme which we consider here. Thus the combination of one laser effecting hyperfine-pumping out of the F=4 level and a second laser resonant with the $F = 3 \rightarrow F' = 3$ transition (linearly polarized along the B-field) which excites π-transitions, the atoms cycle many times until they fall into the only state not being excited, $|F = 3, M_F = 0\rangle$. Although a high degree of pumping has been achieved in thermal beams (Ref. 8) it involves the scattering of many photons, and would cause considerable heating. However a compromise can be achieved by only allowing a few optical pumping cycles since a worthwhile number of atoms are transferred into $M_F = 0$ after scattering up to two photons and those already in this state are not heated at all. Other possible schemes are discussed in (Ref. 8), in particular some which use π-pumping on the F=4 $\rightarrow$ F'=4 transition. The increase in the signal for a given atomic number density will be particularly important if collisional shifts are significant.

4. OPTICAL SYSTEM

The requisite laser beams were obtained using only two independent diode laser systems. Both systems were based on high power Spectra Diode lasers frequency stabilised by an external grating cavity (Ref. 9) with reference to the saturated absorbtion spectrum of caesium. The output in the zeroth order of the grating was about 10mW, which was sufficient for the repumping beams. To obtain more power for the main cooling beams the light was amplified to 100mW by injection locking another diode, and a further increase in power could readily be achieved by using the same master to

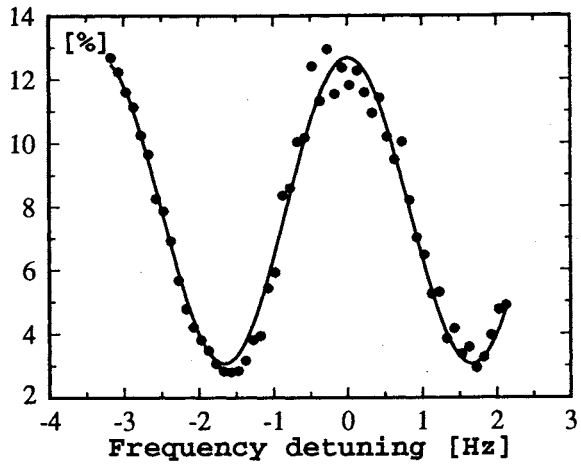


Figure 4: Central Ramsey fringe corresponding to a 11 cm launch above the cavity

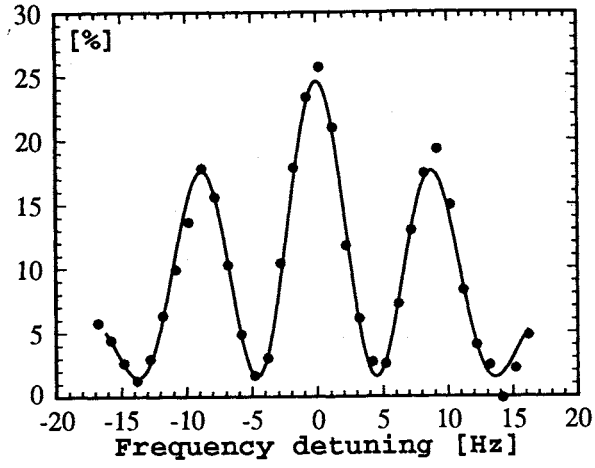


Figure 5: Ramsey pattern obtained with two-laser pumping

control several slave lasers whilst still maintaining the long term reliability required for a clock. The master laser was computer controlled so that its frequency could either be locked at a fixed offset from the F=4 → F=5 saturated absorption line or moved in less than 1 ms to any nearby frequency.

Figure 3 shows the schematic of the lasers and the associated optics. The light from the slave laser is double-passed through acousto-optic modulators to enable a frequency shift to be introduced between the vertical beams and to shut off all the beams quickly. The AOMs shift the frequency of the light, relative to the laser frequency, by 2×80MHz in the horizontal beams (H) or $2\times(80\pm\delta)$ for the vertical beams (V1,V2). In order that these output beams have a small detuning below the required transition we used another AOM at a frequency of 2×75MHz before the absorption cell. The frequency of this AOM is dithered to allow phase sensitive detection of the saturated absorption spectrum in the cell. Part of the repumping laser beam is mixed with the horizontal beam (H) and some was transmitted separately (HR) to the clock for the π-polarized beam required for the second scheme described in section 3. Optical fibres for the vertical beams eliminate possible beam movements from the misalignment of the back reflection through the AOMs and ensure good beams quality.

5. RESULTS

As an example of the possibilities of this new type of frequency standard, we show some Ramsey patterns obtained using this apparatus. The fraction of atoms detected in the level $|F=4, M_F=0\rangle$ is plotted as a function of the microwave frequency detuning in Hz. Figure 4 corresponds to a launch of 110mm above the cavity, corresponding to a width of 1.7Hz, obtained using one-laser pumping (maximum of 14% of atoms in $M_F=0$). Each point is an average of 10 measurement cycles and the S/N is 30. Figure 5 corresponds to a launch of 57mm obtained using the two-laser pumping scheme. To verify the mechanism the signal at the maximum of the central fringe was measured as a function of the angle of polarisation of the repumper relative to the bias B-field (5μT in the horizontal direction) (figure 6). This is a measurement of the number of atoms in the $|F=3, M_F=0\rangle$ level and shows, that the relative number of atoms in this state can be increased by a factor of $\sim$2.5 without significant heating.

We also observed anomalies on some experimental patterns which cannot be ex-

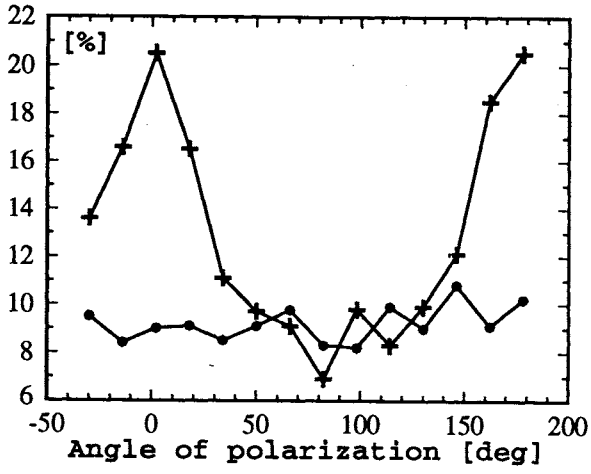


Figure 6: Fringe amplitude as a function of the angle of polarisation of the repumping light (+), and without repumping light (•)

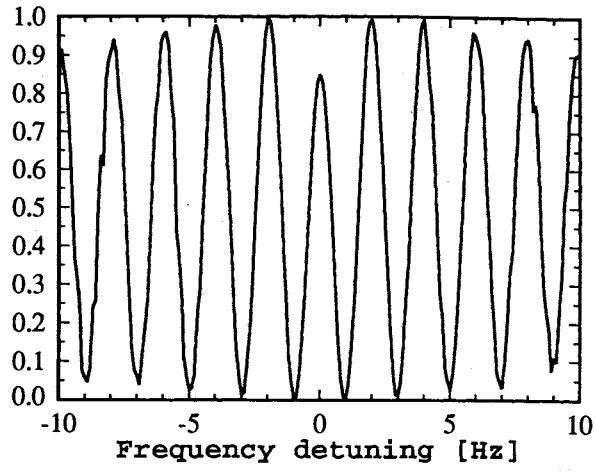


Figure 7: Ramsey pattern showing the effect of a small, constant microwave excitation.

plained using the standard model. We believe that the central fringe amplitude may be affected by leakage of microwave from the waveguide. In order to illustrate this effect, Figure 7 shows a pattern obtained by numerical simulation. For a typical launch ($T = 0.5\text{s}, \tau = 2\text{ms}, b\tau = \pi$ and the velocity distribution is assumed gaussian: $\sigma = 10\text{ms}$) the atoms were submitted to an excitation of the amplitude $0.002 \times b$ during the period T (we assumed a constant amplitude and phase). We clearly see the collapse of the central fringe. This behavior is different to that due to incorrect power level as it affects mainly the central fringe.

This result emphasizes the need for a well designed microwave cavity and further investigation of a possible frequency shift. (ref. 8)

6. CONCLUSION

We have realised a prototype frequency standard based on an atomic fountain of cold caesium atoms and have obtained Ramsey patterns with linewidth below 2Hz. A large part of the observed noise is due to effects of fluctuations of time of arrival (±1%) and the normalisation technique used; this will be reduced by the use of a second probe beam to provide improved normalization of the signal. We have also shown that the ratio of atoms present in the $|F = 3, M_F = 0\rangle$ level can be increased without significant heating. Although no systematic investigations have been done, it has been observed that the S/N is also increased in the same proportion.

This research is sponsored by the National Physical Laboratory. KPZ and AMS are research fellows at Christ Church and Merton College respectively. We acknowledge valuable discussions with A. Clairon, D. Knight, P. Whibberley and C. Salomon.

REFERENCES

1. M. Kasevich, E. Riis and S. Chu, *Rf spectroscopy in an atomic fountain*, Phys. Rev. Lett. 63, 612, 1989.

2. A. Clairon, C. Salomon, S. Guellati and W.D. Phillips, *Ramsey fringes in a Zacharias fountain*, Europhys. Lett., Vol. 16, 1991, pp. 165-170.

3. A. Clairon, C. Salomon, S. Guelati, W.D. Phillips *A laser cooled caesium atomic fountain: Towards a high performance clock* Proceedings of 5th European time and frequency and time forum, Besancon, FR, 1991.

4. A. Clairon, Ph Laurent, A. Nadir, M. Drewsen D. Grison B. Lounis and C. Sa-

lomon *A Simple and compact source of cold atoms for caesium fountains and microgravity clocks*, Proceedings of the 6th European Frequency and Time Forum, Noordjwiik, NL, 1992.

5. Ramsey N.F., Molecular Beams, Oxford University Press, Oxford, 1956

6. E.L. Raab, M. Prentiss, A. Cable, S. Chu, E.L. Raab and D. Pritchard *Trapping of Neutral Sodium with Radiation Pressure* Phys. Rev. Lett. 13,452, 1988.

7. J. Vanier, C. Audoin *The Quantum Physics of Atomic Frequency Standards*, Adam Hilger, 1989.

8. C. Audoin *Caesium beam frequency standards: classical and optically pumped*, Metrologia, Vol. 29, No. 2, 1992.

9. M. de Labachellerie and P. Cerez: *An 850 nm Semiconductor Laser Tunable over a 300 A range* , Optics Communications, Vol. 55, No. 3, 1985, p. 174.

THE FOURTH PRIMARY CESIUM ATOMIC CLOCK OF THE PTB

A. Bauch, B. Fischer, T. Heindorff, R. Schröder

Physikalisch-Technische Bundesanstalt
Bundesallee 100, D-3300 Braunschweig

ABSTRACT

The CS3, the second primary atomic clock with vertical cesium beams, has been put into continuous operation in autumn 1992. Its construction and its overall characteristics are identical with those of CS4. In the first 150 days of operation the CS3 frequency shifts related to the C-field, the second-order Doppler effect and the cavity phase difference, and their temporal stability were measured. In the period of common operation the relative frequencies of CS3 and of CS4 agree within $(25 \pm 9) \cdot 10^{-15}$. The measured frequency instability of CS3 is $8{,}6 \cdot 10^{-12}\sqrt{s/\tau}$ for averaging times τ, 100 s $\leq \tau \leq$ 20000 s.

Keywords: Cesium atomic clock, realization of SI second.

1. INTRODUCTION

More than ten years after the project to construct new PTB cesium atomic clocks was started, its successful end gets nearer with the two clocks, CS3 and CS4, operating since autumn 1992 and summer 1991, respectively. The intention of the project was to include the potential of reducing - compared with CS1 or CS2 - frequency shifts depending upon the velocity of the cesium beam atoms. This goal has been reached, because the mean velocity of the atoms which make the clock transition is about 70 m/s, only two thirds of the value of the CS1 or of the CS2 atoms. As being obvious from Fig. 1 a vertical construction was chosen to overcome problems inherent in the gravitational deflection. As to the mechanical construction, to the electrical supplies, and to the electronic equipment for probing the centre of the clock transition line (Ref. 1), CS3 and CS4 are twin machines. So their properties are nearly identical.

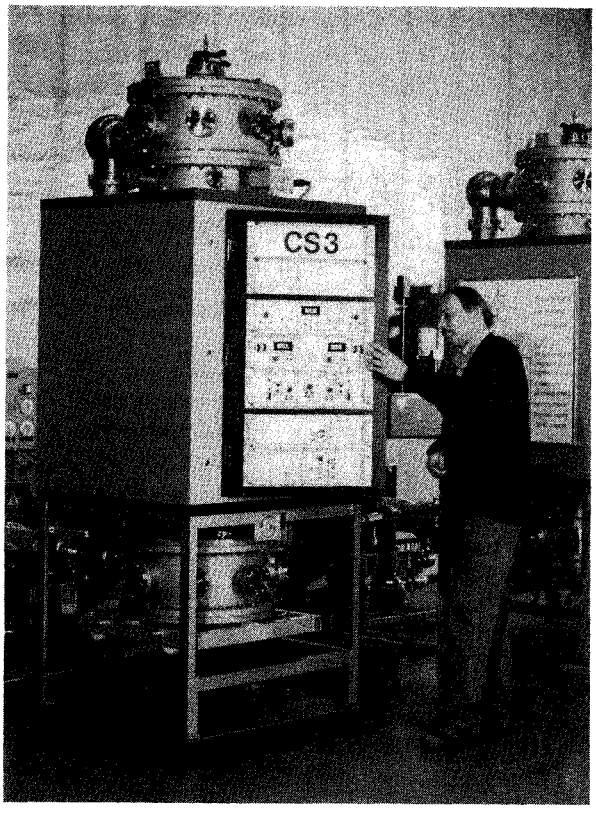


Fig. 1: The PTB primary cesium atomic clock, CS3. The electric and electronic equipment is mounted in the front panel of the square Armco cover. The vacuum pump and its accessories stand at the rear.

2. APPARATUS

The CS3 clock is a two-beam apparatus with a cylindrically symmetric arrangement of the beam tube elements. Its upper half is sketched in Fig. 2. The top and bottom chambers contain an oven and a hot-wire detector with a six-pole state selecting magnet. The copper tube connecting the two chambers carries the C-field solenoid and contains the Ramsey cavity of the CS2 type (Ref. 2). Inside the square Armco cover (Fig. 1) an Armco cylinder and three Mumetal cylinders form the magnetic shield.

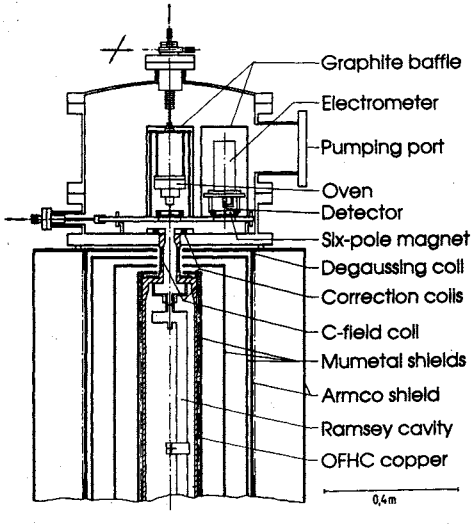


Fig. 2: Simplified sectional drawing of the upper half of CS3.

Because the characteristic data were presented already in detail for CS4 (Ref. 3), only the parameters entering into the frequency measurements are listed in Table 1.

Table 1. Operating parameters of CS3 and of CS4

	CS3	CS4
Total beam current in pA	7	
Flop-out signal of (4,0) atoms in pA	0,7	
Mean velocity of (4,0) down-beam atoms in m/s	69,9	71,6
Mean velocity of (4,0) up-beam atoms in m/s	71,4	72,4
FWHM of velocity distribution in m/s	12	10
Flux density of C-field in μT	7,732	
Length of microwave interaction in mm	23,3	
Drift length in mm	767	
Phase difference of cavity in μrad	51	45
Distributed phase diff. (Ref. 2) in μrad/mm	94	

3. RESULTS

3.1 Frequency and instability

The generation of a SI second pulse at the output of an atomic clock needs the precise knowledge of the systematic frequency shifts. The shifts known for CS3 are caused by the static magnetic field (C-field) in which the clock transition is induced, by the second-order Doppler effect, and by the cavity phase difference. These numbers have been determined step by step and tested again and again for temporal constancy. Then, the C-field flux density has been adjusted so that it compensates for these shifts and for the frequency setting of the digital synthesizer in the microwave synthesis chain. Referencing to the most stable PTB clock, CS2, the time scale difference T(CS3)-T(CS2) shown in Fig. 3 results. For comparison the data of the other primary clocks, CS4 and CS1, are given, too. In the intervals of common operation the relative frequencies of CS3 and CS4 agree within $(25\pm9)\cdot10^{-15}$, all four clocks do it within $(33\pm9)\cdot10^{-15}$. The statistical uncertainties quoted are the values of the instability $\sigma_y(\tau)$ for τ = 150 days measured with CS1 versus CS2 and expected for CS3, CS4, too.

The short-term instability of the 5 MHz signal locked to the CS3 clock transition line centre was studied by using a high-resolution phase comparator with an error multiplication factor of 10^4 (Parzen principle). An active hydrogen maser served as a reference clock whose instability could

be neglected. In the $\sigma_y(\tau)$ diagram of Fig. 4 the CS3 white noise value of $8{,}6 \cdot 10^{-12}\sqrt{s/\tau}$ for averaging times τ, 100 s $\leq \tau \leq$ 20000 s, exceeds the value calculated from the shot noise of the atomic beam and from the Nyquist noise of the electrometer by about 10 percent.

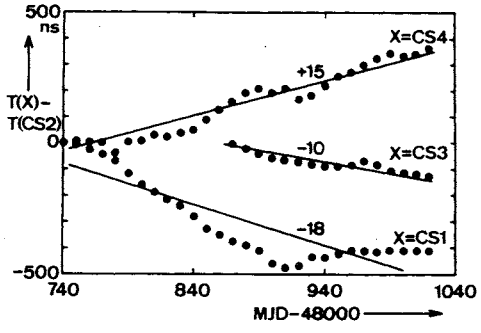


Fig. 3: Time scales of the PTB's primary clocks CS1, CS3, CS4 with reference to CS2, the most stable PTB primary clock, on a basis of 10-day averages. MJD Modified Julian Date. The initial values of each scale have been chosen arbitrarily. At MJD = 48740 the true value of T(CS1)-T(CS2) is -4647 ns. The numbers above the straight lines are the mean relative frequency differences F(X-CS2) in parts in 10^{15}.

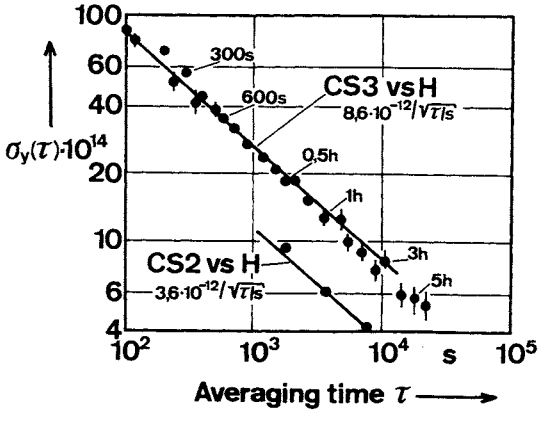


Fig. 4: Two-sample standard deviation $\sigma_y(\tau)$ of CS3 versus a hydrogen maser H. For comparison, data of CS2 versus H are given.

3.2 Frequency corrections and uncertainties

The detailed statements concerning the systematic frequency corrections of CS4 already reported (Ref. 3) are applicable to CS3, too. Therefore, only a short summary will be given below.

The C-field correction is made by measuring the Zeeman splitting. The field inhomogeneity of 3 nT mapped during the assembly of CS3, the instability of the current source supplying the solenoid as well as the thermal instability of the solenoid form lead to an uncertainty of this correction of $3\cdot 10^{-15}$.

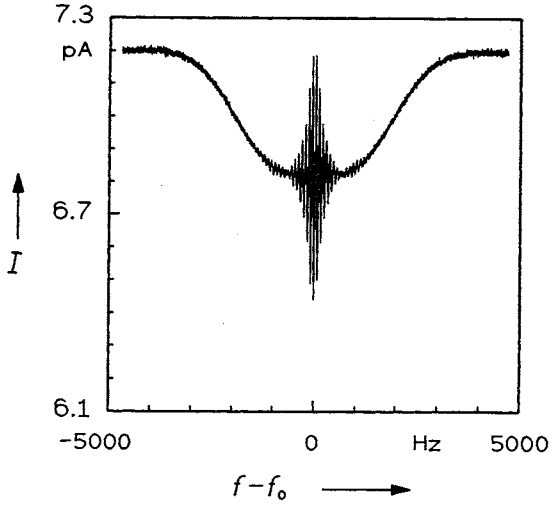


Fig. 5: Resonance pattern of the F=4, m=0 ⇒ F=3, m=0 transition of CS3 at optimum microwave power. I cesium beam current, f frequency of the microwave excitation, f_0 = 9192631770 Hz.

The second-order Doppler correction is calculated taking into account the velocity distribution of the (4,0) atoms and the square-wave modulation scheme used in the frequency control servo. In Fig. 5 the (4,0) ⇒ (3,0) resonance pattern is recorded from which the velocity distribution shown in Fig. 6 was deduced by Fourier inversion. The uncertainty of the Doppler correction is limited by the noise of the data input to the inversion program. Its sensitivity to noise in the input data results in an uncertainty of $1\cdot 10^{-15}$.

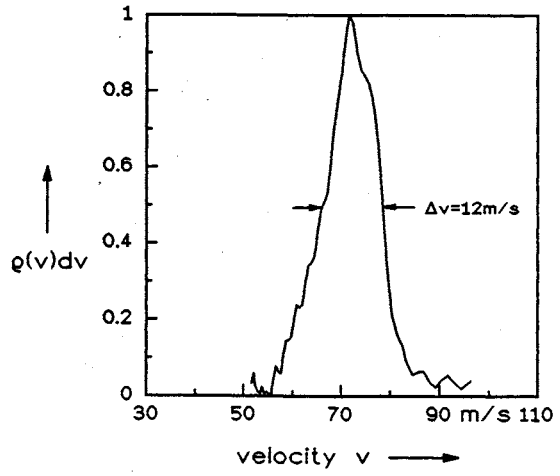


Fig. 6: Normalized velocity distribution of cesium atoms contributing to the (4,0) ⇒ (3,0) clock transition of CS3.

The end-to-end cavity phase difference of 51 μrad and the corresponding frequency shift (see Fig. 7) measured by a reversing of the cesium beam direction (from upwards to downwards and vice versa) are correct to the same extent as the up-beam axis coincides with the down-beam axis. In CS3, because of mechanical tolerances the adjusting procedure gave a maximum error of 0,13 mm. This figure, the distributed cavity phase difference of Table 1, and the mean velocity of the (4,0) atoms result in an uncertainty of the relative phase shift correction of $9 \cdot 10^{-15}$.

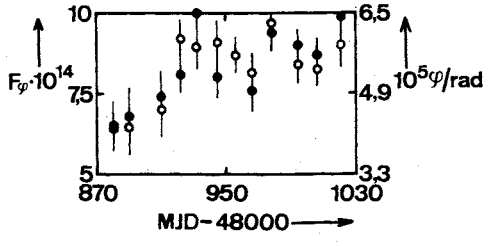


Fig. 7: Phase difference φ and frequency correction F_φ of the CS3 Ramsey cavity, determined by beam reversals. MJD Modified Julian Date. Open circles: CS2 is the reference clock. Dots: An active hydrogen maser is the reference clock.

In Table 2 the corrections applied to the CS3 frequency and to the CS4 frequency and their uncertainties are listed.

Table 2. Systematic frequency corrections and uncertainties (in brackets) in parts in 10^{15}

Type of bias	Correction	
	CS3	CS4
C-field	-277307(3)	
cavity phase difference	84(9)	70(7)
2. order Doppler	28(1)	29(1)

Although the properties of both clocks are very similar, their frequencies seem to differ systematically (see Fig. 3). A reason for this could be the shifting influence of magnetic inhomogeneities over the clocks' frequency - known as frequency shifts caused by Majorana transitions. Because of the manufacturing tolerances of the six-pole magnets as well as of differences of the residual earth's magnetic field the CS3 atoms experience an inhomogeneity of the field between state-selecting magnets and the Ramsey cavity, which is slightly different from that of the CS4 atoms. This fact urges upon an investigation. Of lately Majorana shifts are theoretically better understood, there are instructions how to test their presence and to enhance and to suppress them (Ref. 4). Currently such studies are made at CS4 and will be continued at CS3 next. Only when the measurements will have been finished, a final and total frequency uncertainty of both clocks will be given.

4. FINAL REMARKS

The PTB primary clock CS3 has been undergoing an evaluation being preliminary in the sense that frequency shifting influences like spectral impurities of the microwave, Rabi pulling, Majorana transitions, and servo offsets have not been

studied yet. The pleasant experiences with the other clock of the same design, CS4, let expect, however, that the major part of the uncertainty contributions has been already given in Table 2. Assuming this to be true the total frequency uncertainty should not exceed the 10^{-14} level. Then the relative output frequencies of the four PTB primary clocks CS1, CS2, CS3 and CS4, agree within their uncertainty figures.

ACKNOWLEDGMENT

The authors are greatly indebted to D. Griebsch for solving contructional problems and to Chr. Richter for his contribution in realizing the electronic equipment.

REFERENCES

[1] R. Schröder, "Frequency synthesis in primary cesium clocks", in Proc. 5th European Frequ. and Time Forum, 1991, pp. 194-200.

[2] A. Bauch et al., "CS2: The PTB's new primary cesium clock", IEEE Trans.Instrum.Meas., vol. IM-36, June 1987, pp. 613-616.

[3] A. Bauch et al., "The new PTB primary cesium clocks", Special Issue of the IEEE Trans.Instrum.Meas. for CPEM'92, in press.

[4] A. Bauch, R. Schröder, "Frequency shifts in a cesium atomic clock due to Majorana transitions", submitted to Annalen der Physik, 1992.

PROPERTIES OF AN OPTICALLY PUMPED CESIUM BEAM FREQUENCY STANDARD WITH $\Phi = \pi$ BETWEEN THE TWO OSCILLATORY FIELDS.

C. Audoin, V. Giordano, N. Dimarcq, P. Cérez, P. Petit and G. Théobald

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

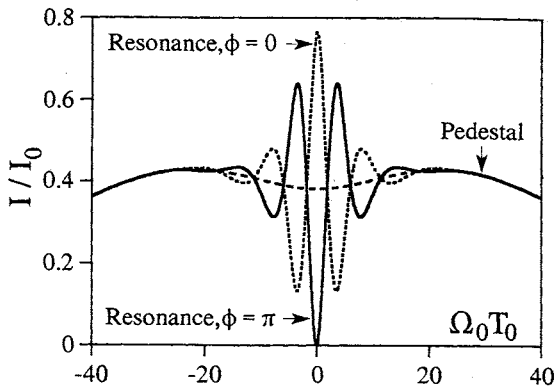
The properties of optically pumped cesium beam resonators showing a phase difference, Φ, of π between the two oscillatory fields are analyzed. They are compared with those occuring with $\Phi = 0$, when the interrogation signal is slowly square-wave frequency modulated. The value of the operating parameters which optimizes the resonator properties is given for $\Phi = 0$ and π, taking into account the effect of the noise sources which perturb the atom detection. It is shown that the design with $\Phi = \pi$ is superior. It provides a better frequency stability and smaller frequency offsets. The conclusion holds for short as well as long tubes.

Keywords : Cesium beam frequency standard, Ramsey cavity, frequency stability, frequency offsets, accuracy, dark resonance.

1. INTRODUCTION

Traditionally, in cesium beam frequency standards, the coupling of the cavity to the microwave supply is designed in order to obtain an even number of longitudinal modes at 9.2 GHz. In that case, the phase difference, Φ, between the oscillatory fields at the two extremities of the cavity is equal to zero. However, the coupling set-up can be arranged to provide an odd number of longitudinal modes in a cavity of appropriate length. We then have $\Phi = \pi$ between the two oscillatory fields that the atoms experience in succession. Such a resonator has been operated recently in cesium beam tubes where state preparation and state detection is accomplished either by magnetic deflection [1, 2] or optical pumping [3].

We shall denote ω and ω_0 the angular frequencies of the interrogation signal and of the atomic transition, respectively. Neglecting momentarily any flux of stray atoms and assuming a flop-in configuration, the flux of atoms detected is equal to zero when we have $\omega = \omega_0$, in the case $\Phi = \pi$ [4]. Then, the resonance pattern shows a minimum at $\omega = \omega_0$, instead of a maximum for $\Phi = 0$, as illustrated in figure 1.



1. Example of computed Ramsey resonance for $\Phi = 0$ and π, assuming $L/\ell = 15$, $b\tau_0 = 1.89$ and $n = 3$. (Notations are that of [4]

It appears that the new design brings a number of advantages, as long as the metrological properties of the cesium beam resonator are concerned. They have been briefly noticed in [1, 3]. In this paper, we deepen the theoretical analysis of the properties of a resonator with $\Phi = \pi$ and we compare them with those occuring in traditional resonators, with $\Phi = 0$.

The results given are derived assuming that the atomic resonance is probed by an interrogation signal which is slowly square-wave frequency modulated, since this modulation waveform appears to be most appropriate for use in advanced cesium beam frequency standards [5]. Furthermore, the velocity distribution of the cesium atoms is assumed that occuring in optically pumped cesium beam tubes, these devices showing promising properties [6]. Results have been computed: a) for long beam tubes, such as Laboratory set-ups, in which the separation L between the two oscillatory fields is much larger than the length ℓ of each interaction region and b) for short beam tubes, like manufactured ones, assuming $L/\ell = 15$.

2. FLUX OF ATOMS HAVING MADE THE ATOMIC TRANSITION

For monokinetic atoms, the probability that a transition occured [7] will be denoted $p(\tau, b, \omega)$. In long laboratory beam tubes we have $L >> \ell$ (e.g. $L/\ell \sim 100$). Then $p(\tau, b, \omega)$ for the central, useful, fringe of the Ramsey pattern is given by :

$$p(\tau, b, \omega) = \frac{1}{2} \sin^2 b\tau \{1 + (\cos \Omega_o T + \Phi)\} \quad (1)$$

Here, b is a measure of the amplitude of the microwave field in each interaction region, expressed in rad s^{-1}. τ is the duration of each interaction with the oscillatory fields and $T = \tau L/\ell$ is the time of flight between these fields. Equation (1) has been used to compute the properties of long tubes. In short tubes, the equation for the probability is more complicated. We have used that given in [4], where it has been set $L/\ell = 15$.

The interaction time τ has a distribution, $f(\tau)$ [4]. In the simplest configuration of an optically pumped cesium beam tube, where a single laser is used, $f(\tau)$ is directly related to the Maxwellian distribution of velocity in a beam, assuming that optical pumping is complete. Results reported in the following apply to that case. It has been verified that the main conclusions given apply as well when $f(\tau)$ refers to the velocity distribution corresponding to the situation where a second laser is used for state detection, via a cycling transition. In both cases, it is convenient to introduce the particular values $\tau_o = \ell / \alpha$ and $T_o = L/\alpha$, where α is the most probable velocity in the cesium oven.

The averaged probability is thus given by :

$$P(b,\omega) = \int_0^\infty f(\tau)\, p(\tau,b,\omega)\, d\tau \quad (2)$$

Therefore, the flux I of atoms having made the transition is equal to $I_o P(b,\omega)$, where I_o is the flux, in the detection region, of atoms which have been prepared in one of the $m_F = 0$ hyperfine ground states.

The outcomes of (2) are expressed using dimensionless parameters. They are : i) $\Omega_o T_o$, ii) $b\tau_o$ which represents another measure of the microwave magnetic field and iii) $\omega_m T_o$ which characterizes the depth of the square-wave frequency modulation.

3. CAVITY PULLING FREQUENCY OFFSET

If the cavity resonant angular frequency, ω_c, is not equal to ω_o, then the microwave amplitude varies from one half-period of modulation to the other. This induced amplitude modulation shifts the measured resonance frequency by $\Delta\omega_c$ given by

$$\frac{\Delta\omega_c}{\omega_o} = \frac{T_c^2}{T_o^2} \Delta_{cav} \frac{\omega_c - \omega_o}{\omega_o} \quad (3)$$

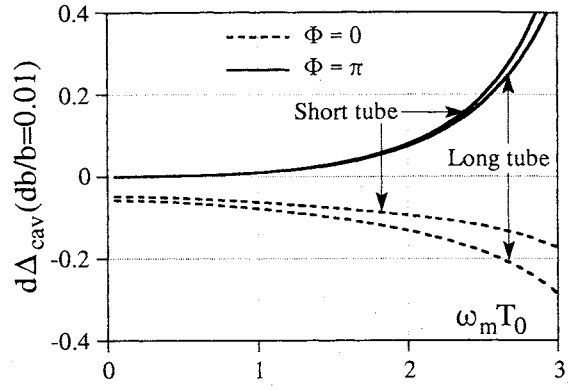
with :

$$\Delta_{cav} = - \omega_m T_o \frac{b\, \partial P(b,\omega_m)/\partial b}{T_o^{-1}\, \partial P(b,\omega_m)/\partial \omega} \quad (4)$$

where the value of the partial derivatives is taken for the values b and $\omega_o + \omega_m$ of the variables. We have $T_c = 2Q_c/\omega_c$ and $\omega_o T_o \sim \pi Q_\ell$, where Q_c and Q_ℓ are the cavity and the atomic line quality factors, respectively.

Variation of the coefficient Δ_{cav} versus b and ω_m has been reported in the case $\Phi = 0$ [8-10] . It can be shown that Δ_{cav} is much smaller for $\Phi = \pi$ than for $\Phi = 0$ when we have $1 < b\tau_o < 2.5$ and $\omega_m T_o < 1.5$, which are the ranges of practical interest. This is related to the fact that the variation of $P(b,\omega_m)$ versus b, around $\omega = \omega_o$, is much smaller for $\Phi = \pi$ than for $\Phi = 0$.

Equation (4) shows that Δ_{cav} is equal to zero, i.e. that the cavity pulling is suppressed, when the flux I of detected atoms goes to an extremum, a maximum in current practice [4, 5,11]. This is one of the major advantages of square-wave modulation.



2. Change of the value of Δ_{cav}, versus $\omega_m T_o$, for 1 % change of the microwave field amplitude.

The condition $\Delta_{cav} = 0$ determines the value of $b\tau_o$ for given $\omega_m T_o$. However, the microwave amplitude may vary around this value. Figure 2 shows the variation of Δ_{cav} for a 1 % change of the microwave amplitude. For $\omega_m T_o < 2$, the variation of Δ_{cav} is smaller for $\Phi = \pi$ than for $\Phi = 0$. It is even much smaller for $\omega_m T_o \to 0$.

The advantage of the case $\Phi = \pi$ is still more pronounced when the frequency offset is related to a spurious amplitude modulation due to imperfections in the electronics.

4. FREQUENCY STABILITY

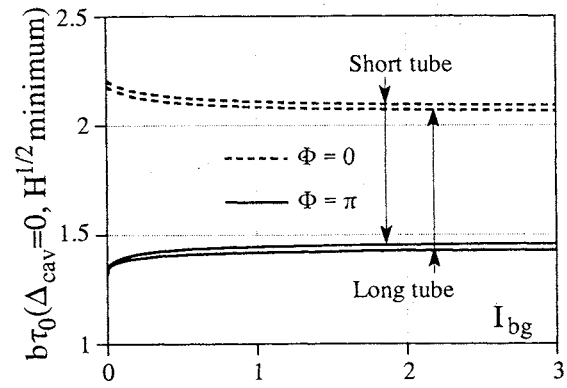
For slow square wave modulation, the white frequency noise limited frequency stability can be expressed from the quantity H, given by [4]

$$H = 2 \frac{I'_b/I_o + P(b, \omega_m)}{\{T_o^{-1}\, \partial P(b,\omega_m)/\partial\omega\}^2} \quad . \quad (5)$$

We have :

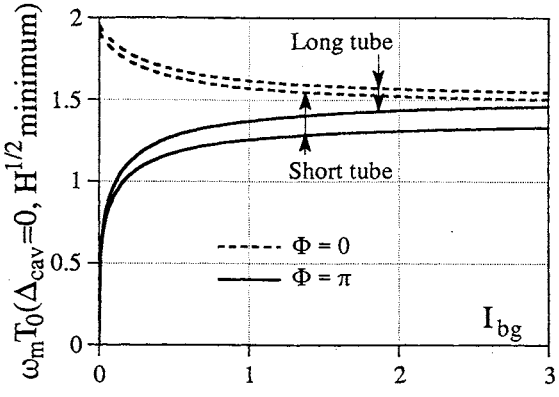
$$\sigma_y(\tau) = \frac{1}{\omega_o T_o} \left(\frac{F}{2I_o}\right)^{1/2} H^{1/2}\, \tau^{-1/2} \quad , \quad (6)$$

where $\sigma_y(\tau)$ is the Allan standard deviation [12] and F the noise factor of the atom detector [13]. In (5), I'_b is a fictitious flux of atoms whose shot noise is equivalent to the noise added by the stray light detected, by the light detector and by any source of extra noise. The atomic beam shot noise is related to the second term of the numerator. I'_b and I_o are calculable from experimental data. We set $I_{bg} = I'_b/I_o$.



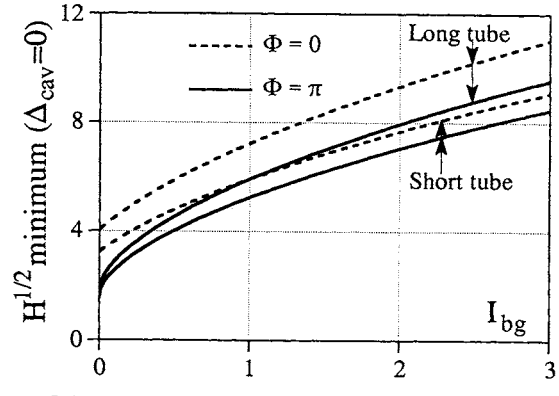
3. Variation, versus I_{bg}, of the microwave field amplitude for which the two conditions $\Delta_{cav} = 0$ and $H^{1/2}$ minimum can be fulfilled.

Assuming, as a first condition, that the microwave amplitude is such that $\Delta_{cav} = 0$, then H depends on two variables only among the three following ones : I_{bg}, $b\tau_o$ and $\omega_m T_o$. If we impose, as a second desirable condition that the frequency stability is optimized, i.e. that $H^{1/2}$ is minimum, then H becomes a function of a single variable, and so does $b\tau_o$ and $\omega_m T_o$. We choose I_{bg} as an independant variable.



4. Variation, versus I_{bg}, of the modulation depth for which the two conditions $\Delta_{cav} = 0$ and $H^{1/2}$ minimum can be fulfilled.

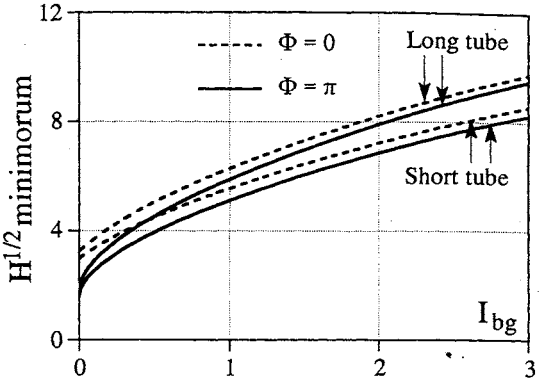
Figures 3 and 4 show the variation, versus I_{bg}, of the values that the parameters $b\tau_o$ and $\omega_m T_o$ must take in order to fulfil the two conditions simultaneously. We see that we have $b\tau_o \sim 1.45$ for $\Phi = \pi$ and $b\tau_o \sim 2.1$ for $\Phi = 0$. For $\Phi = \pi$, $\omega_m T_o$ is also smaller than for $\Phi = 0$ and it tends to zero for $I_{bg} \rightarrow 0$. This can be understood since, for $I_{bg} = 0$, the frequency stability is partly determined by the atomic beam shot noise, which disappears in the dip of the dark resonance pattern, i.e. for $\omega_m = 0$



5. Variation, versus I_{bg}, of the minimum value of $H^{1/2}$ when the condition $\Delta_{cav} = 0$ is fulfilled.

Figure 5 shows the variation of $H^{1/2}$ versus I_{bg}, when the two aforementioned conditions are fulfilled. For a tube of a given type, short or long, the configuration with $\Phi = \pi$ provides the best frequency stability. The advantage is the most marked for small values of I_{bg}. The gain factor can reach a value close to 2.

However, the condition $\Delta_{cav} = 0$, which has been assumed above, constrains the quantities $b\tau_o$ and $\omega_m T_o$ to have values which are not necessarily those which provide the best possible stability. Nevertheless, it happens that, for given I_{bg}, the values of $b\tau_o$ and $\omega_m T_o$ for which $H^{1/2}$ shows an absolute minimum do not differ very much from that of figures 3 and 4, mainly for $\Phi = \pi$.



6. Variation, versus I_{bg}, of the absolute minimum of $H^{1/2}$.

Figure 6 gives the variation of the absolute minimum of $H^{1/2}$, versus I_{bg}. Comparison with figure 5 shows that when the condition $\Delta_{cav} = 0$ is imposed, the frequency stability can be close to the best possible one. With a tube (short or long) having $\Phi = \pi$, the match can be almost perfect.

5. FREQUENCY OFFSETS

Figures 3 and 4 show that the values of $b\tau_o$ and $\omega_m T_o$ must be different for $\Phi = 0$ or π. Therefore, the frequency offsets whose value depend on $b\tau_o$ and on $\omega_m T_o$ [4] will differ in both cases. We shall illustrate this point with two frequency offsets.

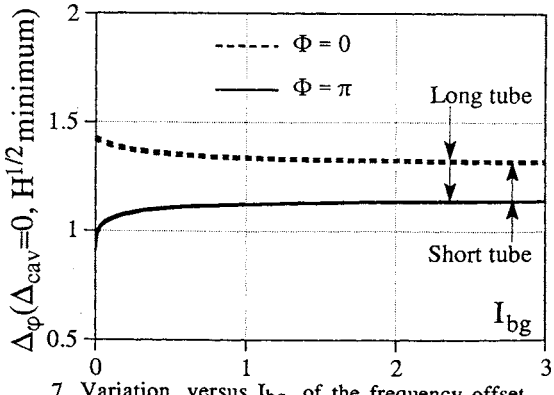
5.1 Frequency offset due to a residual cavity phase shift

If the phase difference between the two oscillatory fields is not exactly 0 or π, then a frequency offset follows. It can be shown that, φ being the spurious phase difference, the frequency offset is given by :

$$\frac{\Delta\omega_\varphi}{\omega_o} = -\frac{\varphi}{\omega_o T_o}\,\Delta_\varphi \quad , \tag{7}$$

with

$$\Delta_\varphi = \frac{\int_0^\infty T^{-1} f(\tau)\partial p(b,\omega_m)/\partial\omega \, d\tau}{\int_0^\infty T_o^{-1} f(\tau)\,\partial p(b,\omega_m)/\partial\omega \, d\tau} \tag{8}$$



7. Variation, versus I_{bg}, of the frequency offset coefficient Δ_φ.

Figure 7 shows the variation of Δ_ϕ, versus I_{bg}, when the two conditions $\Delta_{cav} = 0$ and $H^{1/2}$ minimum are prescribed. The value of the coefficient is smaller for $\Phi = \pi$ and it does not practically depend on the tube length.

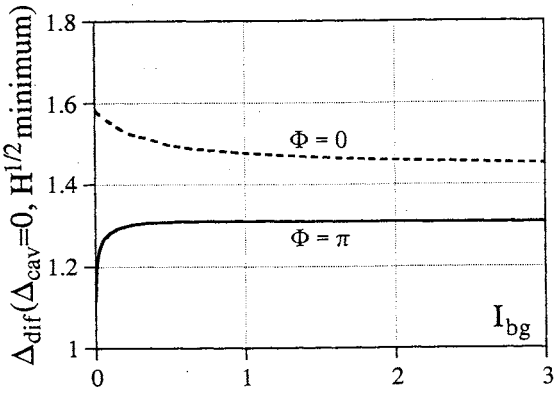
5.2 Frequency offset due to frequency shifts in the microwawe interaction regions

If the atomic transition frequency is not the same in the two microwave interaction regions as in the intermediate zone, then an offset of the observed resonance occurs. The source of this imperfection can be : an inhomogeneous magnetic field, a light shift, or a spectral impurity away from the microwave spectrum of the cesium atom. The frequency shift is given, to first order in ℓ /L, by :

$$\frac{\Delta\omega_{dif}}{\omega_o} = \frac{\ell}{L} \frac{\Delta\omega'_o + \Delta\omega''_o}{2\omega_o} \Delta_{dif} \quad , \qquad (9)$$

where $\Delta\omega'_o$ and $\Delta\omega''_o$ are the frequency offsets in the two oscillatory fields regions. We have :

$$\Delta_{dif} = \frac{2\int_o^\infty f(\tau)\,(1 - \cos b\tau)\,\sin b\tau\,\sin \omega_m T\, d\tau}{\int_o^\infty b\tau\, f(\tau)\,\sin^2 b\tau\,\sin \omega_m T\, d\tau} \qquad (10)$$



8. Variation, versus I_{bg}, of the frequency offset coefficient Δ_{dif}.

Figure 8 shows the variation of Δ_{dif}, versus I_{bg}, when the two conditions specified are verified. Again, the value of the coefficient is smaller for $\Phi = \pi$.

5.3 Other frequency offsets

Frequency offsets due to the 2nd order Doppler effect, to a possible asymmetry of the cesium microwave spectrum (Rabi pulling) and to neighbouring spectral impurities have also been calculated. The advantage of the case $\Phi = \pi$ is even larger for them than for those considered above.

6. CONCLUSION

We have shown that it is more favorable to design a Ramsey cavity with $\Phi = \pi$, instead of $\Phi = 0$, between the two oscillatory fields, as long as the frequency stability and the frequency offsets are considered. In the case of an optically pumped cesium beam tube, this property helps to compensate for the larger atom velocities.

Acknowledgements

We gratefully acknowledge the support of STEI.

REFERENCES

[1] Qin Zhuo, Yuan Xiping and Yang Yuquing. International Symposium on Electromagnetic Metrology, 1989, Pékin (China).

[2] K. Nakagiri. Private communication, 1993.

[3] Petit P., Giordano V., Dimarcq N., Cérez P., Audoin C. and Théobald G. in Proc. 6th European Frequency and Time Forum, ESA Publication Division, ESTEC, Noordwijk (The Netherlands), 1992, 83-86.

[4] Vanier J. and Audoin C. *The Quantum Physics of Atomic Frequency Standards*. Bristol and Philadelphia. I.O.P. Publishing, 1989, 1767 p.

[5] Cutler L.S. and Giffard R.P. in Proc. IEEE Frequency Control Symposium. IEEE. Piscataway, N.J. (USA), 1992, 127-133.

[6] Audoin C. Metrologia, 1992, 29, 113-134.

[7] Ramsey N.F. *Molecular Beams*. Oxford, Clarendon Press, 1956, 466 p.

[8] De Marchi A., Rovera G. D. and Premoli A. IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control, 1987, 34, 582-591.

[9] Rovera G.D., De Marchi A. and de Clercq E. in Proc. Fourth European Frequency and Time Forum, Scientific Communities of Besançon (France) and Neuchâtel (Switzerland), 1990, 161-162.

[10] Audoin C., Dimarcq N., Giordano V. and Viennet J. IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control, 1992, 39, 412-421.

[11] Audoin C., Lesage P. and Mungall A.G. IEEE Transactions on Instrumentation and Measurement, 1974, 23, 501-508.

[12] Barnes J.A., Chi A.R., Cutler L.S., Healy D.I., Leeson D.B., Mc Gunigal T.E., Mullen J.A., Smith W.L., Sydnor R.L., Vessot R.F.C., Winkler G.M.R. IEEE Transactions on Instrumentation and Measurement, 1971, 50, 105-120.

[13] Giordano V., Candelier V., Hamel A., Audoin C., Théobald G. and Cérez P. Optics Communications, 1988, 67, 287-292.

THE HIGH C-FIELD CONCEPT FOR AN ACCURATE CESIUM BEAM RESONATOR

Andrea De Marchi

Dip. di Elettronica, Politecnico di Torino, Torino, Italy

Abstract

The idea is proposed of operating a Cesium resonator at a high value of the magnetic field, chosen in such a way ($B \approx 82$ mT) as to minimize the frequency of the $m_F = -1$ transition. Such minimum frequency turns out to be at $\nu_{-1min} = \nu_{00}\sqrt{15/16} \approx$ 8.900727438257 GHz, and would then be used as the reference. At this field, neighboring transitions are very far from it, and Rabi or Ramsey pulling and Majorana transitions become unimportant.

Magnetic field uniformity and stability requirements for a high accuracy resonator are discussed, and the structure of a prototype realization of a resonator based on this concept is illustrated.

1 Introduction

A number of systematic effects, that are relevant at the level of 10^{-13} in Cesium frequency standards based on thermal beams, must be dealt with in some way before accuracy targets approaching 10^{-14} can be attained [1]. The main concerns in designing toward this goal are:

1. The Zeeman effect, which is of the order of 10^{-10} at the normally employed magnetic (C-field) levels. The obtainable field uniformity along the beam produces an uncertainty in its determination which is often a limit to the accuracy of the standard. Lowering the field helps, but increases effects due to the existence of other energy levels in Cesium.

2. The existence of energy levels which are too close in energy for the two-level model to be valid at the desired level of accuracy. Bias effects related to this are Rabi [2] and Ramsey [3] pulling and effects induced by Majorana transitions [4]. All these can lead to biases of the order of 10^{-13} that are difficult to assess. Increasing the field helps, but makes C-field related problems increasingly difficult. Equal populations in equal $|m_F|$ sublevels would make these effects vanish.

3. The end-to-end phase difference in Ramsey cavities, which is difficult to reduce below few tens of μrad, equivalent to a shifts of the order of 10^{-13} in thermal beams. The accuracy of its evaluation is limited by phase variations across the beam within the microwave cavity. It can be curbed by the use of narrow beams [5], at the expenses of signal to noise ratio, or by resorting [6, 7] to innovative cavity configurations. Reducing uncertainty on this effect below 10^{-14} is always very difficult.

4. The second order Doppler shift, which can be up to several 10^{-13} in thermal beams, and can be reduced only by using slower atoms. Uncertainty reduction on this effect can require accurate velocity distribution measurements.

5. The fact that the average atomic velocity and the velocity distribution are very important both for the line Q and for items 2) to 4). Wide distributions increase the atomic flux, and as a consequence the S/N ratio, but make the second order Doppler shift difficult to determine. A low average velocity decreases all these systematic effects, at the expense of short term stability.

The task of designing for the best accuracy is made complex by the fact that all these aspects of the problem interact with one another, so that it is hard to minimize all bias effects at the same time.

This work is partially funded by:
- Consiglio Nazionale delle Ricerche
- Istituto Nazionale di Fisica della Materia
- Agenzia Spaziale Italiana

2 Established configurations

Two main configurations have been identified over the years, which can approach an accuracy of 10^{-14} with thermal beams [8].

One is characterized by the use of magnetic selection, and is generally refferred to by the qualification of "traditional". In this configuration the second order Doppler effect is minimized by the selection of low velocities, and is quite independent of microwave power because the velocity distribution is narrow. Effects related to neighbouring states, as well as cavity phase shift, are reduced by the increase in line Q gained from the use of low velocities, and a narrow beam [5] coupled to a ring terminated Ramsey cavity [6] can make distributed phase shift very small. On the other hand, because of the magnetic selection process, it is not possible to obtain equal populations and velocity distributions in energy states with equal $|m_F|$. Therefore all effects related to the multilevel character of Cesium are enhanced. This is cured by incresing the C-field, but a limit to this is set by the increasing field uniformity and stability requirements which are then imposed by the need to evaluate and correct the Zeeman bias with high accuracy. A compromise must be found between these conflicting needs. Accuracies in the low 10^{-14}s have been claimed for standards operated at fields of the order of $10\,\mu$T [5].

The other is characterized by optical preparation and detection, and is referred to as the "optically pumped" beam. In this configuration, because the full Maxwellian velocity distribution is used, a high beam flux can be obtained even with narrow beams, together with high symmetry in the population of Zeeman sublevels. The former leads to high short term stability, which is both valuable in itself and a great advantage in the evaluation of accuracy. The flux is so high that even the use of a narrow beam can yield very good short term stability. This allows to reduce phase shift uncertainties by the adoption of a ring terminated Ramsey cavity [9]. The latter makes a standard greatly immune to effects due to neighboring transitions. Higher convenience in the realization of more uniform C-fields has been also claimed to be an advantage of optically pumped configurations over traditional ones. The price paid for these features is an increase in the average beam velocity, which in turn increases the second order Doppler effect. Moreover, the wide velocity distribution makes it power dependent and difficult to evaluate. A number of optically pumped beams have been realized in recent years [9, 10, 11].

3 The high field concept

In this paper a different approach is presented, in which the C-field is chosen in such a way (about 82 mT) as to minimize the frequency of the $m_{F=-1}$ transition, which is then used as the reference. From the Breit-Rabi formula (which takes into account only the hyperfine interaction) such minimum turns out to be at $\nu_{-1min} = \nu_{00}\sqrt{15/16} \approx 8.900727438257$ GHz. High order correction terms, taking into account the influence of other electronic levels, may have to be added [12]. In this configuration, the closest neighboring transition moves 600 MHz away from the reference, and a situation very close to the ideal two-level system is realized. Rabi and Ramsey pulling, and Majorana transitions become then unimportant.

The genesis of this approach takes off from the idea of keeping at least the first (A) magnet for velocity selection purposes, in order to decrease velocity dependent shifts and cavity length. High useful beam flux can then be obtained only by the use of a wide beam. In order to allow this, a number of problems must be solved:

1. A wide useful gap in state selectors is needed. The chosen magnet style is of the parabolic pole piece type [15], with an asymmetry designed to maximize the area where a high transversal field gradient has good uniformity.

2. A wide area detector is needed. Optical detection on a cycling transition is being considered along with the traditional hot wire device.

3. Large beam passing holes are needed in the microwave cavity, along with good phase uniformity across the beam in the cavity mode. These requirements can be met by employing a cylindrical cavity in the TE_{01} mode. In fact, the latter has field nodes in the center of the end caps, and is not significantly disturbed by circular holes up to diameters of more than 10 mm. Furthermore, due to its very low losses, it has a phase of the longitudinal magnetic field component which, at the edge of that circle, is only a few μrad different from the center [16]. A seven half wavelengths Rabi cavity is planned for the prototype, as Doppler related effects, including end-to-end phase shift, are expected to be very small in this configuration [7].

4. Ramsey pulling must be kept small. Opening up the beam passing holes in the cavity allows atoms to experience sizeable microwave field components normal to the C-field. It is then to be expected that Ramsey pulling may become huge. Although no specific calculation

has been carried out for the mode and the case under consideration, it can be guessed by comparison with better known situations that a C-field value of at least 100 μT (about 500 kHz of Zeeman frequency) may be needed in order to separate in frequency neighboring transitions enough to reduce Ramsey pulling below 10^{-14}. Another solution, proposed in this paper, is to work at high field (82 mT) so as to minimize the frequency of the $m_F = -1$ transition. In both cases high stability, uniformity and accuracy are required of the magnetic field.

In the following paragraph an analysis of field specifications which must be met for an accuracy of 10^{-14} is spelled out for both cases of strong "low field" and "high field". It will be shown that high uniformity is required in the second case, higher than would be required in the first, but probably still feasible. On the other hand, field accuracy requirements are much more relaxed in the high field case, and should be quite easy to meet, while they appear awesome in the strong low field case. For this reason the decision was made to try and operate a beam in a high field mode.

The structure of a prototype realization of an experimental beam machine based on this idea is here illustrated, together with a possible architecture for the frequency synthesis system which must both excite the clock transition and measure the C-field on a field dependent line. Functional sub-assemblies are in the development stage. Some have already been realized and are in the process of being tested. Details on the microwave cavity [13] and the uniform high field magnet [14] are given in accompanying papers in this conference. In this paper, interactions among the designs of these subsystems are discussed from the point of view of the accuracy of the whole research prototype.

4 C-field specifications

The accuracy with which the Zeeman effect frequency bias can be evaluated depends on accuracy, uniformity and stability of the field itself, and on its absolute value.

The two cases of high field and strong low field are considered. Closed loop operation with servo control of the field based on frequency measurement of a field dependent transition is assumed, as it will be shown below that in both cases stability and accuracy requirements of the C-field current supply would be too stringent, at the target level of performances, to be satisfied by even a very good supply in open loop operation.

An order of magnitude estimate of biases and uncertainties introduced by field non-idealities is given here. A correction for the bias induced by disuniformity can in principle be calculated once the actual field profile is known. Linear and quadratic terms must be considered in the expansion of the Breit Rabi formula about the field value which minimizes the frequency of the chosen transition. The importance of higher order terms will be discussed.

The low field case

Close to ν_{00} (frequency of the $m_F = 0$ transition at $B = 0$) is:

$$\frac{\nu_0}{\nu_{00}} = \sqrt{1+x^2} \approx 1 + \frac{x^2}{2} - \frac{1}{2}\left(\frac{x^2}{2}\right)^2 \quad (1)$$

$$\frac{\nu_{\pm 1}}{\nu_{00}} = \sqrt{1 \pm \frac{x}{2} + x^2}$$

$$\approx 1 \pm \frac{x}{4} + a_0\frac{x^2}{2} \mp a_0\frac{x}{4}\frac{x^2}{2} \quad (2)$$

with x proportional to B and $a_0 = 15/16$.

Taking only first terms in (1) and (2) and identifying the Zeeman frequency by $f_z/\nu_{00} = x/4$, the Zeeman (C-field) shift $\delta\nu_c$ can be evaluated as:

$$\frac{\delta\nu_c}{\nu_{00}} = \frac{x^2}{2} = 2\left(\frac{2f_z}{\nu_{00}}\right)^2 \quad (3)$$

and deriving (3) it can be found that, if the used value of f_z is for whatever reason incorrect, the error $\delta(\delta\nu_c)$ in determining the shift is given by:

$$\frac{\delta(\delta\nu_c)}{\nu_{00}} = 4\left(\frac{2f_z}{\nu_{00}}\right)^2\frac{\delta f_z}{f_z} \quad (4)$$

For example, the errors induced in the C-field bias evaluation by overlooking higher order terms in (2) can be found from (4). When the value of f_z is taken to be the measured difference $\nu_1 - \nu_0$, as is rather commonly done, the relative error on f_z is $\delta f_z/f_z = -x/8 = -\frac{1}{4}(2f_z/\nu_{00})$. According to (4) the error introduced in the C-field bias evaluation is then given by:

$$\frac{\delta(\delta\nu_c)}{\nu_{00}} = -\left(\frac{2f_z}{\nu_{00}}\right)^3 \quad (5)$$

In the envisioned strong low field case (100 μT, f_z=500 kHz) this amounts to a relative error of 10^{-12}, which means that (3) shouldn't be used as such. First order correction of (3) is to take (5) into account, as is done in [5]. Better it is, at least for strong fields, to calculate $\delta\nu_c$ from a Zeeman frequency obtained as $f_z = \frac{1}{2}(\nu_1 - \nu_{-1})$, so that the quadratic term in (2) disappears and (3) can be used

accurately down to the fourth order term. The error introduced in this way in the considered strong low field case is smaller than 10^{-15}.

In all cases, (4) is still valid to calculate uncertainties in $\delta\nu_c$, and can be used in reverse to figure out the accuracy needed on f_z for a desired accuracy in $\delta\nu_c$. At 100 μT it turns out that it must be $\delta f_z/f_z < 2{\cdot}10^{-7}$ for an accuracy of 10^{-14} in C-field bias. This requires the field to be stable to that level, including current supply and geometrical configuration of the magnetic circuit. The latter appears to be very hard to get as it would imply, if nothing else, a stability of a few mK of the whole apparatus. Servo control of the excitation current is therefore needed. In fact, geometrical effects of temperature variation could then be tracked and corrected, and the necessary current stability can be provided. What's worse is that frequency measurements of the $|m_F| = 1$ field dependent transitions accurate to at least 10^{-11} are needed for this accuracy on f_z. This is due to the fact that f_z is obtained as a difference of microwave frequencies, which are higher than f_z by a factor of $2{\cdot}10^4$ in the considered case. Such accuracy appears very hard to obtain because a number of known biases are much more difficult to evaluate on field dependent transitions than they are on the clock transition. So is for example the phase shift, as it interacts with the end-to-end difference in microwave power level [17].

If this were not the case, and it were in fact possible to measure with very high accuracy the frequency of field dependent transitions, then it would be possible to evaluate $\delta\nu_c$ from the difference between $\nu_1 - \nu_0$ and $\frac{1}{2}(\nu_1 - \nu_{-1})$. It would then be $\delta\nu_c = 16[\nu_0 - \frac{1}{2}(\nu_1 + \nu_{-1})]$. However, because of the inaccuracy on field dependent transitions, the value obtained in this way cannot be expected to agree with (3). In fact, this little calculation might be useful to provide an estimate of the accuracy of ν_1 and ν_{-1}, which may be hard to get otherwise. If the latter were found to be very good, this method could actually be useful for the determination of the C-field bias, because it intrinsically takes into account the difference between linear and quadratic averages of the field along the beam. It is well known that such difference produces a bias in the estimate of $\delta\nu_c$. Given the difficulty of applying the method just described, the only choice left is to get an independent estimate of the field uniformity and use that to infer the bias. Point by point field measurements before the assembly is completed are necessary to this aim, and an uncertainty often remains as to whether the uniformity is the same when the beam is operational.

For the evaluation of this bias, second order terms in the displacements from the average of the field must be kept when expanding the parabolic Zeeman effect. From (1) it can be seen that $\delta\nu_c = k_c B^2$, where the well known formula $x = (g_J - g_I)\mu_B B/h\nu_{00}$ allows to calculate that $k_c = 4.2745{\cdot}10^{10}$ Hz/T^2. It is then, indicating by an overline the spatial average, by ΔB the difference between what is thought to be the average field and the actual average field, and by δB the local difference between field value and average field ($\overline{\delta B} = 0$):

$$\frac{\delta\nu_c}{k_c} = \overline{B^2} = \overline{(\overline{B} + \Delta B + \delta B)^2} = \overline{B}^2\left[1 + 2\frac{\Delta B}{B} + \overline{\left(\frac{\delta B}{B}\right)^2}\right] \qquad (6)$$

where the square of $\Delta B/B$ has been overlooked because much smaller than what was kept. It's easy to recognize that the last term in (6) is the variance σ_B^2 of the relative field variations along the beam. The square root of this we will take as a measure of the uniformity of the field in the relevant volume. From (6) the bias uncertainty can also be written:

$$\frac{\delta(\delta\nu_c)}{\nu_{00}} = \frac{\delta\nu_c}{\nu_{00}}\left(2\frac{\Delta B}{B} + \sigma_B^2\right) \qquad (7)$$

where it can be seen that the first term reproduces (4) by substituting (3) and noting that $\Delta B/B = \delta f_z/f_z$. Comments about this term have already been made. The second term imposes a non-uniformity limit of $5{\cdot}10^{-4}$ for a frequency accuracy of 10^{-14} at the considered strong low field.

A further question mark lurking on accurate correction of large Zeeman effects is that the accuracy level of the Breit Rabi formula is not presently known. Experiments on Rubidium have shown that additional contributions exist, beyond the hyperfine interaction Hamiltonian, due to other electrons in the atom [12]. It will be interesting in this perspective to know with an accuracy of say 10^{-14}, when feasible, the frequency difference between the PTB Cs2 standard, which is operated at a rather strong field, and the new optically pumped standard NIST-7, which is operated at a low field. Consideration of this problem is probably less important in the high field case — apart from the fact that the clock frequency may not be exactly the one calculated by the Breit Rabi formula — because here the system is operated in principle at the clock frequency and not at some (Zeeman) offset which must then be evaluated accurately. In fact, the availability of an accurate value for the minimum frequency of the $m_F = -1$ transition may be of great help in understanding the problem.

The high field case

At B_{min} (field at which $x = 1/4$ and ν_{-1} is minimized) the Breit Rabi formula yields $\nu_{-1min} = \nu_{00}\sqrt{15/16} = \nu_{00}\sqrt{a_0}$ and $\nu_{0min} = \nu_{00}\sqrt{17/16} = \nu_{00}\sqrt{a_1}$. As for numerical values, the former was already given in section 2, while the latter is $\nu_{-1min} = 9.475547941$ GHz. Expansion about $y = x - 1/4 = 0$ gives:

$$\frac{\nu_{-1}}{\nu_{-1min}} = \sqrt{1 + \frac{y^2}{a_0}} \approx 1 + \frac{y^2}{2a_0} \quad (8)$$

$$\frac{\nu_0}{\nu_{0min}} = \sqrt{1 + \frac{y}{2a_1} + \frac{y^2}{a_1}} \approx 1 + \frac{y}{4a_1} + \frac{1}{2}\left(\frac{y}{a_1}\right)^2 \quad (9)$$

It can be noticed by comparing (1) and (8) that in both high and low field cases the Zeeman effect on extremized frequencies at their minimum is quite similar. In fact it is $\nu_{-1} - \nu_{-1min} \approx k'_c(B - B_{min})^2$, with $k'_c = k_c/\sqrt{a_0}$, which amounts to $k'_c = 4.4147{\cdot}10^{10}$ Hz/T^2. Also similar is the sensitivity to the field of the transition used for field measurements, which in this case is ν_0.

The qualifying difference in this scheme is that ideal operation at the desired field would produce no C-field bias, as opposed to producing a bias $\delta\nu_c$ which is then evaluated with an uncertainty $\delta(\delta\nu_c)$. In practice this is not completely true because the limited uniformity of the field along the beam does introduce a bias. A bias is also introduced by the limited ability to set the field exactly at B_{min}. Both will always be positive because of the purely quadratic Zeeman effect. It is:

$$\frac{\delta\nu_c}{k'_c} = (\Delta B)^2 + \overline{(\delta B)^2} \quad (10)$$

However, an estimate of the field accuracy can be given, which will yield an estimate of this positive bias. The uncertainty due to this cause can then be taken to be about half of the bias, and a negative correction of the same amount can be introduced on the frequency. Much along the same line it is in principle possible to evaluate the bias caused by lack of uniformity if the field profile is known. This also will leave an uncertainty $\delta(\delta\nu_c)$ on the bias $\delta\nu_c$, related to the uncertainty $\delta(\delta B)$ with which the field values are known. It is then:

$$\frac{\delta(\delta\nu_c)}{k'_c} = \frac{1}{2}(\Delta B)^2 + \overline{(\delta(\delta B))^2} \quad (11)$$

As a result of this analysis, limits can be stated for the inaccuracy of the average field and the uncertainty of field uniformity measurements, for a given frequency accuracy goal. If the latter is 10^{-14}, the first should be better than $8{\cdot}10^{-7}$ and the second should be better than $5{\cdot}10^{-7}$.

For the first contribution this limit does not seem very different from the one found for the considered low field case. However, due to the much greater ratio f_z/ν_{00}, an accuracy of only $2{\cdot}10^{-8}$ on frequency measurements of the ν_0 line is necessary in this case. This is very reasonable even on field dependent transitions. On the other hand the ratio ν_0/ν_{-1} at B_{min} is known from the Breit Rabi formula to be $\sqrt{17/15}$, and an experimental accuracy of that level on this ratio should not be very difficult to obtain. Furthermore, even if doubts on the validity of the Breit Rabi theory were to raise questions on that ratio, careful tracing of the parabolic curve of ν_{-1} versus ν_0 with stability of the order of 10^{-13} for ν_{-1} and 10^{-8} for ν_0 should be plenty to identify the vertex of the parabola with the desired accuracy. It should be a matter of electronic circuit implementation to stabilize the field on the ν_0 line to that degree. In fact the Q projected for the resonance line with the designed cavity is of the order of $5{\cdot}10^7$.

For the second contribution, the stated limit ($5{\cdot}10^{-7}$) on (rms) field measurement uncertainty is right at the limit of available commercial NMR field measurement equipment. The most difficult aspect of the measurement is the stability of the field for the duration of the acquisition of the whole field profile, as is illustrated in [14]. Furthermore, given a certain measured field profile, there may be uncertainties coming from the calculation process used to figure out the actual frequency bias due to non-uniformity. This may be particularly true in the Rabi cavity planned for the experimental prototype, because the actual time of interaction with the microwaves cannot be considered short compared to the drift time, as it is in a Ramsey cavity. Therefore it should be held wise not to count on an improvement in excess of, say, a factor of 100 from $\delta\nu_c$ to $\delta(\delta\nu_c)$, that is of a factor of 10 from the rms δB to the rms $\delta(\delta B)$. A reasonable scenario might be an rms uniformity of $3{\cdot}10^{-6}$, which would produce a bias of $3{\cdot}10^{-13}$, and an average rms relative uncertainty of $5{\cdot}10^{-7}$, which would yield an uncertainty smaller than 10^{-14} in frequency. This level of accuracy may not be easy, but should be feasible.

5 Experimental realization of the prototype

In fig. 1 a sketch of the experimental configuration is given, which shows the relative position of source and detector, cavity, C-field magnet and deflecting magnets. The whole vacuum system is made with

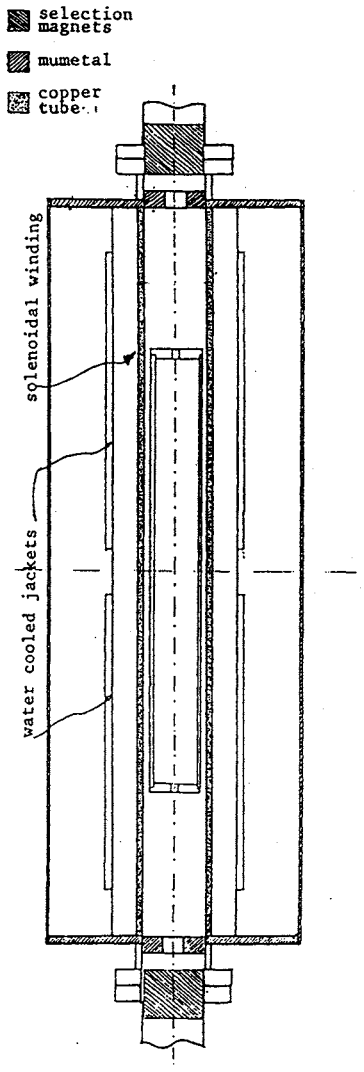


Figure 1: Overview of the experimental apparatus.

standard NW63 CF fittings. Only the 0.7 m long central section, where the C-field is excited, was custom fitted with 5 mm thick copper tubing.

The C-field magnet itself is realized with a multilayer solenoid wound directly on this copper vacuum containment tube. Two 8 mm thick mumetal plates act as magnetic mirrors at the two ends, and a 2 mm thick mumetal tube which encloses the whole magnet constitutes the field return path. The end mirrors are completed within the vacuum envelope by thick mumetal caps with center holes 22 mm in diameter to allow the passage of the NMR probe for the field measurements. At the ends of the magnet, low reluctance magnetic coupling is guaranteed between end plates and inner caps by thin (0.2 mm) stainless steel collars silver brazed to the copper tube.

The mumetal part of the magnetic circuit serves

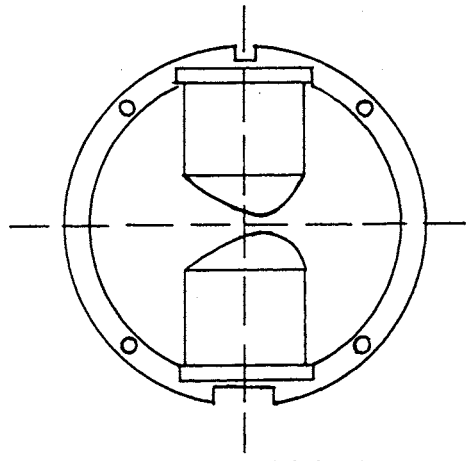


Figure 2: Cross section of deflecting magnets.

also as a shield against external fields. Since the field will be necessarily operated under closed loop, only the effects on uniformity of a limited shielding factor are of particular relevance. It was assumed that these may be smaller than the residual field. If the ratio of the two is about 10, a shielding factor of only 100 should be enough not to spoil the uniformity of the high magnetic field. At any rate, the thickness of the return tube was chosen so that the mumetal would be driven to the maximum permeability point, where μ_r is of the order of $2{\cdot}10^5$. In this way its 2 mm thickness is equivalent to a shield of more than 6 mm at low field.

The design of magnetic circuits was assisted by numerical calculations based on finite elements 2-D aximmetric and 3-D codes for the C-field and deflecting magnets respectively. In fig. 2, a and b, a cross sectional view of the 50 mm long magnets is shown, together with field and field gradient in the middle plane, calculated from the manufacturer's B-H curve for the Alnico 8 drivers.

No special need for mechanical accuracy is present in any part of the apparatus, except for the uniformity of the copper tube's outside diameter, which should be good to reduce final trimming on the field uniformity, and the inside diameter of the cylindrical Rabi cavity, which should be accurate because the mode is used near cutoff.

The cavity resonates at 8.900 GHz, on the extremized ν_{-1} line, with the TE_{017} mode, shown in fig. 3, and at 9.475 GHz, on the ν_0 line, with the $TE_{0\,11\,1}$ mode, which will be used to measure and servo control the C-field value. The same cavity also resonates with the TE_{019} mode close to 9.2 GHz, and can be tuned on the low field ν_0 line for low field operation. An analysis of line shapes and widths expected from this cavity can be found in [7]. A line Q of $5{\cdot}10^7$ is projected in the prototype appa-

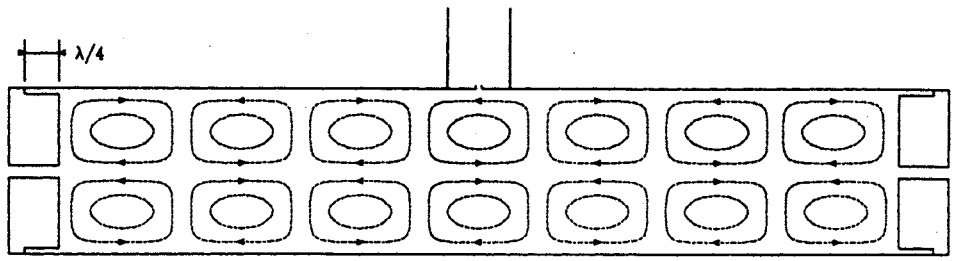


Figure 3: Cross section of the considered cavity with magnetic field lines.

ratus with an average beam velocity of 100 m/s.

6 Short term stability

The whole beam is 1.2 m long, source to detector, which makes for a useful atomic flux of more than 10^{10} atoms/s, in a velocity window of 10% centered around 100 m/s, with a 2 mm diameter source and a square centimeter detector area. Short term stability projections from this are conservatively in the low 10^{-13} in one second, which would be more than adequate for initial accuracy evaluations.

7 Projected accuracy budget

Based on the discussion spelled out in the preceding paragraphs it is now possible to summarize a projected accuracy budget for a system conceived, as illustrated, to realize the high field concept.

In Table I a list of all known effects which may be relevant at the level of 10^{-14} is reported, together with an evaluation of the bias expected from each effect and the uncertainty with which it may be possible to correct it.

Most of these biases are discussed either in this paper or in [13]. For the remaining few, the uncertainty indicated in the table may need comments.

1. *The 2^{nd} order Doppler shift* shown is relative to an average atomic velocity of 100 m/s. Variations in microwave power will change the effective average velocity and in turn this will change the shift. Because of the rather narrow distribution obtained with magnetic deflection, the latter should not change by more than a few percent even for power variations of the order of 1 dB. It should not be necessary to stabilize the power in order to keep uncertainties due to this effect well below 10^{-14}.

2. *The black body shift* shown is calculated at 50°C, where the heat from the resistive magnet will keep the temperature of the cavity in its bore. Such temperature will be controlled by the magnet cooling system for other reasons, and is not expected to vary by more than 1°C under any circumstances. This bias should therefore not vary by more than 1%, as indicated in the Table 1.

3. *Accuracy of the electronics* may be a critical point as the linewidth of the resonator is not expected to be very narrow (≈160 Hz with atoms at 100 m/s). In particular it may be necessary to stabilize the microwave power, because the signal depends rather steeply on power at the power level at which it seems to be best to operate the cavity. However, similar problems have been solved before, and it should be feasible to perfect the electronics so that its accuracy become irrelevant.

8 Conclusions

In this paper the scientific and technical motivations are spelled out for the research program on the accuracy of Cesium frequency standards which is in progress at the Politecnico di Torino. A prototype beam machine is under construction in which a number of non traditional solutions have been adopted. The most relevant ones are the use of a single multiwavelength cavity and the possibility to operate at the high magnetic field of 82 mT, where the frequency of the $m_F = -1$ transition is minimized. The possibility of obtaining with this scheme an accuracy of the order of 10^{-14} is analyzed.

Acknowledgments: The author wishes to acknowledge discussions with many colleagues and encouragement in particular from L. S. Cutler of Hewlett Packard, and D. J. Wineland and R. E. Drullinger of NIST. A special thank goes to the Politecnico di Torino and its Dipartimento di Elettronica for the enthusiastic support granted to this project.

Effect	Bias $\times 10^{14}$	Uncertainty $\times 10^{14}$	Conditions
C-field (accuracy)	< 1	< 1	$\Delta B/B < 5{\cdot}10^{-7}$
C-field (uniformity)	30	< 1	$\sigma_B < 3{\cdot}10^{-6}$ $\quad \sigma_{\delta B} < 5{\cdot}10^{-7}$
Cavity pulling		< 1	with servo tuning
Distributed phase shift		$< .3$	—
Doppler 1^{st} order		< 1	end caps equal to 1%
Doppler 2^{nd} order	5	< 1	magnetic selection
Black body AC Stark shift	2.5	$\ll .1$	temp. stab. 1°C
Electronics		< 1	a pinch of black magic

Table 1: Projected accuracy budget.

References

[1] J. Vanier, C. Audoin, *The quantum physics of atomic frequency standards*, Adam Hilger, Bristol (1989).

[2] A. De Marchi, G. D. Rovera, A. Premoli, "Pulling by neighbouring transitions and its effects on the performance of caesium-beam frequency standards" *Metrologia*, **20** 37–47 (1984).

[3] L. S. Cutler, C. A. Flory, R. P. Giffard, A. De Marchi, "Frequency pulling by hyperfine sigma transitions in cesium beam atomic frequency standards" *J. of Appl. Phys.* **69** 2780–2792 (1991).

[4] A. Bauch, R. Schröder, "Frequency shifts in a cesium atomic clock due to Majorana transitions" submitted for publication on *Annalen der Physik.*

[5] A. Bauch, H. de Boer, B. Fischer, T. Heindorff, R. Schröder, "Planned and realized improvements to the PTB's primary Cesium clocks" *IEEE Trans. on Instrum. Meas.*, IM-40, 153 (1991).

[6] A. De Marchi, J. Shirley, D. J. Glaze, R. Drullinger, "A new cavity configuration for Cesium beam primary frequency standards" *IEEE Trans. Instrum. Meas.*, IM-37, 2 (1988).

[7] A. De Marchi, R. E. Drullinger, J. H. Shirley, "Interference fringes from single-cavity excitation of an atomic beam" *Proc. 44^{th} FCS* (1990). E. Bava, P. Tavella, A. De Marchi, "Doppler effects in a multi-wavelength Rabi cavity for the excitation of atomic beams in microwave frequency standards" *Proc. 46^{th} FCS* (1992).

[8] A. Clairon, A. De Marchi, M. Granveaud, "Accuracy in primary caesium frequency standards" *Metrologia* **28**, 429–430 (1991).

[9] R. E. Drullinger, J. H. Shirley, J. P. Lowe, D. J. Glaze, "Error analysis of the NIST optically pumped primary frequency standard" to be printed on *IEEE Trans. I&M*, April (1993).

[10] S. Ohshima, Y. Nakadan, T. Ikegami, Y. Koga, R. Drullinger, L. Hollberg, "Characteristics of an optically pumped Cs frequency standard at the NRLM" *IEEE Trans. Instrum. Meas.*, IM-38, 533 (1989).

[11] E. De Clercq, G. D. Rovera, S. Bouzid, A. Clairon, "L'etalon primaire de fréquence a pompage optique du LPTF" CPEM-92. To be printed on *IEEE Trans. I&M*, April (1993).

[12] N. Fortson, "New term in atomic Zeeman energy" *Phys. Rev. Lett.* **59**, 988 (1988).

[13] A. De Marchi, E. Bava, P. Tavella, "Accuracy analysis of the atom-microwave interaction in a TE_{017} cylindrical Rabi cavity for a high C-field cesium beam resonator" to be printed on *Proc. 7^{th} EFTF*, Neuchâtel (1993).

[14] A. De Marchi & al. "A high uniformity magnet for a high C-field cesium beam resonator" to be printed on *Proc. 7^{th} EFTF*, Neuchâtel (1993).

[15] P. Kartaschoff, "Caesium beam tubes using confocal conics state selectors" Swiss Pat. No. 552244 and US Pat. No. 3824394.

[16] G. Vecchi, A. De Marchi, "Spatial phase variations in a TE_{011} microwave cavity for use in a Cesium fountain primary frequency standard" CPEM-92. To be printed on *IEEE Trans. I&M*, April (1993).

[17] A. De Marchi, unpublished.

NIST-7, THE NEW US PRIMARY FREQUENCY STANDARD

R. E. Drullinger, J. P. Lowe, D. J. Glaze and Jon Shirley
National Institute of Standards and Technology
Time and Frequency Division
Boulder, Colorado
USA

Abstract

NIST-7, an optically pumped, cesium-beam frequency standard has replaced NBS-6 as the official US primary frequency standard. The present short-term stability of the standard, measured with respect to an active hydrogen maser, is characterized by $\sigma_y(\tau) \approx 8 \times 10^{-13}\ \tau^{-1/2}$. Our first evaluation has resulted in an uncertainty of 4×10^{-14}. An improved servo-electronic system is being developed and this should improve stability and allow for more precise evaluation of the various systematic errors.

1. INTRODUCTION

An optically pumped, thermal atomic-beam frequency standard developed at NIST and known as NIST-7 has officially replaced NBS-6 as the US primary frequency standard. In the preliminary operation reported here, the standard has been shown to have both accuracy and stability that are several times better than NBS-6. Furthermore, these parameters are expected to improve in the near future with improved servo electronics.

The standard has been described elsewhere [1]. Briefly, the Ramsey cavity is 1.55 m long, and the atomic beam is 3 mm in diameter. An axial C-field geometry is employed and the atomic beam goes through the X-band waveguide cavity parallel to the long dimension of the waveguide. The microwave field, therefore, varies as a half-sine wave in the direction of the atomic beam; this results in Rabi lineshapes with very smooth and rapidly damped tails (see Fig. 1). The cavity ends are designed so the Poynting vector vanishes in the center of the atomic beam window [2], thus minimizing distributed-cavity phase shift.

For the evaluation reported here, a single, grating-feedback diode laser (linewidth $\leq$ 100 kHz) was frequency-locked to the $F = 4 \rightarrow F' = 3$ saturated-absorption feature in an external cesium cell. Approximately 0.1 mW in a 3 mm diameter ($2w_0$) light beam with linear polarization was used for state preparation. The optical pumping beam was orthogonal to the atomic beam and was retroreflected with its polarization rotated by 90°. This creates a zone of "randomized polarization" which leads to complete optical pumping ($\geq$ 99.9%) and avoids the problem of coherence trapping [3]. An acousto-optic device was used to synthesize a second light beam with the $\approx$ 450 MHz offset necessary to drive the $F = 4 \rightarrow F' = 5$ cycling transition in the detection region. The power density in this beam was adjusted to scatter approximately 10 photons per atom. The overall fluorescence collection and detection efficiency is $\geq$ 40%.

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

Measurements of frequency shifts reported here were made relative to an active hydrogen maser. At present, the stability of the standard is characterized by $\sigma_y(\tau) \approx 8 \times 10^{-13}\ \tau^{-1/2}$ (Fig. 2) and is limited by phase noise in the microwave radiation. The standard is designed to operate with an oven temperature of 110°C at which point the atomic shot-noise limited stability should be $\approx 3 \times 10^{-13}\ \tau^{-1/2}$. The frequency biases and their associated uncertainties are quoted throughout this paper in terms of fractional frequency change in the standard.

2. EVALUATION

The first evaluation of the standard has been reported in detail elsewhere [4] and will only briefly be presented here with the results summarized in Table 1. The correction for second-order Zeeman effect is made by measuring the first-order Zeeman splitting. Errors caused by field inhomogeneity are less than 10^{-15}. At present, the correction for the second-order Doppler shift has been made by a theoretical calculation which takes into account the velocity distribution, the microwave power, and the modulation parameters. A thermal velocity distribution, weighted by 1/v for detection by a cycling transition, was assumed. It predicts a Ramsey lineshape differing by only 1 or 2 % from the measured lineshape. The quality of this fit together with uncertainties in microwave power, and modulation parameters, leads to an uncertainty in the second-order Doppler correction of $\sim 2 \times 10^{-14}$.

The end-to-end cavity phase shift has been measured by beam reversal. The frequency shift on beam reversal at optimum power was 1.52×10^{-12} with an uncertainty of 3×10^{-14}. A search for distributed-cavity phase shift by alternately blocking one-half of the atomic beam near one end of the Ramsey cavity or the other showed no effect at the 1×10^{-14} level. The cavity Q of 600 and measured mistuning lead to cavity-pulling errors of less than 10^{-14}.

Line overlap shifts have not been evaluated in depth. But in an optically pumped standard with the high spectral symmetry demonstrated in Fig. 1, these effects are expected to lead to errors of less than 10^{-15}. The shift due to blackbody radiation [5] can be calculated to very high accuracy since the temperature of the atomic beam tube of NIST-7 is regulated. The fluorescence light shift in NIST-7

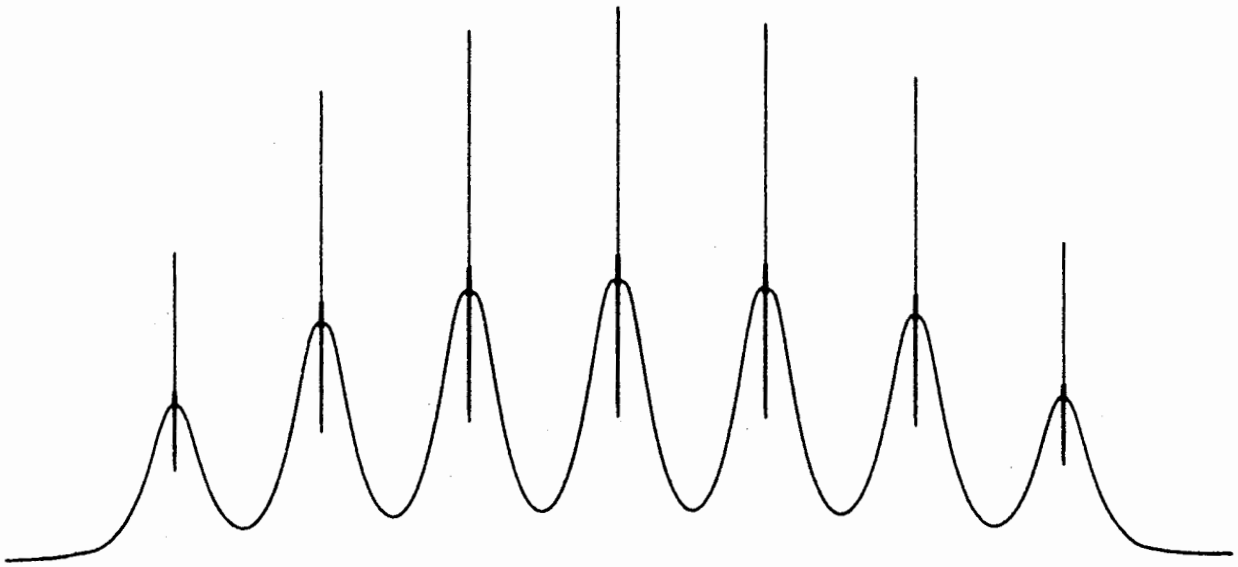


Fig. 1. Zeeman spectrum of the F = 3 to F = 4 transition. The microwave power level is about 7 dB below optimum and the C-field corresponds to a first order Zeeman splitting of 24.2 KHz ($\nu_{0,0} - \nu_{1,1}$).

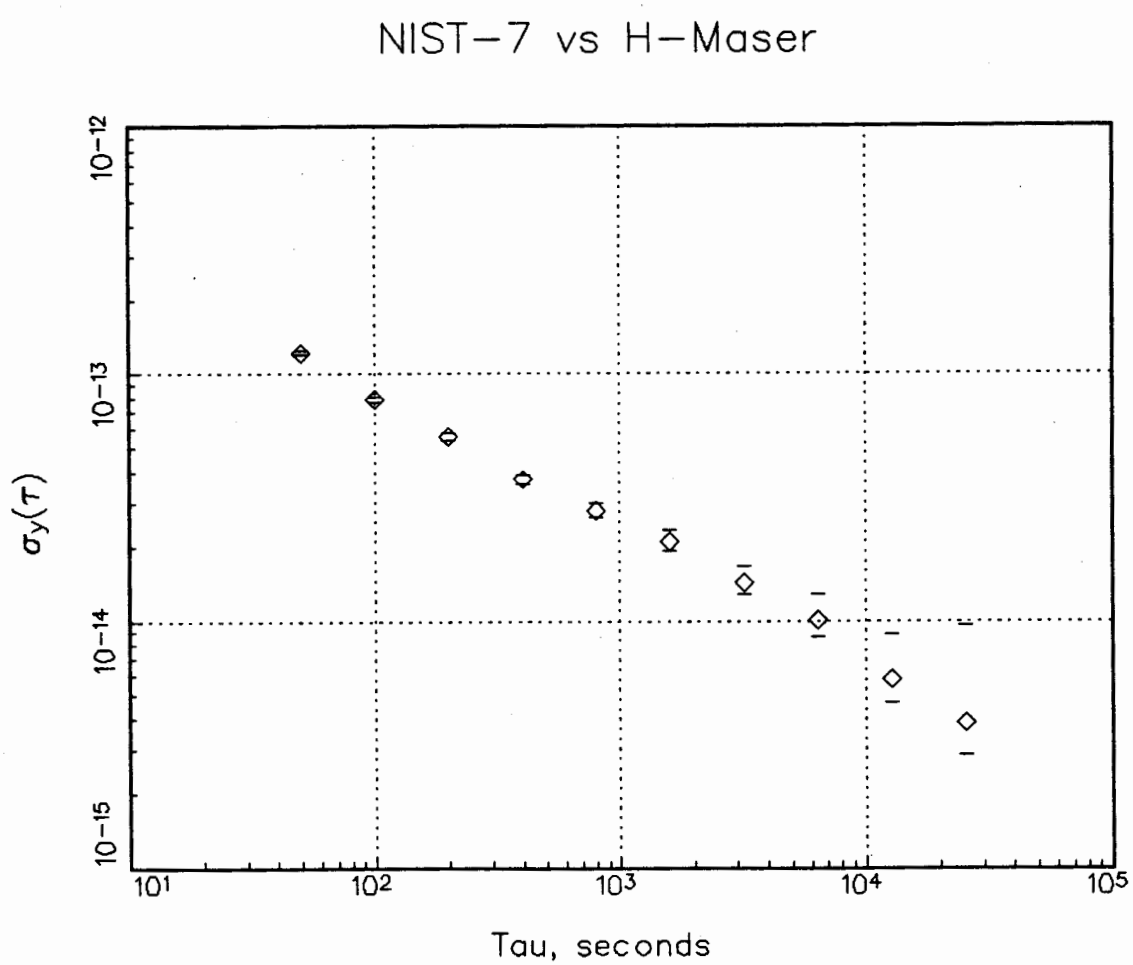


Fig. 2. Frequency stability of NIST-7 measured against an active hydrogen maser. The error bars represent 1σ on the statistics of the measurement.

is expected to be $\leq 10^{-16}$ [6]. Using different optical powers and atomic beam fluxes we have observed no frequency shifts at the 10^{-14} level.

Frequency errors coming in through the electronics and arising from RF spectral purity, modulation distortion or offsets in the integrators or DC gain stages have been shown to be less than 1×10^{-14}.

3. SUMMARY

NIST-7 has undergone a preliminary evaluation and been shown to be several times more accurate than NBS-6. As of January 1, 1993, it has become the official US primary standard for frequency. It will be used to steer the long-term behavior of the NIST time scale while it continues to be developed and improved. As the error budget is reduced, international comparisons will be performed with other high-accuracy standards.

4. ACKNOWLEDGEMENTS

The authors thank Fred Walls and Andrea DeMarchi for their numerous and invaluable contributions throughout this work.

5. REFERENCES

[1] Robert E. Drullinger, David J. Glaze, J. P. Lowe and Jon H. Shirley, "The NIST optically pumped cesium frequency standard," IEEE Trans. Instrum. Meas., vol. 40, pp. 162-164, April 1991.

[2] A. DeMarchi, J. Shirley, D. J. Glaze, and R. Drullinger, "A new cavity configuration for cesium beam primary frequency standards," IEEE Trans. Instrum. Meas., vol. 37, pp. 185-190, June 1988.

[3] G. Théobald, N. Dimarcq, V. Giordano, and P. Cérez, "Ground state Zeeman coherence effects in an optically pumped cesium beam," Optics Comm., vol. 71, no. 5, pp. 256-262, June 1989.

[4] R. E. Drullinger, Jon H. Shirley, J. P. Lowe and D. J. Glaze, "Error Analysis of the NIST Optically Pumped Primary Frequency Standard," to appear in IEEE Trans. Instrum. Meas., April 1993.

[5] Wayne M. Itano, L. L. Lewis, and D. J. Wineland, "Shift of $^2S_{1/2}$ hyperfine splittings due to blackbody radiation," Phys. Rev. A, vol. 25, pp. 1233-1235, February 1982.

[6] Jon Shirley, "Fluorescent light shift in optically pumped cesium standards," in Proc. 39th Freq. Cont. Symp., 1985, pp. 22-23.

Effect	Bias	Measurement Uncertainty (1σ)
2nd-order Zeeman	$\approx$10 000	1
2nd-order Doppler	$\approx$ 30	2
Line overlap		---
AC Stark	1.9	1
Cavity phase	76	3
Electronics		1

Table 1. Summary of systematic errors in NIST-7. The size of the resulting bias is given together with the associated uncertainty (all in units of 10^{-14}).

FILTERS

Chairman: J. Dowsett

EXTENSION OF MONOLITHIC FILTER APPLICATION FIELD AND THEIR MINIATURIZATION

S.A.SAKHAROV, A.V.MEDVEDEV, I.M.LARIONOV , Y.A.SAMSONOV

"FOMOS" Firm, 107076, Moscow-76, P.O.Box 63, Russia

Abstract

The present paper describes results of theoretical investigation aimed at determation of optimum crystal element width to thickness ratio, enabling to fabricate piezoelectric elements with minimum size by maintaining at the same time minimum resistance and high quality factor Q-values. Moreover, the results of the work intended for designing monolithic filters using strong piezoelectrics are presented. The main amplitude-frequency and phase-frequency characteristics of these filters are also given. At an example of 10,7 MHz monolithic filters the possibilities of obtaining different pass bands by using different piezoelectric materials are shown. In particular, for filters having a pass bandwidth 2-2,5%, operating over the temperature range from -10°C to +60°C, the possibility of replacing expensive lithium tantalate by cheaper lithium tetraborate monolithic filters is shown. Lithium tetraborate filters are only insignificantly less stable over the temperature range.

Introduction

At present monolithic filters (MF) are extensively used in radio equipment, in frequency selection apparatus and compared earlier filters with discrete elements have a number of advantages in their miniature size, reliability and easier manufacturing technology.

The most widely used monolithic filters are quartz crystal filters (MCF). These filters operate over a wide frequency range from 2.0 MHz up to 30 MHz at the fundamental mode (filters designed with the use of a reversed mesa structure are not discussed in the paper) and up to 300 MHz at overtones. The pass bandwidth of such filters is 0.008 to 0.3% from the nominal frequency.

Fig. 1 shows in a graphical form the application field of monolithic quartz crystal filters. From the figure it follows that the application field of MCFs is limited by pass bandwidth values. For obtaining filters with a wider pass band it is necessary either to use

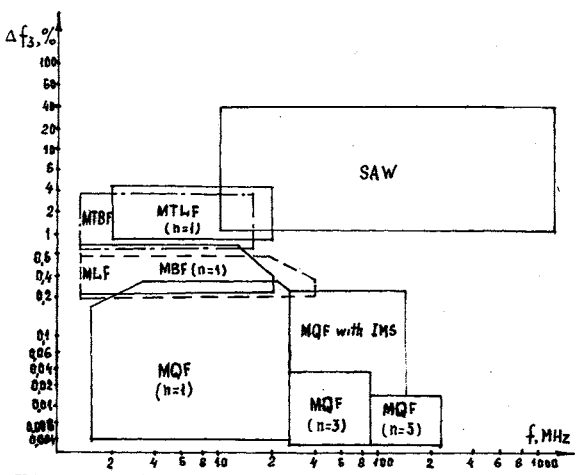


Fig. 1 - Frequency ranges and pass bands of MF using different crystals

added inductance coils, which causes considerable augmenting their size and weight, or to use substrate crystals with high electromechanical coupling factor.

To such crystals used in piezoelectric techniques belong berlinite ($AlPO_4$) and langasite ($La_3Ga_5SiO_{14}$) which we call middle piezoelectrics with the electromechanical coupling factor k_s^2 up to 20% and strong piezoelectrics with k_s^2 more than 20%. These are lithium tetraborate and lithium tantalate. Fig. 1 shows also in a graphical form the application regions of monolithic filters fabricated of these crystals. Table 1 lists the basic piezoelectric characteristics of these crystals. From Fig.1 it is evident, that every crystal has a definite location in this graph which enables to consider the application field of monolithic filters. More detailed information on the use of each of these crystals is described in Section 2,3,4.

1. Monolithic quartz crystal filters

Continuous broadening the filter application field and discovering new fileds, requires further miniaturization of these devices. The progress in this direction during the last tens of years in mainly connected with designing monolithic filters operating by using the "energy trapping" principle.

In accordance with this theory, the distance from the electrode edge to the crystal element edge should not be inferior to a certain value, gouverned by vibration energy losses due to energy reflection from the crystal element edges. As far as the vibration energy transmitted to the neighbouring resonator is concerned, which determines the band of a filtering device, this is really necessary. With regard to the vibration energy propagating in the perpendicular direction, it does not influence the coupling value and it were desirable either to eliminate or to minimize it. In view of the fact that the major part of energy of these vibrations in piezoelectric elements of yzb/35°11'-cut propagates as flexural vibration across the thickness of the crystallographic X-axis, it is likely, that it is possible to find such a value of a crystal element width at which this energy at the edge face will be absent and will not be reflected into the underelectrode region.

It is most likely that these values of a crystal element width will be periodically repeated in certain discrete values of a relative bandwidth, gouverned by the wavelength of flexural vibrations across the thickness, connected with thickness-twist vibrations, which propagate along the crystallographic Z'-axis and are operational ones for the plates of this cut.

This principle was for the first time successfully used by the authors while designing second-order filter at 2.048 MHz for communication systems with pulse-code modulation [1] . The authors are aware of the use of this principle by other scientists [2,3] . Our efforts to use recommendations given in this literature have been unsuccessful, presumably, due to the spread in properties of synthetic quartz and the type of the orientation, used by us and by the above investigators.

The experimental work intended for determination of a relative pass bandwidth of a piezoelectric element which ensures obtaining high Q-value of single resonators with a relative freedom of unwanted responses, showed the real periodicity of piezoelectric element parameters repetition at frequencies 5.0, 8.0, 10.0 and 18.0 MHz (Fig. 2).

In accordance with the relationships found, monolithic filters operating at 5, 8.8, 10.7, 12.5 and 18.5 MHz was designed. The frequency spectrum of one of these second-order sections operating at one of the specified frequencies is given in Fig. 3.

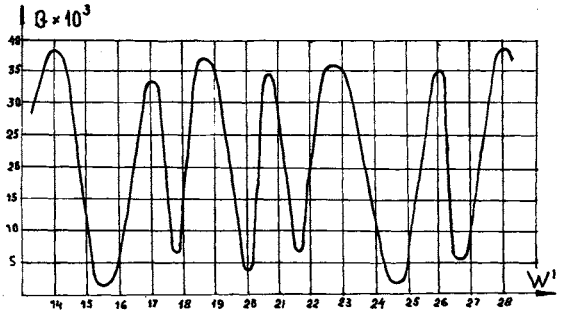


Fig.2 - Dependence of quality factor Q variation of a system consisting of single resonators on the relative piezoelectric element width

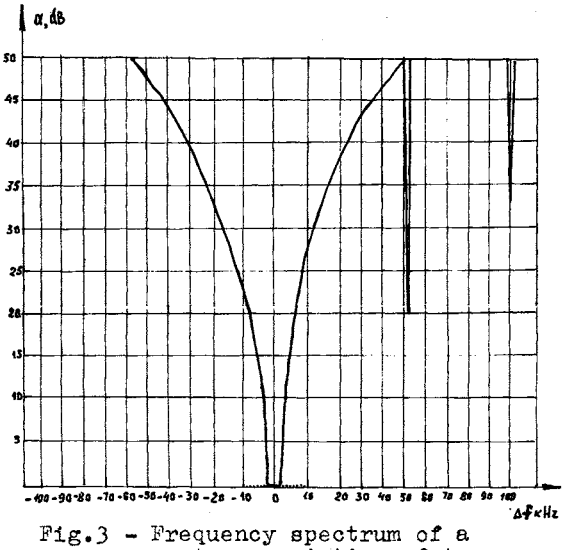


Fig.3 - Frequency spectrum of a system consisting of two coupled resonators with w' = 14

The 4th, 6th, 8th and 10th-order filters at 10.7 MHz packaged in standard microcircuit enclosures are given in Fig. 4.

Frequency spectrum variations from a relative width w' for one of the relationships and Q-value variations versus relative width spread of a crystal element are shown in Fig. 5.

From the analysis of the data obtained one can make the following conclusions:

1. Regions of maximum Q at relative width values of 18.4 and 22.4 make it possible to fabricate piezoelectric elements with the accuracy of existing technological tolerances during edge lapping process. Relative width values of 17.2, 20.6 and 20.8 cannot be used at present due to extreme narrow tolerance range for the width necessary for ensuring satisfactory Q-values.

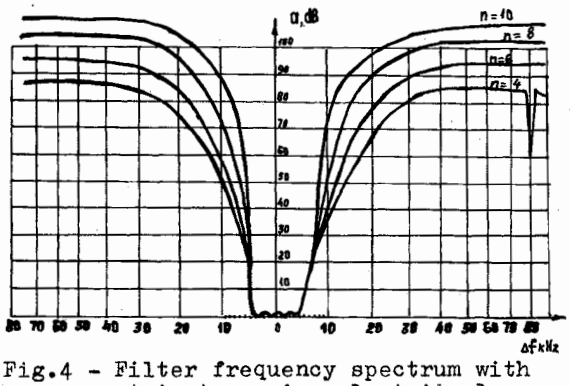


Fig.4 - Filter frequency spectrum with strip-type piezoelectric elements having w' = 22.5

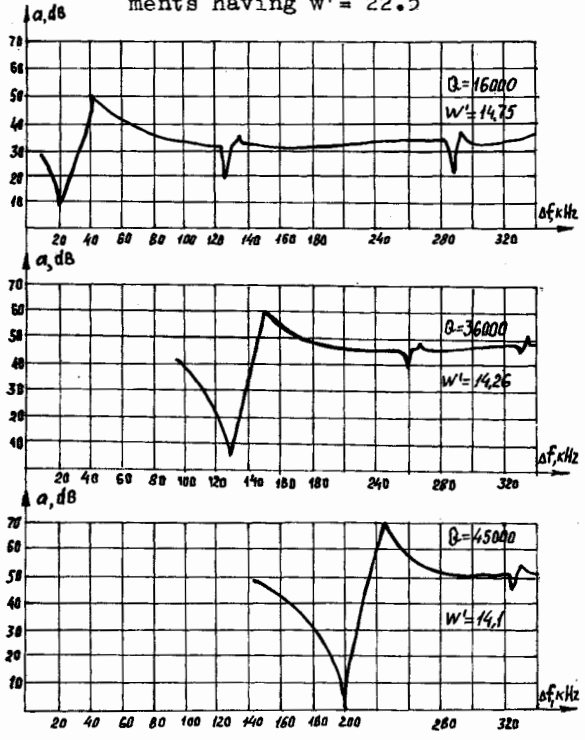


Fig.5 - Dependence of frequency spectrum on w' - value

2. The use of piezoelectric elements with technologically obtainable relative width values enables to decrease the space occupied by piezoelectric element by 3-5 times, which, in its turn, enables to make the volume of a filter 2-3 times smaller and to design higher order filters intended for surface mount technology.

While using new methods of piezoelectric elements manufacturing, excluding mechanical edge treatment (group type technology) further decrease of relative width dimensions is possible.

2. Monolithic langasite filters (MLF)

As it was mentioned above, LGS-crystal belong to medium piezoelectrics with the k_s^2 value = 15%. Their temperature-frequency coefficient (TFC) for thickness-shear mode in approximately three times greater, then that of quartz and equals to 1.6×10^{-6} 1/°C for the wide temperature range from -60°C to +85°C. The frequency, vs temperature dependence has the form of the quadratic parabola, the inflection point of which can be shifted with respect to 25°C. This is shown in Fig. 6.

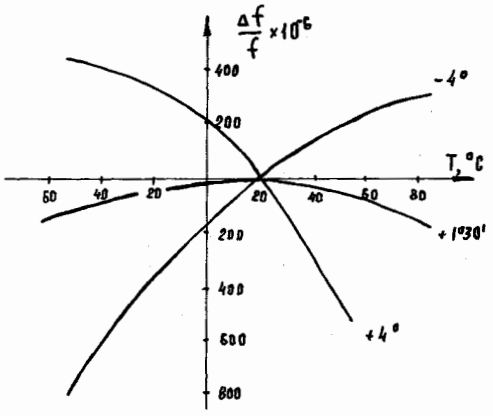


Fig.6 - Frequency vs temperature dependence for cut angle range from -4° to +4°

The frequency constant has lower value then for quartz and equals to 1380 kHz mm for yxl/1°30'-orientation [4].

The characteristics of LGS-crystals determine the application field of monolithic langasite filters (MLF).

The experiment showed, that the best results when designing MLFs operating at the fundamental frequency can be obtained over the frequency range from 2 MHz to 25 MHz, the maximum pass bandwidth obtained was 0.9%. It is necessary to note, however, that the lower frequency range from 2.0 MHz to 10.0 MHz is considered to be the preferred one. This is caused by frequency constant values, for example, piezoelectric element dimensions of monolithic filters made of quartz are equal to 32 x 16, those of lithium tantalate 24 x 16 and those of LGS 16 x 8.

While designing MLFs the piezoelectric effect's influence cannot be neglected.

LGS-crystal application in monolithic filters can serve as an improvement of a number of outdated filters with discrete elements and also to develop principally new filters having characteristics which could not be obtained in monolithic filters.

Table 2 lists comparative parameters of MLF and MCF having similar nominal frequency- and pass band values. The advantage in size of MLF can be seen from the table.

Fig. 7 and 8 show centre frequency vs temperature and time characteristics of MLF which were held at +85°C.

The use of LGS-crystals enables to enlarge considerably MF filter application field with relative pass band values up to 0.9%, with maintaining satisfactory temperature stability at the same time.

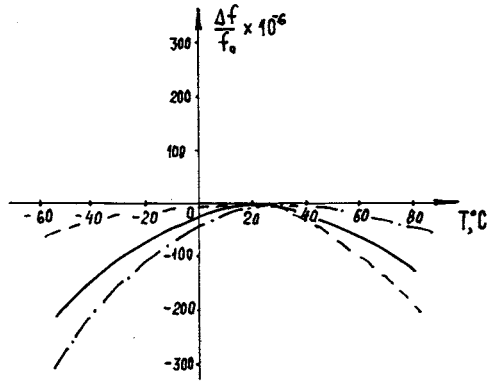


Fig.7 - Dependence of LGS-filter centre frequency and cut-off frequencies f_{c1}, f_{c2} on temperature

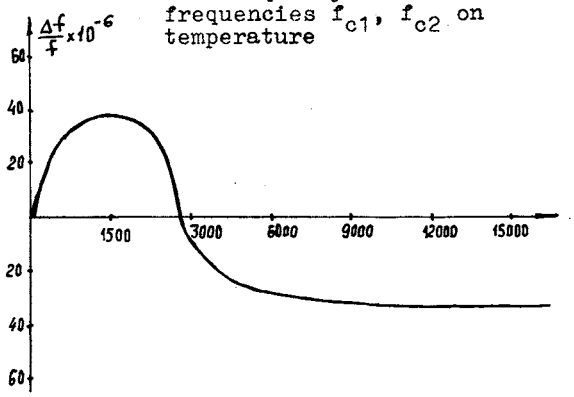


Fig.8 - Dependence of centre frequency vs time of 18.5 MHz LGS-filter at +85°C

3. Monolithic lithium tantalate filters

The use of lithium tantalate crystals of xys/-49°-cut as a substrate enables to fabricate monolithic filters operating at the fundamental frequency of thickness-shear mode with the relative pass band value up to 4.5%.

High value of the frequency constant, N= 2120 kHz.mm, and a sufficiently large crystal element thickness as a result, gives the possibility to apply lithium tantalate at frequencies of 27-30 MHz without utilization of special plate treatment technologies.

The pass bandwidth realizable within this frequency range is from 0.5 to 1.0%.

Due to considerable cut-off frequency deviations in the low temperature region(-60°C) $\Delta f/f = 800 \times 10^{-6}$, the application of lithium tantalate crystals in filters with pass bands less than 0.5% is not recommended.

Fig.9 shows the spectrum of MLTF designed for the input channel of a multi-channel FM radiostation at 27.21 MHz. This filter is packaged in a ceramic enclosure, v= 0.3 cm^3, which is designated for SMT.

The application of lithium tantalate for filters enabled to replace a number of filters with discrete elements by monolithic ones, by improving, thus, considerably both amplitude- and phase-frequency characteristics, size and weight of crystal filters.

Figure 10 shows the amplitude- and phase-frequency characteristics of MLTF of the 12th order, v= 5 cm^3, at 12 MHz, having Tchebyshev attenuation characteristic with the pass band of 150 kHz.

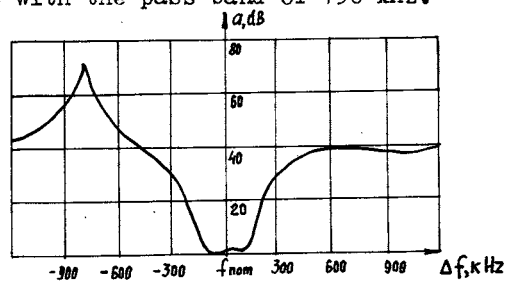


Fig.9 - Frequency spectrum of MLTF at 27.21 MHz

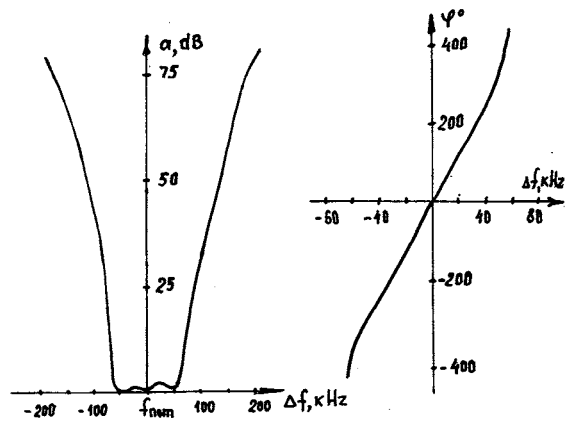


Fig.10 - Frequency spectrum and phase characteristic of MLTF

Fig.11 shows the ageing characteristic of the filter after 10 000 hours. It is necessary to note that lithium tantalate has a number of disadvantages. The main of them is its high price.

We have made a number of experiments in order to fabricate crystal elements of smaller size. A miniature MLTF operating over the frequency range 5 to 27 MHz has been designed as a result with v = 0.5 cm^3 for the 4th order filter and v= 1.4 cm^3 for the 8th order filter. These filters operate over a wide temperature range.

Moreover, we have investigated the possibility of replacement of lithium tantalate by another cheaper piezoelectric crystal. Lithium tetraborate can be considered as such material.

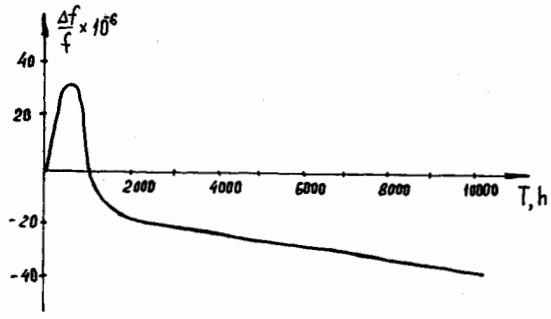


Fig.11 - Dependence of centre frequency vs time of 12.0 MHz MLTF of the 12th order at +85°C

4. Monolithic lithium tetraborate filters

Ten years ago lithium tetraborate has been called one of the promising piezoelectric materials which could enlarge MF application field [5]. A number of papers appeared during the last years describing investigations of BAW resonators and SAW filters using this material.

Publications appeared in Russia and in other countries describing investigations of monolithic lithium tetraborate filters are known to the authors. Some results of our investigation of monolithic lithium tetraborate filters operating at the fundamental frequency of the thickness-shear mode are presented.

In accordance with the data given in [6] for a trial-made monolithic filter we took Y/51°-cut and determined data, necessary for designing monolithic filters: the frequency constant, the electromechanical coupling factor etc. The maximum theoretical pass band value of monolithic lithium tetraborate filter(MTBF) with the Butterworth characteristic is 1 to 3.5% and with that of Tchebyshev is 3 to 4% (the ripple value being of 0.1 dB). The frequency spectrum of MTBF at 10.7 MHz with the pass band value of 200 kHz is shown in Fig.12.

Fig.13 shows the centre frequency vs temperature characteristic of this filter.

In this figure the extremum of FTC lies at the temperature T = -30°C.

It is evident that the use of a double-rotated cut will enable to shift the extremum of FTC to the region of normal temperatures. Having the FTC of this kind, the centre frequency deviation of 200 - 250x 10^{-6} over the temperature range from -10°C to +60°C, which is comparable to the temperature stability of MLTF, can be obtained.

Thus, it is possible to replace costly MLTF operating over the temperature range from -10°C to +60°C by MTBF, which is, according to our estimation, 15-20 times cheaper.

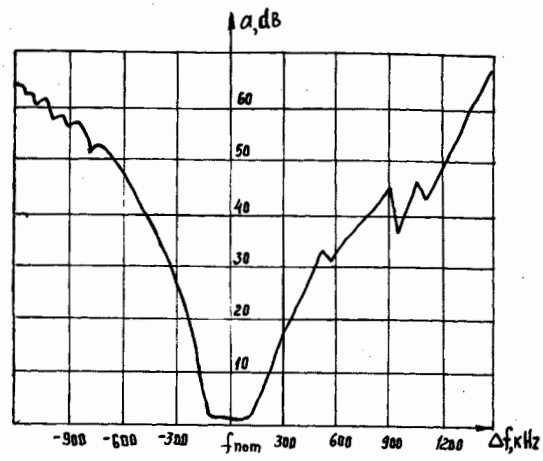


Fig.12 - Frequency spectrum of MTBF of the 2nd order

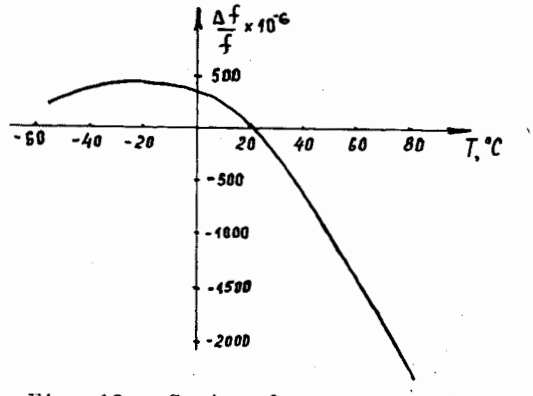


Fig. 13 - Centre frequency vs temperature characteristic of MTBF

5. Conclusion

The investigations connected with MCF-miniaturization and the use of the following piezoelectric crystals: langasite, lithium tantalate, lithium tetraborate, the main characteristics of which are given in Table 1, allowed us to decrease considerably their size, to increase reliability and to decrease labour-consumption in the filter manufacturing process.

Table 2 gives the main parameters of 10.7 MHz monolithic filters using different piezoelectric materials and having different pass bandwidths, these filters being packaged in an enclosure with the volume v= 1.4 cm^3.

At present the authors investigate crystals having gallium-germanate structure and also continue their developments of monolithic filters intended for SMT.

Table 1

Main properties of piezoelectric crystals

Crystal	Electromechanical coupling factor, %	Relative resonance spacing, %	Qx 10^3	Frequency-temperature coefficient, $10^{-6}/°C$
SiO_2	7.0	0.25	100	0.5
$La_3Ga_5SiO_{14}$	13.8	0.8	30	1.6
$LiTaO_3$	47.0	7.0	2-4	4.0
$Li_2B_4O_7$	25.0	2.5	10	6.0

Table 2

10.7 MHz monolithic filter parameters

Crystal	Pass bandwidth		Ripple	Stop bandwidth		Insertion attenuation	Guaranteed attenuation	Order	Operating temperature range
	dB	kHz	dB	dB	kHz	dB	dB		°C
SiO_2	3	±3.5	1.5	80	±9.0	3.0	90	8	-60 to +85
SiO_2	3	±7.5	1.5	85	±25	3.0	90	8	-60 to +85
$La_3Ga_5SiO_{14}$	3	±15.0	1.5	80	±40	3.0	85	8	-60 to +85
$La_3Ga_5SiO_{14}$	3	±25.0	2.0	80	±65	3.0	85	8	-60 to +85
$LiTaO_3$	3	±65.0	1.0	80	±190	3.0	85	8	-60 to +85
$Li_2B_4O_7$	3	±100	1.0	40	±250	2.0	50	8	-10 to +60

References

1 E.G. Bronnikova, I.M. Larionov et al, "Monolithic filters for telecommunication systems with pulse-code modulation. Design features of LF-monolithic filters", Electronic Engineering, Series "Radiodetails and Components", issue 1(54), 1984, p.59-61.

2 M. Onoe, M.Okazaka, "Miniature AT-cut strip resonators with tilted edges",Journ. Acoust. Soc.,1977, v.3, p 368-375.

3 S. Yamashita,et al, " New frequency-temperature characteristics of 4.19 MHz beveled rectangular AT-cut quartz resonators",Proc.33rd Annual Frequency Control Symposium,1979, p.277-285.

4 S.A.Sakharov, I.M.Larionov, A.V.Medvedev, "Application of langasite crystals in monolithic filters

References(continued)

operating in shear modes", Proc. of the 1992 IEEE Frequency Control Symposium,p.713-723.

5 R.C.Smythe, M.D.Howard, "Current trends in crystal filters",Proc. of the 37th Annual Frequency Control Symposium, 1983, p.349-353.

6 M. Adachi, T.Shiosaki, H.Kobayashi, O. Ohnishi and A.Kawabata," Temperature compensated piezoelectric lithium tetraborate crystal for high frequency surface acoustic wave and bulk wave device applications", 1985 Ultrasonics Symposium, p.228-232.

DESIGN OF ENERGY TRAPPING INTEGRATED FILTERS

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

*FRANCE TELECOM CNET Laboratoire de Bagneux, dept PAB/BAG/MCT/CRL
196, Av.H.Ravera BP 107, 92225 Bagneux France.
**L.M.C.P. Université Paris VI, Paris VII et CNRS 4, place Jussieu
75252 Paris-Cedex France.

Abstract:

Bulk acoustic wave filters are very attractive for filtering in the equipements of radiocommunication systems due to their very good selectivity. Recent technologies allows to obtain plates having fundamental frequencies in the range from 70 to above 300MHz. At these high frequencies, it is possible to obtain filters with extremely reduced dimensions by the integration of several resonators or monolithic cells on the same plate. In order to optimize the design of such filters a model based upon the equations governing the essentially thickness modes was established. We consider structures in which two electroded regions are separated by a "coupling" region. The electroded and the coupling regions are surrounded by an external unelectroded region having an higher cut-off frequency to ensure that the energy can be trapped in the device. Depending on the distance between the electrodes and some other parameters, the resonators can be from strongly (monolithic filters) to negligibly acoustically coupled. The latter case seems to be very interesting at very high frequencies. A semi-algebraical method of solution is employed, making use of series of exact solutions of the equations and of a discretization of the continuity conditions between the different regions. This permits to obtain the eigen frequencies and the eigen-modes. The forced modes of the structures are then computed as linear combinations of the eigen modes. Using the forced modes, the terms of the admittance matrix of the 4 ports structure are calculated so that, the elements of the lattice equivalent circuit are obtained. A computer program was implemented and applied to different types of structures.

1/INTRODUCTION

Up to the recent years, the devices, presenting the most important degree of integration on a single piezoelectric substrate were the acoustically coupled (monolithic) filters. Most of the applications of these filters were situated in the range from 10 to 40 MHz.

One of the most important application of crystal filters is constituted by their use in the radiocommunication equipments. In this field, major evolutions are currently occurring with the progressive replacement of analog systems by digital ones that have much better spectral efficiency and also with the use of higher and higher frequencies. These evolutions induce new requirements for the filters used in such systems which require larger bandwidth, higher centre frequencies and also nearly linear phase responses. Other important trends also related to this application are the research of very small dimensions and of reduced cost.

Most of the systems being currently installed or planed for the next ten years use frequencies in the low gigahertz region. So that, thin film resonators (TFR) filters may be a solution for front end filtering in these systems. Intermediate filters for these systems must have centre frequencies situated in the low hundred megahertz frequency domain. New techniques allow the obtentions of plates vibrating in this frequency range, where the filters have relative bandwidth (0.5% to 5%) fully adapted to this application. Furthermore, due to the reduced dimension of the resonators at these frequencies, it is possible to integrate several resonators and hence complete filters on a plate with very small lateral dimensions. Recently, several very high frequency filters, integrating several acoustically isolated resonators were described and it was shown that such devices have very interesting properties relatively to achievable bandwidth, type of responses, and ease of adjustement [1]. Our purpose in this communication is to elaborate a model of such devices that can permit their optimization either for UHF application (TFR filter) or I.F. filters at VHF frequencies.

2/PRINCIPLE OF THE METHOD:

Several models of monolithic filters were presented sometimes ago [2], [3], [4], [5]. The most recent among these models are based upon finite elements or finite differences solutions of the equations obtained from the theory of the essentially thickness modes established by H.F Tiersten and co-workers [6],[7] or from theories of the same nature [8].

We have shown, for single resonators, that a semi-algebraical method of resolution of the Tiersten equations has the very interesting property to lead to very accurate solutions with a reduced amount of computation [9]. This method uses solutions in the form of series of exact solutions of the equation that verify the continuity and/or boundary conditions at a discrete number of points. The accuracy of the method is the result of a fast alternate numerical convergence of the solutions as the number of terms of the series is increased (Figure 1). Furthermore, this accuracy is obtained for nearly any electrode geometry of practical interest so it was possible to find optimal geometries for several applications [10].

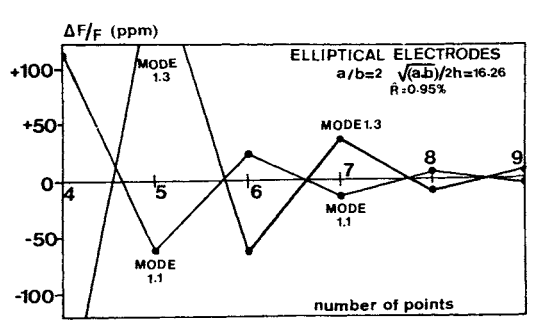


Figure 1: Convergence of the eigen frequencies observed for resonators

In this communication, we consider structures made on a piezoelectric plate of constant thickness having two electroded regions characterised by a mass loading $\hat{R}_2$ and a cut-off frequency f_{ce2} that are separated by a "coupling" region characterised by cut-off frequency f_{ce1} (and a mass loading $\hat{R}_1$ lower than $\hat{R}_2$). The external region surrounding the two previous ones is unelectroded and has an higher cut-off frequency to ensure that an energy trapping behaviour can occur.

Depending on the distance between the electrodes and several other parameters, the resonators can be from strongly to negligibly coupled. This corresponds respectively to the limiting cases of the two poles monolithic filters and to the case of acoustically isolated resonators integrated on the same plate. In the following, we consider mostly structures that are symmetrical with respect to the two perpendicular axis situated in the plate that lead to the cancellation of the mixed partial derivatives in equation {1} of §3 [7]. In this case only one fourth of the structure and the modes having the correspondent symmetry are to be considered in the analysis. A structure which is rather of the first kind is represented in figure 2, whereas the structure of figure 3 is rather of the second kind. For less symmetrical structures, the principle of the analysis remains very nearly the same, but we need to consider one half or the whole structure and less symmetrical modes.

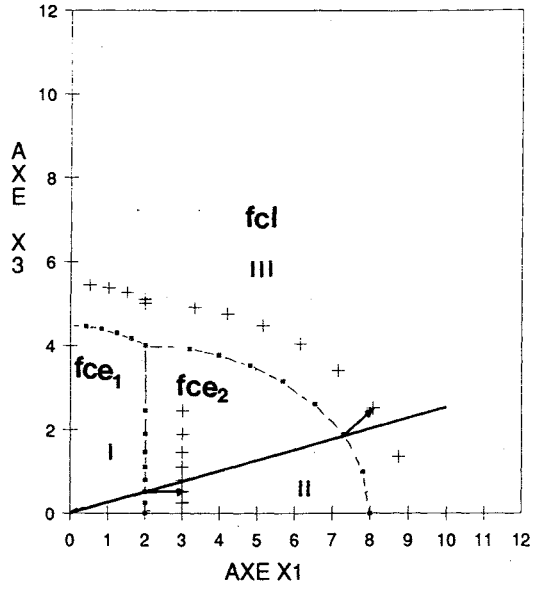


Figure 2: 2 poles structure with a large acoustical coupling.

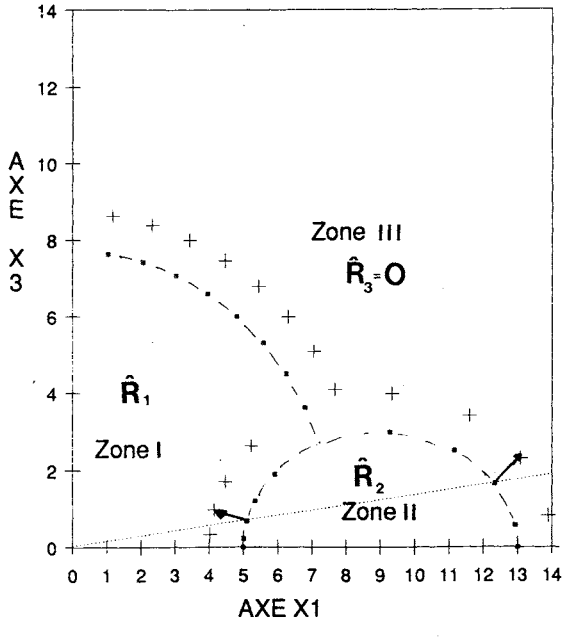


Figure 3: 2 poles structure with a weak acoustical coupling.

Since, a good electrical modelization of integrated structures using either weakly acoustically coupled resonator or monolithic filters having electrode geometries quite different from the usual rectangular one requires also a great accuracy in the description of the mode, we have set up a model using the same principles as for the resonators. As in this latter case we use the theory of essentially thickness modes, so that in each region the lateral dependence of the dominant component of the displacement u_1 is governed by equation {1} (§ 3). For each region, specific values of the mass loading and cut-off frequencies are considered in this equation.

3/ANALYSIS OF 2 POLES STRUCTURES

As basis of the model we use the scalar equation governing the lateral dependence of the anharmonics in the vicinity of one overtone (noted n) of the thickness shear. This equation was first established for y rotated plates of class 32 crystals [6] (plates with a 2-monoclinic symmetry), then for plates of arbitrary symmetry [7].

$$M_n \frac{\partial^2 \bar{u}_1^n}{\partial x_1^2} + P_n \frac{\partial^2 \bar{u}_1^n}{\partial x_3^2} - \frac{n^2\pi^2 c^*}{4h^2}\bar{u}_1^n + \rho\omega^2 \bar{u}_1^n = \rho\omega^2(-1)^{(n-1)/2}\frac{e_{26}4V_0e^{j\omega t}}{c^{(1)}n^2\pi^2} \qquad \{1\}$$

In this equation, $\bar{u}_1^n(x_1,x_3,t)$ is the lateral dependence of $\hat{u}_1^n(x_1,x_2,x_3,t) = \bar{u}_1^n \sin(n\pi x_2/2h)$. $\hat{u}_1$ arises from a transformation of the usual component of the displacement u_1 and of ϕ the potential , made to replace the inhomogeneous boundary conditions $\phi = \pm V_0 e^{j\omega t}/2$ at $x_2 = \pm h$ by the homogeneous conditions $\hat{\phi} = 0$. The inhomogeneous term of the equation results from this transformation.

c* is either $\bar{c}^{(1)}$ for the unelectroded regions or $\hat{c}^{(1)}$ for the electroded regions. $\bar{c}^{(1)}$ is the stiffened elastic constant relative to the corresponding one dimensional mode (Eigen-value of the Christoffel Matrix see Table I). $\hat{c}^{(1)} = \bar{c}^{(1)}(1 - 8k^{(1)2}/n^2\pi^2 - 2\hat{R})$ is a constant that includes the electrical and inertial effects of the metallization. $4k^{(1)2}/n^2\,\pi^2$ is approximatively the relative frequency lowering due to the electrical effect of the metallization, $\hat{R}$ is the mass loading. $c^{(1)} = \bar{c}^{(1)}(1 - k^{(1)})^2$ is a "pseudo ordinary elastic constant" which is equal to C_{66} for the Y rotated cuts. $k^{(1)}$ is the coupling coefficient of the corresponding one dimensional mode, this quantity is equal to k_{26} (noted k in Table I) for the Y rotated cuts. M , P are functions of the material constants and of the plate orientation that can be obtained as described in reference [6] and [7].

The important parameters of equation {1} for the fundamental mode of the AT cut of the quartz like materials were computed [11] from the published constants corrected for berlinite and GaPO4 by experimental results as described in [11]. They are given in table I. For the gallium phosphate the coefficients are to be considered as tentative values, since the constants used were taken from different sources. Much probably a better approximation for these coefficients can be obtain from the constants recently measured by Krempl et al. [12].

Coefficients of the equation

	M1	P1	$\bar{C}$	M1/P1
Quartz	11.008	6.922	2.921	1.261
AlPO4	8.718	5.084	2.285	1.309
GaPO4	7.663	3.805	2.310	1.419

Table I: Energy trapping constants for the AT cuts of the quartz like materials.

EIGEN SHORT CIRCUIT MODES.

The eigen modes (at V=0) of the structure are obtained by the simultaneous solution of the equations for the electroded and unelectroded regions of the structure, taking into account the continuity conditions at the electrode edge. The forced modes can then be computed as a linear combination of the eigen modes verifying the inhomogeneous equation.

At V=0 the homogeneous form of the approximate equation for the three regions reduces to anisotropic Helmoltz equations:

$$M_n u^n_{1,11} + P_n u^n_{1,33} + 4\pi^2 \rho (f^2 - f^{*2}) u^n_1 = 0 \qquad \{2\}$$

with :

$f^* = f_{ce1} = \frac{n}{4h}\sqrt{\frac{c^{(1)}_1}{\rho}}$ for region I. Or:

$f^* = f_{ce2} = \frac{n}{4h}\sqrt{\frac{c^{(1)}_2}{\rho}}$ for region II. Or:

$f^* = f_{cl} = \frac{n}{4h}\sqrt{\frac{\bar{c}^{(1)}}{\rho}}$ for the unelectroded region III.

Using the coordinate transformations :

$$X_1 = \left(\frac{c_{66}}{M_n}\right)^{1/2} \cdot \gamma x_1; \qquad X_3 = \left(\frac{c_{66}}{P_n}\right)^{1/2} \cdot \gamma x_3$$

$$X_1 = r\cos t; \qquad X_3 = r\sin t$$

with: $\gamma^4 = \frac{n^2\pi^2}{(2h)^4}$

We obtain the isotropic Helmotz equation:

$$u_{1,rr} + \frac{1}{r}u_{1,r} + \frac{1}{r^2}u_{1,tt} + A u_1 = 0 \qquad \{3\}$$

where:

$$A = \frac{n\pi}{4}\left(\frac{f^2 - f^{*2}}{f^2_{66}}\right) \qquad \{4\}$$

This partial derivative equation is separated into two ordinary differential equations by setting $u_1(r,t) = R(r).T(t)$:

$$r^2R'' + rR' + R(r^2A - \upsilon^2) = 0 \qquad \{5\}$$

$$T'' + \upsilon^2 T = 0 \qquad \{6\}$$

The separation constant υ has to be an integer since the solution of the second equation has to have a periodicity of $2\upsilon\pi$.

υ=m

The (symmetrical) solutions bounded at zero and at infinity are respectively for the different regions:

Region I:

$$u_1 = \sum_{m=0}^{\infty} A_m Be_m(r\sqrt{A})\cos mt \qquad \{7\}$$

Region II:

$$u_1 = \sum_{m=0}^{\infty} \{B'_m J_m(r\sqrt{A}) + B''_m Y_m(r\sqrt{A})\}\cos mt$$

$$= \sum_{i,m} B_i Bs_m(r\sqrt{A})\cos mt \qquad \{8\}$$

Region III:

$$u_1 = \sum_{m=0}^{\infty} C_m K_m(r\sqrt{A'})\cos mt \qquad \{9\}$$

in these relations m is even (for the symmetry). The antisymmetrical solutions with respect to x_1 contains sin(mt) also with m even, the antisymmetrical solutions with respect to x_3 contains cos(mt) with m odd.

J_m, Y_m, I_m, K_m are respectively the Bessel functions of first and second kind and the modified Bessel function of first and second kind, of order m. Be_m is either I_m or J_m depending on the sign of A (relation {4}). Bs_m is successively J_m and Y_m.

Boundary and continuity conditions :

The traction free conditions on the major free surfaces, normal to the thickness are automatically verified as a consequence of the method used to establish equation {1}. On the surfaces limiting the different regions of the structure we have to specify the continuity of u_1 and of its normal derivative.

In this paper we suppose that the amplitude of the vibration mode near the edges of the plate are negligible so that no boundary conditions at the plate edge are taken into account. It was previously shown that, in this case, their consideration induces a negligible modification of the eigen modes and resonances frequencies. In coherence with this, the solutions bounded at infinity were chosen in region III {9}.

The continuity of u_1 and of its normal derivative on the surface limiting the regions can be approximated [13], by the expression of these conditions at only a discrete number of points (figure 4).

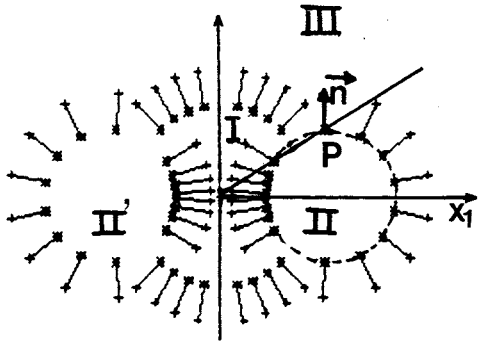


Figure 4: Discretization of the continuity conditions.

For example at a point $P(x^P_1 x^P_2 x^P_3)$ on the boundary between regions I and II, we have:

$$u^I_1(x^P_1 x^P_2 x^P_3) = u^{II}_1(x^P_1 x^P_2 x^P_3) \qquad \{10\}$$

$$n_1\frac{\partial u^I_1(P)}{\partial x_1} + n_3\frac{\partial u^I_1(P)}{\partial x_3} = n_1\frac{\partial u^{II}_1(P)}{\partial x_1} + n_3\frac{\partial u^{II}_1(P)}{\partial x_3} \qquad \{11\}$$

Different schemes can be use to express these conditions at the different region boundaries and to truncate the series, only one will be described here after:

We suppose that we use ngap points at the boundary between regions I and III, and nres points between regions I and II and also between regions II and III. This leads to 2ngap+ 4nres conditions. If we chose, as in figure 4, 2ngap=2nres=q, we obtain 3q linear relations between the coefficients A_0, A_2, A_{2q-2}, B_0,

B_2, B_{2q-2}, C_0, C_2, C_{2q-2}. To obtain a solution we have to truncate the series solutions for the three regions to q terms.

These 3q relations in 3q unknown constitute an homogeneous linear system in the 2q A and C coefficients. The determinant of this system has to vanish to have a non trivial (0) solution in these coefficients. This condition constitutes a frequency equation that is solved numerically for the eigen frequencies. For each eigen frequency ($f_{n\mu}$ where n is the rank of the overtone and μ the rank of anharmonic), the A, B and C coefficients of the eigen modes are found as a function of one of them, as the solution of the homogeneous linear system.

On the whole, the eigen modes for the three region of the structure are given by :

$$\hat{u}_1^n = sin(n\pi x_2/2h)\bar{u}_1^n \qquad \{12\}$$

$$\bar{u}_1^{n\mu} = e^{j\omega t}\sum_{i=0}^{i=q-1} D_{2i}^{n\mu} Be'_{2i}\left(r.\sqrt{A_e^{n\mu}}\right)cos(2it) \qquad \{13\}$$

Where the D coefficients are either A, B, or C and the real coefficients of r in the argument of the relevant Bessel or modified Bessel functions (Be') are now given either by the square root of:

$$A^{n\mu} = [n\tfrac{\pi}{4}(f_{n\mu}^2 - f^{*2})/f_{66}^2]$$

or $A'^{n\mu} = -A^{n\mu} = \left[n\tfrac{\pi}{4}(f^*2 - f_{n\mu}^2)/f_{66}^2\right]$ depending on the sign of A.

The eigen solution must verify the orthogonality relation:

$$\int\int\int_{V_t} \hat{u}_1^{n\mu}.\hat{u}_1^{k\lambda}dV = \delta_{nk}\delta_{\mu\lambda}N^{(n)(u)}$$

where V_t is the total volume of the structure and the displacement, is as in the following, defined by piece on the different regions.

COMPUTATION OF THE FORCED MODES:

The modes forced by the potential $V_{rs}e^{j\omega t}$ applied between the electrodes r and s can be obtained by a linear combination of the short circuit eigen modes. For the potential a supplementary term must be added [14]

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

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

Where :
$$u_1 = \hat{u}_1 - \frac{e_{26}x_2 V_{rs}e^{j\omega t}}{2hc^{(1)}}$$

u_1 being the forced conventional displacement.

Accounting for the fact that the orthogonality of sin $(n\pi x_2/2h)$ has already been expressed in the forms of equation {1} that are relevant for the three regions we substitute in these equations the lateral dependence of $\hat{u}_1$. For regions I and III, $\hat{u}_1$ is not different of u_1. For region (II) which has electrodes on which a potential is imposed, equation {1} contains a forcing term. After this substitution, we multiply these relations by $\bar{u}_1^{k\lambda}$ and integrate respectively on the volumes corresponding to the three regions. The summation of these relations gives (in the case of the Y rotated cuts of class 32 crystals where $c^{(1)}=C_{66}$ and $k^{(1)}=k_{26}$) :

$$\sum_n\sum_\mu H_{rs}^{n\mu}\frac{\pi^2}{4h^2 f_{66}^2}(f^2 - f_{n\mu}^2)\int\int\int_{V_t}\bar{u}_1^{n\mu}\bar{u}_1^{k\lambda}dV$$

$$= (-1)^{(n-1)/2}\frac{\rho\omega^2 e_{26}4V_{rs}e^{j\omega t}}{C_{66}^2 n^2\pi^2}\int\int\int_{V_{ers}}\bar{u}_1^{k\lambda}dV$$

Equivalent scheme for 2 resonators integrated structures:

Electrical response :

Using the constitutive equation for D_2 and the definition of the admittance Y_{rs} between electrodes r and s:

$$D_2 = e_{26}u_{1,2} - \varepsilon_{22}\phi_{,2}$$

$$D_2 = \frac{\varepsilon_{22}}{2h}V_{rs}e^{j\omega t}(k_{26}^2 + 1) + \sum_n\sum_\mu(-1)^{(n-1)/2}H_{rs}^{n\mu}\bar{u}_1^{n\mu}\frac{e_{26}}{h}$$

$$Y_{rs} = \frac{\partial I_{rs}}{\partial\Phi_{rs}};$$

Expressing the current I from the time derivative of the charge and hence from the normal component of D (D_2) we obtain:

$$Y_{rs} = \frac{-1.}{V_{rs}e^{j\omega t}}\int\int_{Sers} D_2 dS$$

So that the expression of the admittance Y_{rs} is:

$$Y_{rs} = j\omega\left[\frac{S_{ers}.\varepsilon_{22}}{2h}(k_{26}^2 + 1) - \sum_n\sum_\mu\frac{f^2 4k_{26}^2\varepsilon_{22}h\left(\int\int_{S_{ers}}\bar{u}_1^{n\mu}dS\right)^2}{(f^2 - f_{n\mu}^2)2n^2\pi^2 N^{(n)(\mu)}}\right]$$

The identification with the admittance of the equivalent scheme (that includes in parallel, one serial arm with the "static" capacitance and as many LC serial (dynamically) arms as there is modes), leads to :

$$C_{Zrs} = \frac{S_{ers}\varepsilon_{22}^S}{2h}(k_{26}^2 + 1) \quad \text{is the capacitance at f = 0} \qquad \{17\}$$

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

the inductance of the forced mode $(n\mu)$ is:

$$L_{rs}^{n\mu} = \frac{1.}{(C_{rs}^{n\mu}4\pi^2 f_{n\mu}^2)}$$

As observed for the resonators the parallel capacitance contains C_{Zrs} and an infinite sum of the opposite of the dynamic capacitances of all the modes :

$$C_0 rs = C_{Zrs} - \sum_n\sum_\mu C_{rs}^{n\mu} \qquad \{19\}$$

For symmetrical structures:

Using the transformed form of the orthogonality relation accounting for the orthogonality of the sinusoidal dependence in x_2:

$$\int\int\int_{V_t}\bar{u}_1^{n\mu}\bar{u}_1^{k\lambda}dV = 2h\int\int_{S_t}\bar{u}_1^{n\mu}\bar{u}_1^{k\lambda}dS = 2\delta_{nk}\delta_{\mu\lambda}N^{(n)(\mu)}$$

We obtain:

$$H_{rs}^{n\mu} = (-1)^{(n-1)/2}\frac{f^2 he_{26}}{(f^2 - f_{n\mu}^2)n^2\pi^2 C_{66}}\frac{4V_{rs}e^{j\omega t}\int\int_{Sers}\bar{u}_1^{n\mu}dS}{2N^{(n)(\mu)}} \qquad \{16\}$$

where S_{ers} = electroded surface (rs)
:
S_t = total surface

Only the symmetrical and the antisymmetrical parts of Y_{13} (Ya and Ys) are required (Bartlett theorem):

$$Y_{13} = j\omega\left[\frac{S_e.\varepsilon_{22}}{2h}(k_{26}^2 + 1) - \sum_n\sum_\mu\frac{f^2 4k_{26}^2\varepsilon_{22}h\left(\int\int_{S_e}\bar{u}_1^{n\mu}dS\right)^2}{(f^2 - f_{n\mu}^2)2n^2\pi^2 N^{(n)(\mu)}}\right]$$

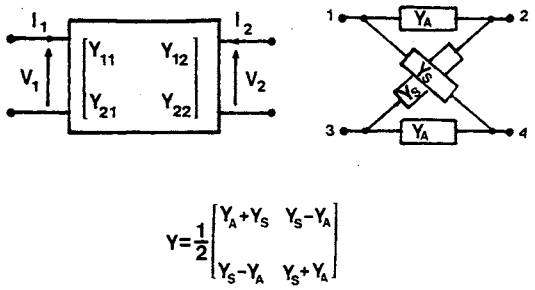


Figure 5: Symmetrical Equivalent Lattice

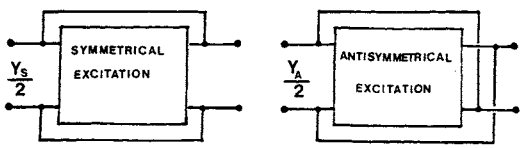


Figure 6: Symmetrical and antisymmetrical excitation.

With summation over the symmetrical modes (then $Y_{13}=1/2.Y_s$) or the antisymmetrical modes (with respect to the axis of the coupling region) (then $Y_{13}=1/2.Y_a$).

As in the general case for Y_{rs}, Y_{13} and hence Y_s and Y_a can be represented by the scheme of figure 7 where the capacitance and inductance have the same expression as for Y_{rs} but with summation over either the symmetrical or the antisymmetrical modes. We obtain thus a total equivalent scheme to the 2 poles structure under the form of a symmetrical lattice already indicated by R.C.Smythe [15].

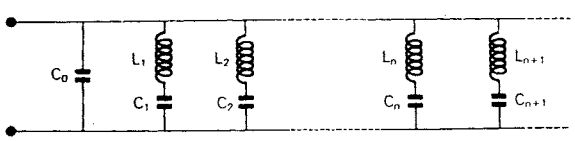


Figure 7: equivalent scheme of one arm of the lattice equivalent to the 2 pôle.

Discussion:

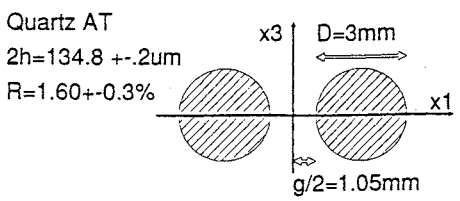
To obtain a good insulation between two resonators put on the same plate:

In the symmetrical case: Ys-Ya has to be very small, so that the symmetrical and antisymmetrical [half] modes have to be as similar as possible.

In the non symmetrical case a great advantage is the possibility to have different resonance frequencies for the 2 resonators, this gives another mean to obtain small crossed admittances Y_{ij} on a large frequency range.

Comparison of experimental and computed results:

A computed program was implemented. The experimental results were compared to computed ones for different 2 pôles structures having in most cases a weak coupling between resonators. A quite good agreement was observed in nearly all cases. A typical example is given in figure 8 where the parameter used are displayed. The rate of convergence of the resonnance frequency of the symmetrical mode with the increase of the number of points appear to be somehow slower than that was observed for the single resonators. In all cases, the detailed examination of the numerical coefficients of the truncated series for the different regions has indicated that the rate of convergences of these series is very different between regions, so that we can expect that different choices in the discretization and truncation schemes will permit to increase greatly the accuracy by giving a better representation to the region "containing" most of the energy.



	f.svm.(MHz)
experimental	12.258
computed 4x4pts	12.138
computed 4x6pts	12.277
computed 4x8pts	12.306

Figure 8: Comparison of experimental and computed resonance frequencies in a typical case.

Conclusion

The principle of a method of modelization of the two poles integrated filters was described. A computer implementation of this model was written and was applied to acoustically isolated two poles filters, the agreement between experimental and computed results is raisonably good. Further work is needed to increase the numerical accuracy and to obtain better convergence properties for the solutions. It is expected that improved schemes of discretization of the continuity conditions will lead to a much better accuracy of the computed resonances frequencies and modes.

Acknowledgements: The authors acknowledge the assistance of MM. Caumont and LeThuault for the preparation of the many samples used in this study.

References:

[1] J.P.Michel, M.Zafany,T.Soubirous, Y.Billon, D.Mouton, M.Berthe, L.Schwartz, S.Lechopier,J.L.Le Corre Proc. 7th European Frequency and Time Forum (1993) (these proceedings).

[2] H.F.Tiersten J.Acoust.Soc. Amer. vol.62 no 6 p.1424-1430 (1977).

[3] H.Seikimoto Eectron. Communication in Japan vol. 65A,n°12 p10-19 (1982).

[4] L.N.Dworsky Proc. 40th Annual Frequency Control Symposium p.161,167 (1986).

[5] H. Carru, R.Lefèvre, J.P.Aubry, S. LeChopier. Proc.42nd Annual Frequency Control Symp. p.101 (1988).

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

[7] D.S Stevens, H.F.Tiersten J.Acoust.Soc. Amer.vol.79 n°6 p.1455 (1986)

[8] R.C. Peach IEEE Trans. Ultrason. Ferroel. Freq.Control vol.35 no 5 p.593-611 (1988)

[9] J.Détaint, J.Schwartzel, C.Joly, E.Philippot Proc. 41st Annual Frequency Control Symposium p.314-324 (1987)

[10] J.Detaint, H.Carru, J.Schwartzel, C.Joly, E.Philippot, Proc. 5th.European Frequency and Time forum p.30-40 (1991)

[11] J.Detaint, J.Schwartzel, C.Joly, B.Capelle, A.Zarka, Y.Zheng, Y.Toudic, E.Philippot Proc. 6th European Frequency and Time Forum p.223-227 (1992)

[12] W.Wallnöffer, J.Stadler, P.Krempl. Proc. 7th European Frequency and Time Forum (1993) (these proceedings).

[13] K.Nakamura, M.Shimizu Proc.1976 IEEE Ultrasonics Symposium p.606 (1976)

[14] R. Holland, E.P.Eernisse Design of Resonant Piezoelectric Devices MIT-Press (1969).

[15] R.C.Smythe in Precision Frequency Control E.A Gerber, A.Ballato editors p.202 vol.1 Academic Press (1985).

NEW TECHNICS USED FOR 200 MHZ WIDE BAND FILTERS MANUFACTURING

J.-P. MICHEL, M. ZAFRANY, T. SOUBIROUS, Y. BILLON, D. MOUTON, M. BERTE,
L. SCHWARTZ, S. LECHOPIER, J.-L. LE CORRE

THOMSON / CEPE
44, av. de la Glacière - BP 165
95105 ARGENTEUIL - FRANCE

SUMMARY

The market for wide band IF filtering is oriented toward higher frequencies (for example 200 MHz with a 200 kHz bandwidth for DCS radiotelephone). Facing this new market leads to increase both manufacturing yield and output. This paper will show the various improvements which have been implemented at CEPE for that purpose. Work has been performed both on the quality level with an extension of the statistical process control capabilities, and on the technical and industrial level, particularly on the equipments. This is especially the case for the equipment for dynamic frequency adjustment of the resonator which has been newly redesigned in order to achieve simultaneously a good electrode shape definition, an accurate frequency tuning, and a high manufacturing output. New technics have been introduced to deal with the very small electrode dimension (0.1 mm diameter). A dry process technic has been chosen. The efficiency concerning the tuning of high frequency resonators (70 to 200 MHz) is presented including tuning accuracy, inductance reproducibility and manufacturing speed. Further results show the promizing application of this technic for the tuning of multi-resonators per plate.

1./ INTRODUCTION

The communication market nowadays requires wider bandwidths for signal transmission. This outcomes from two reasons: the transmitted information is no more only voice but now more frequently data which need a wider pass-band, and, on another hand, the data are more often time rather than frequency multiplexed. This leads the actual bandwidth to be generally closed to 200 kHz. In addition, the improvements in the signal processing ability towards higher frequencies, namely the improvements in integrated circuits capability, make it possible to manage higher intermediate frequencies, having in mind that higher frequencies are a preferential option regarding the lesser occurence of fold up products. The best compromize between integrated circuits availability and the requirement for the highest possible frequency is about 70 MHz for nowadays device and will be around 200 MHz for the next generation. This assertion may actually be verified on various kinds of systems for civilian as well as for militarian applications. The major example is the european digital radiotelephone system where IF frequency for GSM bases stations is 70 MHz, while for the coming DCS 1800, it is approximately 200 MHz. Regarding filter design and manufacturing, this leads to very stringent conditions, because the need for a wide bandwith prevents any recourse to overtones for the crystal resonators involved in the filter. However, the large extension of the digital radiotelephone make the market for IF filters dedicated to that application moving towards a quite large volume market where both manufacturing rates and cost becomes even more stringent parameters than electrical characteristics. That is why various improvements have been achieved on the manufacturing conditions and control.

2./ DESCRIPTION OF THE FILTER

Typical specifications are the following:

- centre frequency: 70 to 200 MHz
- 3 dB bandwidth: > 180 kHz
- insertion loss (50 ohms) < 5 to 8 dB
- attenuation at +/- 200 kHz > 20 to 25 dB
- attenuation at +/- 600 kHz > 50 to 60 dB
- ultimate rejection > 75 dB
- group delay variation < 1us over 100 kHz

This specification lead to a four pole filter. The requirements for out of band rejection, especially for spurious rejection, leads to the choice of a Jauman structure. The relative bandwidth of the filter is 10^{-3}; it is to wide a bandwidth to be fitted by quartz overtone resonators. The lithium tantalate which is the material which is often invoked for high bandwiths is not driveable on overtone modes because of its poor frequency versus temperature behaviour. Working on the fundamental mode cannot be avoided. The best compromize between a low shape factor and a good phase linearity is achieved by using a SLP synthesis whose essential advantages are a flat

group delay (about 65% of the 3dB bandwidth), a small shape factor (lower than Chebyshev) and a lower iterative impedance (-10% compared to Chebyshev). Even so the iterative impedance remains at a quite high level leading to the need for high quality coils and transformers.

3./ MANUFACTURING IMPLICATIONS

The thickness of the crystal plate at 200 MHz fundamental frequency is approximately 8 microns. It cannot be achieved by conventionnal mechanical means. It requires either ionic or chemical etching with a very fine accuracy on the thickness of the plate. In another hand, when working at a closed to 200 MHz frequency, the size of the electrodes of the resonator is very small (about 0.1 mm for the side dimension of a square electrode). The size of the electrode for a filter is always a compromize between the global need for a lower level of impedance for the whole filter - which would lead to a larger electrode - and a lower propensity for spurious generation implying smaller dimensions. Finally the dimensions will be the largest possible for no spurious generation; the consequence is a strong sensitivity to any eventual small increase of the size of the electrode or to any small misalignment between the upper and lower electrode. Regarding the dimension of the whole electrode (i.e. 0.1 mm) the estimate accuracy on the electode shape definition and superposition should be about 0.05 mm.This order of magnitude is no more consistent with the achievable accuracy on conventionnal tuning equipments acting by automatically controled plating through masks.

Another cause of complexity comes from the economical and throughput constraints. A filters requires - in a major case - four resonant functions. So the number of resonant functions to manufacture is four times as big as the number of filters, namely approximately 40,000 per year. Producing 40,000 plates with a very low thickness and very stringent specifications on flatness and thickness accuracy, whatever may be the industrial channel used (ionic or chemical etching), is an important charge for each step of the manufacturing line and is an important part of the cost of the assembled filter. The most efficient solution to palliate this problem is the gathering of resonant functions on the same plate. However, once more, this solution leads to face such problems as the tuning of the resonant functions, especially the selective tuning on the same plate, and the need for a high yield on the plate manufacturing in order to avoid any redhibitory global yield due to the mechanism of yield multiplication on the same plate.

In order to face the whole of the manufacturing problems work has been performed over the following points: 1/ setting up of systematic means to control the quality of the device at any step of the manufacturing process in order to avoid any lost of yield and to allow the gathering of resonant functions on the same plate; the solution which has been chosen is a statistical process control performed at any step. 2/ the development of a new tuning equipment particularly suitable for a high electrode shape accuracy and for selective tuning of resonators on the same plate.

4./ STATISTICAL PROCESS CONTROL

4.1./ STATISTICAL PROCESS CONTROL APPROACH

A Statistical Process Control program has been launched almost two years ago on the resonator fabrication line. The approach we have followed could be summarised in a few questions :

-"what goes on in the process ?
-"what are the major problems of the process ?"
-"what are the causes of the major problems?"

The process flow chart describes what are the main steps of the process (Fig.1); calibrating procedures and control procedures are usually made on equipements along each step while decision points made on the product itself are located at the end of a step.
The determination of the decision points is the first difficulty met when implementing SPC in quartz crystal manufacturing since almost all of the resonator characteristics are measurable at the end of the process or only when mounted on oscillator. The final characteristics of a quartz resonator are for most applications:

1 Nominal frequency
2 Resistance
3 Inductance
4 Spurious (frequency,attenuation,amplitude)
5 Temperature behaviour (frequency changes, inversion point, dips,...)
6 Aging
7 Drive level
8 Transit time
9 Mechanical behaviour (vibrations,acceleration sensitivity)

A product analysis is performed in order to identify the links between the final performances and some "substitution characteristics" directly measurable during the process. The resistance value is, for example, dependant on the blank surface quality, radius of curvature, etc ...
The identified substitutive characteristics are given in table I.

The systematic control of all these parameters along the process represents a huge amount of informations that would be quite difficult to handle. We need, at this step, to answer the second question in order to concentrate our efforts on parts of the process. The non-quality costs approach has been used for the localisation of the major problems; Figure 2 shows that the blanks processing and the automatic tuning are

Final characteristics	substitutions characteristics
1 Nominal frequency	blank thickness
	electrode thickness
	tuning precision
	stabilization cycles
	sealing process
2 Resistance	surface quality
	parallelism of the faces
	radius of curvature
	electrode size precision
3 Inductance	electrode size precision
	parallelism of the faces
	radius of curvature
	electrode misalignment
4 Spurious	electrode size precision
	parallelism of the faces
	radius of curvature
	electrode misalignment
5 Temperature behaviour	cut angle
	radius of curvature
	mounting structure
6 Ageing	Cleanlyness
	Stabilization
	Material
7 Drive level	Cleanlyness
8 Group delay	electrode size precision
	parallelism of the faces
	radius of curvature
	electrode alignment
9 Mechanical roughness	mounting structure

TABLE 1 - SUBSTITUTION PARAMETERS FOR RESONATORS PROCESS CONTROL

gathering more than 50% of the non quality costs (the automatic tuning difficulties have motivated the introduction of a new technology; see below: "tuning equipment"). The important part of the final control (36%) evidences that some substitution characteristics are still not well controled at earlier steps.

Pareto charts used at every step of the process give an insight into the "causes of the major problems". Figure 3 shows the current causes that produce loss of yield at the ion beam milling or at manual tuning steps.

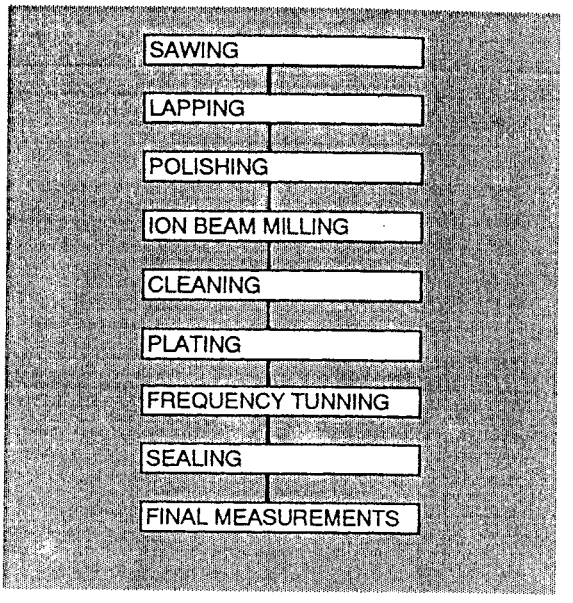


FIGURE 1 - MAJOR STEPS OF THE PROCESS FOR RESONATORS MANUFACTURING

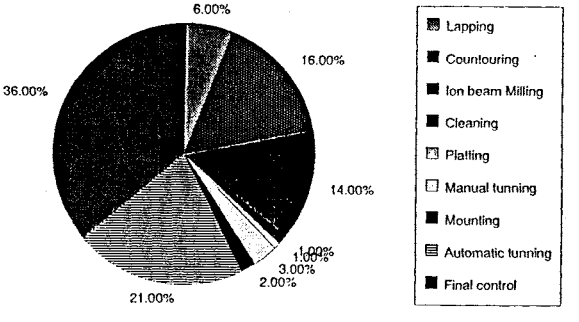


FIGURE 2 - NON QUALITY COSTS DISTRIBUTION FOR RESONATOR MANUFACTURING

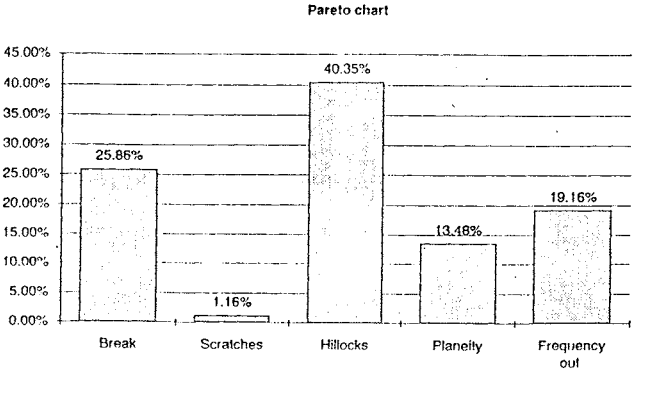


FIGURE 3 - IONIC ETCHING STEP: DISTRIBUTION OF THE CAUSES OF DEFECTS

4.2./ BLANK PROCESSING

Blank processing includes sawing, lapping and ion beam miiling. The orientation of quartz plate has

influence on several final characteristics of the resonators, consequently, for every batch of sawing, each plate is measured using an X-ray goniometer and histogrammes are plotted (Fig.4). The difference between the target angle and the average angle of the batch as well as the dispersion are plotted on a control chart to ensure a high degree of reproducibility.

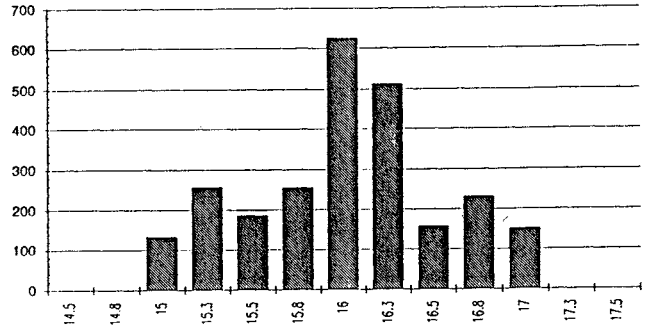


FIGURE 4 - CUT ANGLE HISTOGRAMME AFTER SAWING (HORIZONTAL UNITS: MINUTE)

The same method has been followed for the lapping process which is performed under automatic lapping control. Each batch of lapping is caracterised by an average frequency and a dispersion recorded by an air gap electrode technic. Figure 5 shows the control chart of the difference between target frequency and the average batch frequency for 150 runs; the dispersion is typically 100 KHz.

A special emphasis has been given at the ion beam milling step. For each of the causes presented on figure 3 a given number of causes/effect relations have been evidenced and corrective actions have been conducted :

1/ Frequency out of tolerance : the etching rate of each equipment is plotted on a control chart for each run and the ion beam current density is measured before starting a run.

2/ Planeity : the dispersion of each run is plotted on a control chart and the spacial homogeneity of the beam is checked by a specific apparatus.

3/ Hillocks at the surface of the blank, due to pollution during the etching cycle, is one of the thoughest problem that we met. The only solution up to now has been an intensive maintenance to preserve cleanliness of the equipment.

4/ Scratches at the blank surface and breaking were partly due to an inadequat frequency control equipment that has been modified.

The output is a reduction of the contribution of the first two points and elimination of scratches; although hillocks proportion has been reduced, it still remains to improve.

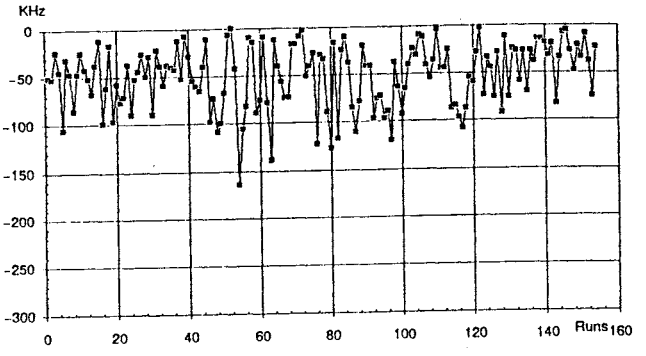


FIGURE 5 - CONTROL CHART FOR AUTOMATIC LAPPING CONTROL

4.3./ RESONATOR PROCESSING

Proportion of non conformity charts are used for all the minor steps (minor in the non-quality costs sense) as shown figure 6 for the plating operation. The advantage of this kind of control charts is to deal with non measurable parameter, e.g. visual control of planeity, and to be easely implemented.

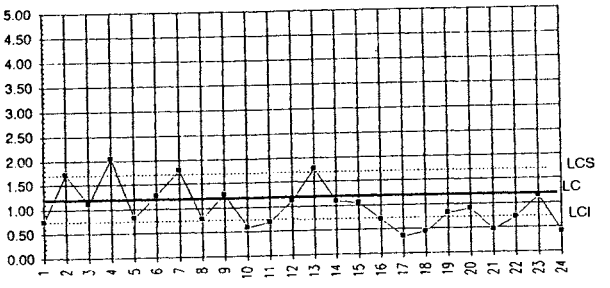


FIGURE 6 - TYPICAL CONTROL CHART FOR NON MEASURABLE PARAMETERS QUANTIFIED BY THE PERCENTAGE OF REJECTED PIECES

Statistical Process Control for quartz crystal industry need some deviation from standard SPC for three main reasons :

1/ A substantial number of controls are completly visual and always ambigous to deal with limits of acceptance.

2/ Some important steps of the process are manual; this introduces an operator dependant factor.

3/ The quantity of products having exactly the same characteristics is, in most cases, rather low. Any statistics is thus quite difficult to analyse .

The approach we have followed is based on non quality costs to identify where the problems are; Then we needed histogramms to evidence where these problems came from and finally control charts or non conformity charts on critical parameters helped us to initiate some solutions.

5./ TUNING EQUIPMENT

Two major process have been used for long for the tuning of piezoelectric devices. Each of them has inconveniences regarding the to-day application :

- electrolytical tuning. The major inconveniences of this process comes from the fact that it cannot easily be performed on a automatical way because the resonators cannot be accurately enough measured when immersed in the bath.

- Automatic plating. In that process plating is achieved by any possible mean (evaporation, sputtering, ...) under vacuum. The metal deposits on the plate through masks according to a shape defined by the mask openings. The frequency of the resonator is permanently measured; when it reaches the target value shutters are closed preventing any adding deposition. This method is applied for single resonators or for monolithic cells. The major inconveniences regarding our aim are: the need for accurate mechanisms working under vacuum, a quite important consumption of metal and a lack of accuracy for the electrode shape definition due to the spreading of evaporated metal caused by the unavoidable distance between the masks and the plate (to give an order of magnitude, the distance between the mask and the plate is hardly less than 1 mm while the electrode size is 0.1 mm).

In order to face the inconveniences of this to-day state of the art method a new process has been conceived. Firstly it has been established that the process should benefit of the best possible precision for the electrode shape definition: this is given by the plating through contact masking. During contact masking, there is no possibility for tuning, however the process can be achieved collectively which is cost saving as for as both working time and metal consumption. Consequently, the tuning method has to move from the previous adding method to a substractive method. The first studied substractive method has been plasma etching. It has demonstrated to be very promizing regarding tuning accuracy, ability for selective tuning of several resonators on the same plate, ability for automatization. However, it failed on the important subjects of speed of tuning and tuning amplitude: in order to compensate for plate eching spread, the tuning range must extend over 10^{-3} of the resonant frequency. But the plasma tuning appeared to be limited towards a few 10^{-5}. The method which has been finally chosen is the ionic etching [1]. According to this method, the

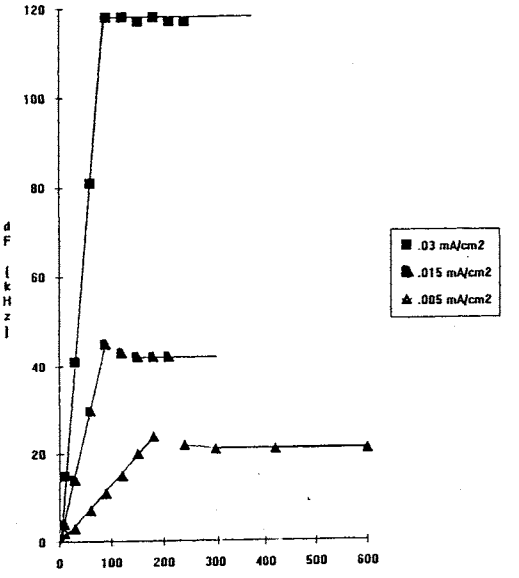


FIGURE 7 - VARIATION OF THE RESONANT FREQUENCY OF A 70 MHz RESONATOR DURING TUNING UNDER ION BEAM MILLING, FOR DIFFERENT VALUES OF THE ION CURRENT DENSITY.

resonator to be tuned is placed in an ion flow generated by a small 3 cm ion gun. The tuning is performed by the metal etching under ion bombardment. When the required frequency is reached, the accelerating tension of the gun is simply turned off. The process has demonstrated to be satisfactory on both technical and economical requirements.
Several experiments have been performed in order to demonstrate the capability of this solution. The main results are presented.

5.1./ TUNING RANGE, SPEED AND ACCURACY

Figure 7 shows the variation of the resonant frequency of a 70 MHz resonator during tuning for different values of ion current density.
The noticeable observations are the following: 1/ the range of tuning is very important: under a 0.03 mA/cm^2, a 120 kHz change in frequency has been achieved. This amplitude is consistent with the need for compensation for the thickness spread of the plates. 2/ The speed of tuning is high: a 120 kHz shift has been achieved in hardly more than one minute. Even higher rates have

been demonstrated to be feasible. It is fully consistent with a high throughput on a manufacturing line. 3/ The variation of the frequency is very linear versus exposure time. This a great advantage considering the automatic control of the tuning: the actual frequency of the resonator under tuning can easily be predicted on a time basis and only time to time measurements are necessary to control the ratio of tuning. 4/ The tuning rate can be modified simply by acting on the ion current density. It is of particular interest for accuracy of tuning: the tuning rate can be made slower when approaching the target when a very fine accuracy is required. On the other hand, for resonator only requiring a poor accuracy of tuning, the tuning rate can be made faster. 5/ The process also demonstrates a good control of the frequency at stopping point: the angular points on the curves are due to the stop command given to the process. It can be seen that there is no extra variation of frequency after the stop command has been given. The slight variations which are observable on the lowest current density curves are due to measurement unaccuracies.

5.2./ WARMING EFFECT

One expected limitation of the method was the warming of the resonator during exposition to the ion beam. A demonstration has been made on the following way: a 10 Mhz Y-cut resonator has been exposed at various beam intensities for one minute, then released. The frequency is measured over two minutes from the beginning of the exposure. Due to the linear frequency versus temperature behaviour of the Y-cut (1 kHz /°C), the observation of the frequency shifts gives a combined information on the temperature effect and the etching effect. Figure 8 gives a typical set of curves for an electrode diameter of 1.9 mm.
On the first 60 seconds the frequency is increased both by temperature increase and etching effect. When the beam is stopped, the etching effects is stopped and the frequency decrease is only caused by the temperature recovery of the resonator. On these curves, the temperature effect is much bigger than the

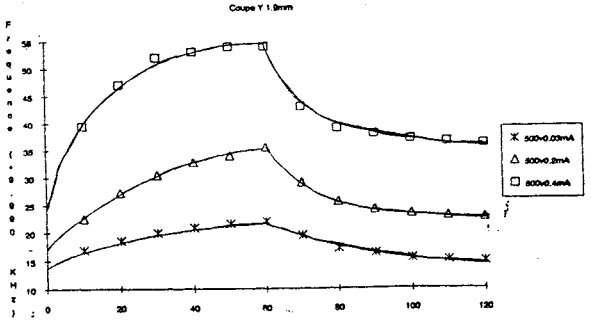


FIGURE 8 - VARIATION OF FREQUENCY FOR A 10 MHZ Y-CUT RESONATOR VERSUS TIME FOR DIFFERENT BEAM ENERGIES. THE RESONATOR IS EXPOSED TO THE BEAM FOR 60 SECONDS THEN RELEASED FOR 60 SECONDS (ELECTRODE DIAMETER: 1.9 MM).

etching effect: it is due to the high sensitivity of the frequency versus temperature for an Y-cut (100 ppm/°C). When the frequency shift is converted in temperature values, the temperature increase vesus beam energy is given by figure 9.

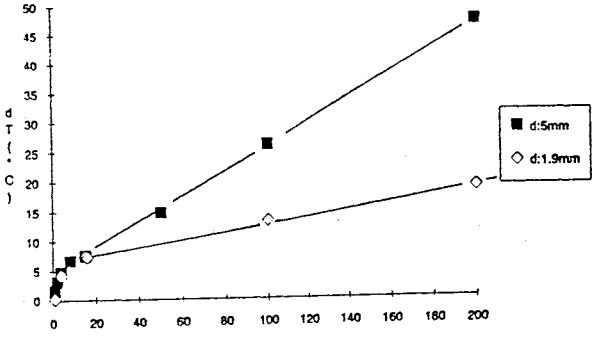


FIGURE 9 - TEMPERATURE INCREASE VERSUS BEAM ENERGY FOR A 10 MHZ RESONATOR EXPOSED TO A ION BEAM, FOR 2 DIFFERENT DIAMETERS OF THE ELECTRODE.

The following comments can be made: 1/ The temperature increase reaches a limit due the balance between the energy coming from the beam and the radiated energy from the plate (fig.2). 2/ For energy values below 100 mW/cm^2, the temperature increase remains moderate (fig.3: about 10°C for a usual electrode size). Applied to an AT-cut resonator, this temperature shift leads to a negligible 1 ppm frequency shift. In addition, the energy involved in that experiment must be compared to the energy actually needed to achieve a fast enough tuning: figure 1 shows that a convenient tuning rate is got with 15 mW/cm^2. 3/ In case of high beam energies, the recovery time by radiating effect is lower than 1 minute (fig.8). So it is possible to achieve a two times tuning: first tuning is performed at a very high rate for a rough approach, then the resonator is exposed at a lower rate for fine tuning. A one minute delay is required between the two steps, but it is of no consequence when working with batches: firstly, each resonator of the batch is exposed to the higher rate, then, when coming back to the first resonator for fine tuning, the spent time has been sufficient for its recovery. A considerable reduction of processing time may result of this method for high ranges of tuning.

5.3./ VARIATION OF THE ELECTRICAL CHARACTERISTICS

Another point of interest is the evolution of the electrical characteristics of the resonator during etching. It is well known that such tuning

methods as plasma etching are spatially inhomogeneous leading to variations of removal rates of metal on the extent of the electrode. Excess of plasma etching duration may even lead to perforate the electrode. Consequently, the electrical characteristics of the resonator such as self and resistance gradually evoluate during plasma etching operation. Here is the origin of the limit on the tuning range associated to this method. As for as ion beam tuning, it has been demonstrated that no noticeable variations of the electrical parameters occur for very high tuning ranges. Actually, the etching rate is perfectly homogeneous over the electrode extent and the electrical characteristics remain unchanged until the sudden disappearence of the whole electrode. Figure 10 shows the variations of the electrical inducatnce for various frequency shifts on a 70 MHz resonator.

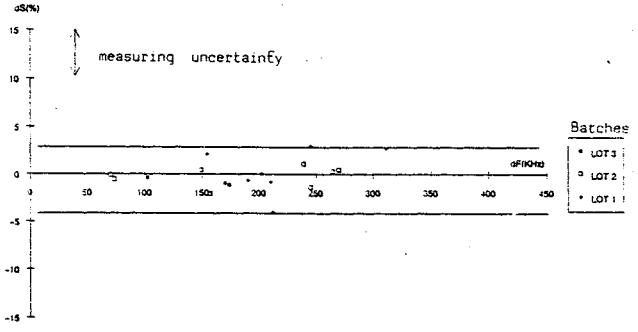


FIGURE 10 - VARIATIONS OF THE INDUCTANCE OF A 70 MHZ RESONATOR FOR VARIOUS FREQUENCY SHIFTS

5.4./ TUNING OF MULTI-RESONATORS PER PLATE

As it has been stated above, a very important requirement for the tuning equipment is the ability to separately tune multiple resonant functions gathered on the same plate. In order to face this problem, a special method has been developped [2]: the electrode of the resonator for which

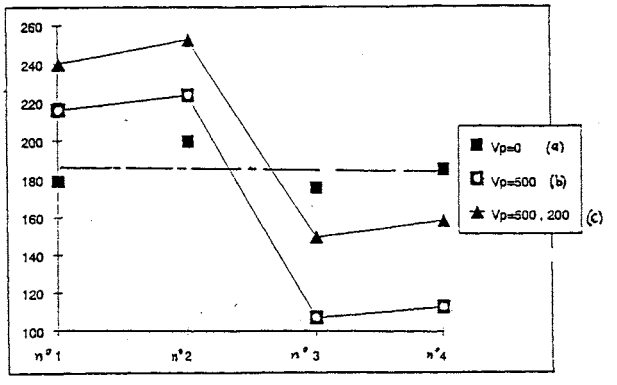


FIGURE 11 - VARIATION OF FREQUENCY FOR A FOUR 70 MHZ RESONANT FUNCTIONS ON THE SAME PLATE EXPOSED TO A 50 MW/CM^2 ION BEAM, FOR3 DIFFERENT POLARIZATION SCENARII.

tuning must be stopped is connected to a positive electrical potential whose value is part or the whole of the acceleration potential of the beam. When so polarized the electrode is no more attractive for ions and etching is stopped. Figure 11 illustrates this method. The samples are composed of 4 resonant functions on the same plate (n. 1 to 4 in absissa) at 70 MHz. The samples are exposed to a 50 mW/cm^2 ion beam accelerated by a 500 V potential, according to the following description: (a) The whole plate is exposed to the beam during 2 minutes. The mean frequency shift is about 190 kHz, the spread is about 10 kHz. (b) The whole plate is first exposed during 1 minute, then the resonant functions n.3 and 4 are set to a potential of 500 V, and the whole plate is exposed to an adding 1 minute: the total shift on resonant functions 1 and 2 is 220 kHz while it is only 110 kHz for n. 3 and 4. It demonstrates that during the time where they are polarized, electrodes n. 3 and 4 have not been etched. (c) The whole plate is first exposed to the beam during 1 minute without polarization, then it is exposed to an adding 70 seconds with the following polarization sequence for resonant functions n. 3 and 4: 30 seconds under a 200 V polarization, then 40 seconds under a 500 V polarization. The total shift on resonant functions 1 and 2 is 240 kHz, while it is only 160 kHz for n. 3 and 4. This demonstrates the linearly dependancy of the tuning rate on the polarization value.

6./ CONCLUSION

In order to face the communication market requirement for wide band filter manufacturing (typically 200 kHz bandwidth at 200 MHz), and the need for quite high volumes, the whole manufacturing approach had to be changed. The main improvements have been oriented toward yield and throughput increase. Two of them have been here selected in order to demonstrate this new approach: 1/ The statistical process control applied to the resonator manufacturing is quite unavoidable to achieve high yield resonators having tiny tolerance requirements due to the very low thickness of the plate and the small size of the electrodes. 2/ The usual means of tuning appeared to be widely unaccurate for the range of specification: it has been demonstrated that a new method involving ionic etching of the plating is far more suitable regarding accuracy over the whole electrical characteristics and tuning speed. In addition, separate tuning of different resonant functions gathered on the same plate can be achieved by that mean.

7./ BIBLIOGRAPHY

[1] Castellano, Hokanson
[2] patent pending

NEW ATOMIC FREQUENCY STANDARDS

Chairman: C. Audoin

LASER-COOLED METASTABLE XENON: PROGRESS TOWARD AN OPTICAL FREQUENCY STANDARD

A. Witte, S.L. Rolston, W.D. Phillips

Atomic Physics Division
National Institute of Standards and Technology
Gaithersburg, MD 20899, USA

H.J.L. Megens

Eindhoven University of Technology, Eindhoven, The Netherlands

and M. Walhout

University of Maryland, College Park, Maryland 20742, USA

Abstract

We report on our efforts to realize an optical frequency standard based on a two-photon transition in metastable xenon. Optical frequency standards offer the possibility of stabilities that are orders of magnitude higher than in current state-of-the-art cesium standards. To realize the potential accuracy of 10^{-18} of this system, very slow atoms are required. Metastable xenon is ideal for laser-cooling, and atomic velocities as low as 1 cm/s can be reached. We report on the first laser-cooling and magneto-optical trapping of metastable xenon atoms.

Keywords: Frequency standards, laser-cooling, atomic fountains

1. Introduction

The motion of atoms is one of the major sources of imprecision and inaccuracy in atomic frequency standards. This has been one of the primary motivations for the development of the field of laser cooling of atoms. In contrast to trapped ion standards, neutral atom frequency standards cannot have the nearly infinite observation time available to trapped ion standards, so the linewidth of the clock resonance in a neutral atom standard will not generally be as narrow as in a trapped ion standard. On the other hand, neutral atom standards may have significantly better signal-to-noise performance, since they are not limited to a few atoms. Recent advances in the field of laser-cooling now allow interaction times of the order of one second for free atoms, allowing Q-values of about 10^{14} with optical transitions.

A two-photon-transition in metastable xenon was recently identified (Ref. 1) as a promising candidate for an advanced optical frequency standard. The two-photon transition between the $^3P_2$ and the $^3P_0$ metastable states has a predicted linewidth of about 2 Hz, yielding a Q of 10^{14}. The wavelength of the necessary radiation is 2.19 μm, which may be supplied from a number of different sources, such as a CW optical parametric oscillator.

One difference between laser cooling of neutral atoms and of ions is that there are no deep traps for neutral atoms, so they must be cooled while in flight, typically in an atomic beam. They can then be further cooled while confined in optical molasses (Ref. 2) or a shallow magneto-optical trap (MOT) (Ref. 3). Since the neutral atoms cannot be trapped by virtue of a net charge, as are ions, traps for neutral atoms are formed by imposing spatially dependent perturbations on internal atomic energy levels. Such perturbations of the energy levels are usually unacceptable for a frequency standard, so most proposed neutral atom standards use free atoms, possibly released from a trap.

Metastable xenon is ideal for laser-cooling, possessing many of the same properties as cesium. Because the attainable flux of metastable atoms is considerably lower than that obtainable with ground-state atoms, efficient cooling is required to get a high flux of cold metastable atoms, which is essential for a good signal-to-noise performance of the frequency standard.

In the first part of this paper we outline a future frequency standard using metastable xenon. In the second part, we

describe our techniques for the efficient preparation of a cold sample of metastable xenon atoms.

2. A frequency standard based on metastable xenon atoms

Metastable xenon has been identified as an excellent candidate for an optical frequency standard (Ref. 1). A partial diagram of the energy levels for xenon is shown in Fig. 1. The first excited state is extremely metastable, with a calculated lifetime of 150 s (Ref. 4). For the purpose of the clock and the laser-cooling we will consider this $^3P_2$ state to be the ground state. The laser-cooling transition with a cycling transition at 882 nm is ideal for a Titanium:sapphire laser. The lifetime of this transition is 30 ns. The degenerate ground state with J=2 allows the use of polarization gradient cooling methods (Refs. 5, 6), with a minimum attainable rms velocity v_{rms} of 3 photon recoils in one dimension, which is v_{rms} = 1 cm/s in the case of xenon.

The optical clock transition (see Fig. 1) is a two-photon transition connecting the $^3P_2$ state with the $^3P_0$ state that is 9129 cm^{-1} higher. The $^3P_0$ state decays to the $^3P_1$ state via a magnetic dipole transition, and has a calculated lifetime of 0.08 s (Ref. 4), which corresponds to a linewidth of 2 Hz. The $^3P_1$ state in turn decays in 4 ns to the atomic ground state. The wavelength of the radiation necessary for the two-photon transition is 2.19 μm. A possible source for this radiation is a CW optical parametric oscillator pumped by a Nd:YAG (Ref. 7).

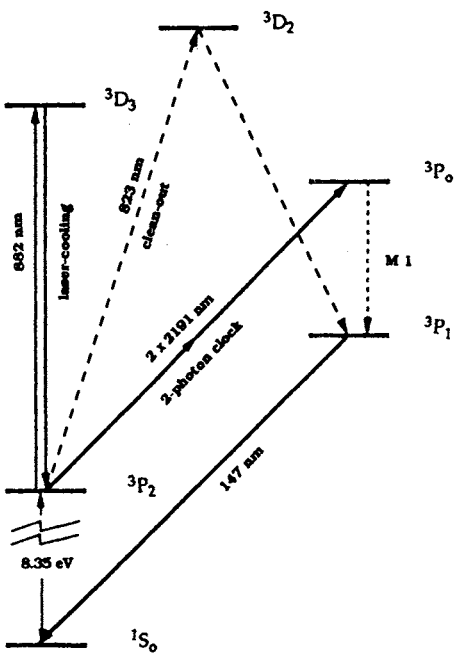


Figure 1: Energy levels and relevant transitions in xenon (not to scale).

The metastable $^3P_0$ "clock state" has 9.5 eV of excitation energy, which allows the possibility of metastable detection. Quantum efficiencies in the range of a few to fifty percent are possible (Ref. 8). Of course any metastable detector would also be sensitive (although somewhat less) to the lower lying metastable state (with 8.3 eV of excitation energy), but this state can be completely depopulated by the application of radiation from a laser diode at 823 nm that drives an allowed transition (see Fig. 1) with a 26% branching rate for decay to the $^3P_1$ state. Application of 100 $\mu W/cm^2$ of this radiation for one ms will completely remove any atoms that did not make the clock transition with an approximately 10^6 rejection ratio, allowing an essentially background-free method of detection. Furthermore, there are numerous possibilities for optical detection discussed at length in Ref. 1.

To realize long interaction times with the light fields interrogating the clock transition, we envisage using an atomic fountain. The atomic fountain was first proposed and attempted by Zacharias (Ref. 9) and has been studied by other authors (Ref. 10). Atoms are launched upward, and pass twice through the same interaction region (a microwave or optical cavity), once on the way up and once on the way down which constitutes the two interactions separated in time to produce Ramsey fringes (Ref. 11), with a resolution determined by the time between the two interactions. This can be a very powerful technique, because the interaction time can be very long (a 30 cm fountain gives a 0.5 sec interaction time) in a relatively compact apparatus. In practice the interaction region would be a high finesse optical cavity, which gives considerable power buildup, so as to lessen the power requirements on the optical source and assures a good balance between the optical waves propagating in opposite directions. The fountain geometry has an additional advantage in that the two Ramsey interactions occur in the same cavity, which will help minimize some sources of error.

To generate the sample of metastable xenon atoms to be launched to form a fountain, we decelerate the metastable xenon atoms in an atomic beam using the Zeeman cooling method and load them into a MOT. The temperatures achieved in a MOT are generally not as low as those in molasses, but the trap can be turned off, and molasses with appropriate detuning and intensity turned on for a short period of time. Because typically only about 10^{-5} of the atoms in the atomic beam are metastable, special care has to be taken, to achieve a high flux of cold metastable atoms. Our method for the cooling and trapping of the

metastable atoms is described in detail in section 3.

To estimate the stability of the xenon frequency standard, we will assume 10^7 atoms in the MOT. Since the lifetime of the upper metastable state is only about 0.1 s, there is no advantage to be gained in constructing a fountain with a flight time much more than about twice this lifetime (the Ramsey method produces a ground-excited state coherence that decays at half the rate of the excited state). In this case we will consider a 5 cm fountain. Note that a shorter fountain, with less time for the atoms to spread out, is well suited to a MOT with its small volume, and an optical clock with a small interaction region. We further assume 20% of the atoms to be in the m=0 ground state, 50% to be lost due to spreading and other effects, and a one second load time. Because of spontaneous emission from the $^3P_0$ state, the signal will be reduced by a factor of three during the 0.2 s free fall. We find the Allan variance to be $\sigma(\tau) = 2 \times 10^{-17}$ at one second and better than 2×10^{-18} at 100 seconds if the stability were limited by shot noise and a 10% detection efficiency. This is a considerable improvement over present day laboratory Cs standards (Ref. 12), which have $\sigma(\tau) = 1\text{-}5 \times 10^{-12}$ at one second. It should be mentioned, that presently no "flywheel" oscillators exist with the stability required.

Unfortunately, this phenomenal stability is not easily matched in accuracy in the neutral atom optical clock. One source of inaccuracy is the AC-Stark shift (light shift) caused by the light fields driving the two-photon transition. The AC-Stark shift has a very similar form to the two-photon Rabi frequency, since it is essentially the difference between the energy shifts of the ground and excited states due to two-photon transitions that couple through intermediate states back to the same initial state. For the specific example of metastable xenon, preliminary calculations have been performed in (Ref. 1) using calculated oscillator strengths (Ref. 13), and summing over 6 intermediate states. The result is, that for a 1 ms interaction time, the intensity required to produce a $\pi/2$-pulse causes a 0.6 Hz AC Stark shift, which corresponds to a 5×10^{-15} shift in the frequency. Control of the intensity of the radiation to 1% should lead to an instability in the shift of 5×10^{-17} and to realize the 10^{-18} long term stability implied by the previous analysis will require controlling the intensity to .02%, certainly a challenge, although not an impossibility. If we choose to reduce the optical power and accept a loss in signal-to-noise, so that the stability reaches 10^{-18} in one day instead of 100 seconds, the instability due to the AC Stark shift can be reduced to the 10^{-18} level and would only require 0.6% amplitude stabilization.

What is often the primary source of inaccuracy in a microwave standard is the distributed cavity phase shift due to intensity imbalance in the two travelling waves that form the standing wave in the cavity. This produces a residual first order Doppler shift which can be seen to arise from a phase shift across the cavity. This particular phase shift is totally absent in the two-photon optical clock, since the Doppler-free two-photon transition requires the near-simultaneous absorption of two photons from opposite directions, and is insensitive to intensity imbalance in the counterpropagating beams. There is, however, an important shift due to an angular misalignment of the optical waves in the cavity (or mismatch in the wavefront curvature of the counterpropagating waves), which allows some residual first-order Doppler shift to appear.

The Doppler-free nature of the two-photon transition relies on the cancellation of the wave vectors due to their opposite directions of propagation. If the transverse velocity of the atoms in the fountain is 1 cm/s, the first-order Doppler broadened line would be approximately 10 kHz wide, which is about 10^{-10} of the optical frequency. This effect will inhomogeneously broaden the line. To reduce this effect to less than the natural width will require the wave vector mismatch to be $\delta k/k < 10^{-3}$. If we assume the transverse velocity distribution is symmetric to within 10%, a reduction in the shift in the center frequency to less than 10^{-16} will require $\delta k/k < 10^{-5}$. For this reason it will be important to use a high finesse optical cavity, probably in the confocal configuration. The mode of such a cavity is insensitive to mirror misalignment, and should allow a reduction of the wavevector mismatch to a level much better than the inverse of the finesse of the cavity. A simple estimate of the mismatch for a plane-plane cavity would be $\delta k/k \cong (w_0/L)1/F$, where w_0 is the mode size on the mirror, L is the cavity length and F is the cavity finesse. For a cavity with $F = 10^4$, $\delta k/k$ could be about 10^{-6}, and should be much better for a confocal cavity which is insensitive to angular misalignment. As with the AC Stark shift, this source of error can also be reduced at a cost of signal, in this case by collimating the launched atoms to reduce the transverse velocity.

Collisions have been demonstrated to produce sizable frequency shifts in a frequency standard based on a cesium fountain (Ref. 14) and might finally limit the attainable accuracy of this standard.

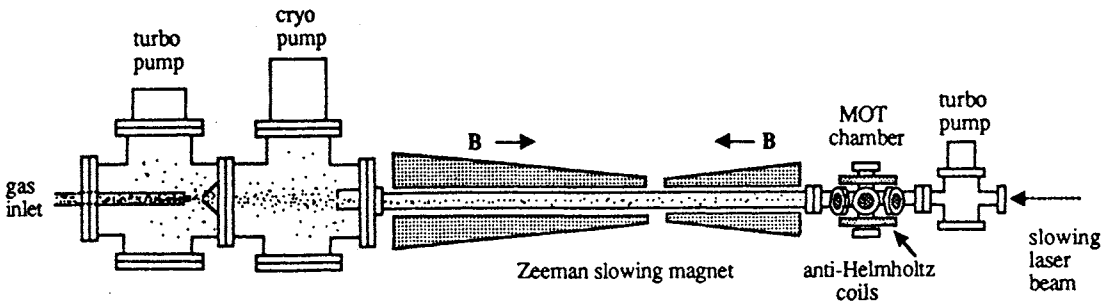


Figure 2. Apparatus for decelerating and trapping a beam of metastable xenon.

The effect of collisions on the xenon standard remains to be investigated. Because the interatomic potentials relevant in these processes are not known to sufficient accuracy, these investigations will have to be performed experimentally.

Finally, the gravitational red shift plus the second-order Doppler shift produce a fractional shift (Ref. 11) of $1/6\ v_0^2/c^2$, which for the 5 cm fountain we are considering would be 2×10^{-18}.

3. Experimental set-up for cooling and trapping

In this section we describe our method to generate cold xenon metastable atoms at high flux rates. We decelerate a beam of xenon atoms in the metastable $^3P_2$ state using the familiar Zeeman slowing technique (Ref. 15), in which the changing Doppler shift of the resonance frequency of the atoms is compensated by a changing Zeeman shift. The slow atoms are then collected in a magneto-optical trap. Our method of Zeeman deceleration resembles that used for calcium in Ref. 16. This method generates a nearly monoenergetic beam of slow atoms with velocities in the capture range of a MOT, so that essentially all decelerated atoms can be loaded into the MOT. This method furthermore uses a magnetic field, that passes through zero and a corresponding red detuning of the cooling laser beam, which results in a frequency offset between the cooling laser beam and the MOT beams. At a sufficiently large offset, the trap can be placed on axis of the atomic beam and the cooling laser beam, because the atoms in the trapping region are far out of resonance with the cooling laser beam. This geometry allows very efficient loading of the trap.

Our apparatus is shown schematically in Fig. 2. The metastable source, which is similar in design to that of Fahey et al. (Ref. 17), is a quartz chamber filled with xenon gas (in natural isotopic abundance) to a pressure of 100 to 250 Pa. Atoms escape from the source with an average speed of about 300 m/s through an aperture 0.14 mm in diameter. The beam is collimated by a 1-mm skimmer aperture placed 1 cm from the source and by a similar aperture 30 cm further downstream. As the gas leaves the source, it is excited by a dc discharge running between the skimmer and a cathode filament inside the quartz chamber. Atoms that emerge from the source in the $^3P_2$ state are accessible for laser cooling. All of the laser beams in the experiment are derived from a single Titanium:sapphire laser, which is locked to the $^3P_2$ - $^3D_3$ cooling transition at 882 nm using saturated absorption in a dc discharge cell. Different frequencies for the slowing, trapping and probe laser beams are generated by acousto-optic modulators (AOM's).

The slowing magnet is composed of two tapered solenoids placed end to end and producing fields in opposite directions. The field along the atomic beam decreases monotonically in the first solenoid from 16 mT near the last collimating aperture to zero at a point 1.1 m downstream. In the second solenoid the field continues to decrease from zero to -6.5 mT over a distance of 0.4 m. A 10 mW, circularly polarized laser beam is directed against the atomic beam and detuned 135 MHz below the atomic resonance frequency. As the atoms travel the length of the magnet, they scatter photons from this slowing laser beam and decelerate. The changing Zeeman shift compensates for the changing Doppler shift, so that resonant scattering is maintained. At the end of the slowing region, the magnetic field changes rapidly from -6.5 mT to zero, and the atoms go out of resonance with the slowing laser beam, having reached an

average velocity of 20 m/s with a spread of about 5 m/s.

The slow atoms are captured in a MOT positioned 15 cm beyond the end of the slowing magnet. The trap is formed at the intersection of three orthogonal pairs of counterpropagating, oppositely circularly polarized laser beams, and at the center of a quadrupolar magnetic field produced by a pair of anti-Helmholtz coils. The field gradient along the MOT coil axis is typically 0.5 mT/cm, and the MOT beams are 2 cm in diameter. The laser power in each beam is normally about 15 mW, and the trap is brightest for a MOT laser detuning of about 3 MHz, or approximately half a linewidth, below resonance.

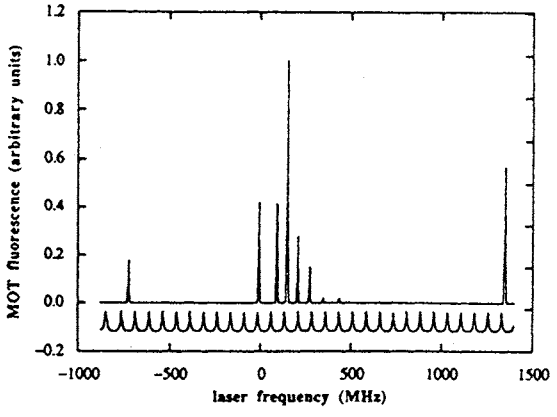


Figure 3. The fluorescence from the MOT exhibits nine peaks, corresponding to trapping of the nine xenon isotopes, as the frequencies of the slowing and trapping lasers are scanned simultaneously. From left to right, the peaks are identified as ^{131}Xe, ^{136}Xe, ^{134}Xe, ^{132}Xe, ^{130}Xe, ^{128}Xe, ^{126}Xe, ^{124}Xe, and ^{129}Xe. Also shown is the periodic transmission of the laser light through the Fabry-Perot etalon with a free spectral range of 75 MHz.

We have used our trap to capture all nine stable isotopes of xenon. These include ^{124}Xe and ^{126}Xe, each of which has less than 0.1% natural abundance (see Fig. 3). We estimate the mean atomic density in the MOT to be $2\text{-}3\times10^{10}$ atoms/cm^3. This value is higher than densities obtained in traps for other metastable rare gases, which are more strongly limited by Penning ionization. With He* and Ar* MOT's, for example, the highest measured densities were in the ranges of 10^8 atoms/cm^3 and 10^9 atoms/cm^3, respectively (Refs. 18, 19). Collisions between trapped atoms contribute strongly to trap loss. Typically we find that the fluorescence decreases to half its initial value in a few tenths of a second. For the more abundant isotopes, we consistently see a non-exponential decay of the MOT fluorescence at shorter times, indicating a density dependent loss mechanism due to collisions between trapped atoms.

As a first application of the trapped atoms, we measured the isotope shifts of the cooling transition and the hyperfine constants for ^{131}Xe with increased accuracy (Ref. 20). Presently we are in the process of measuring the lifetimes of the metastable $^3P_2$ and $^3P_0$ states by detection of the VUV-photons emitted from these states. The lifetime of the $^3P_0$ state will finally limit the attainable Q-value of the clock.

This work was supported by the Office of Naval Research. A.Witte thanks the Alexander-von-Humboldt foundation for its support.

References

1. S.L. Rolston and W.D. Phillips, Proc. IEEE **79**, 943 (1991).

2. See e.g.: P.D. Lett, W.D. Phillips, S.L. Rolston, C.E. Tanner, R.N. Watts, C.I. Westbrook, J. Opt. Soc. Am. B **6**, 2084 (1989); D.S. Weiss, E. Riis, Y. Shevy, P.J. Ungar, S. Chu, J. Opt. Soc. Am. B **6**, 2072 (1989).

3. E. Raab, M. Prentiss, A. Cable, S. Chu, and D. Pritchard, Phys. Rev. Lett. **59**, 2631 (1987).

4. N. Small-Warren and L. Chow Chiu, Phys., Rev. A**11**, 1777 (1975).

5. J. Dalibard, C. Cohen-Tannoudji, J. Opt. Soc. Am. B **6**, 2023 (1989).

6. P.J. Ungar, D.S. Weiss, E. Riis, S. Chu, J. Opt. Soc. Am. B **6**, 2058 (1989).

7. R. Byer, private communication.

8. F. Dunning, R. Rundel, and R. Stebbings, Rev. Sci. Instrum. **46**, 697 (1975).

9 J. Zacharias, Phys. Rev. **94**, 751 (1954).

10. See e.g.: R. Beausoleil et al., Phys. Rev. A**33**, 1661 (1986); A. DeMarchi, Metrologia **18**, 103 (1982); M. Kasevich et al., Phys. Rev. Lett. **63**, 612 (1989); A. Clairon et al., Europhys. Lett. **16**, 165 (1991).

11. N. Ramsey, Phys. Rev. **76**, 996 (1949).

12. J. Vanier and C. Audoin, *The Quantum Physics of Atomic Frequency*

Standards, Adam Hilger, Bristol (1989) vol. 2, p. 782.

13. C. W. Clark, private communication.

14. K. Gibble and S. Chu, Phys. Rev. Lett. **70**, 1771 (1993).

15. W.D. Phillips, J.V. Prodan, and H. Metcalf, J. Opt. Soc. Am. B **2**, 1751 (1985).

16. A. Witte, T. Kisters, F. Riehle, and J. Helmcke, J. Opt. Soc. Am. B **9**, 1030 (1992).

17. D.W. Fahey, W.F. Parks, and L.D. Schearer, J.Phys. E**13**, 381 (1980).

18. F. Bardou, O. Emile, J-M. Courty, C. Westbrook, and A. Aspect, Europhys. Lett. **20**, 681 (1992).

19. H. Katori and F. Shimizu, Jap. J. Appl. Phys. **29**, 2124 (1990).

20. M. Walhout, H. Megens, A. Witte and S. Rolston, submitted to Phys. Rev. A.

FREQUENCY MEASUREMENT OF THE HYPERFINE SPLITTING OF $^{171}YB^{+}$ IONS STORED IN A PAUL TRAP

A. Bauch, D. Schnier, Chr. Tamm

Physikalisch-Technische Bundesanstalt, Braunschweig, Germany

ABSTRACT

A precision measurement of the ground state hyperfine splitting frequency of $^{171}Yb^{+}$ ions is described in which buffer gas cooled ion clouds stored in a Paul trap and a reference oscillator locked to an active hydrogen maser are used. The final value is f_0 = 12 642 812 118,471 Hz. The sum of systematic and statistical measurement uncertainties (1σ) is evaluated to be 9 mHz. The uncertainties are mainly caused by the determination of the frequency shifts resulting from the second-order Doppler effect, the quadratic Stark effect, and collisions with buffer gas atoms.

Keywords:

Trapped ions, ytterbium, optical and microwave spectroscopy

1. INTRODUCTION

High resolution spectroscopy of ions stored in a trap is regarded as a promising method to realize advanced atomic frequency standards. Among other candidates, the ytterbium ion has found wide attraction as it offers several possible "clock transitions" in the optical, infrared and the microwave frequency domain /1/. In previous experiments systematic uncertainties of 1,2 Hz /2/ and 0,4 Hz /3/ were reported for a measurement of the hyperfine splitting frequency of $^{171}Yb^{+}$ at 12,6 GHz, but the measurement results differed by 4,6 Hz. The present measurement is essentially free of two uncertainty contributions which appeared to be dominating in the previous investigations as the absolute frequency calibration and the quadratic Zeeman effect due to the ambient magnetic field do not contribute more than 0,4 mHz to the systematic uncertainty. The remaining sources of systematic uncertainties are relativistic Doppler effect, quadratic Stark effect and collisional frequency shift.

2. EXPERIMENTAL SET-UP

• Ion Trap

Clouds of approximately 10^4 ions are stored in a radio frequency trap in a Helium buffer gas atmosphere /4,5/. Isotope enriched material is used which contains 89 % ^{171}Yb and 8 % ^{172}Yb. Typical data of the trap and storage parameters are listed in table 1.

Table 1 Trap and storage data

r_0 = 3 mm, z_0 = 2,4 mm
U_0 = 16,5 V, V_0 = 380 V
$\Omega / (2\pi)$ = 2,12 MHz
a_z = -0,04, q_z = 0,53

The theory of Garboriaud et al. is occasionally used to determine the absolute number of ions from electronic detection of the stored ions /6/. The ion number given above may be in error by a factor of 2 - 4 as the line shape of the detection signal is distorted as a result of anharmonicities of the trap potential /7/. The background pressure is measured with a commercial quadrupole residual gas analyzer. After switching off the ion pump and filling the recipient with helium buffer gas, typically the residual gas composition is given by 90% helium, 5% hydrogen and 5% carbon monoxide, the amount of the latter two constituents increasing with time. No absolute calibration of the pressure scale has been made yet which entails some systematic uncertainty in the evaluation of the collisional frequency shift.

The trap and its supporting structure is made of non-magnetic material. Two layers of magnetic shielding and three sets of orthogonal coil pairs permit the adjustment of the magnetic flux density at the trap centre to very small values, typically to 0,25 μT.

• Microwave probing signal

The 5 MHz output signal of an active hydrogen maser, CH1-75 /8/, is multiplied to 12,6 GHz. This signal is mixed with the

output frequency of a dielectric resonator oscillator, DRO, at f_p = 12 642 812 xxx Hz. The signal at the difference frequency is mixed with the output of a commercial synthesizer. The beat signal is used to control the DRO. Computer controlled stepping of f_p with a resolution of 1 mHz is possible. The relative frequency instability of the probing signal as a function of the averaging time τ is estimated from the maser performance data to $3 \cdot 10^{-14}$ at τ=10s, $4 \cdot 10^{-15}$ at τ = 100s, $2 \cdot 10^{-15}$ at τ = 1000 s. The average maser frequency is known to $2 \cdot 10^{-14}$ with respect to PTB's primary clock CS2 which realizes the SI second with an uncertainty of $1{,}5 \cdot 10^{-14}$ /9/.

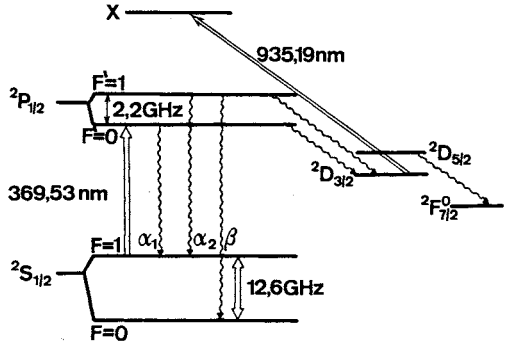


Fig.1 Hyperfine structure of the $S_{1/2}$ and the $P_{1/2}$ energy levels in $^{171}Yb^+$, other relevant energy levels are shown, too. "X" denotes the $4f^{13}5d6s(^3D)^3[3/2]^0_{1/2}$ level.

• Lasers and optical detection

Optical detection and state preparation (optical pumping) of the trapped ions is performed by driving the $^2S_{1/2}$ - $^2P_{1/2}$ resonance transition of Yb^+ at a wavelength of 369 nm (see Fig. 1). 5μW of narrow-bandwidth ultraviolet (uv) radiation are generated at this wavelength by frequency doubling of a continuous-wave single-mode dye laser. An auxiliary excitation at λ_{FL}=935 nm is used to deplete the metastable $^2D_{3/2}$ level which is populated during optical excitation of the resonance transition /10/. The λ_{FL}=935 nm infrared (ir) radiation is generated by a Nd^{3+}-doped fiber laser pumped by the fundamental dye laser radiation /11/. The efficient depletion of the $^2D_{3/2}$ level avoids population trapping in the extremely long-lived $^2F_{7/2}$ level which is otherwise observed in the presence of a buffer gas /12/.
The fluorescence light leaves the trap through a hole in the upper endcap, passes an interference filter and is detected with a photomultiplier. The overall detection efficiency is estimated to 10^{-3}. The background count rate (stray light and dark counts) is about 80 s^{-1}. The experimental set-up is sketched in Fig.2.

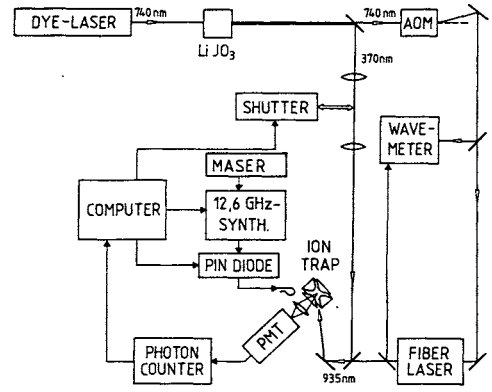


Fig.2 Experimental set-up

3. DOUBLE-RESONANCE SPECTROSCOPY

• Quasi-cyclic excitation

It is desirable to use an optical excitation scheme which permits a large number of photons per ion to be emitted after each microwave transition. This condition is met if the uv frequency is tuned close to the α_1 transition (see Fig.1). At a suitable laser tuning the ions undergo an average of a few hundred excitation cycles on the α_1 transition before they are optically pumped to the F=0 state. This large number of excitation cycles is only possible because after spontaneous decay $^2P_{1/2}$ F'=0 - $^2D_{3/2}$ F=1 and subsequent ir excitation, most ions are returned to the $^2S_{1/2}$ F=1 state. The actual number of excitation cycles on the α_1 transition is mainly determined by the overlap of the Doppler profiles of the α_1 and the α_2 transition /4/.
To obtain a high average number of photons per ion all Zeeman sublevels of the F=1 state must be excited in order to avoid optical pumping between the Zeeman sublevels. Using linearly polarized optical excitation, in our case a mixed polarization state is obtained by an oblique orientation of the applied static magnetic field.

• Sequential mode

In the sequential mode the microwave radiation is switched off during optical excitation. Within less than 1 second the ions are completely transferred to the F=0 level of the ground state. The gate time of the photon counter is set to a sligthly smaller value and a fluorescence count number S is obtained. While the uv laser is blocked by a mechanical shutter, a microwave π pulse of duration t_w is applied. The laser is switched on again and S is registered as a function of the setting of f_p during the preceding t_w interval. Repeating this sequence permits to determine the line shape and the centre frequency of the microwave transitions.

• Simultaneous mode and determination of the kinetic energy

If the laser fields and the microwave field are applied simultaneously, S is registered either as a function of the uv laser frequency (see Figs.4 and 5) or of the microwave probing frequency. The line width of the microwave resonance signal obtained in this mode is determined by the optical excitation rate which can be measured in this way.
It has been shown that the kinetic energy of trapped ions can be derived from the line width of an optical transition which in most cases is governed by the first-order Doppler effect /13/. This information is therefore obtained from a scan of the uv laser frequency with f_p tuned to resonance. However, at the prevailing experimental conditions, both position and width of the fluorescence lineshape of the α_1 transition are affected by competing excitation of the α_2 transition and by the resulting hyperfine pumping (see above). A rate equation treatment of the simultaneous action of both laser fields and of the microwave excitation predicts that the resulting signal height, the position of the fluorescence peak and its line width depend on the parameters microwave power, laser excitation rate, and kinetic energy of the ions /4/. Fig. 3 shows the calculated relation between the line width of the fluorescence signal, $\Delta\nu_I$, of the α_1 line and the Doppler width $\Delta\nu_D$.

4. EXPERIMENTAL RESULTS

• Spectroscopy of optical transitions

Depopulation of the $D_{3/2}$ state by the ir excitation was observed to be possible for two slightly different wavelength settings of the fiber laser. The frequency difference $\Delta\nu = c\cdot(1/\lambda_1 - 1/\lambda_2) = 2{,}5$ GHz might be related to the hyperfine splitting of the $4f^{13}5d6s(^3D)^3[1/2]^0_{1/2}$ level of $^{171}Yb^+$. For the two settings of the wavelength λ_{FL}, Figs. 4a and 4b show the signals obtained when the uv radiation is scanned. One signal peak which is present in both scans can be attributed to the α_1 transition of $^{171}Yb^+$. If the ir wavelength λ_2 is selected, a second fluorescence peak with larger optical line width is observed which can be attributed to $^{172}Yb^+$ as it is insensitive to the microwave excitation.
It is reasonable to assume that the line width of the $^{172}Yb^+$ signal corresponds to the Doppler width $\Delta\nu_D$ while the $^{171}Yb^+$ signal is affected by hyperfine pumping, thus showing a reduced line width $\Delta\nu_I$. The line widths of the two peaks were determined using a cloud of ions at various buffer gas pressures. Fig. 3 shows that the measured line width data are in good agreement with the dependence of $\Delta\nu_I$ on $\Delta\nu_D$ as expected from ab-initio rate equation calculations.

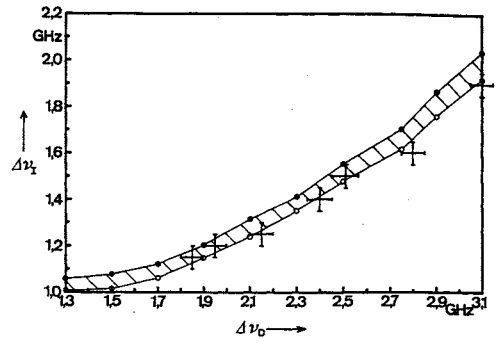


Fig.3 Optical line width of the fluorescence signal of the α_1 transition of $^{171}Yb^+$, $\Delta\nu_I$, as a function of the Doppler width $\Delta\nu_D$; $\Delta\nu_I$ is reduced compared to $\Delta\nu_D$ as a result of hyperfine pumping. $\Delta\nu_I$ is calculated to be in the range indicated by "•" and "O", using independently determined experimental parameters and their possible variations. "+": experimental data obtained by identifying the line width of the fluorescence signal of $^{172}Yb^+$ with $\Delta\nu_D$.

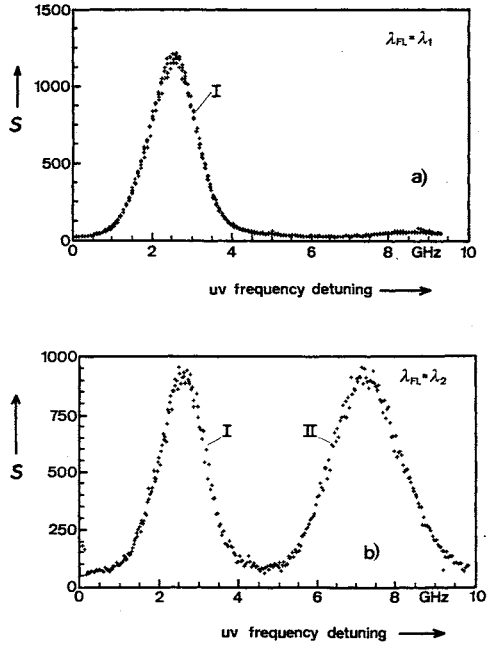


Fig.4 Uv frequency scan during continuous resonant microwave excitation for two slightly different wavelengths of ir excitation λ_{FL}; a): $\lambda_{FL}=\lambda_1=935{,}1880(30)$ nm, b): $\lambda_{FL}=\lambda_2=935{,}1953(30)$ nm; signal *I*: excitation of the α_1 line of $^{171}Yb^+$, signal *II*: excitation of $^{172}Yb^+$; S: fluorescence count number

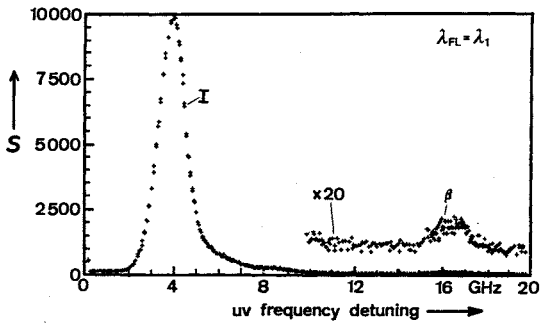


Fig.5 Uv frequency scan during continuous resonant microwave excitation, illustrating the ratio of the double resonance signals arising from excitation of the α and the β lines; S: fluorescence count number

A wider uv frequency scan allows to also detect the β transition of $^{171}Yb^+$ (Fig. 5). Only a few photons per ion are emitted during the excitation of the β transition due to hyperfine pumping.

• Spectroscopy of the microwave transitions

Using the sequential mode, the magnetic field-dependent transitions F=0, m_F=0 ⇒ F=1, m_F = ±1 were recorded with interaction limited line widths of less than 1 Hz /5/. From the frequency separation to the m_F=0 ⇒ m_F=0 clock transition, the magnetic field averaged over the ion cloud could be determined with a precision of 0,01 nT.

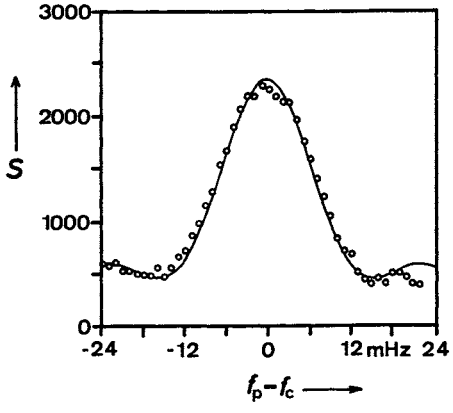


Fig.6 F=0, m_F=0 ⇒ F=1, m_F=0 clock transition signal, recorded with an interaction time of t_w=60 s; solid line: line shape fitted according to eq.(1); the uv laser was tuned to the fluorescence maximum of the α_1 line and the ir laser to $\lambda_{FL} = \lambda_1$; S: fluorescence count number

The clock transition was recorded with a maximum interaction time t_w of 60 s. Together with the data points in Fig.6 a curve following the theoretical line shape

$$P(f_p) = A\cdot(4b^2 / \Omega^2)\cdot\sin^2(\Omega\, t_w /2) + B \qquad (1)$$

with $\Omega = (4b^2 + (2\pi(f_p-f_c))^2)^{1/2}$

is shown where t_w = 59,85 s and $b = \pi/(2t_w)$ are assumed as fixed parameters and where the amplitude A, the background B and the line centre f_c are fitted. The experimental curve is only slightly wider than the theoretical one, but a line Q of $9{,}4\cdot10^{11}$ is obtained.
For routine measurements (e.g. Fig.7) an interaction time of t_w = 9,85 s was chosen. It turned out that with t_w= 9,85 s the line centre can be determined from a single frequency scan lasting 400 s with a precision of 0,8....2,5 mHz, depending on the signal strength.

5. EVALUATION OF THE UNPERTURBED HYPERFINE SPLITTING FREQUENCY

• Relativistic Doppler effect

Given the Doppler width $\Delta\nu_D$, the correction to be applied in account of the relativistic frequency shift of the clock transition is

$$\delta f_D = 5{,}199\ \text{mHz}\cdot(\Delta\nu_D / \text{GHz})^2. \qquad (2)$$

ν_D is determined from the optical line width $\Delta\nu_I$ of the $^{171}Yb^+$ fluorescence peak, as shown in Fig. 4a, using the $\Delta\nu_I(\Delta\nu_D)$ relation shown in Fig.3. Assuming a systematic uncertainty of 0,1 $\Delta\nu_D$, a systematic uncertainty of

$$\delta(\delta f_D) = 1{,}04\ \text{mHz}\cdot(\Delta\nu_D / \text{GHz})^2 \qquad (3)$$

has to to be attributed to the clock transition line centre data given below. It may be noted that even at the highest buffer gas pressures the kinetic energy does not fall below 0,19 eV ($\Delta\nu_D$= 1,7 GHz). This is higher by a factor of 4 compared with findings by Cutler et al. /14/.

• Stark effect

The relative frequency shift due to the quadratic Stark effect is described by

$$\delta f_S/f_0 = -\beta_S < E^2 > \qquad (4)$$

where $< E^2 >$ is the squared electrical field strength averaged over the ions' density distribution. The expected order of magnitude of β_S follows from calculations for Ba^+, β_S= $3{,}41\cdot10^{-17}$ $(cm/V)^2$, and Hg^+, β_S=$1{,}43\cdot10^{-18}$ $(cm/V)^2$ /15/. In the present investigations, the microwave resonance

frequency was measured at fixed helium pressure as a function of the number of ions stored in the trap. The relative ion number N_n was derived from the area of the $^{171}Yb^+$ Doppler peak (Fig.4a) taking into account the influence of optical pumping on the signal strength. In Fig.7 the results of measurements at a buffer gas pressure of $9 \cdot 10^{-4}$ Pa are shown after correction for the second-order Doppler effect. For the case that the shape of the density distribution of the ion cloud does not change with ion number, $\langle E^2 \rangle$ should vary proportional to $N_n^{2/3}$. Extrapolation to $N_n=0$ yields the resonance frequency $f(N_n=0)$ free of the Stark effect shift.

The Stark effect coefficient $\beta_S(Yb^+)$ was coarsely determined in the following way. For one ion cloud, for which a -8 mHz frequency shift was measured and attributed to the Stark effect, the laser focus was moved across the cloud. A spatial separation of the points of half fluorescence maximum of 0,5 mm was observed. This width was interpreted as the spatial extension of a spherical cloud of Gaussian shape. From the calculation of $\langle E^2 \rangle$ for this cloud, one would derive $\beta_S(Yb^+) = 2{,}25 \cdot 10^{-17}$ (cm/V)^2.

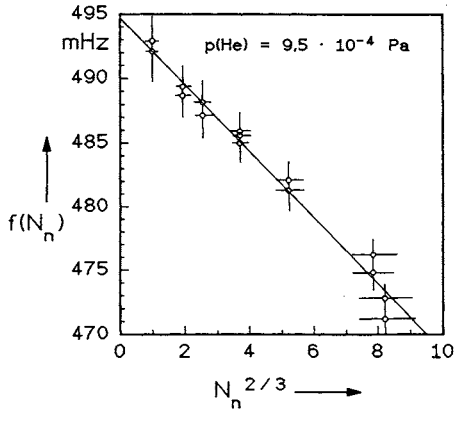


Fig. 7 Clock transition frequency, 12 642 812 118 Hz + $f(N_n)$, corrected for second-order Doppler effect according to (2) as a function of the normalized number of trapped ions, N_n. Each data point was obtained from a fit of a line shape according to (1) to data recorded with t_w = 9,85 s. The error bars indicate 1σ statistical uncertainties.

• Collisions with the buffer gas atoms

Extrapolated values $f(N_n=0)$ were obtained for background pressures in the range between 10^{-4} Pa and $2{,}5 \cdot 10^{-3}$ Pa. The data are collected in Fig. 8. A relative collisional

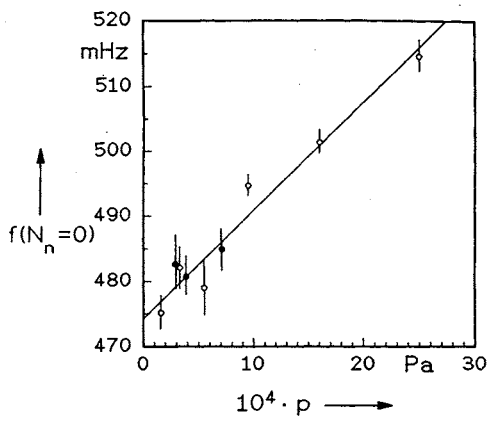


Fig. 8 Clock transition frequency, 12 642 812 118 Hz + $f(N_n=0)$, as a function of the helium buffer gas pressure; the data points "○" were obtained from linear extrapolation to $N_n=0$ of data as shown in Fig.7. The data "•" were obtained from a linear extrapolation to $D_z = 0$ (see text) and were not included in the linear extrapolation to p=0. The error bars indicate 1σ statistical uncertainty.

frequency shift of $(1{,}31 \pm 0{,}18) \cdot 10^{-9}$/Pa is obtained. Here the quoted uncertainty is the 1σ statistical uncertainty of the linear fit. No systematic uncertainty taking into account the absolute pressure calibration and the composition of the background gas can be given up to now. The observed pressure shift coefficient is considerably larger than the values reported for the alkali-like ions $^{137}Ba^+$, $5{,}3 \cdot 10^{-11}$/Pa /16/, and $^{199}Hg^+$, $4{,}3 \cdot 10^{-11}$/Pa /17/, yet comparable with the value reported for neutral cesium, $1{,}09 \cdot 10^{-9}$/Pa /16/.

• Magnetic field

The magnetic flux density in the ion trap was determined daily from the difference between the (F=0)-(F=1,m_F=1) and the (F=0)-(F=1,m_F=0) transition frequencies. The mean value was B=0,237 μT. The correction to be applied is therefore δf_B = -1,75 mHz /5/. The systematic uncertainty, $\delta(\delta f_B)$, due to possible temporal variations of the field does not exceed 0,1 mHz.

• Frequency reference

The average maser output frequency was lower by $1{,}1 \cdot 10^{-13}$ compared with the primary clock CS2. The correction to be applied is thus δf_R = -1,36 mHz. The uncertainty of the absolute frequency calibration of the maser, $\delta(\delta f_R)$, is less than 0,3 mHz.

• Final result

The final value of the unperturbed hyperfine splitting frequency f_0 is obtained from $f(N_n=0)$, extrapolated to p = 0 (cf. Fig. 8), adding the two corrections δf_R and δf_B. The result is

$$f_0 = 12\ 642\ 812\ 118{,}471\ \text{Hz}.$$

The sum of the contributions to the systematic uncertainty is estimated to be $\delta(\delta f_D)+\delta(\delta f_R)+\delta(\delta f_B)$ = 6 mHz. It is dominated by the contribution $\delta(\delta f_D)$ of 5,7 mHz (3), assuming $\Delta\nu_D$ = 2,35 GHz. The statistical uncertainty of the extrapolation to p=0 in Fig. 8 is 3 mHz. Here equal statistical weight has been given to all data in the linear fit. Our final result for f_0 is lower by 1,1 Hz and 5,7 Hz compared with the results reported in /3/ and /2/, respectively.

• Alternative procedure

In the pseudopotential model ion storage is described by a harmonic potential well which in general has different depths in the axial and radial dimensions of the trap /13/. The storage parameters given in Tab.1 correspond to an axial well depth of D_z = 17,9 eV. The clock transition frequency was also determined as a function of D_z at fixed values of a_z and q_z, using three values of helium buffer gas pressure. The relativistic Doppler effect was corrected according to (2) and resonance frequency values extrapolated to D_z = 0 were determined. The three data are included in Fig. 8, showing a reasonable consistency of both extrapolation procedures in the evaluation of f_0.

6. CONCLUSION

The hyperfine splitting frequency of the $^{171}Yb^+$ ion was measured with an uncertainty which is considerably smaller than that of previous investigations. The experimental system allowed absolute frequency measurements with minimum random uncertainty, based on a highly stable reference frequency source. The electrical and mechanical design of the trap permitted adequately stable magnetic field conditions. Further improvements of the present set-up should include absolute pressure calibration and active pressure stabilization. Further investigation should show whether reduced anharmonicities of the trap potential lead to smaller kinetic energies of the ion cloud. In future experiments, also an imaging system will be employed to determine the actual density distribution of the trapped ion cloud.

7. REFERENCES

/1/ R. Blatt et al., in: "Frequency Standards and Metrology", ed. A.de Marchi (Springer, Berlin, 1989), p. 306.

/2/ R. Blatt, H. Schnatz, G. Werth, Z.Phys. **A 312** , 143 (1983).

/3/ R. Casdorf et al., Annalen der Physik **48**, 41 (1991).

/4/ D. Schnier, PhD-thesis, PTB-Bericht **PTB-Opt-37** (1992).

/5/ D. Schnier et al., Proc. 6th EFTF, 415 (1992).

/6/ M.N. Garboriaud, M. Desaintfuscien, F.G. Major, Int. J. Mass Spectr. Ion Phys. **41**, 109 (1981).

/7/ K. Sugiyama, J. Yoda, Appl. Phys. **B 51**, 146 (1990).

/8/ N.A. Demidov et al., Proc. 6th EFTF, 409 (1992).

/9/ A. Bauch et al., IEEE Trans **IM 40**, 153 (1991).

/10/ Chr. Tamm et al., Phys. Rev. **A44**, R20 (1991).

/11/ Chr. Tamm, D. Schnier, Opt. Comm. **87**, 240 (1992).

/12/ A. Bauch, D. Schnier, Chr. Tamm, J. Mod. Opt. **39**, 389 (1992).

/13/ M. Schubert, I. Siemers, R. Blatt, Phys. Rev. **A 39**, 5098 (1989).

/14/ L.S. Cutler et al., Proc. 41st Ann. Symp. Freq. Control, IEEE Cat. No. 87CH2427-3, 12 (1987).

/15/ W.M. Itano, L.L. Lewis, D.J. Wineland, Phys. Rev. **A 25**, 1233 (1982).

/16/ J. Vetter, M. Stuke, E.W. Weber, Z. Phys. **A 273**, 129 (1975).

/17/ L.S. Cutler, R.P. Giffard, M.D. McGuire, Appl. Phys. **B 36**, 137 (1985).

REPEATABILITY OF THE Mg FREQUENCY STANDARD

A. Godone, C. Novero and P. Tavella[1]
K. Rahimullah[2*]

[1] Istituto Elettrotecnico Nazionale G.Ferraris, Torino, Italy
[2] Aligarh Muslim University, Aligarh, India

* At the time of this work at Istituto Elettrotecnico Nazionale G. Ferraris supported by a grant from the "ICTP Programme for Training and Research in Italian Laboratories, Trieste, ITALY"

ABSTRACT

We report in this paper the data corresponding to the Mg frequency measurements with respect to UTC (IEN) performed during one year, showing a repeatability of $\pm 3 \times 10^{-13}$, mainly limited by the short term frequency instability of the standard itself.
The possible implications of our measurements on the temporal stability of the fundamental constants which are involved in the comparison between a standard based on a fine-structure transition (Mg) and a reference based on a hyperfine one (Cs) are discussed as well.

Keywords: frequency standards

1 - INTRODUCTION

A detailed description of the submillimetric frequency standard based on the $^3P_1 - {}^3P_0$ $\Delta m_j = 0$ transition in the metastable triplet of ^{24}Mg is reported in [1] where the stability and accuracy levels achieved are reported and discussed also from the theoretical point of view.
The short-term frequency instability observed in the experiments, evaluated through the Allan standard deviation, turns out $\sigma_y(\tau) = 1 \times 10^{-11} \tau^{-1/2}$ in the white frequency noise region $1\ s \leq \tau \leq 1000\ s$ and reaches the value of 2×10^{-13} for counting times of 3000 s (Fig.1, curve a)).
An improvement of an order of magnitude is possible as it has been demonstrated in [2] through optical pumping. A better control of the environmental conditions is necessary to reduce the frequency drift, observed for $\tau > 3000$ s , below the flicker floor not yet reached.
The value of the clock transition frequency, measured with respect to UTC (IEN), is: ν_{10} = 601277157869.1 ± 0.6 Hz (relative uncertainty $\pm 1 \times 10^{-12}$). The corresponding accuracy budget is reported in Tab.I and turns out limited, at the moment, mainly by the residual 1st order Doppler shift; the accuracy level of 1×10^{-13} appears attainable as widely discussed in [1].

	BIAS (relative)	RELATIVE UNCERTAINTY x_i
2nd order ZEEMAN	+3.18 Hz (+5.29 10^{-12})	3 10^{-13}
2nd order DOPPLER	-2.32 Hz (-3.86 10^{-12})	3 10^{-13}
CAVITY PHASE SHIFT	--------	1 10^{-13}
1st order DOPPLER	--------	7 10^{-13}
FREQUENCY INSTABILITY	--------	3 10^{-13}
ELECTRONICS	--------	5 10^{-13}
OTHERS	--------	1 10^{-13}
		$(\Sigma x_i^2)^{1/2} = 10 \times 10^{-13}$

Tab. I Present accuracy budget of the magnesium frequency standard. [1]

The above reported results have been made possible also by an all-solid-state frequency multiplication chain from 5 MHz up to 600 GHz with spectral purity and output power level suitable to

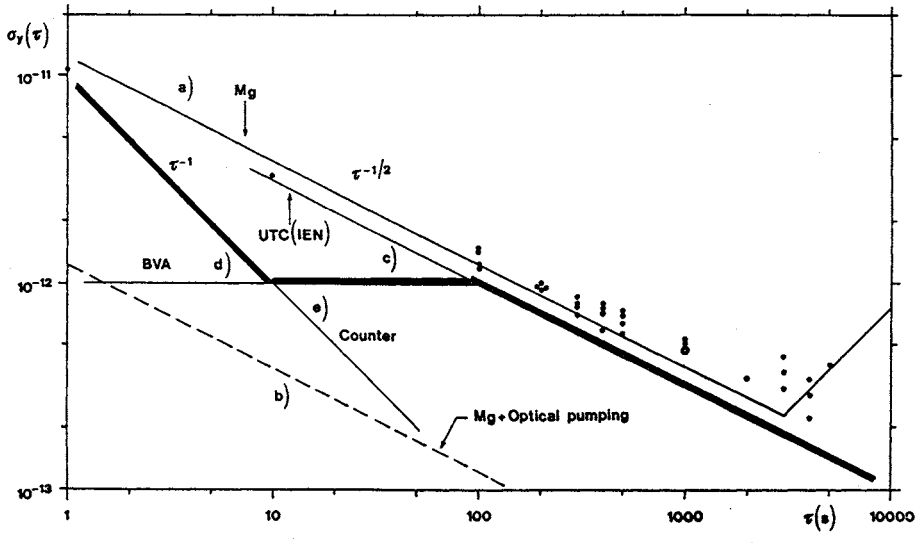


Fig. 1 Allan standard deviation of the **a)** Mg standard without optical pumping; **b)** Mg standard with optical pumping (estimated); **c)** UTC (IEN); **d)** BVA quartz; **e)** frequency counter with the difference-frequency multiplier. The last three curves define the sensitivity of the set-up used.

excite the Mg fine-structure transition [3].
We report in this paper the data corresponding to the Mg frequency measurements with respect to UTC (IEN) performed during one year, showing a repeatability of $\pm 3 \times 10^{-13}$. Moreover we also discuss the possible implications of our measurements on the time stability of the fundamental constants involved in the comparison between a fine-structure transition (Mg) and a hyperfine-structure one (Cs).

2 - EXPERIMENTAL SET-UP AND FREQUENCY MEASUREMENTS

The set-up used to measure the Mg frequency with respect to UTC (IEN) is shown in Fig. 2.

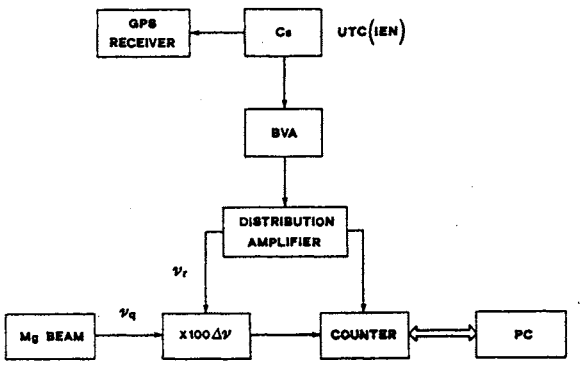


Fig. 2 Set-up for the Mg-Cs comparison.

The measurement system limit is reported in Fig. 1 where the contributions of the counter with the difference-frequency multiplier, of the BVA oscillator and of UTC (IEN) are displayed. The BVA oscillator, locked with a 100 s time constant to the Cs signal, has been used to improve the short term stability of the measurement system. The main parameters characterizing the Mg atomic beam and the operating conditions are summarized in Tab. II.

Discharge current	I=300 mA
Servo-loop time constant	T≤ 100 ms
Modulation wave-form	square-wave
" frequency	f_M=128 Hz
" depth	$\Delta\nu_{pp}$=1.4 kHz
Ramsey linewidth	$\Delta\nu$=1.2 kHz
Ramsey-cavity finesse	F=140
Ramsey length	L=30 cm
RF power	P_{RF}~160 μW
C-field	B=91 μT
Atomic beam divergence	2θ=10 mrad
Oven temperature	T_{oven}~530 °C

Tab. II Main parameters and operating conditions of the Mg beam.

We have performed the measurements of the Mg frequency during a period of one year from 48596 MJD to 48923 MJD, approximately every month, checking every time the parameters which can affect the output frequency of the standard through the main physical effects considered in the accuracy budget.
The frequency of the Mg clock transition is given by [1]:
(1)

$$\nu_{Mg} = N\nu_r + \delta_m$$

where N is the effective multiplication order of the frequency synthesizer;

$\nu_r = \nu_{Cs}$ is the reference Cs frequency (5 MHz);

$\delta_m = N(\nu_q - \nu_r)$ is the frequency difference between the quartz locked on the Mg transition and the reference used, at the submillimetric output frequency.
In Fig. 3a we report the observed δ_m values versus the date MJD of the measurement; every value is the mean value of the frequency measures performed in the day.

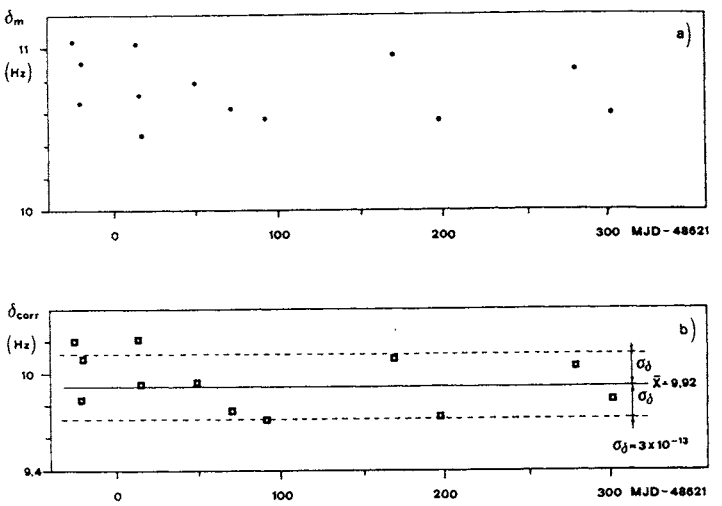


Fig. 3 Mg offset frequencies with respect to UTC (IEN) versus the MJD: **a)** measured values; **b)** corrected values.

The data reported in Fig. 3b are obtained from Fig. 3a introducing the corrections corresponding to the 2nd order Zeeman and Doppler effects and taking into account the effective operating values of the standard, such as RF power level, oven temperature, C field, etc..
The reported uncertainty bar of 3×10^{-13} corresponds to the standard deviation of the mean as deduced from the Allan standard deviation of the measures; we have used counting times of τ = 100 s or τ=1000 s, in order to operate in the white frequency noise region where the Allan variance coincides with the classical one (see Fig.1).
The reference frequency is in all these cases UTC(IEN), which is obtained by a commercial Cs clock continuously corrected by a phase-microstepper to mantain agreement with UTC. Finally in Fig. 4 we refer our measurements to UTC (PTB). The comparison is effectuated by the GPS common view method. The uncertainty on each common view between IEN and PTB is estimated to be 10 ns (1 σ) and the daily frequency comparison uncertainty lies between 1×10^{-13} and 10×10^{-13} (1 σ) according to the number of available common views. This variable uncertainty added to the measurement uncertainty of 3×10^{-13} is reported on each point of Fig. 4 representing a daily measurement.

3 - DISCUSSION OF THE RESULTS

The standard deviation σ_δ of all the measurements reported in Fig. 3 is $\sigma_\delta = 3 \times 10^{-13}$ and we can assume this value as the repeatability of the Mg standard; a comparison with the curve shown in Fig. 1 reveals that the estimated repeatability is probably limited by the frequency instability of the standard itself.
For the data referred to UTC (PTB) (Fig. 4) we have also determined the linear best fit with a weighted least square

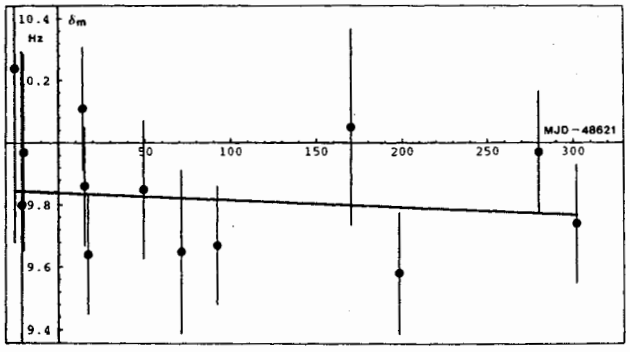


Fig. 4 Mg offset frequencies with respect to UTC (PTB) versus the MJD.

procedure, obtaining for their slope the following value(normalized to ν_{Mg}):
(2)

$$\frac{1}{\nu_{Mg}}\frac{d}{dt}\delta_m = \frac{\delta_m'}{\nu_{Mg}} = -1.44 \pm 3.4 \times 10^{-13}/y$$

This result allows us to put a limit to the possible time variation of some fundamental constants, in fact the measurement of the Mg frequency with respect to the Cs reference standard implies a comparison between a fine-structure transition frequency and a hypefine-structure one.
It can be shown [4] that the transition frequency of the ^{24}Mg $^3P_1$ - $^3P_0$ fine-structure transition is given by the following relation:
(3)

$$\nu_{Mg} \approx \frac{cR_\infty \alpha^2 Z_s^4}{6n^3}$$

where :

c is the speed of light in vacuum;

R_∞ is the Rydberg constant for infinite nuclear mass;

α is the fine-structure constant;

Z_s is the effective nuclear charge ($Z_s \approx 4.867$ in our case);

n is the principal quantum number ($n = 3$ in our case).

Equation (3) does not account for the Landè anomaly, finite nuclear mass, relativistic effects and quantum electrodynamic corrections, whose relative total contribution has been estimated to be of the order of 10^{-2}.
The hyperfine-structure transition in the ground state $6s\ ^2S_{1/2}$ of Cs may be expressed, at the same order of approximation considered above through the Fermi-Segre relation, and turns out:
(4)

$$\nu_{Cs} \approx \frac{32}{3}\frac{cR_\infty \alpha^2 Z_s^3}{n^3} g_I \frac{m_e}{m_p}$$

where:

g_I is the nuclear gyromagnetic ratio of Cs;

m_e/m_p is the ratio of the electron-to-proton mass;

$n = 6$ and $Z_s \approx 13.75$ (in this case).

Being interested in the time stability of ν_{Cs}/ν_{Mg} , numerical constants and quantum numbers can be disregarded and from (3) and (4) we have:
(5)

$$\frac{d}{dt}\ln\frac{\nu_{Cs}}{\nu_{Mg}} = \frac{d}{dt}\ln\left[g_I\frac{m_e}{m_p}\right]$$

If we consider this time derivative and take into account (1) and (2), we finally obtain:
(6)

$$\frac{d}{dt}\ln\left[g_I\frac{m_e}{m_p}\right] = -\frac{\delta_m'}{\nu_{Mg}} = +1.44 \pm 3.4 \times 10^{-13}/y$$

which represents a limit to the possible time variation of $g_I m_e/m_p$.
Moreover, astrophysical observations of a Mg II electronic line and of a hydrogen hyperfine line emitted from the radio source (AO 0235+164) [5] have led to the following limit:
(7)

$$\left|\frac{d}{dt}\ln\left(\alpha^2 g_p \frac{m_e}{m_p}\right)\right| \le 2 \times 10^{-14}/y$$

where g_p is the proton gyromagnetic ratio.
Recently one year of frequency comparisons between H masers and Cs standards have been reported [6]. Since both standards are based on a hyperfine transition, from this comparison we can deduce the following limit:
(8)

$$\left|\frac{d}{dt}\ln\left(\frac{g_p}{g_I}\right)\right| \le 5.5 \times 10^{-14}/y$$

Since:
(9)

$$\left|\frac{d}{dt}\ln\alpha^2\right| \le \left|\frac{d}{dt}\ln\left(\alpha^2 g_p\frac{m_e}{m_p}\right)\right| + \left|\frac{d}{dt}\ln\frac{g_p}{g_I}\right| + \left|\frac{d}{dt}\ln g_I\frac{m_e}{m_p}\right|$$

introducing in (9) the reported limits and their uncertainties we obtain finally for the fine-structure constant:
(10)

$$\left|\frac{d\ln\alpha}{dt}\right| \le 2.7 \times 10^{-13}/y$$

The limit deduced in (10) improves previous estimations reported in [7] and [8] by approximately one order of magnitude.

4 - CONCLUSIONS

In this work we have reported the frequency comparison between the Mg standard and UTC (IEN), performed during one year, which shows a repeatability of $\pm 3 \times 10^{-13}$ for the Mg prototype put in operation at IEN; this value is mainly limited by the present frequency instability of the standard itself. Moreover, from the comparison against UTC (PTB) we have also deduced a limit for the time stability of the combination of fundamental constants involved in the comparison, that is: $|d/dt \ln[g_I m_e/m_p]| \le 4.8 \times 10^{-13}/y$, which may imply for the fine-structure constant the following limit: $|d/dt[\ln\alpha]| \le 2.7 \times 10^{-13}/y$ (central value + 1 σ).

REFERENCES

[1] A. Godone and C. Novero: "The Magnesium frequency standard" (Submitted to *Metrologia*)
[2] A. Godone, C. Novero and E. Bava: "Optical pumping of the metastable states of Mg" *IEEE Trans. Instr. Meas.* **IM 40**, 149 (1991)
[3] A. Godone and C. Novero: "Solid state frequency synthesis up to 600 GHz: an application to the Mg frequency standard" *Proc. 6th Europ. Freq. Time Forum*, Noordwijk, NL, March (1992) pp. 471-476
[4] K. Rahimullah, P. Tavella, C. Novero and A. Godone: "A semi-Theoretical calculation of 3s3p fine-structure splittings in Mg I using asymptotic value of quantum defect for 3snp orbits" (Submitted to *Il Nuovo Cimento D*)
[5] A. M. Wolfe, R. L. Brown and M. S. Roberts: "Limits on the variation of the fundamental atomic quantities over cosmic time scales" *Phys. Rev. Lett.* **37**, 179 (1976)
[6] N. A. Demidov, E. M. Ezhov, B. A. Sakharov, B. A. Uljanov, A. Bauch and B. Fischer: "Investigations of the frequency instability of CH1-75 hydrogen masers" *Proc. 6th Europ. Freq. Time Forum*, Noordwijk, NL, March (1992) pp. 409-414
[7] N. Bahcall and M. Schmidt: "Does the fine-structure constant vary with cosmic time?" *Phys. Rev. Lett.* **19**, 1294 (1967)
[8] J. P. Turneaure and S. R. Stein: "An experimental limit on the time variation of the fine-structure constant", *in "Atomic Masses and Fundamental Constants-V"*, ed. J. H. Saunders and A. H. Wapstra (Plenum Press, London , 1976) pp. 636-642

DIODE-LASER PUMPED, RUBIDIUM CELL FREQUENCY STANDARDS

C. Szekely, R. E. Drullinger, F. L. Walls, J. P. Lowe, and A. Novick
Time and Frequency Division
National Institute of Standards and Technology
Boulder Colorado, USA

Abstract

We report on our project to experimentally investigate the limits to the short-term stability achievable in optically pumped, rubidium cell frequency standards. Theory indicates that the atomic limited stability is several orders of magnitude better than what has actually been achieved. The difference is related to the complex spectral output of the lamps used in conventional standards as well as phase noise in the interrogating microwave radiation. Using a novel filter technique, we have built a microwave synthesis chain with phase noise capable of supporting a standard with $\sigma_y(\tau) \leq 4 \times 10^{-14} \tau^{-1/2}$. With extended-cavity, diode lasers we have demonstrated measured line Q and signal-to-noise ratios commensurate with $\sigma_y(\tau) = 4 \times 10^{-13} \tau^{-1/2}$ in a commercial, buffer gas type standard and $\sigma_y(\tau) = 2 \times 10^{-13} \tau^{-1/2}$ in an evacuated, wall-coated cell. Both of these numbers can be improved significantly in a fully engineered and optimized standard.

Key words: rubidium; optically pumped; diode laser

1. INTRODUCTION

Optically pumped, rubidium-cell frequency standards have theoretical, atomic shot-noise-limited, short-term stability that is *several orders of magnitude* better than what is actually achieved. Qualitatively the potential performance can be understood by considering *only* the number of atoms in a cell and how often each of them can contribute to the signal. Take, for example, a 1 cm^3 cell at 50°C. At this temperature the vapor density of rubidium is about 10^{11} cm^{-3} and the spin exchange rate is about 100/s [1]. Thus, we might expect something of the order of 10^{13} "clock" transitions per second and a linewidth of order 100 Hz. This corresponds roughly to a stability of $\sigma_y(\tau) \approx (Q \times S/N)^{-1} \approx 5 \times 10^{-15} \tau^{-1/2}$. Detailed calculations [2] predict similar limits.

The reason for the tremendous difference between potential and achieved performance is, in part, related to the complex spectral output of the lamp used for the optical pumping in all commercial standards. The non-monochromatic nature of the lamp output has two deleterious effects on the short-term stability of the standard. The presence of light that does not directly contribute to the optical pumping adds a background signal and, hence, noise to the detected signal. Additionally, nearly resonant light actually leads to some active depumping, thus limiting the use of the available atoms.

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

Our preliminary experiments with diode lasers in place of lamps support this argument. The amount of light reaching the detector is typically a factor of 30-100 less than that in the lamp-pumped standards, while at the same time, the "clock" signal in terms of absolute amount of light change at the detector is larger, in some cases by a factor of 10.

In addition to changes in the optical pumping light source, reaching the high stability available in rubidium cell standards will require exceptionally low phase noise in the interrogating microwave radiation [3]. In fact, the requirements are so severe that they cannot be met with state-of-the-art crystal oscillators and noiseless multiplication. Fortunately, however, the requirement is not on broadband noise. The system is most sensitive to noise centered at a frequency twice the modulation frequency either side of the carrier. We have taken advantage of this fact to suggest a special filter with a notch at both the upper and lower $2\omega_m$ sidebands [4]. This will substantially reduce the noise where it counts and allow improved short-term frequency stability.

We report here on our project to investigate the limits to the short-term stability attainable in a rubidium cell type standard when pumped with a diode laser. The use of diode lasers to pump rubidium standards is not a new idea. There have been several previous investigations of short-term stability [5-10]. All have used commercial rubidium standards and solitary diode lasers without explicit AM or FM noise control. None has reported measured stability better than the original lamp-pumped standard. In the work reported here, we have used extended-cavity diode lasers with their inherently lower FM noise. With these lasers, we have been able to show that at least a large part of the theoretical, short-term stability of rubidium cell frequency standards is experimentally achievable in both buffer-gas cells and evacuated, wall-coated cells.

2. EXPERIMENTAL WORK

We have used an extended-cavity, grating feedback laser as the optical pumping source for two different physics packages. The laser operates on the D_1 line (794.9 nm). The cavity length is servo-controlled to lock the laser frequency to a saturated-absorption feature in an external, evacuated ^{87}Rb cell.

The first experiments were done with the same commercial standard used in Ref. 5. The standard uses a TE_{111} cavity, and the cell contains a natural mixture of ^{87}Rb and ^{85}Rb. Because of the presence of ^{85}Rb, we obtained the best optical pumping on the F = 1 to F' = 2 transition. At the normal operating temperature of the standard($\approx$70°C), the relative absorption dip ($\Delta I/I$) was 10% whereas it was

0.04% when pumped with its lamp. The measured short-term stability of the standard, without AM control on the laser, was 4 x 10^{-12} $\tau^{-1/2}$; see Fig. 1. This is a factor of 3 improvement over the performance of the same standard when lamp-pumped. Limitations in the electronics package and restrictions in the physics package prevented further improvements in short-term stability while lack of AM control on the laser resulted in the relatively poor performance beyond 30 seconds. In changing from the lamp to the laser, the light falling on the detector decreased by a factor of 80 while the absolute magnitude of the microwave-induced change in light signal increased by a factor of 3. The microwave resonance linewidth was essentially unchanged.

The second physics package consisted of a TE_{011} cavity and an evacuated, wall-coated cell with ^{87}Rb [11]. The microwave cavity was surrounded by two magnetic shields and the temperature of the cell and cavity could be controlled separately from the rubidium reservoir. The diameter of the spherical cell was about 3.5 cm. The laser beam had an elliptical cross section of about 0.25 cm^2. This device was never operated as a frequency standard because a microwave electronics package with adequate phase noise was not yet available; see discussion section. Rather, the parametric dependence of line Q and signal-to-noise ratio (S/N) with optical power, microwave power, and spin density was investigated. At room temperature and low microwave power, the resonance is a pure Dicke-narrowed, symmetric, Lorentzian line with a FWHM of 87 Hz. However, on heating of the cell and cavity from room temperature to 340 K, the line Q decreased by a factor of about 2. This was attributed to magnetic effects of the heater but the issue is not resolved at this time. The results quoted are for the actually measured linewidth. At a cell wall temperature of 340 K and for rubidium densities ranging from 1.9 x 10^{10} to 2.6 x 10^{11} atoms/cm^3 we evaluated the short-term stability dependence on laser at near optimum microwave power; see Fig. 2.

In addition to the work with laser optical pumping, we have obtained two crystal filters with 1 dB on insertion loss at 10 MHz and more than 25 dB attenuation at 10 MHz $\pm$ 75 Hz as shown in Fig. 3. The residual phase noise added by these filters 10 Hz from the carrier is less than -152 dBc/Hz and less than -162 dBc/Hz at the notch frequency. These limits are probably due to the measurement system and not the resonators. Further measurements are planned. We have also fabricated multiplier chains which have residual noise referenced to 10 MHz of -165 dBc/Hz. Together, these components should allow our new microwave synthesizer to support operation of a standard with $\sigma_y(\tau) \leq 3 \times 10^{-14}\ \tau^{-1/2}$.

3. DISCUSSION

It now seems clear that at least a large part of the stability potential suggested by theory is actually available in a properly engineered, laser-pumped rubidium standard. However, AM and FM noise in the laser as well as phase noise in the microwave radiation must be addressed.

Extended-cavity diode lasers allow solutions to the necessary AM and FM control. Simple, solitary laser diodes have linewidths of the order of 50 MHz, and the output wavelength is an extremely sensitive function of temperature and current: $d\nu/dT \approx 30$ GHz/K and $d\nu/di \approx 3$ GHz/mA. For grating-feedback lasers operating in the high feedback limit, the linewidth is reduced 2-3 orders of magnitude. The tuning with current and temperature is reduced by a similar amount [12]. With the wavelength set by the grating angle and cavity length (a piezo-mounted mirror), the current and temperature become almost free parameters. In our laser, comparatively crude temperature control is used to adjust the band gap and help with mode selection while the laser injection current is used to set and stabilize the laser output power.

AM noise can be easily controlled to the shot-noise limit in these extended-cavity lasers. Only a very small fraction of the laser output power is needed for the microwave-optical double resonance. Half of the laser power can be used as a monitor in a power-stabilization and noise-controlling servo. This allows the laser output to be AM-controlled within 3 dB of the shot-noise limit [13]. If, for example, 1% of the laser power is used to probe the rubidium cell, the shot noise-limited S/N ratio on this beam will be 10 dB less than that for the full output of the laser. Hence, controlling the full laser power within 3 dB of its shot-noise limit insures that the probe beam used in the clock will be shot noise limited.

Microwave radiation with phase noise adequate to support a standard operating with outstanding short-term stability can be generated from a conventional crystal, local-oscillator/multiplier chain if precautions are taken to reduce the noise at the second harmonic of the modulation frequency. We have demonstrated that such a result can be produced with crystal notch filters designed to work at $\pm 2\omega_m$ from the carrier.

4. SUMMARY

We have built a microwave synthesis chain that incorporates a crystal notch filter concept and is capable of supporting operation of a standard with $\sigma_y(\tau) \leq 3 \times 10^{-14}\ \tau^{-1/2}$. Our optical pumping results are summarized in Table 1 along with those of other researchers. The measured stability of a commercial standard operated with our line-narrowed laser is a factor of 3 better than with its original lamp and is limited by phase noise in the commercial microwave synthesis package. With an AM loop on the laser and our new microwave synthesizer, our measurements indicate that a short-term stability of $\sigma_y(\tau) \approx 4 \times 10^{-13}\ \tau^{-1/2}$ can be achieved in this standard. Optimizing the physics package for this new system should yield still further improvements. In an evacuated, wall-coated cell in a TE_{011} cavity we have measured line Q and S/N ratios commensurate with a $\sigma_y(\tau) = 2 \times 10^{-13}\ \tau^{-1/2}$. With improved magnetic field control and a refined AM loop on the laser, we think this can be reduced to $\sigma_y(\tau) \leq 5 \times 10^{-14}\ \tau^{-1/2}$.

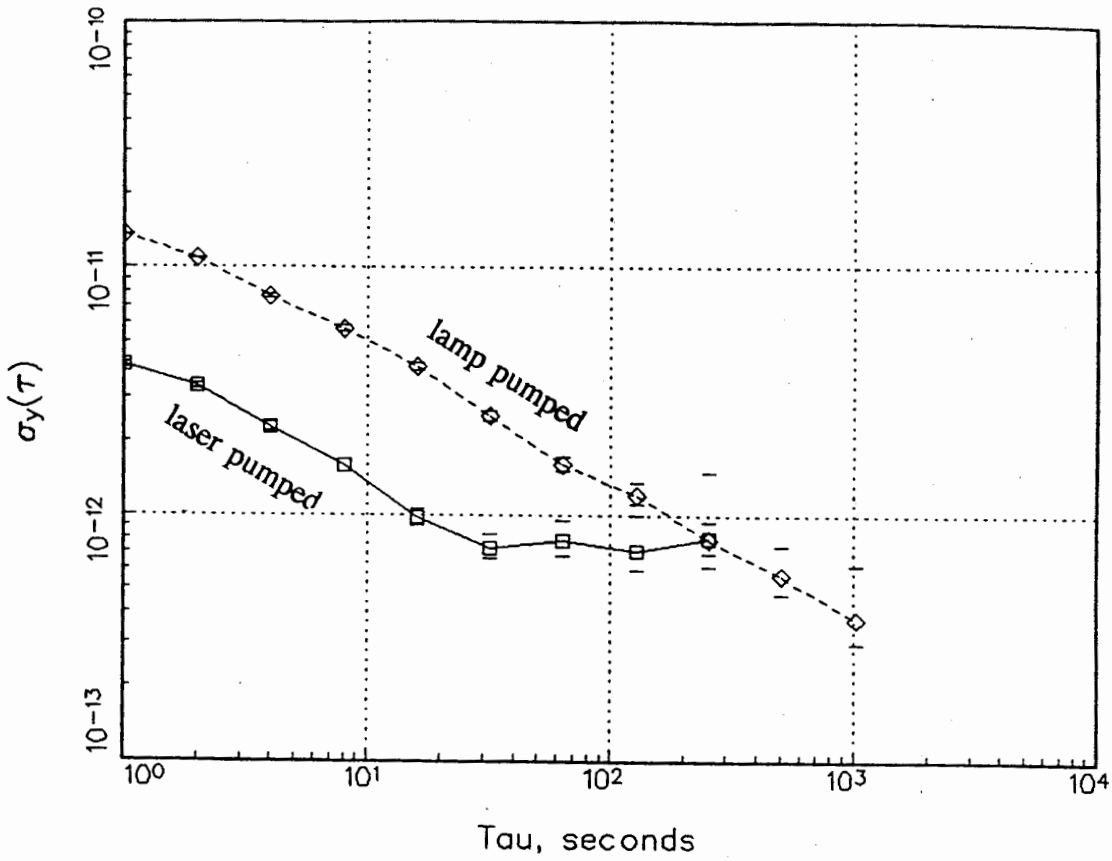


Fig. 1. Frequency stability of commercial standard with lamp pumping and laser pumping. The flicker floor seen in the laser data at times greater than about 20 s is the result of unregulated laser amplitude.

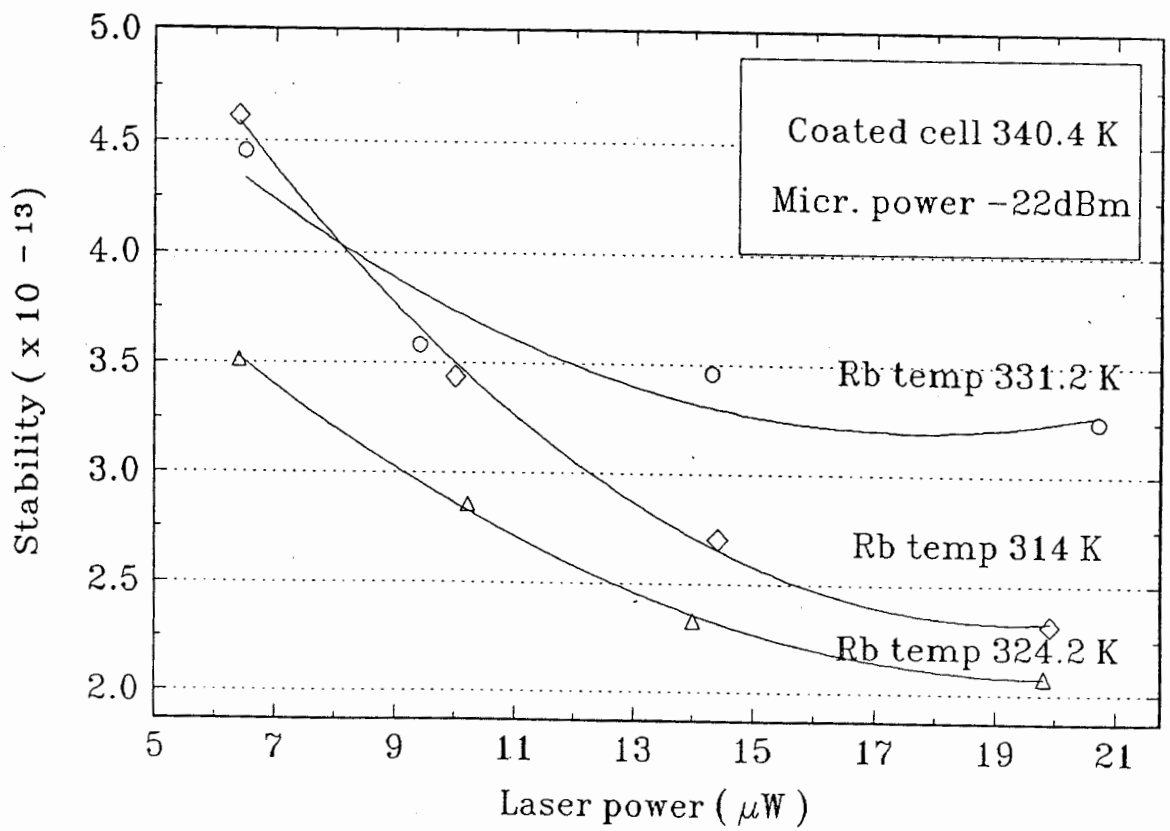


Fig. 2. Evaluated stability for evacuated, wall-coated cell standard. The evaluation is based on experimentally measured parameters and the relationship $\sigma_y(\tau) \equiv (1/\nu_0\sqrt{2})(i_n/D)\tau_{-1/2}$ [14].

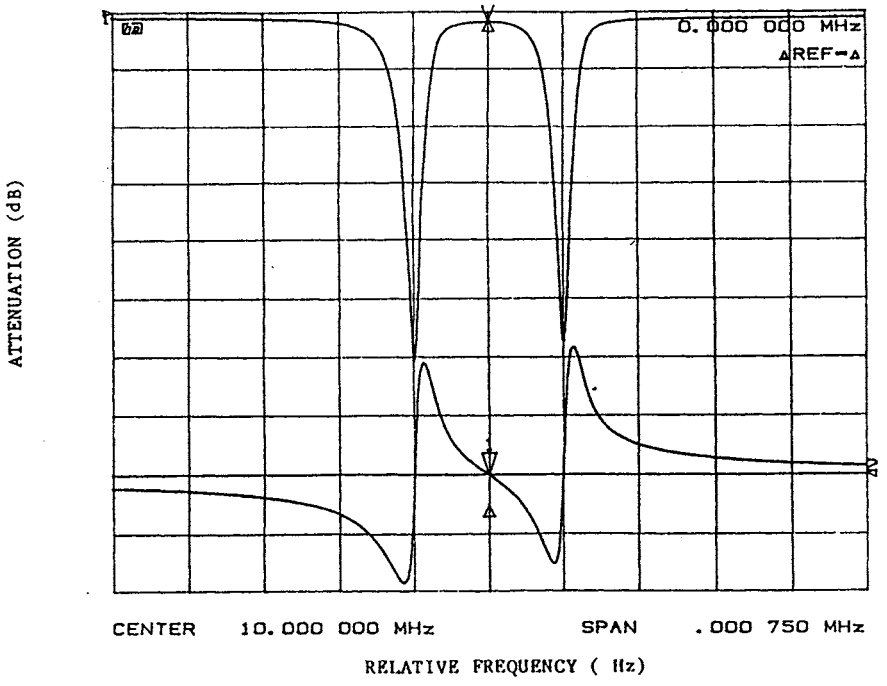


Fig. 3. Transfer function of filter to be used to notch out noise of local oscillator at the $2\omega_m$ sidebands above and below the carrier. The upper trace is the attenuation (5 dB/div) on transmission while the lower curve is the phase shift (50°/div).

Table 1. Summary of experimental results on laser pumped rubidium cell standards.

Source	laser, linewidth, lock technique	cavity and cell type	Measured $\sigma_y(\tau)$	Projected $\sigma_y(\tau)$ based on measured Q and S/N
ref. 5	A	TE_{111}, Buffer gas, Mixed isotopes	$\approx 2 \times 10^{-11}\tau^{-1/2}$ $\tau \leq 100$ s	
ref. 6	B	TE_{111}, Buffer gas, Mixed isotopes	$\approx 10^{-10}\,\tau^{-1/2}$ $\tau \leq 10$ s	
ref. 7	A	Buffer gas, ^{87}Rb		[a] $\approx 2 \times 10^{-11}\tau^{-1/2}$
ref. 8	A	TE_{011}, Buffer gas, mixed & ^{87}Rb		[a] $\approx 3 \times 10^{-12}\tau^{-1/2}$
ref. 10	A	same as ref 8		[b] $\approx 8 \times 10^{-13}\tau^{-1/2}$
this work	C	same as ref 5	$\approx 4 \times 10^{-12}\tau^{-1/2}$ $\tau \leq 30$ s	[c] $\approx 4 \times 10^{-13}\tau^{-1/2}$
this work	C	TE_{011}, Evacuated, wall coated, ^{87}Rb		[d] $\approx 2 \times 10^{-13}\tau^{-1/2}$

A: Solitary laser, D_2 line (780 nm), linear absorption lock, $\Delta\nu_{laser} \approx 40$ MHz
B: Solitary laser, D_2 line, free running frequency, $\Delta\nu_{laser} \approx 400$ MHz.
C: Extended cavity laser, D_1 line (795 nm), saturated absorption lock, $\Delta\nu_{laser} \approx 0.1$ MHz.
[a] Based on measured line Q and S/N combined through the relationship $\sigma_y(\tau) = 0.2/[Q \cdot S/N]$.
[b] Based on measured line Q and S/N combined through the relationship $\sigma_y(\tau) = 0.11/[Q \cdot S/N]$.
[c] Based on measured stability rescaled for available AM noise control in the laser.
[d] Based on measured noise and discriminator slope $[\sigma_y(\tau) \equiv (1/\nu_0\sqrt{2})(i_n/D)\tau^{-1/2}]$ [14].

5. REFERENCES

1. Jacques Vanier and Claude Audoin, The Quantum Physics of Atomic Frequency Standards (Adam Hilger, Bristol, 1989), p. 1277.

2. J. C. Camparo and R. P. Frueholz, "Fundamental stability limits for the diode-laser-pumped rubidium atomic frequency standard," J. Appl. Phys. vol. 59, pp. 3313-3317, 1986.

3. See for example C. Audoin, V. Chandelier and N. Dimarcq, "A Limit to the Frequency Stability of Passive Frequency Standards," IEEE Trans. Instr. Meas., vol 40, pp. 121-125, 1991.

4. J. P. Lowe, F. L. Walls and R. E. Drullinger, "Ultra-High Stability Synthesizer for Diode Laser Pumped Rubidium," in Proc 1992 IEEE Frequency Control Symp., pp. 183-187, Hershey, PA 1992.

5. L. L. Lewis and M. Feldman, "Optical Pumping by Lasers in Atomic Frequency Standards," in Proc. 35th Ann. Freq. Control Symposium, pp. 612-624, Ft. Monmouth, 1981.

6. C. H. Volk and J. C. Camparo, "Laser Diode Pumping in Atomic Clocks," in Aerospace Report No. ATR-82(8498)-1, pp. 55-60, 1982.

7. J. T. Liu, P. Thomann, L. Zhang, and G. Busca, "Studies of a laser diode pumped Rb standard," in Proc. 4th European Frequency and Time Forum, Neuchatel, March 1990.

8. Motoichi Ohtsu, Minoru Hashimoto and Hidetaka Ozawa, "A Highly Stabilized Semiconductor Laser and Its Application to Optically Pumped Rb Atomic Clock," in Proc. 39th Ann. Freq. Control Symposium, pp. 43-54, Philadelphia 1985.

9. Minoru Hashimoto and Motoichi Ohtsu, "Experiments on a Semiconductor Laser Pumped Rubidium atomic Clock." IEEE J. Quant. Elect., vol. QE-23, pp. 446-451, 1987.

10. Minoru Hashimoto and Motoichi Ohtsu, "Modulation Transfer and Optical Stark Effect in a Rubidium Atomic Clock Pumped by a Semiconductor Laser," J. Opt. Soc. Am. B, vol. 6, pp.1777-1789, 1989.

11. This cell and cavity were kindly lent to us by Hugh Robinson of Duke University.

12. R. W. Fox, H. G. Robinson, A. S. Zibrov, N. Mackie, J. Marquardt, J. Magyar and L. W. Hollberg, "High-Sensitivity Spectroscopy with Diode Lasers," in SPIE proceedings 1837: Conference on Frequency Stabilized Lasers, November 16-18, 1992, Boston.

13. L. Hollberg, R. Fox, N. Mackie, A. S. Zibrov, V. L. Velichansky, R. Ellingsen, and H. G. Robinson, "Diode Lasers and Spectroscopic Applications," in Tenth International Conference on Laser Spectroscopy, ed. M. Ducloy, E. Giacobino and G. Camy, pp. 347-352, World Scientific, 1992.

14. William J. Riley, "A Rubidium Clock for GPS," Proceedings of the 13th Precise Time and Time Interval (PTTI) Applications and Planning Meeting, pp. 609-630. 1981.

SENSORS

Chairman: P. Krempl

LAMB WAVES PRESSURE SENSOR

N. Tirole[1], A. Choujaa[1], D. Hauden[1], G. Martin[1], P. Blind[2], M. Froelicher[2], J.C.Pommier[3], A. Cachard[3]

[1] Laboratoire de Physique et Métrologie des Oscillateurs du CNRS
associé à l'Université de Franche-Comté-Besançon,
32 av. de l'Observatoire-25044 Besançon Cedex-France

[2] CETEHOR, 39 av. de l'Observatoire-25000 Besançon-France

[3]Laboratoire de Traitement du Signal et Instrumentation
23 rue du Dr. Michelon-42023 St-Etienne Cedex 2-France

Abstract

Pressure sensors using Lamb waves are very sensitive because they are propagating in a thin membrane and because their lower frequency than Rayleigh waves. Previously such a sensor was made from a silicon membrane and a ZnO piezoelectric layer [1].

We built an anisotropically etched silicon membrane (10 microns) in a <100> wafer with an AℓN thin piezoelectric layer (1 to 2 microns). This device was converted into an oscillator by addition of a loop amplifier locked on the first Lamb symmetrical mode. The thicknesses of the membrane and of the piezoelectric layer were optimised numerically in order to obtain the highest electromechanical coupling factor.

The temperature behavior (between 0 to 80° C) and the pressure sensitivity have also been computed and compared to experimental results.

Keywords.
Lamb waves, pressure sensors, temperature compensation

Introduction

Surface acoustic waves are well known for signal processing applications like oscillators, filters, convolvers. Furthermore, they are also studied for sensors where the frequency variations of the output signal is the sensitive quantity. Quartz crystal or $LiNbO_3$ substrates, where Rayleigh waves are propagating, are the most popular materials for these sensors. In that family, pressure sensors using quartz were previoulsy realized with a good sensitivity [2-3]. Lamb waves propagating in very thin membranes are more sensitive to stresses due to pressure than Rayleigh waves [1]. In this paper, we present a Lamb wave pressure sensor. It consists of a nitrid aluminium piezoelectric layer (AℓN) upon a 10 to 15 microns silicon membrane. We discuss about the electromechanical coupling factor according to the AℓN thickness for the first antisymetrical mode (A0) and the first symmetrical mode (S0). Then, we describe the behavior of a sensor using the S0 mode according to the temperature and the pressure. At last, we discuss about the improvement of the temperature behavior and we present some compensation solutions.

1 - Lamb waves and coupling factor

The structure of the sensor is illustrated on figure 1.

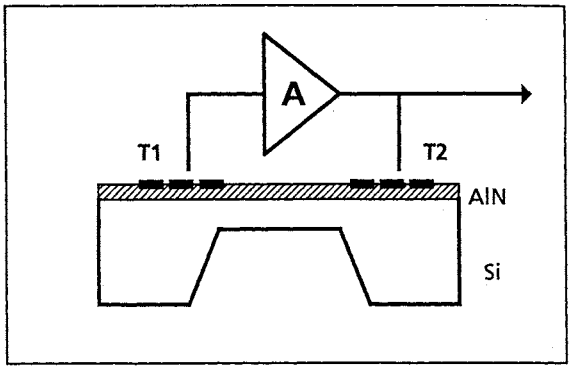


Fig. 1 : *Lamb wave sensor structure*

It consists of a chemically etched silicon membrane covered with an AℓN piezoelectric layer. Two interdigital transducers T_1 and T_2 excite and detect Lamb waves, respectivelly. We obtain an oscillator by addition of a feedback amplifier A. We can write the phase condition in the form :

$$\Phi_{ext} + 2\pi f \frac{L}{v} = n2\pi$$

where :

Φ_{ext} = phase shift caused by the IDTs and the amplifier.
L = centre to centre separation of the IDTs
v = Lamb wave's phase velocity
f = frequency

The phase shift Φ_{ext} is much smaller than the phase shift in the propagation path. Thus, it exists a proportionnal relation between f and v. Due to the small thickness of the membrane, only two modes can propagate : the A0 and the S0 modes. The theoretical dispersion curves of these 2 modes are shown on Fig. 2 [4].

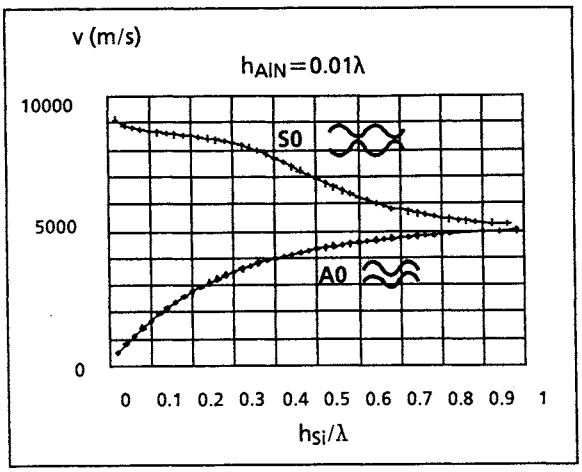


Fig. 2 : *Velocity of the A0and S0 modes versus* h_{Si}/λ

For $h_{Si}/\lambda \ll 1$, the S0 mode is low dispersion and the velocity of the A0 mode is very low that could be interesting for an immersed sensor (chemical or biological applications).

For these 2 modes and for a given membrane thickness (10 % of wavelength), the coupling factor has been computed for different transducers location along the piezoelectric layer thickness. The results are presented on Fig. 3.

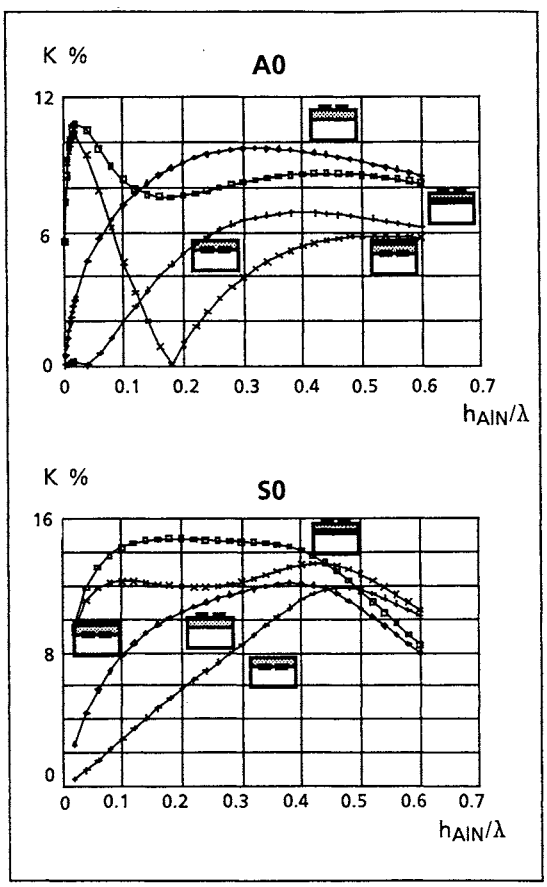


Fig. 3 : *Coupling factors of the A0 and S0 modes* ($h_{Si}=0.1\ \lambda$).

For the 2 modes and for $h_{A\ell N}/\lambda \ll 1$, the best coupling is obtained when the IDTs are on the top surface and the metallization is at the interface. In our application, we choose to use the S0 mode because of its higher coupling and its greater facility to excite in practice than the A0 mode.

2 - Device fabrication and test

Silicon membranes are collectively etched in a <100> 3" wafer, N type, resistivity 0,26-0,52 Ωcm. The different fabrication steps are as follows :

- Silicon oxyde : A 1,2 µm thick silicon dioxyde layer is thermally grown at 1 050° C.
- Membrane fabrication : 4×8 windows are patterned in the oxyde layer by using a BHF/H_20 solution. Then, the silicon is etched with a KOH/H_20 (10 mol/ℓ) solution at regulated temperature. The etching is stopped when the membrane thickness is at the defined value (10 or 15 microns).
- AℓN deposition : AℓN is deposited by planar magnetron sputtering of a high purity aluminium target in a nitrogen D.C. discharge. During the process, the pressure in the chamber is maintained at 10^{-2} mbar and the temperature of silicon substrate is 600°C. The discharge power is adjusted to produce a deposition rate of 45 nm/mn.
  AℓN films have hexagonal 6mm structure. Using the above mentionned deposition parameters, the films growth with a columnar texture, the c-axis being perpendicular to the Si substrate (standard deviation is less than 1°).

After the AℓN deposition, the membrane is made thinner by an isotropic etching. The solution is a chemical mixture of $HF/HNO_3/CH_3COOH$ which etches the silicon without attacking the AℓN. Moreover, the isotropic etching allows to make round the angles of campling corners and reinforce the membrane bending (Fig. 4).

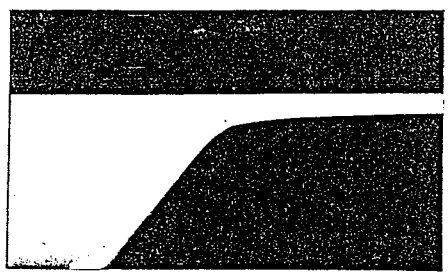


Fig. 4 : *Photo of the membrane clamping corner*

- Transducers fabrication : 1000 Å of aluminium are deposited on the thin layer of AℓN and interdigital transducers are patterned with a lift-off process. The IDTs have a period of 100 µm and 12 finger pairs. The aperture is 3mm and the centre to centre separation of the IDTs is 2,2mm.
- Electronic circuitry : The complete structure of the oscillator is constituted by an oscillating loop and an output amplifier.

Each amplifier has been designed with 2 VHF-MOSFET double gate transistors. A precise control of the feedback gain is possible and the input/output electric separation is in the same order of this one of a cascode amplifier.

3 - Experimental results

The results presented in this part concern a pressure sensor with a 10 microns silicon membrane and a 1 micron AℓN layer. The network analyser response of the S0 mode is presented on Fig. 5.

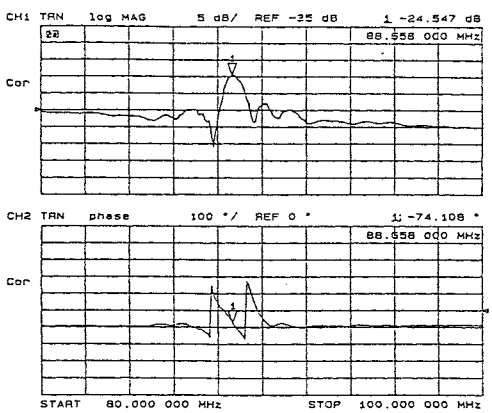


Fig. 5 : *Transfer fonction of the S0 mode*

The velocity of this mode is 8800 ms^{-1} compared to 8685 ms^{-1} predicted by the simulation program. The difference between these 2 values is caused by the residual stresses due to the AℓN layer.

3.1 - Temperature behavior

Temperature is an important parameter for a pressure sensor. We observed the frequency variation of our device with the temperature between 0 and 80° C. The results are given on Fig. 6.

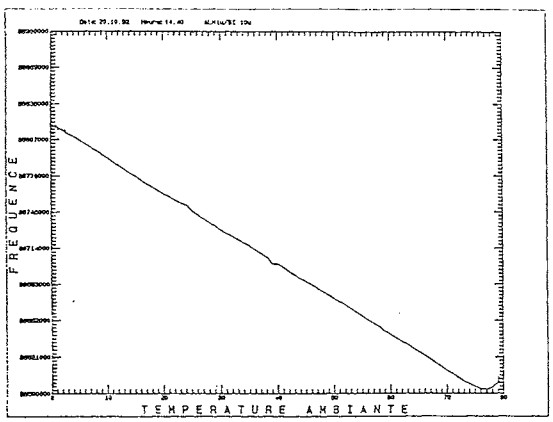


Fig. 6 : *Experimental frequency variation versus the temperature.*

The frequency temperature sensitivity is about -22ppm/° C. For pressure sensor applications, this sensitivity must be avoided.

3. 2 - Pressure sensitivity

The test device is presented on Fig. 7.

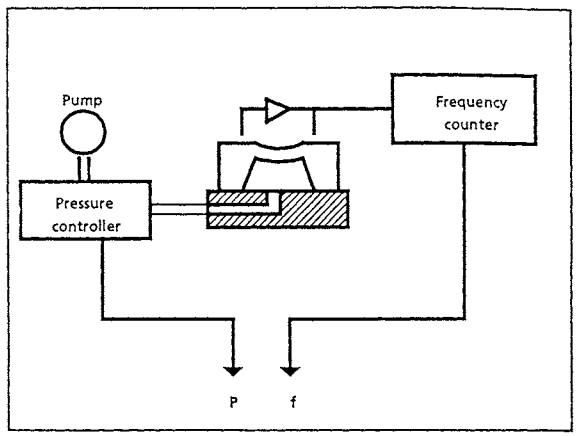


Fig. 7 : *Pressure test set-up*

The pump and the pressure controller fix the pressure and the oscillation frequency is accurated with a frequency counter.

This device allows to draw directly the frequency variation according to the applied pressure as shown on Fig. 8.

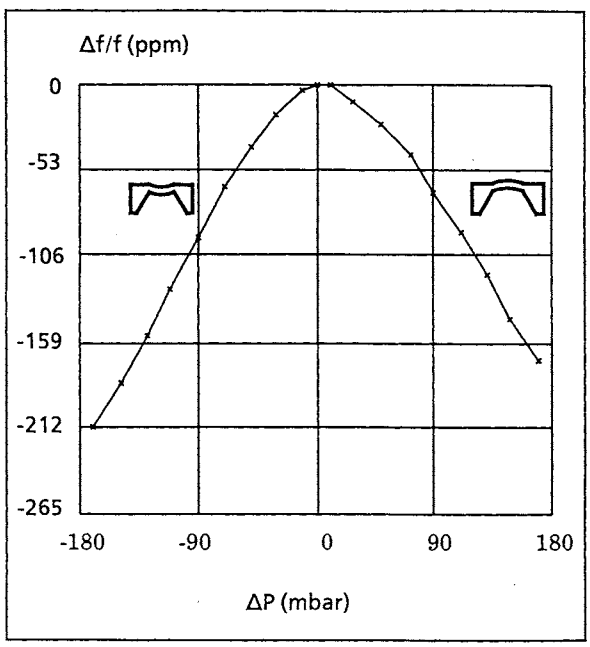


Fig. 8 : *Frequency variation measured versus the pressure applied to the membrane*

The sensitivity for $\Delta P<0$ or $\Delta P>0$, around the atmospheric pressure, is approximatively the same and is 100 Hz/mbar or 1Hz/Pa.

4 - Discussion

The results presented above are particularly interesting for the pressure sensitivity. But the very high temperature sensitivity is a disadvantage that will be resolved by two different ways.

4. 1 - Temperature compensation of the AℓN/Si structure

AℓN and silicon have opposite temperature coefficients as presented on table 1 (for K.Tsubouchi *et al.* [5]).

The principle of this first compensation solution is to compute the relative thicknesses of AℓN and silicon to obtain a compensated structure. For that, we used the Adler *et al* program and we computed the velocity of the S0 mode for an AℓN/Si structure according to the temperature.

Material	AlN	Si
c_{11}	$3,45.10^{11}$	$1,66.10^{11}$
c_{33}	$3,95.10^{11}$	
c_{12} (N/m²)	$1,25.10^{11}$	$0,639.10^{11}$
c_{13}	$1,20.10^{11}$	
c_{44}	$1,18.10^{11}$	$0,796.10^{11}$
α_{11} (l/°C)	$5,27.10^{-6}$	$2,6.10^{-6}$
α_{33}	$4,15.10^{-6}$	
$TC(c_{11})$	$0,8.10^{-4}$	$-0,68.10^{-4}$
$TC(c_{33})$	1.10^{-4}	
$TC(c_{12})$	$1,8.10^{-4}$	$-1,0.10^{-4}$
$TC(c_{13})$	$1,6.10^{-4}$	
$TC(c_{44})$	$0,5.10^{-4}$	$-0,44.10^{-4}$

Table 1 : *Material constants of AℓN and silicon (for K.Tsubouchi et al. [5]). cij: stiffness, aij : thermal expansion coefficient, TC(cij) : temperature coefficient of stiffness.*

This computation was made by changing the C_{ij} values, the density, the thicknesses and the wavelenght for each temperature between 0 and 80° C.

For example, for a membrane of 10 microns of silicon and 1 micron of AℓN, we obtained the curve presented on Fig. 9.

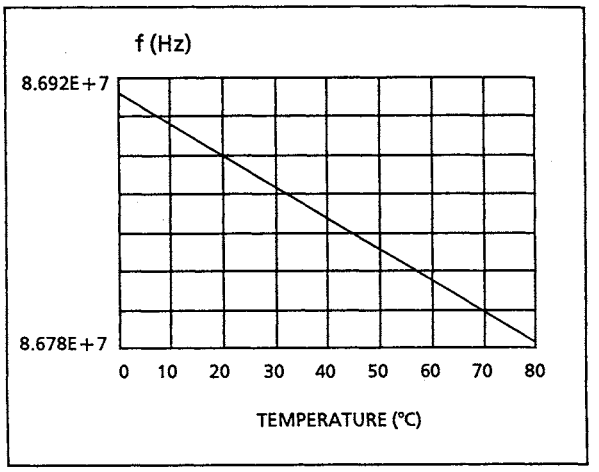


Fig.9 : *Calculated frequency variation versus temperature for a silicon (10µ)/AℓN (1 µm) stucture*

The sensitivity is 18 ppm/° C to compare with the experimental value of 22 ppm/°C. The difference between these 2 values is due to the incertitude of the temperature coefficient values and to the fact that this simulation does not take into account the stresses at the interface caused by the thermal expansion.

Figure 10 shows the calculated variation of the temperature coefficient according to the AℓN thickness.

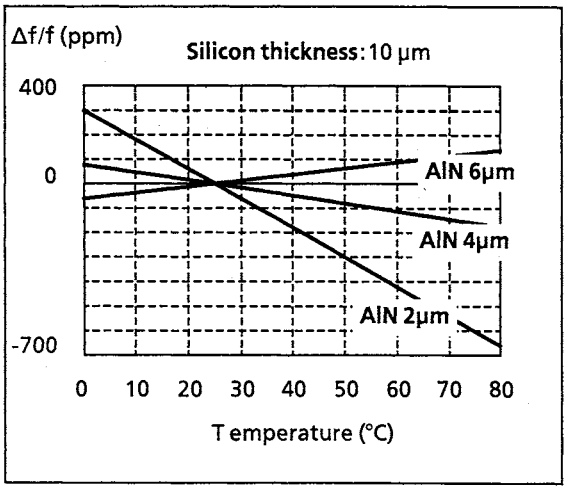


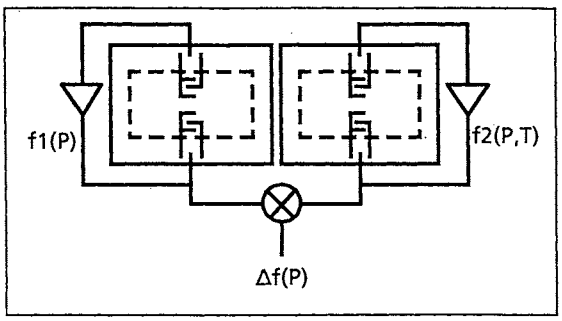
Fig. 10 : *Computed frequency shift between 0 and 80 °C for a AℓN/Si (10µm) structure where AℓN thickness is respectively 2, 4, 6 microns.*

It results from this computation that a 10 microns silicon membrane needs about 5 microns of AℓN to be compensated.

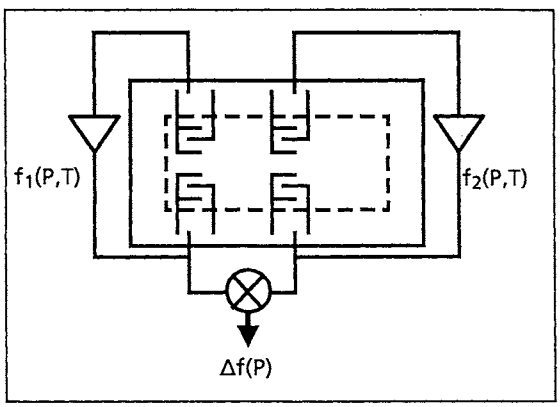
This compensation solution appears hard to implement because of the difficulty to growth a thick AℓN film. Nevertheless, with 2 microns of AℓN, it could be possible to etch thinner the silicon membrane (about 4 to 5 microns thick).

4.2. -Differential technics

A classical method to obtain a well compensated structure is to use a differential technique. Two solutions are possible as presented on Fig. 11.



(a)



(b)

Fig. 11 : *Differential compensation of the temperature sensitivity..*

The first solution (Fig. 11-a) needs two devices ; the first one acts as a reference oscillator while the second is the sensor itself. The output signal is theoretically independent of the temperature if the AℓN and silicon thicknesses are exactlly the same in the both devices.

The second solution (Fig.11-b) needs only one membrane with 2 IDTs deposited at the centre and at the edge of the membrane. A such structure realise a well temperature compensation.

The oscillation frequency f_1 is practically independent of the pressure because of the low deflection of this part of the membrane. So, the pressure sensitivity depends only on the oscillation frequency f_2 and is equal to the sensitivity obtained with a single membrane and a single IDT.

5 - Conclusion

We realized a Lamb waves sensor by using a AℓN/Silicon multilayer structure. The temperature and the pressure behavior have been studied and some suggestions have been proposed to improve the temperature behavior. Experimental tests of the temperature compensated structure are in progress for comparison.

References

[1]. S.W. Wenzel and R.M. White.
Analytic comparison of the sensitivities of bulk-wave, surface-wave, and flexural-wave ultrasonic gravimetric sensors.
Appl.Phys.Lett. **54**(20), 15 may 1989.

[2]. T.M. Reeder, D.E. Cullen and M. Gilden.
SAW oscillator pressure sensors.
US 1975, p. 264.

[3]. D. Hauden.
Miniaturized bulk and surface acoustic wave quartz oscillators used as sensors.
IEEE Trans.Ultrasonics,Ferroelectrics,Freq. Control, Vol UFFC-**34**(2), March 1987.

[4].E.L. Adler,J.K. Slaboszewicz, G.W. Farnell and C.K. Jeon.
PC software for SAW propagation in anisotropic multilayers.
Ultrasonics, Ferroelectrics, Frequency Control, vol. **37**(2), May 1990.

[5]. K.Tsubouchi,K.Sugai and N.Mikoshiba.
Zero temperature coefficient surface-acoustic-wave devices using epitaxial AℓN films.
Ultasonics symposium 1982, p. 340.

FORCE-FREQUENCY TRANSDUCER USING A MULTIMODE DOUBLY ROTATED QUARTZ RESONATOR

B. DULMET[①], R. BOURQUIN[①], N. SHIBANOVA[②*]

[①] Laboratoire de Chronométrie, Electronique et Piézoélectricité
Ecole Nationale Supérieure de Mécanique et des Microtechniques
La Bouloie, Route de Gray, 25030 Besançon Cedex, France.

[②] Institute of Physics and Technology of Moscow
141700 Dolgoprudny, Moscow Region, Russia

I. ABSTRACT

This paper discusses the usability of disk plate resonators operating in thickness-shear modes as force sensors providing with frequency output. The main limitation occuring in the case of single mode operation comes from its thermal sensitivity. Thus, we suggest a drastic reduction of this limitation by the use of two vibrating modes operation. This can be done because both of force and temperature sensitivities of thickness modes can be predicted. Then the frequency variations of both employed modes provide with a system of two equations with two unknowns : force and temperature. If the two modes are properly chosen, we can expect better results solving this system than separately measuring temperature and then correcting the output, in case of monomode operation. It is shown that simultaneously using anharmonic modes presents significant advantages in the suggested kind of sensors. The discussion includes experimental results obtained from SC-cut contoured resonators.

II. INTRODUCTION

High stability thickness-shear resonators essentially operate on trapped energy modes. So, the amplitude of vibration is absolutely negligible at the circular edge. Because of that, if a tensile or compressive force is exerted on a diameter of the disk, the corresponding change of the boundary conditions does not directly modify the frequency. Instead, the induced static stresses slightly change ultrasonic propagation properties in the volume of the resonator. Hence, a slight frequency variation occurs from the non-linearity of the stress-strain relationship of the material. This effect, often called acoustoelastic effect, was first investigated by Ratajski [1]. The frequency shift can be directly estimated from the knowledge of the static stresses at the plate center, where the trapped mode usually exhibits its maximum amplitude [2]. A more refined and precise calculation [3] takes into account the repartition of static stresses and vibration in the whole bulk of the resonator. The frequency shift is then explicitely given by a perturbation integral [4], related to the theory of small dynamic fields superimposed on a quasi-static bias [5]. Since models are available to describe as well the vibration as the static stresses, the force-frequency effect can be predicted in a general way for the whole family of quasi-thickness modes, including the so-called anharmonic modes, which exhibit nodal lines in the plane of the disk. Those modes are not of common use in resonators for clocks application, but they show an interesting feature for force-sensitivity application : their amplitude maximums are not located at the plate center. So, they can exhibit lower or higher force-sensitivity than usual metrological modes.

> Work carried during stay at LCEP/ENSMM*

All quasi-thickness modes are sensitive to temperature, although up to different extents. This sensitivity has been studied for years, in such way that it can be precisely predicted, especially in widely known crystal cuts such as SC-cuts. So it does make sense to have a resonator operate on two simultaneous modes to extract temperature and force at once, or at least to numerically compensate thermal sensitivity that may degrade force output.

The first part of the paper explains how to do that in a doubly rotated cut (for instance SC-cut). We suggest to use one C-mode, sensitive to the exerted force and relatively not very sensitive to temperature, and one B-mode, very sensitive to temperature, and less sensitive to force. A set of theoretical and experimental data are obtained for SC-cut resonators, to help understanding and guide decisions. The second part of this paper is devoted to the specifics of the use of anharmonic modes, from a practical point of view. A specific design of electrodes for multimodes operation is suggested. As a matter of fact, if both modes have not their amplitude maximum in the

same region of the disk, it is a good idea to attempt driving them with separate electrodes, thus simplifying the driving electronics, and also optimizing motionnal elements of their equivalent circuit. This idea grew at LCEP in the late seventies [6], but was then aimed at the achievement of thermal compensated crystal oscillators. Early experiments showed feasability of suitable electrodes that could help obtaining interesting Q factors for anharmonic modes, provided that the radius of contour was designed to allow sufficient trapping for the vibration.

III. CONSTRUCTION PARAMETERS FOR TWO-MODES SENSORS

Let us consider a plano-convex resonator with a disk shape, cut in a doubly rotated thin plate. The model of Stevens and Tiersten [7] predicts that, in the general case of doubly rotated cuts, the symmetry axes of the modes patterns are oblique with respect to the usual doubly rotated axes. We denote by α_n this third rotation in the plate plane. It depends on the overtone number n, i.e. the number of nodal planes along the thickness, and on the kind of thickness propagation (thickness extensional A, fast shear B, slow shear C). Experimental and theoretical values of α_n for B and C modes, 3rd and 5th overtone of SC cut resonators are shown on the following table [8] :

modes	α_n from [7]	α_n from [8]	experiment
C3	21°	25°	~50°
C5	29°	22°	~45°
B3	32°	32°	~40°
B5	-37°	-38°	~ -37°

Table I

The in-plane distribution of the quasi-thickness modes is given by analytical solutions of the following partial derivatives equation, written in the third rotation axes, for the considered overtone and thickness family :

$$M'_n \frac{\partial^2 u}{\partial x^2} + P'_n \frac{\partial^2 u}{\partial z^2} + \varrho\omega^2 \frac{\bar{c}\, n^2 \pi^2}{e_0^2}\left(1 - \frac{x^2+z^2}{Re_0}\right)u = 0 \qquad (1)$$

M'_n and P'_n are dispersion constants that only depend on the material constants, the overtone number n and the thickness mode family. e_0 is the maximum thickness at the center of the plate, R is the radius of contour, $\bar{c}$ is the effective stiffness for thickness propagation.

The behavior of solutions is separately given along x" and z" axis by Hermite's functions. Then the modes can be identified by their belonging to a given thickness propagation family (A,B,C), the overtone number n, and the number of nodal directions, i.e. the number of roots of the proper Hermite function, along the α_n -rotated symmetry axes in the plane of the plate. For instance, a usual metrological mode is $(C,3,0,0)$, exhibiting a gaussian repartition (no nodal line) along both in-plane axes. We call anharmonic mode any (n,m,p) mode with m or p, or both m and p, non zero. The trapping effect is separately governed along both in-plane axes by the scaling factors:

$$s_x = D[\sqrt{(n\pi)}\,(4\,\bar{c}\,/M'_n Re_0^3)^{1/4}]$$
$$s_z = D[\sqrt{(n\pi)}\,(4\,\bar{c}\,/P'_n Re_0^3)^{1/4}] \qquad (2)$$

where D is the diameter of the plate. Trapping effect increases with values of s_x and s_z. For given values of those factors, the trapping decreases when anharmonic numbers m and p increase.

The parameters of first interest for the study of force-frequency effect is the azimut ψ of the acting force. Azimut origin is generally chosen as the X' axis of the doubly rotated frame. We summarize all angles of importance on Fig.1. It should be noticed that, in this field, all steps of the resonator production must be known in order to avoid errors in the sign of angles. α_n can be ignored to classically calculate the current drained by circular electrodes and to predict the force-sensitivity factor $K_f = (\Delta f/f^2)\,(nD/F)$ (where F is the acting force) from the values of static stresses only at the plate center, which is a rather reasonable restriction for the metrological modes $(n,0,0)$, since they actually show their maximum at the plate center. The same remark holds for the radius of contour R, acting on trapping effect. Given some mode, if R is not small enough, trapping is not sufficient and the mode exhibits poor Q factor and stability, and then, it may be not of practical use. Of course, we did take α_n and R into account in a previous paper [3], to calculate the force-sensitivity from Tiersten's perturbation formula :

$$\frac{\Delta\omega}{\omega} = \frac{1}{2\omega^2}\,\frac{\int_V c_{ijkl}\,\bar{u}_{l,k}\,\bar{u}_{j,i}\,dV}{\int_V \varrho\,\bar{u}_i\,\bar{u}_i\,dV} \qquad (3)$$

$$\bar{c}_{ijkl} = (\delta_{jl}c_{ikmn} + c_{ijklmn})\,\bar{S}_{mn} + c_{ijsk}W_{l,s} + c_{ipkl}W_{j,p}$$

in which the integral is actually performed over the whole bulk of the resonator. In this formula, u is the dynamic displacement, $\bar{S}$ is the static strain, and w the static displacement.. We only tested all $(n,0,0)$, $(n,0,2)$ and $(n,2,0)$ modes, 3rd and 5th overtones of B and C modes in SC cut. A comparison of results for force-

frequency effect of some of those modes is given on Table II. In order to reduce errors due to the strong thermal sensitivities of B modes, the frequency was continuously tracked along time. A corresponding comparison of theoretical and experimental data for thermal sensitivities of some modes is given on Table III. a, b, c classically denote the first, second and third order temperature coefficients of frequency. The theoretical values are obtained using thermal derivatives of effective elastic coefficients referred to the natural stress-free state of material, taken at 25°C [9]. The numerical values are taken from Ref [10].

modes	Ψ	$K_f x10^{14}$	Ψ	$K_f x10^{14}$	Ψ	$K_f x10^{14}$
C300	50	1.205	140	-1.15	64.1	1.065
C500	50	1.235	140	-1.18	xxxx	xxxx
C520	50	1.212	140	-0.98	61.0	1.130
B300	50	0.464	140	-1.11	xxxx	xxxx
B302	50	-0.324	140	-1.16	61.0	0.0
B500	50	0.466	140	-1.14	64.1	0.0

Table II

Theoretical values of Kf coefficient

SC-cut resonator ϕ=-22.42° θ=34.11°

e_0=1.093mm R=250mm

	First order $\times 10^6$		Second order $\times 10^9$		Third order $\times 10^{12}$	
mode	a th	a exp	b th	b exp	c th	c exp
C300	0.659	0.684	-12.891	-12.57	59.380	53.40
B302	-31.022	-27.48	-32.452	-13.83	-32.25	-299.7
B500	-29.541	-26.57	-21.981	-8.92	-22.522	-234
C520	0.042	0.0565	-15.308	-15.16	54.104	48.82

Table III

Temperature coefficients for various modes

Now, we consider a resonator simultaneously operating on a first mode which is sensitive to force and relatively non sensitive to temperature, and another mode which is less sensitive to force and more sensitive to temperature. In a SC-cut, we can find a couple of C and B modes matching this behavior, provided that the azimut angle is correctly chosen. Thus we can write :

$$\begin{aligned} \sigma_c &= KF + a\Delta T + b\Delta T^2 + c\Delta T^3 \\ \sigma_b &= LF + \alpha\Delta T + \beta\Delta T^2 + \gamma\Delta T^3 \end{aligned} \qquad (4)$$

where ΔT is the temperature change, and σ_c and σ_b are the relative frequency shifts of C and B modes, respectively measured from the frequency without any exerted force, at a reference temperature, usually 25 °C. From (4), it is rather obvious to obtain the following equations:

$$L\sigma_c - K\sigma_b = (La - K\alpha)\Delta T + (Lb - K\beta)\Delta T^2 + (Lc - \gamma K)\Delta T^3 \qquad (5)$$

$$(La - K\alpha)F = a\sigma_B - \alpha\sigma_c + (b\alpha - a\beta)\Delta T^2 + (c\alpha - a\gamma)\Delta T^3$$

Because α is far larger than a (see Table III), the function of ΔT at the right hand member of (5-a) is monotonous in the [0-100°C] operating range. We give on Fig.2 a plot of the defect of linearity of this function in the [0-100°C] temperature range, for the couple of modes $(C,3,0,0)$ $-(B,5,0,0)$. Because this effect is relatively small, ΔT can be obtained as a polynomial in terms of powers of $(1\sigma_c - K\sigma_B)$. From Eq.(5-a), we got :

$$\Delta T = t_1 x + t_2 x^2 + t_3 x^3 + t_4 x^4 + O(x^5) \qquad (6)$$

$$t_1 = 1/(La - K\alpha) \qquad x = L\sigma_c - K\sigma_B$$
$$t_2 = a_1 t_1^3 \qquad a_1 = -Lb + K\beta$$
$$t_3 = (2a_1 t_1 + a_2) t_1^4 \qquad a_2 = -Lc + K\alpha$$
$$t_4 = 5t_1^6 a_1 (a_1^2 t_1 + a_2)$$

The value of F itself is obtained by substitution of (6-a) into Eq.(5-b). Since our goal is not to discuss the design of a temperature sensor, the ΔT information itself is not required. So we directly put the F information as a series expansion in powers of x :

$$(La - K\alpha)F = a\sigma_b - \alpha\sigma_c + \sum_{n=2}^{\infty} f_n x^n \qquad (7)$$

$$f_2 = t_1^2 A \qquad A = b\alpha - a\beta$$
$$f_3 = 2t_1 t_2 A + t_1^3 B \qquad B = c\alpha - a\gamma$$
$$f_4 = (2t_1 t_3 + t_2^2)A + 3t_1^2 t_2 B$$
$$f_5 = 2(t_1 t_4 + t_2 t_3)A + 3(t_1^2 t_3 + t_1 t_2^2)B$$
$$f_6 = (2t_2 t_4 + t_3^2)A + (3t_1^2 t_4 + 6t_1 t_2 t_3 + t_2^3)B$$

Here, the f_6 coefficient omits some terms because we use restricted expansion of ΔT to the fourth order, thus we neglect the cross product of t_1 and t_5 coming from ΔT into f_6. The optimum order of (7) expansion depends on the desired operating range of the sensor, the couple of C and B modes, and also on the chosen value for azimut Ψ. In this field, two different strategies make sense. The first one consists in chosing it for maximum force sensitivity of C mode. The second one consists in chosing it for zero force sensitivity of B mode. As an example, we show on Table IV some theoretical results obtained from expanding (7) up to different orders in powers of x, in the

case of boths strategies (columns *a)* and *b)*, respectively) for azimut Ψ, and for the combination *(C,3,0,0)-(B,5,0,0)*, with the same resonator geometry as for data in Table III (thickness 1.093mm, diameter 15mm, radius of contour 250mm).

		Order of expansion					
F	ΔT	4a)	4b)	5a)	5b)	6a)	6b)
20	-25	1.7e-3	-1.1e-3	1.5e-4	2.5e-5	7e-6	2e-5
20	25	-1.6e-3	1.3e-3	1.7e-4	2.1e-5	6e-6	1e-5
20	55	-7.8e-2	7.8e-2	2.1e-2	2.3e-3	2e-4	1e-3
20	75	-3.6e-1	-4.0e-1	1.5e-2	1.2e-2	-2e-3	4e-3
5	25	-1.6e-3	1.3e-3	1.7e-4	2.1e-5	6e-6	1e-5

Table IV

error on F calculation vs order of (7) expansion
columns a) $\Psi=50°$ $|K_{fc}|$ *max*
columns b) $\Psi=64.1°$ K_{fB} *zero*

The precision of series expansion does not depend on the value of F itself, but only depends on temperature. This is rather obvious, since we used in (4) developments linear versus F and cubic versus ΔT, the latter quantity being approximately substituted into (5-b). For the considered couple of modes, chosing Ψ for $L=0$ ($\Psi\approx64.1°$) provides with better results than $\Psi\approx50°$ (K_{fc} maxi), when associated with developments up to the 5-th order in (7). Conversely, the results of Table IV show equivalent errors for both ways of chosing Ψ, when the development is limited to the 4-th order.

Technically, it seems better to choose the $\Psi\approx50$ value since it ensures an extremum of σ_ψ, thus minimizing the dispersion of results due to misalignment of ψ. If we adopt this choice, we find that, in the considered case, any combination of temperature and force variations in the range [0-100°C][-20,+20N] is covered by the frequency shifts ranges [-275,+160Hz] for *(C,3,0,0)* mode and [-22kHz,+7kHz] for the *(B,5,0,0)* mode. Figure3 shows a tridimensionnal plot of F versus both frequency shifts. We give on Fig. 4 a comparison of predicted [3] and measured values of K_f for this combination of modes. Agreement is better for the C mode than for the B one. The discrepancies between theoretical and measured values should not be regarded as impeaching the effectiveness of an actual sensor : In fact, it is a little difficult to precisely measure L without thermal compensation, as we did. But an actual achievement of the discussed sensor would provide with precise value of L since fitting static frequency temperature curves by cubic polynomials has proven to be fairly reliable and it is easy to produce calibrating forces with help of gravity and some known masses.

The presented analysis shows that, in case of large temperature ranges, it would be better to center the expansions (5-7) around the middle temperature of the range, instead of 25°C, thereby minimizing the orders of development. As a whole, the exposed method seems sufficient to seriously think of designing class 2 or 3 force sensors in the [0-80°C] temperature range.

IV. SPECIFICS FOR USE OF ANHARMONIC MODES

From a practical point of view, a main limitation of the above-proposed design lies in the use of a single pair of electrodes, since both modes *(C,n,0,0)* and *(B,n,0,0)* have their maximun at the center of the plate: thus both signals at different frequencies are superimposed, and both needed oscillators must be able to separate them.

A better situation occurs if the sensor is designed in such way that two pairs of electrodes are used, each pair being devoted to the drive of one mode. This can be done by using anharmonic modes since they exhibit different vibration patterns. By properly chosing two anharmonic modes which have their maximum of amplitude in different regions, and the geometry of electrodes, it is possible to substantially reduce the level of the signal coming on an electrode from the second mode of vibration simultaneously present in the dual-mode sensor.

The chosen anharmonic modes must match the following conditions :

- quasi orthogonality of the mode patterns,
- existence of an azimut ψ of the exerted force for which the above-mentionned force-sensitivity coefficient K is high while the L sensitivity is low,
- good trapping of modes.

In the general case, the current through the k-th pair of electrodes is given by :

$$I_k = Y_{0k} V_k + \sum_{q=1,2} Y_{kq} V_q, \quad k=1,2 \qquad (8)$$

$$Y_{0k} = j\omega \varepsilon_{22} S_k / e_0$$

$$Y_{kq} = -8j \frac{(-1)^{(n_k+n_q-2)/2} \hat{e}_{26})_k \hat{e}_{26})_q Q_q}{c' n_q^2 \pi^2 e_0 L_q} \int_{S_k} v_q \, ds \int_{S_q} v_q \, ds$$

$(n_q\ m\ p) \rightarrow q$

$L_q = \pi^2 2^{m+p+1} mlpID^2/(s_x s_y)$

V_k denotes the voltage applied on the pair of electrodes of surface S_k. Q_q is the Q-factor of the q-th modes. Here, the components of the mechanical displacement are its projections v_q onto the normalized orthogonal eigensolutions of the Christoffel's problem. For more details about this point and the definition of coefficients appearing in (8), cf Ref [7].

The best combination seems to use the $(C,5,2,0)$ and the $(B,3,0,2)$ modes since they have their modes shapes quite perpendicular (See Table I). In addition, the $(B,3,0,2)$ mode is preferred to $(B,3,2,0)$ because the trapping is better along the α -rotated X" direction than along the Z" one (M'_n = 67.5 GPa, P'_n =28.8 GPa.)

The Fig.5 gives theoretical and experimental plots of K_f factors for these modes vs ψ. The Table V indicates the error associated to the calculation of the F force for different orders of expansion (7). From this point of view, the expected performances of a sensor using anharmonic modes would be very similar to those of a sensor that would operate on the usual modes, as presented in the previous Section. The Fig.6 presents some suggested electrodes design for simultaneous two anharmonic modes operation.

		Order of expansion					
F	ΔT	4a)	4b)	5a)	5b)	6a)	6b)
20	-25	-1.3e-4	-2.8e-3	1.5e-4	-6.4e-5	6e-5	2e-5
20	25	5.1e-4	3.5e-3	1.7e-4	-9.6e-5	5e-5	2e-5
20	55	4.1e-2	2.0e-2	2.1e-2	-1.3e-2	4e-3	3e-3
20	75	2.4e-1	1.05	0.133	-0.10	2e-2	2e-2
5	25	5.1e-4	3.5e-3	1.7e-4	-9.0e-5	5e-5	2e-5

Table V

error on F calculation vs order of (7) expansion

columns a) Ψ=50° $|K_{fC520}|$ max

columns b) Ψ=61° K_{fB302} zero

References

[1] J.M. Ratajski, " The force sensitivity of AT-cut crystals", Proc. 20th A.F.C.S., pp 33-49, April 1966.

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

[3] R. Bourquin, B. Dulmet, "Force sensitivity of trapped energy vibrations in a contoured resonators", Proc. 41st A.F.C.S., pp289-294, 1987.

[4] H.F. Tiersten, "Perturbation theory for linear electroelastic equations for small fields superposed on a bias", J.Ac.Soc.Am, V64,n°3,pp832-837, 1978.

[5] J.C. Baumhauer, H.F. Tiersten, "Non linear electroelastic equations for small fields superposed on a bias", J.Ac.Soc.Am, V33, n°2, Feb 1961.

[6] A. Thirard, "Etude topographique des vibrations de cisaillement d'épaisseur de résonateurs à quartz", Thesis work, n°105, Besançon 1980.

[7] D.S. Stevens, H.F. Tiersten, "An Analysis of doubly rotated quartz resonators utilizing essentially thickness modes with transverse variation", J.Ac. Soc. Am, V79,n°6, pp1811-1826, 1986.

[8] B. Dulmet, R. Bourquin, "3D Analysis of thickness vibrations in doubly rotated contoured quartz plates", Proc. IEEE Ultrasonics Symposium, pp463-468, 1991.

[9] B. Dulmet, R. Bourquin, "Influence of the definition of material constants in the study of crystal plates thermal behavior", Proc. 3rd EFTF, pp305-312, 1989.

[10] B. Dulmet, " Contribution à l'étude des vibrations à énergie piégée des résonateurs piézoélectriques à ondes de volume", Thesis Besançon, n°243, 1991.

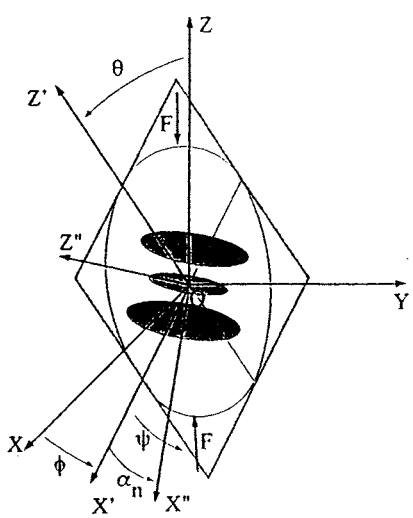


Fig. 1

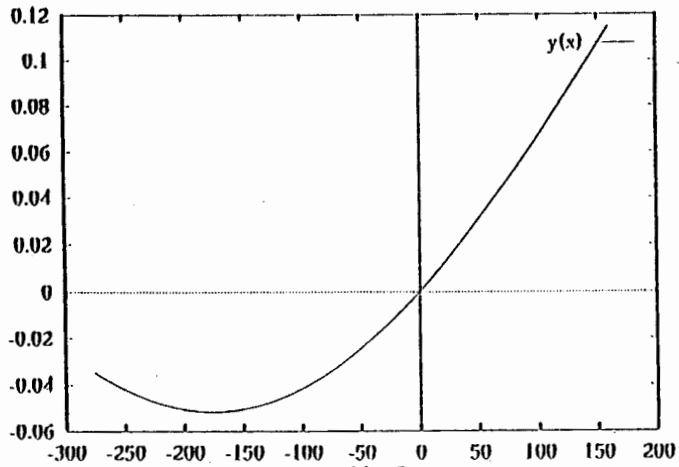


Fig. 2

Nonlinearity of the right-hand member of Eq(5a) vs. x

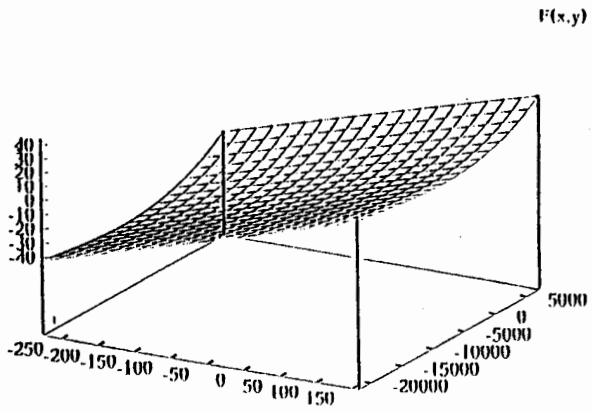


Fig. 3

$F = F(\delta f_c, \delta f_p)$ $\psi=50°$

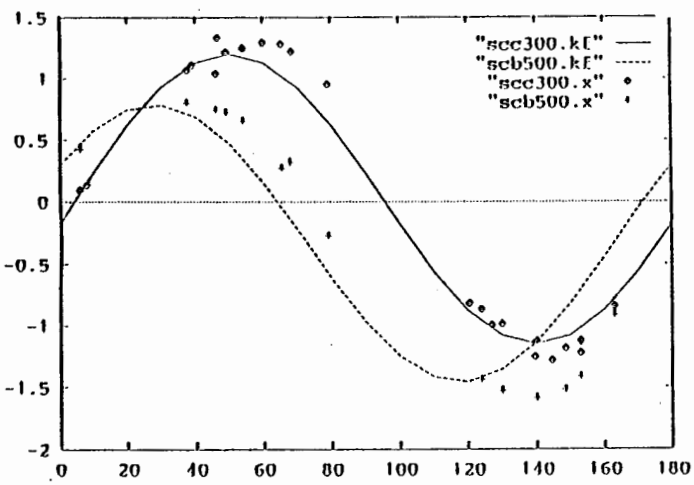


Fig. 4

K_t for $(C,3,0,0)/B(5,0,0)$ vs. ψ

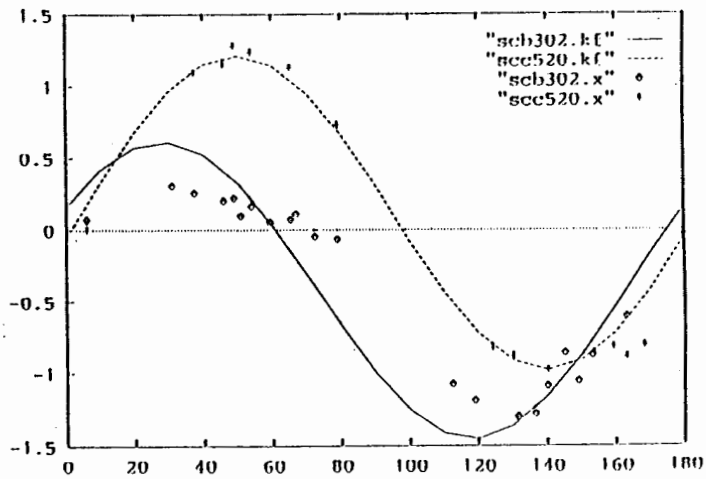


Fig. 5

K_t for $(C,5,2,0)/B(3,0,2)$ vs. ψ

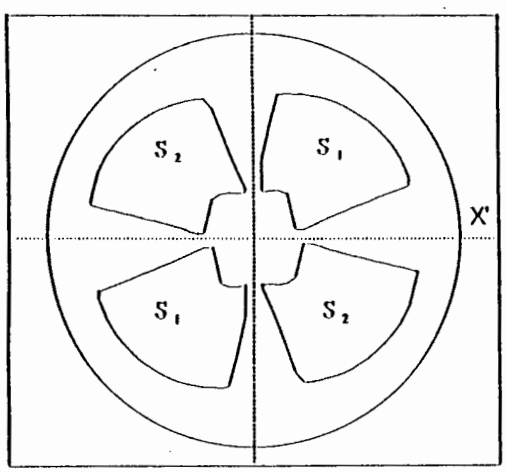


Fig. 6

Suggested electrodes shape for anharmonic modes

QUARTZ CRYSTAL ROTATION RATE SENSORS

S. Dalla Piazza, M. Christen

ASULAB S.A., R & D Laboratories of the SMH GROUP
Passage Max-Meuron 6, CH-2001 NEUCHATEL (Switzerland)

The rotation rate sensors presented in the present paper are based on quartz tuning forks, vibrating in their first "in-plane" flexural vibrating mode.

When such a vibrating structure experiences a rotation, the Coriolis force induces on each material point of the structure a displacement which is orthogonal to both the rotation axis and the velocity of the structure.

The sensors are obtained by a process similar to the one used in watch crystal manufacturing, so the potential cost is very low.

Different designs will be presented and experimental results will be given.

A FLOW SENSOR BASED ON SHEAR HORIZONTAL ACOUSTIC WAVE OSCILLATOR

Corinne Déjous(1), Dominique Rebière(1), Jacques Pistré(1), Moussa Hoummady(2)

(1) IXL - Université de Bordeaux I (URA CNRS N°846)
351, Cours de la Libération, 33405 Talence Cedex (France)
(2) LPMO - Université de Franche-Comté - (CNRS)
32, Avenue de l'Observatoire, 25000 Besançon (France)

ABSTRACT

This paper concerns the description of a Shear Horizontal Acoustic Plate Modes (SH-APM) flow sensor. Interdigital transducers on a Quartz-plate were used to generate and detect acoustic waves. Using a simple rotated crystal cut, with the propagation normal to the X cristallographic axis, only SH-APM waves are generated. The experimental set-up consisted in an SH-APM delay-line connected into the feedback path of a radiofrequency amplifier, and a computer controlled bench which monitored the frequency perturbation due to gas flow injection. The responses of this oscillator to flow variations are described and experimentally characterized. These measures were performed using the temperature dependance of SH-APM waves. Respective performances of SH-APM and SAW devices are compared.

Keywords : **flow sensor, acoustic waves, shear horizontal, plate mode, interdigital transducers, delay line.**

1. INTRODUCTION

For some years, acoustic devices are getting increasing interest for many sensing applications (Ref. 1-2). Acoustic waves can be devided into various types, including Surface Acoustic Waves (SAW), Lamb waves and Shear Horizontal Acoustic Plate Modes (SH-APM). Depending on the desired sensing application, some of them are more suitable than others, since they have different propagation characteristics. During the last decade, the studies on these topics have mainly concerned SAW devices, especially for gas, temperature and pressure sensors. An acoustic wave sensor is based on a delay line, whose delay time is sensitive to at least one external parameter. This delay line consists in a piezoelectric crystal, along which an acoustic wave can propagate. The delay time variation can be evidenced by introducing the line into the feedback path of an amplifier and measuring the frequency variation of the resulting oscillator. As crystals properties are temperature dependant, we can expect that an acoustic wave oscillator can be used for gas flow detection and measurement.

2. THE DELAY LINE

2.1. SH-APM possibilities

Our device is based on a quartz crystal plate, with a simple rotated cut. An acoustic wave is generated and then received by two interdigital transducers sputtered and etched on the crystal. For several years we have been studying acoustic devices and particularly gas sensors using Surface Acoustic Waves (SAW) oscillators (Ref. 3). However Rayleigh waves are too much attenuated when the quartz surface is in contact for example with a liquid (Ref. 4-5). On the other hand, Shear Horizontal Acoustic Plate Mode (SH-APM) waves have a particle displacement normal to the propagation direction x_1 (fig. 1) and parallel to the plate surface, i.e. only the x_2 component exists. Therefore, no energy radiates out of the quartz (Ref. 4) and the presence of a liquid medium at the surface does not attenuate the wave propagating along the plate. These waves present other advantages in comparison with surface acoustic waves; the most interesting is the possibility to dissociate wave generation and detection on one surface, and wave perturbation on the other one, due to the fact that the wave propagates throughout the plate thickness and not just beneath the electroded surface.

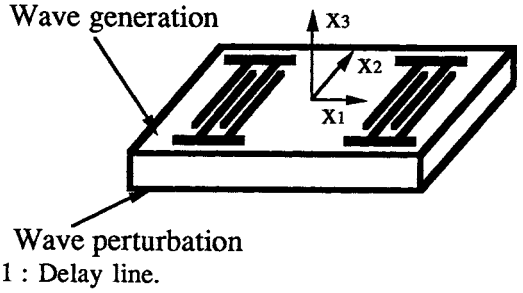


fig.1 : Delay line.

2.2. Waves generation

Different kind of waves can be generated by an

interdigital transducer, according to the plate cut and the propagation direction. In the case of a simple rotated quartz plate, there are two particular propagation directions : along the diad X cristallographic axis, and normal to it (Ref. 6). In the first case, four types of waves are generated : the Rayleigh wave and three different bulk waves (shear vertical, shear horizontal and longitudinal waves). If the propagation direction is normal to the X axis, only shear horizontal waves can be piezoelectrically generated.

When a sine wave, at frequency f, is applied to an interdigital transducer, a shear horizontal bulk wave is generated if the waves due to individual pairs add constructively (Ref. 7-8). So, if λ is the transducer periodicity, there is a bulk wave travelling in a direction θ_e from the surface, if :

$$f = \frac{V(\theta_e)}{\lambda \cdot \cos(\theta_e)}$$

where $V(\theta_e)$ is the bulk wave phase velocity (fig. 2). This equation shows that coherent bulk wave generation can only occur above a cut-off frequency f_c (Ref 4) :

$$f > f_c \text{ and } f_c = \frac{V(0)}{\lambda} = \frac{C_a}{\lambda}\sqrt{\frac{C_{44}}{\rho}}$$

with $C_a^2 = \left(\frac{C_{66}}{C_{44}}\right) - \left(\frac{C_{46}}{C_{44}}\right)^2$

where C_a is the anisotropic factor, ρ is the piezoelectric material density and C_{44}, C_{66} and C_{46} are stiffnesses in the (x_1, x_2, x_3) orthogonal reference system, represented on fig.1.
This wave propagates under the surface at $\theta_e = 0$, and is called the Surface Skimming Bulk Wave (SSBW).

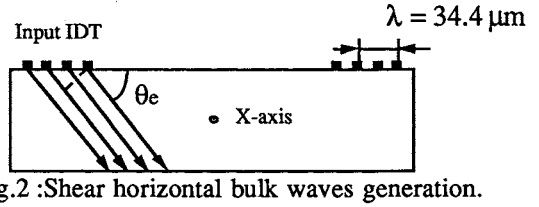


fig.2 :Shear horizontal bulk waves generation.

When the upper and lower quartz faces are parallel, free and optically smooth, these bulk waves can reflect on them. With a surface acoustic wave device, only the surface skimming bulk wave is present, as the bottom face of the quartz plate is not optically smooth and is coated with an acoustic absorbent. As the shear horizontal bulk waves reflect on a surface without polarization change, they can propagate in the plate by successive reflections on the two faces (fig.3).

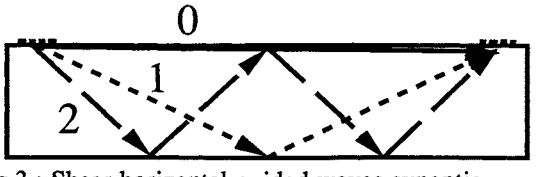
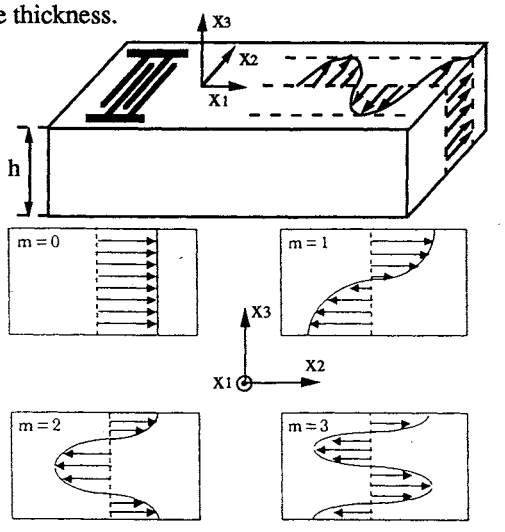


fig.3 : Shear horizontal guided waves synoptic

For a thin plate, incident and reflected waves interfere with each other. It results a wave propagating along the plate, with a particle displacement throughout the plate thickness : these waves are called shear horizontal acoustic plate mode waves. The propagation of these modes was studied first by Auld (Ref. 4), and then by others teams (Ref. 5), by assuming an isotropic material. With this assumption, such modes have a particle displacement sinely dependant between surfaces with a maximum on them, the mth order mode having m nodes in the (x_2,x_3) plane (fig.4). It was also demonstrated that in a delay line, such as described, even with an anisotropic material, the mth mode propagates at the unperturbed frequency F_m :

$$F_m = f_c \cdot \sqrt{1 + \left(\frac{m \cdot \lambda}{2 \cdot C_a \cdot h}\right)^2}$$

where f_c is the cut-off frequency and h is the quartz plate thickness.



First modes
fig.4 : Shear horizontal acoustic plate mode waves

These shear horizontal acoustic plate mode waves can be seen on the device transmission response given fig.5 :

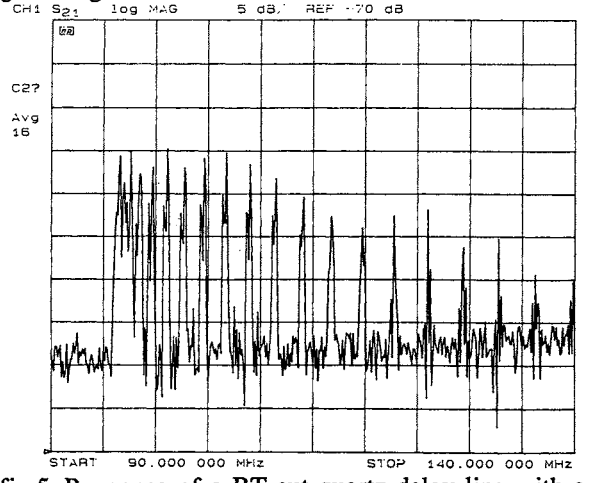
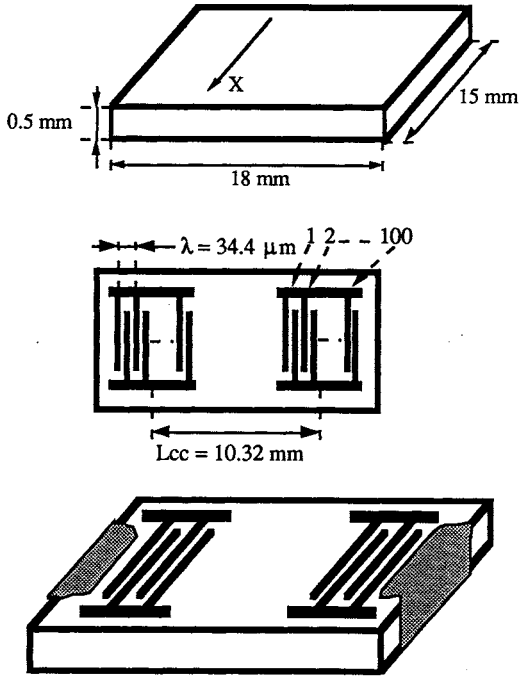


fig.5 :Response of a BT-cut quartz delay line with a propagation direction normal to the X axis.

3. DEVICE DESCRIPTION

3.1. The delay line

The mechanical characteristics of the studied delay line are (fig.6) :
- quartz crystal plate, BT-cut and propagation direction normal to the X axis (Euler angles are : $\varphi = 0\,°$, $\theta = 31.5\,°$, $\psi = 90\,°$), 18 mm long, 15 mm wide, 0.5 mm thick ;
- the interdigital transducers have 100 pairs of electrodes, a periodicity λ equal to 34.4 µm and a center to center distance L_{cc} equal to 10.32 mm ;
- an acoustic absorbent was deposited at each end of the plate to avoid parasitic reflections.



Viscoelastic elastomer : electrical insulator

fig.6 : Delay line mechanical characteristics.

3.2. The test device

This delay line was mounted in a test device (fig.7) : a platinum resistor was screen-printed onto an alumina substrate for quartz crystal heating. The quartz crystal plate was sticked on this substrate with the electroded face on the top and spaced by two alumina wedges. An elastomer seal along the two wedges was used.
An inlet and an outlet were sealed at each end to ensure a gas flow under the plate; the airtightness was obtained with an elastomer sealed all around the plate. A thermocouple was embedded in this elastomer, its end in contact with the quartz surface.

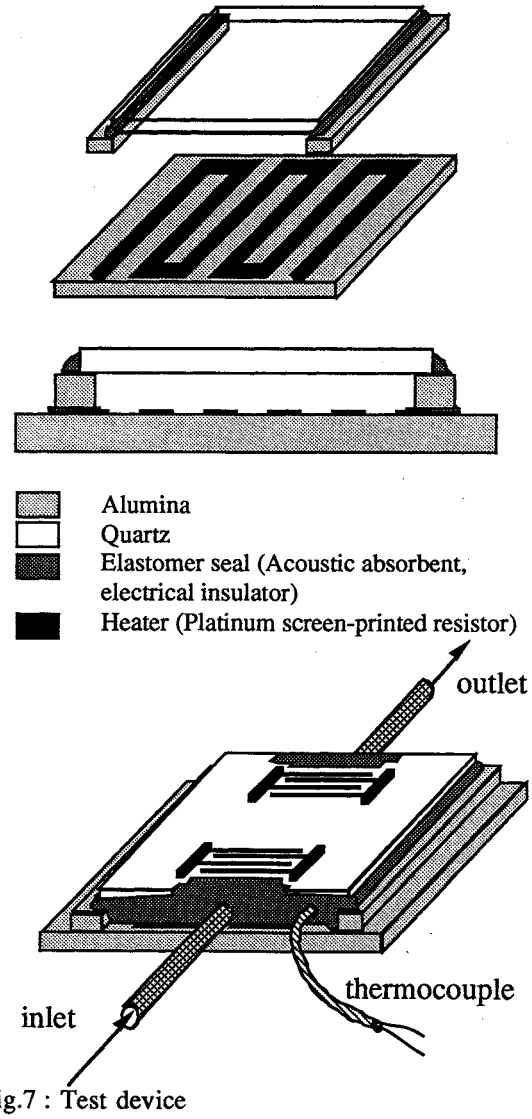


fig.7 : Test device

3.3. Oscillator configuration

This device was then placed in a test cell to avoid external perturbations. Finally this global device was connected into the feedback path of a radiofrequency amplifier (fig.8) The assembly was almost the same as for a surface acoustic wave device : a 28 dB amplifier (ANZAC AM-182) with a variable attenuator (ANZAC AT-101) and a coupler (ANZAC CH-132) allowed respectively to satisfy the gain oscillation condition and to measure the oscillator frequency. Matching networks (LC filters) were placed on the two delay line electrical ports to minimize reflections for the chosen mode. The mode used for these experiments was the fifth, which frequency was about 99.7 MHz. The most important difference with a surface acoustic wave configuration (Ref. 9), is the presence of a variable phase shifter and a tunable bandpass filter (K&L 3BT-95/190-5O). The first one allowed to adjust the oscillation phase condition ; the phase was controlled by a variable

continuous voltage. The voltage controlled phase circuit has been made at the IXL laboratory. The bandpass filter avoided parasitic oscillations.

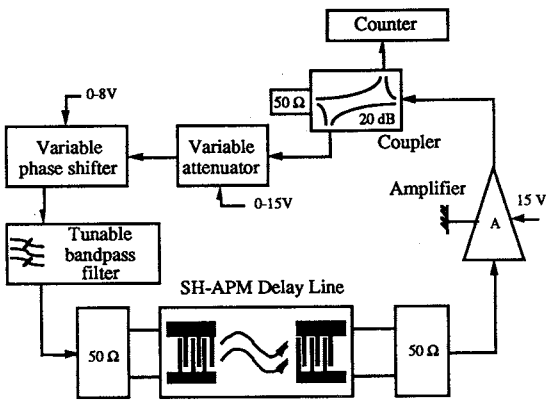


fig.8 : SH-APM oscillator synoptic.

A computer controlled bench monitored the temperature and the frequency perturbation due to gas flow injection, which was itself supplied by a computer controlled gas-line.

4. EXPERIMENTS AND RESULTS

4.1. Frequency variations with an air flow injection

Series of measurements were done, with three heating power levels : 0.9 W, 2.0 W and 2.6 W. For each of them, we measured the frequency variations caused by air flow injections. The computer controlled gas-line flowmassmeter allowed to control the gas flow between 100ml.min^{-1} and 700ml.min^{-1}.

The power applied to the heater induced a temperature T'_2 on the alumina substrate, and a temperature T_2, slightly lower, on the quartz surface (fig.9). The gas flow modified the quartz delay line temperature : the difference between the gas flow temperature T_1 and the quartz unperturbed temperature T_2 induced forced convection. The frequency variation should be all the more important since (T_2-T_1) and the rate flow are large.

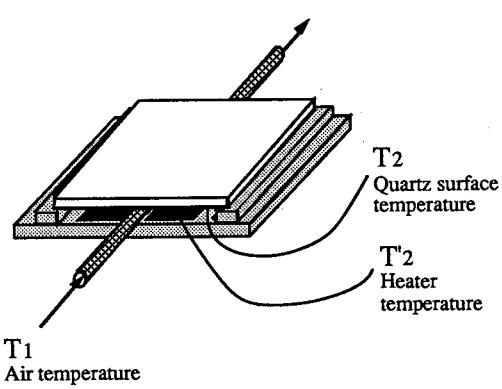
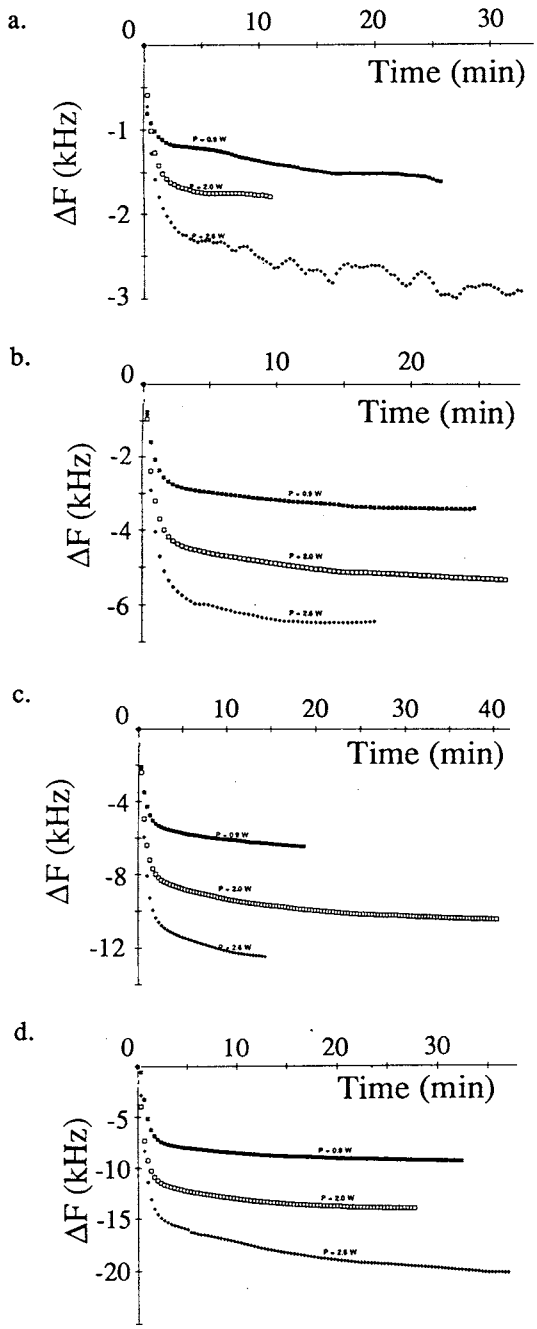


fig.9 : Temperatures to take into account.

The following curves were obtained, showing the oscillator transient responses for different air flow injection levels and three heating powers (fig.10.a-e). As it could be expected, the frequency variation is larger for greater flow injections, with the same heating power; it is also larger, for a fixed gas flow, when the injected power is increased. The sensitivity to air flow variation depends on the power applied to the screen-printed resistor. This frequency variation goes from -1.6 kHz for a power of 0.9 W and a flow equals to 100ml.min^{-1} (fig.10.a), up to -20.7 kHz for 2.6 W and 700ml.min^{-1} (fig.10.e).



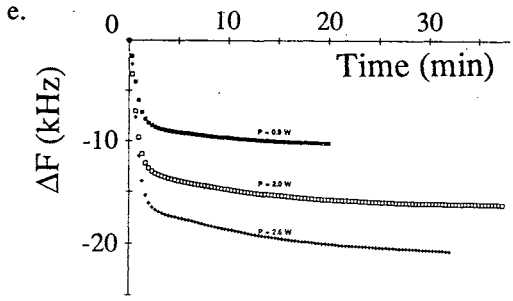


fig.10.a-e : Frequency variation versus air flow injections, at three heating powers (■ $P_1 = 0.9$ W, □ $P_2 = 2.0$ W, ◆ $P_3 = 2.6$ W).
a. : 100 ml.min^{-1}, b. : 200 ml.min^{-1}, c. : 400 ml.min^{-1}, d. : 600 ml.min^{-1}, e. : 700 ml.min^{-1}.

With the previous curves, we can represent the frequency variation versus gas flow, with power as a parameter : the frequency value is obtained in steady state. The variations reported here are assumed to be linear (fig.11).

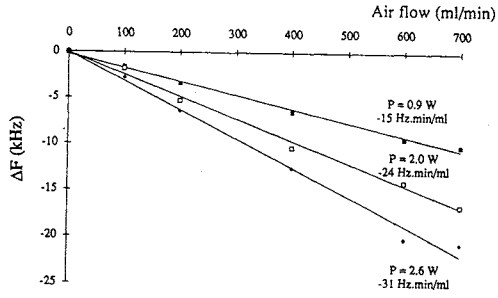


fig.11 : Frequency variation, in steady state, with an air flow injection, at three heating powers.

4.2. Frequency dependence with temperature

Furthermore, this oscillator frequency was measured versus the quartz lower surface temperature (T_2), without any flow. For this experiment, the voltage applied to the heater was controlled so that the temperature under the quartz remained at a value programmed by the computer. This temperature was increased (fig.12) from 20 °C up to 110 °C by steps (3.2 °C, 3 minutes), then decreased in the same way. As the temperature difference between successive steps was small enough (for a step during 3 minutes), the quartz crystal plate could be considered in equilibrium, as well as the oscillator frequency. This assumption was confirmed during the temperature decrease for which no hysteresis was observed.

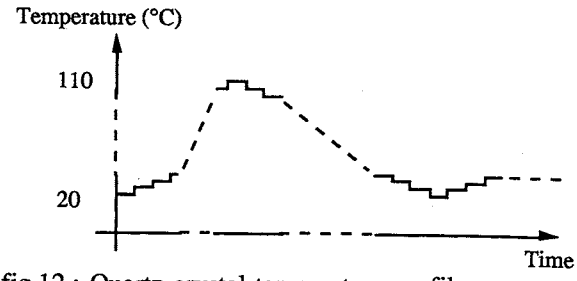


fig.12 : Quartz crystal temperature profile.

On the resulting curves (fig.13), there is a frequency discontinuity around 60 °C. This phenomenon can be explained by the fact that the delay line transmission response was not perfect : even with acoustic absorbent on the two plate ends, resonant frequency peaks were not pure and phase variation was nonlinear. There must be then a critical temperature above which a frequency jump occurs. This phenomenon needs more investigation.

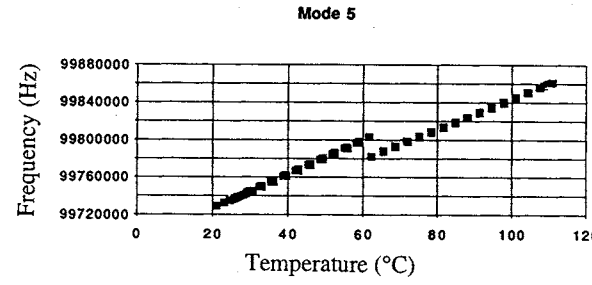


fig.13 : Oscillator frequency versus temperature

5. CONCLUSION

These two types of experiments will help us to understand precisely the mechanisms of this flow sensor operation. At each couple flow-power, indeed, corresponds a temperature at the quartz crystal surface which was measured, and when a flow is injected with a constant power applied to the heater, four types of phenomena are to be taken into account : radiation, conduction, and natural and forced convections. The theoritical study of these parameters and how they act in the above experiments are under investigation. This has already be done for a surface acoustic wave device with ST-cut, with which similar experiments were carried out and far lower frequency variations were obtained (Ref. 9). The main result with this shear horizontal acoustic plate mode device, using a quartz crystal, is that it is more sensitive to the temperature.

References

1 R.M. WHITE, Surface Elastic Wave, *Proc. of the IEEE, Vol. 58, No 8, August 1970*, pp. 1238-1277.

2 R.M. WHITE, Surface Acoustic Wave Sensor, *Proc. Ultrasonics Symp., San Francisco, CA, USA, Oct. 1985, Vol. 1*, pp. 480-494.

3 D. REBIÈRE, M. HOUMMADY, D. HAUDEN, J.-L. AUCOUTURIER, J. PISTRÉ, P. CUNIN, R. PLANADE, Surface Acoustix Wave (SAW) NO_2 Sensors : Theoritical Studies, Measurements and Design, *Proc. 6th Int. Conf. Solid-State Sensors and Actuators IEEE (Transducers '91), San Francisco, CA, USA, June 24-27, 1991*, pp. 351-354.

4 B.A. AULD, Acoustic Fields and Waves in Solids, *Second Edition, Robert E. Krieger, Publishing Company, 1990, Vol. 2*, Chapters 10 and 12.

5 S.J. MARTIN, A.J. RICCO, T.M. NIEMCZYK, G.C. FRYE, *Sensors and Actuators, 20, 1989*, pp. 253-268.

6 M. HOUMMADY, B.A. AULD, D. HAUDEN, E. BIGLER, Excitation et détection par des transducteurs interdigités des ondes transverses horizontales guidées dans une plaque de quartz, *1er Congrès Français d'Acoustique 1990, Colloque C2, supplément au n°2, Vol. 51, Feb. 1990*, pp.1245-1248.

7 D.P. MORGAN, Surface-Wave Devices for Signal Processing, *Studies in electrical and electronic engineering, Vol. 19, Elsevier Science Publishers B.V., 1985*, Appendix F.

8 E. DIEULESAINT, D. ROYER, Ondes élastiques dans les solides, Application au traitement du signal, *Masson et Cie, 1974*, chapter 7.

9 D.REBIÈRE, Capteurs à ondes acoustiques de surface - Application à la détection des gaz, *Thesis 826, Université de Bordeaux I, France, 1992.*

SPACE 1

Chairman: S. Starker

EUROPEAN COMPLEMENT TO GPS: PRESENTATION OF THE CONCEPT AND EXPERIMENTAL SYSTEM

Jacques BARBIER and Thierry TRÉMAS

Centre National d'Etudes Spatiales
Toulouse

ABSTRACT

This twofold article presents the European Complement to GPS (CE-GPS) and the relevant experimentation.
First part of this article deals with CE-GPS mission and system. Advantages and drawbacks of GPS are briefly recalled. The need and the reference mission of a complement to GPS is then introduced. The configuration of CE-GPS system is presented. System requirements are mentioned, specially those relevant to time/frequency fields.
The second part of the paper starts with an overview of the experimental system designed by CNES with a description of the powerful time oriented functions of the CE-GPS ground station models. The principles of the frequency and time servo control loops techniques implemented in the ground stations are presented. The final system makes it possible to assess the performances of six different methods of time transfer introduced at the end of the paper.

Keywords: GPS, CE-GPS, Experimentation, servo control, time transfer.

1- DESCRIPTION OF THE EUROPEAN COMPLEMENT TO GPS

1.1- The Global Positioning System

1.1.1- Advantages of the system

During the nineteen seventies, the American Department of Defense (DoD) undertook the installation of the Global Positioning System. This system consists of a constellation of 24 LEO satellites (orbit inclination of 55°, altitude of 20,000 km), controlled by a ground segment whose stations are distributed over the Earth's surface. The satellites, equipped with atomic clocks, permanently transmit radioelectric signals generated within their own payloads. These signals offer the possibility of two types of service: the Precise Positioning Service (PPS) and the Standard Positioning Service (SPS). The first service is more efficient than the second, but a system of encoding information (Anti Spoofing or A-S) limits access to it to a restricted number of users who have been granted DoD approval. The rest of this article will only refer to the SPS.
The user segment consists of receivers which are capable of receiving signals from visible satellites, of decoding the signals and of extracting or deducing various information of which the most important are: the moment at which a signal is transmitted from the satellite, the exact position of the satellite at the moment of transmission, the moment at which the signal is received by the receiver. By simultaneously picking-up the signals from four satellites, a receiver is able to know the distances between itself and the satellites and hence, to determine its position (in 3D) while at the same time synchronizing its internal clock with those of the GPS satellites which are synchronized among themselves.

Because it achieves a set of exceptional characteristics and performances, the GPS system opens the way to numerous applications in the areas, among others, of navigation and time/frequency. Among its most used qualities should be noted:

- the global coverage of the system
- the location accuracy (29 m in 2D, 95% of the time, SPS service) (Ref. 1)
- the synchronization accuracy with respect to UTC (363 ns, 95% of the time) (Ref. 1).

Since the satellites are equipped with Cæsium clocks, the system broadcasts a reference time with stability characteristics which are among the best known available.

Finally we should add that the nominal GPS performances can be improved, according to the techniques or technologies used to receive and/or exploit signals transmitted by the satellites (kinematic techniques, differential corrections, common views for synchronization purposes, hybridizations with existing systems, etc.).

1.1.2- Drawbacks of the system

Even though the GPS alone provides a significant set of efficient functions, it does have drawbacks which are more or less penalizing depending on the applications intended.

For instance the use of the GPS as sole navigation means for civil aviation (up until the non-precision approach) would entail strict requirements concerning the integrity and availability of the system. Whereas, at the present time, the GPS does not fulfil these conditions.

For example a GPS satellite might transmit false information to several users, without them being aware, thus leading them to mistake their position, for a period which might be as much as 2 hours, without any alarm being given. Now, for a non-precision approach phase the Civil Aviation Authority prescribes an alarm time of less than 10 seconds between the moment when the error is detected in the navigation solution and the moment when the pilot is informed.

In addition the constellation does not have enough satellites to ensure a really global and/or really permanent coverage of the whole of the Earth's surface (it should be remembered that in 3D, four satellites have to be permanently in the user's line of sight).

Apart from these technical weaknesses there are problems of a political nature. The system is entirely and solely

controlled by the United States. In order to prevent a hostile power using the GPS and thus threatening the national security, the performances have been deliberately downgraded (DoD authorization is required in order to have the use of nominal performances). This downgrading (referred to as Selective Availability or SA and for which, of course, the characteristics remain secret) is divided into SA ephemerides and SA clock. This characteristic affects navigation and time/frequency applications. Various methods may be used to free oneself of SA, but only to a certain extent. In general however, these techniques improve some performances to the detriment of some of the other qualities of the system, or risk no longer being valid if the characteristics of the downgrading are altered.

1.2- Upgrading the GPS, reference mission.

All of these drawbacks have led the civil community to look for ways of completing the GPS. The requirements differ according to the groups of users and the solutions proposed are not identical.

Since the middle of the nineteen eighties CNES has undertaken work relating to the improvement of GPS performances. Since several groups of users are involved (maritime transport, terrestrial transport, civil aviation, etc.), CNES has focused in particular on adapting GPS to the requirements of civil aviation concerning navigation quality (which has the corresponding time/frequency implications dealt with in 1.3.2). Given that these requirements are among the most demanding, it is possible to postulate that a system which would satisfy them would also satisfy, to a great extent, the needs of most of the other groups of users.

The reference mission adopted by CNES, following exchanges with the Civil Aviation community (Refs. 2, 3) and given the current state of technology, is as follows:

- achieve the requirements of a non-precision approach phase
- enable GPS to be used as a sole navigation means.

This mission shall first of all cover a zone including the whole of Europe. Later stages will enable the mission to cover the whole Earth.

For various reasons, only the quality of 2D navigation, determined using the GPS, has been considered in the reference mission.

The first reason is that the altitude measurement used on any aircraft is based on atmospheric pressure whereas the altitude supplied by the GPS is based on the centre of a model of the terrestrial geoid (the WGS 84). It would be catastrophic for aircraft to fly together in the same airspace if they did not have the same altimetric references. We should however specify here, that even if it is not directly used by the pilot, the vertical component of the navigation solution established by means of the GPS is of fundamental importance in the accuracy of the horizontal location.

The second reason is that determination of the clock bias is not used on board the aircraft. Its accuracy however, affects the horizontal components of the navigation solution. To sum up, even though the reference mission only deals with 2D navigation, it is still necessary to determine the four components of the navigation solution.

The concrete effects of this reference mission mean that:

- a pilot must be informed of any non-integrity in his navigation position within a time of 10 seconds. This means in particular that a pilot must be informed of failure on a GPS satellite within 10 seconds (this is referred to as the **monitoring mission**),
- the system availability must increase by eliminating coverage gaps (this function is referred to as the **navigation mission**).

Moreover it shall be possible to improve navigation accuracy by broadcasting differential corrections to be applied to the pseudo-ranges of GPS satellites.

While studying possible solution for improving GPS system performances, CNES has taken the following requirements into account:

- the GPS system shall on no account suffer interference
- dependency on the operators of the GPS system shall be reduced to a minimum
- the complementary system shall have maximum availability (which will directly affect the integrity of the navigation function)
- the complementary system shall not be very vulnerable
- the complementary system shall be open-ended and capable of evolving greatly
- the complexity of the user segment (receivers) shall not be increased.

CNES moreover believes that this complement should, as far as possible, continue to take advantage of a satellite-based navigation system (in other words, should not introduce the "conventional" restrictions of ground navigation systems for example).

These are the main elements which influenced CNES' choice.

1.3- The European Complement to GPS

1.3.1- The System (See figs. 1 and 2)

GENERAL CONFIGURATION OF CE-GPS SYSTEM

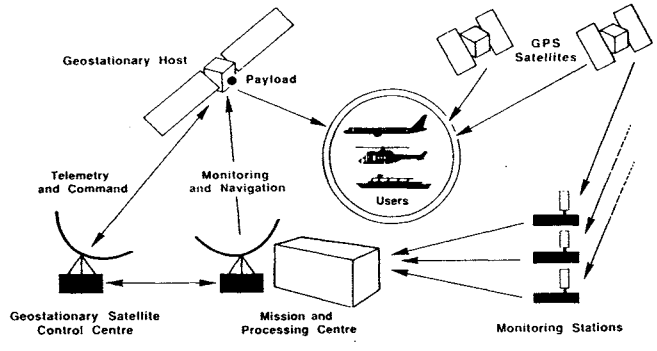
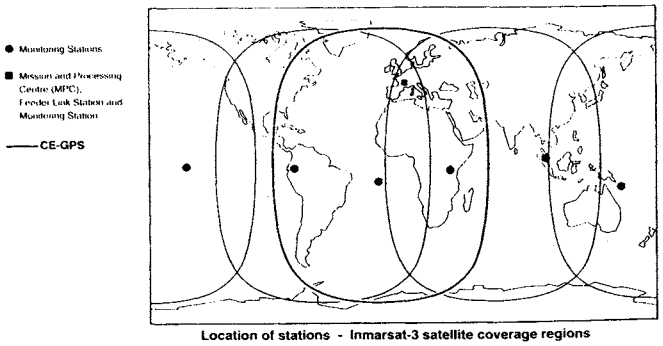


Figure 1



Location of stations - Inmarsat-3 satellite coverage regions

Figure 2

1.3.1.1- Information to be transmitted to Users

Information transmitted by the complementary system is of two kinds:

- information enabling a user to judge the integrity of his position
- navigation information.

Integrity information may be broadcast on any radiocommunications channel. In order not to increase the complexity of the receiver module carried by the aircraft, CNES has chosen to transmit on the same frequency as the GPS, at 1575.42 MHz (L1 band).

The navigation function, on the other hand, involves more constraints on the signal. The latter, in order to enable measuring pseudo-ranges within GPS tolerance, must be spread by modulating it with a Gold code, not reserved for the GPS, at a rate of 1.023 Mchips (which introduces a band occupancy with a 2 MHz spectrum).

In order to minimize modifications of the GPS receivers, which would be adapted to receive this complementary signal, the structure of the data message (and the navigation data itself) must remain similar to that of the GPS messages.

It should however be noted that the GPS message consists of data transmitted at 50 bits/sec, organized into 5 sub-frames of 300 bits each. In the context of a complementary system fulfilling the mission described above, the data to be transmitted might just as well refer to the state of health of GPS satellites as provide the information necessary for using signal for navigation purposes. In addition the message shall include differential type corrections for the pseudo-ranges of each satellite monitored, thus helping the user to define the integrity of his position. The volume of data will therefore be greater than a traditional GPS message. Consequently the data transmission rate will be greater than 50 bits/sec. However the subdividing into 300 bit sub-frames could be maintained.

1.3.1.2- The Space Segment

The extent of the mission's zone of application (and of its future extension) combined among other things, with the wish to preserve the advantages of a space system over a ground-based system resulted in the choice to broadcast information to users via a satellite.

Furthermore, due to certain reasons, among others, such as the necessity of rapid availability, of reducing costs to a minimum, of having redundancy for resources and also because of the proposals of satellite operators, it was agreed that the space segment would be provided by the Inmarsat-3 geostationary satellites (which it will be possible to complete, in the future, with other geostationary or LEO satellites). A payload adapted for the mission will be carried on each of these satellites. The satellites will broadcast information to provide a large geostationary coverage (covering an area of approximately 150° in both latitude and longitude).

1.3.1.3- The Ground Segment

The Feeder Link Station

In order that a user may incorporate the message coming from a geostationary satellite into his navigation solution, the signal will have to have similar characteristics to the GPS signals. However, a fundamental difference between and GPS satellite and a geostationary communications satellite is that the latter's payload does not include a high precision clock, synchronized with that of the LEO satellites. Moreover the onboard oscillator is subject to frequency fluctuations due to the space environment. This poses several problems:

- a solution will have to be found to transmit the navigation message in a synchronized way with similar messages broadcast by all of the GPS satellites (which in fact means creating a virtual clock on board the geostationary satellite).
- a frequency of 1575.42 MHz will have to be maintained at output from the geostationary payload (GP).
- the slippage between the signal carrier and the code will have to be maintained within acceptable limits for the receivers. No such problem occurs for the GPS satellites, since both code and carrier are generated on board the satellite from a single clock.

All of these points (and the corresponding specifications mentioned in 1.3.2) require that a servo-control signal be sent to the satellite by a station. The servo-control loop is necessary in order to compensate on the uplink for Doppler effects, ionospheric and tropospheric delays, the stability of the on board oscillator, etc. This station is referred to as the feeder link station.

Monitoring Stations

The monitoring mission and its availability (Ref. 4), combined with the desire to optimize the monitoring network (with as a criteria the minimizing of the cost/performance ratio) led to choosing a network of 7 stations, called monitoring stations distributed along the equator. These stations will continuously receive the signals from all of the GPS satellites in their line of sight. The whole constellation will thus be monitored at all times. Likewise this network will monitor the GPs which may be used in navigation solutions just as it does with the GPS satellites.

The first stage of the programme plans for the two main missions to be achieved solely above the zone defined by the coverage of a geostationary satellite above Europe. Only this zone will therefore receive information on the integrity of navigation satellites. It is important to note that in such a configuration, a user located on the edge of this coverage area would be likely to locate himself with GPS satellites which would be outside of the zone in which this information is broadcast. This justifies the fact that the monitoring stations should be placed beyond the zone above which the mission is achieved. Moreover such a configuration would make it possible to predict ahead of time, the state of health of a satellite, before it becomes visible to a user located in the zone covered by the GP. These stations would also make it possible to collect the data necessary for determining the precise orbitography of the GPS and geostationary satellites, which is indispensable for defining correction of the differential type used in the integrity mission.

Processing Centre

All of the data collected by the monitoring stations will be sent to a computing centre (the Processing Centre). This centre will recalculate and update the orbitography for each of the satellites monitored. Using the differences between the pseudo-ranges measured by the stations and the pseudo-ranges estimated from this orbitography, the computing centre will judge the state of health of the satellites monitored. This comparison will also make it possible to calculate the corrections to be made to the satellites' pseudo-ranges.

The computing centre is responsible for generating the message containing this information (including the GP's ephemerides for the navigation mission) and for transmitting it to the feeder link station which will then transmit it to the GP.

1.3.2- Time/Frequency specifications for the System (Ref. 5)

1.3.2.1 - Specifications resulting from the navigation mission

Synchronization with respect to GPS satellites

As has previously been indicated, the messages to be broadcast by the CE-GPS should be transmitted from the GP at the same moment as the messages (with an identical structure) transmitted from the GPS satellites. Such a statement implies that the clocks on the GPS satellites are synchronized among themselves, to the extent permitted by the SA and existing technological limits. On the basis of the level of horizontal precision defined in (Ref. 1), in other words, 38 m of UERE, 95% of the times in the presence of SA, a deduction was made that the standard deviation of the dating difference between an event whose references were the output signal code from the GP and between an equivalent event on output from a GPS satellite should be less or equal to 120 ns.
In the absence of SA, this synchronization accuracy drops to 20 ns. These values take into account the corrections of clocks broadcast in GPS messages. It is moreover obvious that these figures are likely to vary as a function of changes in the SA model.

Since the GP's virtual clock will therefore have to satisfy the same criteria as those for GPS satellites, the automatic processing of code to be implemented should enable the same statistical quality of clock synchronization between a GP and a healthy GPS satellite, as between two GPS satellites in a good state of operation. It should be insisted that such a specification does not take into account the quality of synchronization of GPS satellites with GPS time in Colorado Springs. If there is a bias between the time broadcast by the GPS satellites and the Colorado Springs time (which has not yet been observed), this bias will affect the time broadcast by the GP. This will not affect the CE-GPS mission in any way (see 1.3.2.2).

Frequency stability

The absolute value of the instantaneous deviation on output from the GP shall not exceed 500 Hz. The purpose of this specification is to facilitate locking by receiver loops.

The frequency of the carrier received by a stationary receiver shall have a stability better than 3.10^{-10} per 1 s.

The frequency stability specifications are not really restrictive. This is mainly due to the fact that it is not intended to make measurements of pseudo-speed with the GP during the CE-GPS navigation mission.

Code/Carrier Slippage

This specification is specifically for the transmission of a GPS type message via a GP. Indeed the GP's local oscillator will induce a frequency translation in the carrier without affecting the code. Moreover on the uplink, the ionosphere will have a non-linear affect on the code and the carrier. For a stationary receiver this slippage shall not have a speed greater than 0.2 m/s. Likewise and in order that the receiver be able to smoothe the code by means of the carrier, the variation in pseudo-ranges which will be measured on the code and carrier respectively shall not be greater than 10 m over a time of 1000 secs (this expression is specific for certain types of receivers).

1.3.2.2- Specifications resulting from the Monitoring mission

Relative synchronizations

The uncertainty about the time difference between two stations in view of the same geostationary satellite shall be less than 10 ns. This specification represents the degree of uncertainty on the pseudo-distance measurement due only to the desynchronization of the station clocks. Two stations which are not visible to the same GP shall have a synchronization uncertainty of 10 ns, i.e. 15 ns. A specification greater than 15 ns would lead to an uncertainty on the pseudo-distance which would affect the CE-GPS' performances.

Indeed measurements of pseudo-ranges by two stations on a common satellite must be consistent. A desynchronization between the stations' clocks would result in an increase in the uncertainty on the pseudo-ranges (an asymptotic case would involve considering two monitoring stations in the same place, sighting on the same satellite. If their respective clocks are desynchronized by a bias ΔT, the two pseudo-ranges which are supposed to be identical would be different by $c \cdot \Delta T$, without it being possible to determine whether at least one of the two measurements is correct).

Likewise when a satellite is lost by one station and relayed by another it is necessary to maintain continuity in the pseudo-distance measurement. This continuity cannot be ensured if the stations do not have the same time base.

Thus, as previously specified, the monitoring of navigation satellites is done by comparing the pseudo-ranges measured at the monitoring stations with the pseudo-ranges estimated from the precise orbitography of each of the GPS satellites (it should be remembered that this orbitography is calculated and continually updated at the Processing Centre using measurements taken of GPS satellites by the monitoring stations. In a steady state, this orbitography is independent of the orbitography broadcast by the GPS satellites and this has a fundamental effect on the reliability of integrity information). The pseudo-ranges measured at the stations take into account the ephemerides of the GPS satellites whose time reference is the GPS time generated at Colorado Springs. The estimated pseudo-ranges are calculated using ephemerides which have been calculated in the GPS system's own time base, which differs from the first by an unknown bias ΔT (since no-one has direct access to the Colorado Springs time). The pseudo-ranges measured and estimated will therefore all be systematically different by $c \cdot \Delta T$. This difference shall be included in the corrections to be made to the pseudo-ranges of each of the satellites and transmitted as such to the users. Since this pseudo-distance bias is common to all of the satellites, it will be perceived as a clock bias between the receivers' time reference and that of the GPS time. This will in no way trouble the pilot's navigation as he does not need this bias in order to navigate. The estimated orbitographies of the satellites can therefore be established in a time base which is not necessarily the Colorado Springs GPS time, but which must however have stability characteristics which are at least as good as those of the GPS time itself (available from the GPS satellites).

Synchronizations between stations shall be obtained by means independent of the GPS and shall not engender too great a complexity (and hence a cost overrun) in the system.

2- DESCRIPTION OF THE CE-GPS EXPERIMENT

2.1- The Need for an Experiment

Different technical solutions have been proposed to achieve the functions to be provided by the CE-GPS. Some of these solutions have come directly from other achievements and have been transposed into the CE-GPS environment, while others have been entirely created in terms of the systems' requirements.

In both cases the technical performances to be implemented are poorly known in the CE-GPS context. Studies have been undertaken but have not enabled clear conclusions to be made concerning error estimations for certain items. Whereas it is essential to know the performances and limits of possible solutions ahead of time as they will have a direct effect on the capacities of the future system.

These reflections have led to the creation of an exploratory programme which will enable testing and confirming the feasibility of the proposed solutions in an environment which is as close as possible to that of the future system.

2.2- Experimental Objectives

The main objectives of the CE-GPS experiment are :

- to confirm the possibility of recreating, by servo-control from the ground, a signal with characteristics similar to that of a GPS signal, as output from a transponder on a geostationary satellite,
- to demonstrate the feasibility of synchronizing a virtual clock on board the same satellite with GPS time.
- to demonstrate the capability of GPS receivers to receive and process these signals transmitted by the geostationary satellite,
- to evaluate the UERE (User Equivalent Range Error) wrt the geostationary satellite and its composition (error in restituted time at Monitoring stations, servo-control loop error, error in restitution of ephemerides, etc.),
- to synchronize with great accuracy two and then three stations which may be very far away from each other.

2.3- Description of the experimental system

2.3.1- General system architecture

The experimental system set up by CNES consists (figure 3), for the space segment of:

- the INMARSAT IIF2 geostationary satellite of which part of a transponder in the payload has been made available, free of charge, to CNES for the duration of the experiment,
- the GPS satellites in the NAVSTAR constellation,

and for the ground segment:

- of a network of CE-GPS ground station models (STA) whose development has been entrusted to the IN-SNEC,
- of a computer processing station for collected data (STE) located at the Toulouse Space Centre, and whose software development has been entrusted to the DATAID-Eurosoft company.

The CNES experiment has been divided into three progressive stages:

- a first stage in April 1993 with a network of two stations both located in Caen on the IN-SNEC site which will enable system running-in and validation of time transfer techniques (possibility of direct measurement of the shift between the clocks of two stations),
- a second stage from May to July 1993 with a station installed on the CNES site in Toulouse and another installed on the LPTF site in Paris,
- a third stage from September to December 1993 with a nominal network of three stations located in Toulouse, Kourou (Guiana) and Hartebeeshoek (South Africa).

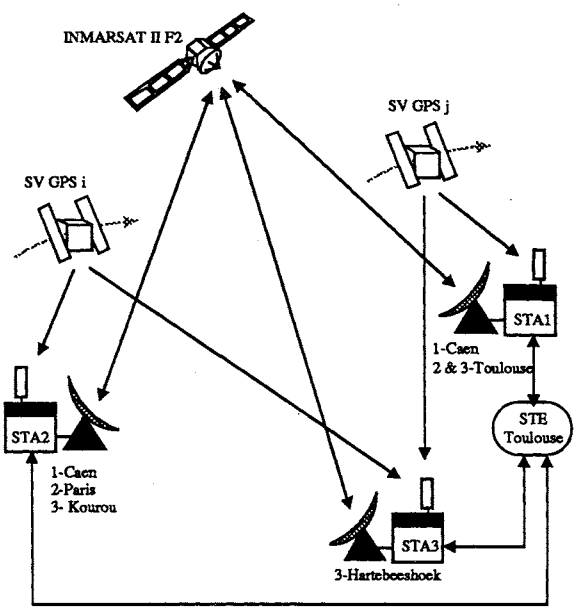


Figure 3

2.3.2- Ground stations physical architecture

All of the experimental stations have been designed with identical physical architecture as shown in figure 4.

The Signal Processing Drawer mostly provides the following functions:

* generation of two GPS type signals (CE-GPS signal) with L1 frequency:
  - a first signal whose code phase is controlled by an external synthesizer with phase continuity (Direct Digital Synthesizer - DDS1); this signal created by each of the stations is transmitted towards the geostationary satellite following transposition and amplification in C band; the Gold code assigned to this signal (chosen from the GPS system's reserved codes) identifies the station from which it was transmitted in the same way as a C/A code identifies a GPS satellite.
  - a second signal whose code phase is synchronized on the external 1 pps impulse delivered by the cæsium clock; this signal is switched towards the two NR 106 GPS receivers thus enabling on the one hand, relative synchronization, and on the other hand,

linking their internal clocks to the cæsium clock which is the station's time reference,

* conversion from L1 to C-band frequency of the GPS type signal to be transmitted towards the geostationary satellite; the uplink frequency is controlled by an external synthesizer with phase continuity (DDS2),

* conversion of the frequency of CE-GPS signals relayed in L-band by the INMARSAT II satellite into the L1 band used for GPS receiver operation (the fixed transposition frequency is derived from the 10 MHz reference),

* switching of L1 band signals (CE-GPS and GPS) towards the NR106 GPS receivers; the switches are remotely controlled by the station's main computer and enable all possible combinations of signals coming from simulators and antennae (L band and GPS) towards the NR106 receivers (cf. figure 5).

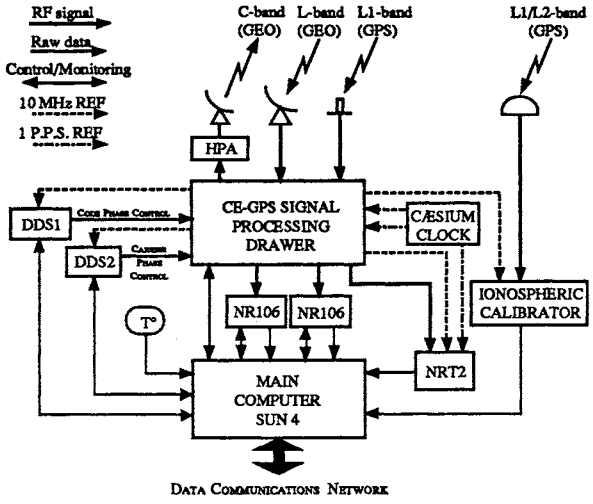


Figure 4

The GPS SERCEL NR106 receivers are 10 channel navigation receivers whose software has been slightly modified to accept the CE-GPS signals generated by the experimental stations. These CE-GPS signals only differ from the NAVSTAR-GPS signals in the content of their message and their acquisition characteristics (signal coming from a geostationary satellite and hence having a greater downlink propagation time and a smaller Doppler effect than for a NAVSTAR-GPS satellite).

The GPS SERCEL NRT2 receiver has a PCB for dating external events (time-oriented receiver) and four GPS channel PCBs; it is used in particular to determine the time deviations between the Cæsium reference time (1 pps) and the GPS time restituted from a satellite in the NAVSTAR constellation.

The TEC-100-A-D ionospheric calibrator is a 10 channel codeless receiver. It is used to measure ionospheric delays along the axes of GPS satellites. These measurements will be subjected to special analysis in order to estimate the ionospheric delay along the axis of the geostationary satellite.

The principal computer (SUN4 work station) is responsible mainly for commanding the station equipment, controlling its operation, developing the CE-GPS message to be transmitted and collecting the data measured.

2.3.3 - Ground station Time features

In the field of Time/Frequency applications, each station offers the possibility (figure 5):

- of dating the transmission of the CE-GPS signal towards the geostationary satellite,
- of dating the reception of the same signal following transit via the geostationary satellite's payload,
- of dating reception of the CE-GPS signals transmitted by other stations and retransmitted by the geostationary satellite,
- of dating reception of the GPS signals coming from the NAVSTAR constellation,
- of linking all of these dates to the station's reference clock by means of the dating in the NR106 receivers of the GPS type signal synchronized with the physical pps of the Cæsium clock (in this way a navigation receiver is transformed into a time-oriented receiver)

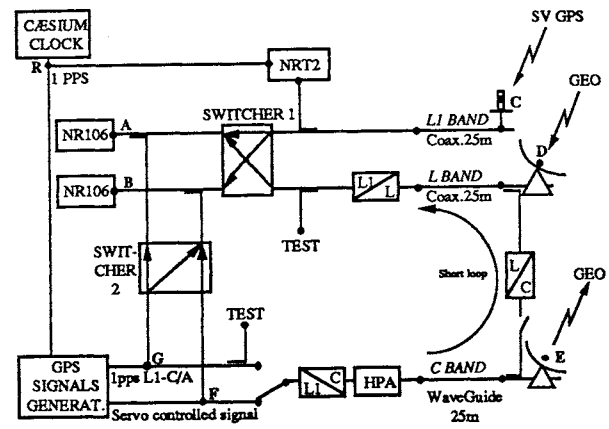


Figure 5

The transmitted signal may be either the servo-control signal or the signal synchronized on the station pps.

Given the accuracy objectives defined for the orbitography of the geostationary satellite and the time transfers, a system for calibrating internal station delays along the station-satellite round trip trajectory has also been installed.

2.4- Description of trial techniques in the Time/Frequencies field

2.4.1- Generating of a CE-GPS reference time

The CE-GPS system's reference time, called CE-GPS time, is the GPS time restituted from measurements taken by the network established to monitor the NAVSTAR constellation.

One of the experimental objectives is to test methods for restituting GPS time to the feeder link station.

This restitution will not be done in real time in the experimental framework but later using data recorded in the stations and collected in a data base managed by the STE.

The analysis will consist of developing and comparing different methods of restituting GPS time by means of information delivered by the NR106s. The NRT2, a time-oriented receiver whose use and performances are well-

known to the experimenters, will be used to analyse the results obtained.

It will be possible to compare the results obtained with the NRT2 time-oriented receiver and the NR106 receivers by means of the pseudo-SV GPS dating signal synchronized with the physical 1 pps signal from the Cæsium clock which will be permanently received by the two NR106s in each station.

2.4.2- Time and frequency servo-control techniques

2.4.2.1- Generating of a clock synchronized with CE-GPS time as output from a geostationary satellite

The experimental station playing the role of the connection station (STA1=Toulouse station) should be able to provide this function.

The objective is to demonstrate that it is possible to control the moment of transmission of a signal from this station in such a way that on leaving the geostationary satellite (in other words in the phase centre of the geostationary satellite's transmitting antenna) it be synchronized with the CE-GPS reference time available to this station.

Definition of reference time

Since one of the experimental objectives is precisely to define the best algorithms for determining CE-GPS time (cf. 2.4.1) this optimal reference time will not be available in real time in the experimental stations.

A temporary solution was therefore adopted, based on the Cæsium clock which offers a stability comparable with that of a GPS satellite.

Two reference times were defined:

- REF1: the time physically delivered in the form of pps by the station's atomic clock (Cæsium BNC pulse) which will have been synchronized to the GPS time restituted by the GPS NRT2 receiver when the system was initialized,

- REF2: the time delivered by the internal clock of a NR106 receiver which will have been synchronized, when the system was initialized, on the GPS time restituted from a block 1 satellite (this time synchronization is done by the software in the station's main computer).

It is important to emphasize that the time deviation (or desynchronization) between REF1 and REF2 for the same station is constant by construction since these two clocks are derived from the same 10 MHz reference (the 10 MHz distribution was designed to ensure excellent Group delay stability).

The system developed offers the possibility of using REF1 or REF2 for the servo-control signal.

Servo-control of the code phase from the feeder link station : principle

The servo-control of the code phase developed by the SNEC for the first stage of the experiment is based on the simplifying hypothesis of a symmetry between the up and downlinks between the station and the geostationary satellite.

It therefore consists of keeping the transmission moment symmetrical with the reception moment with respect to the reference time.

In reality the uplinks and downlinks are not altogether symmetrical and it is theoretically necessary to introduce a shift on the transmission moment in order to obtain the required accuracy for absolute restitution of time on board the geostationary satellite.

This asymmetry is mainly due :

- to the group delay in the station equipment (off-line bias estimated according to the equipment calibration results carried out during factory acceptance and during the collection phase),
- to the group delay in the GP,
- to the rotation of the earth (similar to the Sagnac effect, this results in a constant bias of 46 ns in the geostationary satellite's virtual clock which is generated by the Toulouse station),
- to the relative movements of the geostationary satellite with respect to the Earth (a negligeable effect),
- to the ionosphere whose effect on the round trip cannot be neglected (in fact it is inversely proportional to the square of the frequency used and may reach a few score of ns within an hour - the ionospheric delay). Special attention will be paid to the ionospheric effect and it will be evaluated during analysis of the results, using the ionospheric delay measurements taken by the station's ionospheric calibrator.

These corrections will not be introduced in real time into the servo-control loop developed by the IN-SNEC, but will be calculated and analyzed later using information which will have been collected in the stations (the global error in restitution of the reference time leaving the geostationary satellite has been theoretically calculated for the two REF1 and REF2 options).

Servo-control of the code phase by the other stations

The objective of controlling the code by servo-mechanism in the other stations (STA2 or STA3) is to minimize the cross-correlation noise between codes interpreted by the GPS receivers (Ref. 7).

This objective is achieved in the station j (STAj;j≠i) by making the moments of reception of its own CE-GPS signal coincide with that coming from the station 1 which is playing the role of the feeder link station (STA1=Toulouse).

Since the downlink trajectories of the two codes are then absolutely identical, the servo-mechanism tends to synchronize, at output from the geostationary satellite, the transmission of the code generated by STAj (j≠i) with that generated by STA1.

The STA1 station which transmits continually towards the geostationary satellite may therefore considered from the point of view of the servo-control loops as a "master" station permanently broadcasting its reference time which will be used by the other "slave" stations STA2 or STA3 for automatic adjustment when they transmit.

Comment: each of the stations may be configured by software either as a feeder link station or as a monitoring station, which might be interesting for certain Time/Frequencies applications in which each of the stations would generate on board the geostationary satellite a clock synchronized with their atomic clock.

2.4.2.2 - Control of carrier frequency by servo-mechanism

The GPS satellites in the NAVSTAR configuration generate constructed signals whose consistency between the Gold

code phase and the carrier phase is perfect (this consistency is however slightly altered when crossing the ionosphere).

In the case of the Experiment, this consistency, provided by construction, in the output from the transmitting station's GPS signals generator, is lost on board the geostationary satellite (L1 to C and C to L-band frequency transpositions). The two aforementioned phases are therefore submitted to slippage in relation to each other.

In order not to affect receiver operation (Ref. 8), a servo-mechanism was developed to try and minimize this slippage and was installed in the experimental stations.

The code/carrier slippage used as a voltage error in the carrier's servo-loop is calculated using raw phase and code measurements made by the GPS receivers on the signal coming from the geostationary satellite. The computer then estimates the command to be applied to the DDS2 synthesizer which acts on the carrier in such a way as to reduce this slippage to a minimum.

In order to take into account the specificities of the geostationary satellite used for the CE-GPS experiment, two different quantifications of this slippage were defined:

- a conventional formula in L1 band: Fp - 154O Fc where Fp represents the carrier frequency and Fc the frequency of the code received by a GPS receiver,
- a new formula which may express the equivalent slippage in the L band: $(Fp-F_{TR}) - F_L / F_{C/A} \bullet Fc$ where F_{TR}, F_L and $F_{C/A}$ are constants representing respectively the transposition frequency on reception (L->L1), the frequency of the L band and the frequency of the C/A code.

The fixed transposition inserted into the reception chain means that it is not possible to ensure at one and the same time the code/carrier consistency in the phase centre of the geostationary satellite's transmitting antenna and on reception by the L1 band receiver.

It may be shown (Ref. 6) that the new approach to quantifying slippage in L band makes it possible to ensure code/carrier consistency for the satellite but with the new consistency ratio corresponding to the L band.

It is also seen that the equivalent slippage in the L band is nil in all of the stations (allowing for ionospheric effects and imperfections in the servo-control loops) which is again an analogy with the GPS satellites.

Each station may be configured by software into one or other of these two carrier servo-mechanism modes which will be compared when the results are analyzed.

2.4.3- Time transfers

The objective of the relative synchronization between the monitoring stations is to transfer the times with an accuracy of 10 ns (1 σ) following corrections, calibration and modelling.

Comment: time transfer is taken to mean the determination with an accuracy better than 10 ns of the time deviation between the clocks of two experimental stations.

Six synchronization methods will be evaluated during the experiment:

* 2 methods based on the GPS constellation:

Method 1: GPS common view
Method 2: GPS "global" view

* 4 methods using the GPS signals simulated on board the geostationary satellite:

Method 3: Synchronizing by means of servo-control loops
Method 4: "Two way" time transfer
Method 5: "One way + ranging" method
Method 6: Orbitography and relative synchronization

2.4.3.1- GPS common view

This is the conventional method which involves using a GPS satellite visible to two stations for which one wants to determine the desynchronization of the reference clocks.

The principle of the method is to determine the shift between the clock of this GPS satellite (Space Vehicle k - SVk) and the reference clock of this station and then to calculate the difference :

$$\Delta T_{H_i-H_j} = \Delta T_{H_i-SV_k} - \Delta T_{H_j-SV_k}$$

This method has the advantage of eliminating a certain number of biases, in particular the satellite clock bias (clock correction and GPS satellite group time).

2.4.3.2- GPS "global" view

This method was developed by the SERCEL company and has been described in its report on the study of the monitoring network for the European Complement to GPS, done in 1988 (Ref. 9). It has since been corrected and refined by the BIPM (Ref. 10).

The method consists of restituting, in each of the experimental stations, the best possible approximation of GPS time using all of the NAVSTAR constellation satellites which are visible for each stations (whether or not they are visible at the same time to the two stations with which we are concerned). Contrary to the previous method, it is therefore necessary to take into account the clock corrections broadcast in the GPS messages. The determination of the time deviation $\Delta T_{H_i-GPS_i}$ between the reference time of the i station and the restituted GPS time in this station is based on Kalman filtering techniques with bad measurements being rejected.

The desynchronization between the reference clocks of the i station and the j station would then be determined by:

$$\Delta T_{H_i-H_j} = \Delta T_{H_i-GPS_i} - \Delta T_{H_j-GPS_j}$$

which is consequently a generalization of the previous method.

A similar time deviation could also be calculated at the central station, which would group together the measurements corresponding to all of the CNES experimental stations. These time deviations would have a common clock as reference if the relative synchronization of the stations were carried out.

It is hoped to be able to reduce the influence of the SA ephemeride by means of such an averaging. The time deviation thus defined represents the reference time for an RGIC system (Ranging GPS Integrity Channel) such as the CE-GPS.

2.4.3.3- Servo-control loop synchronization method

This synchronization method is derived from the code phase servo-control loop techniques used in the experimental stations.

Each of the stations generates as output from the geostationary satellite a virtual clock identified by its Gold code. This clock is materialized by the moment of presence, $(t_{Geoi})_{H_i}$ of the particular code event which it has transmitted, in the phase centre of the geostationary satellite's transmitting antenna :

$$t_{Geoi} = \frac{t_0^i + t_i^i}{2} + \varepsilon_i$$

with t0i and tii respectively being the moments of transmission and reception of the code from the station i and εi being an error term (cf. 2.4.2.1).

When several experimental stations transmit, the servo-control loop principle binding the slave stations to the master station tends to synchronize the transmission moments of the codes from the 1 and j stations on board the geostationary satellite:

$$(t_{Geo1})_{H_j} = (t_{Geoj})_{H_j}$$

When the servo-control loops have converged and after regrouping the data in the central station, it is then possible to calculate the desynchronization between the clocks H1 and Hj of the stations 1 and j by means of the following formula:

$$\Delta T_{H_1\text{-}H_j} = (t_{Geo1})_{H_1} - (t_{Geoj})_{H_j} + \varepsilon_1^j$$

where ε1j represents an error term including:

- the estimation error of tGeo1 in H1,
- the estimation error of tGeoj in Hj,
- the servo-control loop error of tjj on t1j in the station j.

Comment: an alternative to this method consists of configuring the two experimental stations in feeder link station mode. Each station may then estimate, on the basis of its own measurements, the shift in its reference clock with respect to that of the other station.

2.4.3.4- "Two way" time transfer

This method has been used for several years now either in experiments or in operations as in the United States. It has been described in numerous articles and consists of quasi-simultaneous transmission of a signal (C/A code) from the station i towards the station j (and vice-versa) via the GP.

However the two channel method in the CE-GPS experiment is a variant on the conventional method to the extend that the moments of signal transmission are servo-controlled in each of the stations in order to compensate the uplink Doppler (servo-control of the code phase). Excluding servo-control loop errors, the CE-GPs signals of the transmitting stations are in fact synchronized and relayed at the same moment by the geostationary satellite, which reduces the dating errors due to the intercorrelation between the Gold codes in the GPS receivers.

It has been demonstrated by the time equations (Ref. 6) that it is always possible to obtain the desynchronization between the two stations without having to know the position of the geostationary satellite.

This variant on the two channel method will be tested and validated during the first stage of the Experiment in Caen (stations in the same place).

2.4.3.4- "One way + ranging" method

The "One way" method simply consists of sending a reference signal from station i to station j. Given the dates of departure and arrival of the signal, and the total delay of the signal from STAi to STAj, one may then deduce the desired time deviation $\Delta T_{H_i\text{-}H_j}$ between the two clocks of the two stations i and j.

In order to know the total delay from i to j, ranging measurements (round trip) are taken at the same moment. The ranging measurement is obtained in each station using the dating made by the GPS NR106 receivers of the moments of transmission and reception following the station-satellite round trip of its own CE-GPS signal.

The advantage of this method is that it is not necessary to know the Group delay of the GP which is eliminated in the time equations (Ref. 6).

2.4.3.6- Orbitography and relative synchronization

Two global methods of simultaneous restitution of the orbitography of the geostationary satellite and of the desynchronizations between stations will also be tested:

- the first method, implemented by ONERA, is based on the use of a Kalman type sequential filtering, hence well suited to real time problems,
- the second method, implemented by CNES, is based on a weighted least error squares method (ZOOM orbitography programme).

Once the geostationary orbit has been accurately determined, it will be possible to test the common view methods with the CE-GPS signals of the geostationary satellite. It will therefore be possible to restore later the desynchronizations of the CE-GPS signals between the receiving stations.

In the future system, when the geostationary satellite's ephemerides will be transmitted in the feeder link station's CE-GPS message, the scientific community will be able to use a permanent common view means of the GPS type without voluntary downgrading (no SA on the CE-GPS).

3- CONCLUSION

CE-GPS system will provide civil community with an independent means of GPS monitoring. It's also a first step towards a civil Global Navigation Satellite System. CE-GPS has been designed to meet civil aviations requirements, that are amongst the strictest in the civil GPS users' world. In order to reach this goal, new techniques in the front-line of progress have to be designed and developed.
An experimentation of the system has been set up to confirm some design hypotheses and to find optimized solutions. This test bed makes it possible to compare and assess the performances of several techniques belonging to time/frequency field and on which is built CE-GPS: servo-control loops, GPS time restitution and long basis time transfer. The results of this experimentation will be of great interest for many applications beyond the scope of GNSS.

ACKNOWLEDGEMENT

The authors wish to thank all the CNES management and technical staff and consequently the BIPM, LPTF, IN-SNEC, SERCEL and ONERA for their precious help and participation to the design and development of the CE-GPS experimental system.

REFERENCES

[1] OTAN - Technical characteristics of the NAVSTAR GPS- June 91.

[2] OACI: 10th Air Navigation Conference Report - Montreal, 5-20 Sept. 91 - AN-CONF/10.

[3] RTCA: "Minimum Operational Performance Standards (MOPS) for airborne supplemental navigation equipment using GPS" - Do28

[4] CNES: J.-M. Durand "Pre-operational European Complement to GPS (CE-GPS) - Preliminary mission specifications" - Ref. AP/RC/IS/MF 92/162.

[5] CNES: J.-M. Durand "CE-GPS-Spécifications préliminaires de temps/fréquence"
Ref. AP/RC/IS/MF 91/718 Ed.1 of 29 June 1992.

[6] CNES- J. Barbier, P. Brun, M. Brunet, J.L. Issler, N. Suard, T. Trémas, C. Vallorge, C. Yven
"Définition détaillée de l'Expérimentation CE-GPS"
Ref.: CNES/442/SDS/SM/ES - 28 October 1992

[7] IN-SNEC: "Expérimentation CE-GPS: Compatibilité des émissions 1 pps au travers de la charge utile geéostationnaire"
Ref.: SNEC/1274/DPT/0014 - 17 June 1992

[8] CNES: J.-L. Issler, J.-P. Lefevre "Mesurement of the RGIC Signal Code/Carrier Coherency with a GPS Receiver"
ION GPS 92 - Proceedings

[9] SERCEL: "Complément Géostationnaire - Etude du réseau de surveillance"
Contrat 832/CNES/87/4879/00 - 1988

[10] BIPM: Mrs C. Thomas "Real-Time Restitution of GPS Time through a Kalman Estimation"

RIRT'S CONTRIBUTION TO THE SOLUTION OF METROLOGICAL SUPPORT PROBLEM WITH REGARD TO NATIVE RADIO SYSTEMS AND AIDS INCLUDING GLONASS

Prof. Yuri G. Gouzhva
Dr. Pyotr P. Bogdanov
Irina G. Pushkina
Prof. Arvid G. Gevorkyan
Dr. Sergey N. Klyushnikov

Russian Institute of Radionavigation and Time (RIRT)
2, Rastrelli square, St.-Petersburg, 193124, Russian Federation

Abstract The gist of the problem of metrological support is to provide the users with dimensions of time/frequency units having the prescribed accuracy level. It is the Russian State Time and Frequency Service that gives a technical basis to solve this problem. RIRT's role is to realize the prescribed accuracy of frequency/time measurements at the second hierarchic level of metrological service for radio systems and aids. While being responsible for frequency/time support with respect to various radio systems, RIRT deals with design of quartz oscillators and atomic frequency standards (AFS). Using these devices, RIRT creates time standards, secondary reference complexes, reference standards and operational metrological aids for users, as well as aids to transmit frequency/time data (FTD).

Introduction

The main elements of frequency/time support for various radio systems and aids are quartz oscillators and AFS, as well as time standards using these devices. There is no way to provide a proper functioning of radio systems without recourse to metrological monitoring of parameters of frequency standards and clocks. RIRT's equipment and aids for FTD transfer to users cover the range of metrological accuracies from 10^{-10} to 10^{-14}.

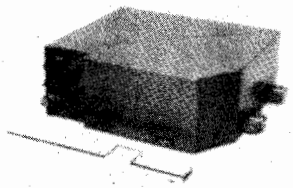
Within the framework of the Russian State Time and Frequency Service RIRT is responsible for sequential monitoring of metrological parameters of frequency standards through all the stages of its development and application for various hardware complexes and radio systems as a whole. This monitoring is performed using RIRT's operational reference, monitoring/measurement equipment and common time system's aids including GLONASS aids, thus enabling to transfer FTD to users at the prescribed accuracy level. In so doing, RIRT realizes a metrological support for various radio systems, among which are navigation and timing ones.

Metrological support for frequency/time measurements of radio systems and aids

One of the most important tasks for metrological service is to transfer the dimensions of time/frequency units (i.e. 1s and 1 Hz) from Russian State Reference to users. If a user has an equipment that enables to reproduce the dimensions of units with the relevant accuracy (corresponding with that of operational aid, reference standard or reference), this problem is solved through measurement of actual value of oscillator output signal frequency by comparing it with that of operational aid, reference standard or reference. The parameters to be measured are fractional fluctuation of frequency actual value and frequency instability of the investigated oscillator. The unification of measurements (i.e. obtaining measurement results in unified system of physical parameters) is largely aided by similar ways of measurement performing using unified hardware support.

A frequency/time support for radio systems and aids is realized on the basis of metrological monitoring of frequency standards (FS) and time standards with those FS using secondary reference complexes, operational references, reference standards and user operational metrological aids, as well as aids for FTD transfer.

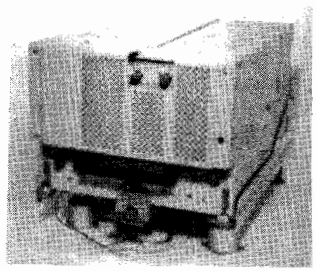
The metrological monitoring of frequency/time standards is performed in a step-by-step manner, from the very instant of its release from works (Fig. 1) up to its use in radio systems and hardware complexes as a whole, on the basis of RIRT's operational reference, monitoring/measurement equipment, as well as using such common time system aids as TV synchronization equipment, meteor-burst synchronization equipment, transportable clock. Used also are standard time signals emitted by VLF-, LF- and HF- radio stations of various radio systems (see Fig. 2).



AMETHYST - Rb AFS for space vechiles
$\sigma = \pm 2 \cdot 10^{-11}$ per day



GEM - Cs AFS for GLONASS board
$\sigma = \pm 5 \cdot 10^{-13}$ per day



ETALON - Rb AFS for aircraft
$\sigma = \pm 2 \cdot 10^{-11}$ per day



SAPPHIRE - Cs AFS, terrestrial
$\sigma = \pm 3 \cdot 10^{-13}$ per day

Fig. 1.

Examples of AFS produced by RIRT experimental works

The most accurate means of synchronization on a global scale is GLONASS, that is the global satellite navigation system which enables to perform clock synchronization with an error no more than 20 ns via user equipment (in particular, ASN-16) operating in differential mode. As a result, an oscillator frequency actual value may be determined with an error no more than 2.10^{-13}

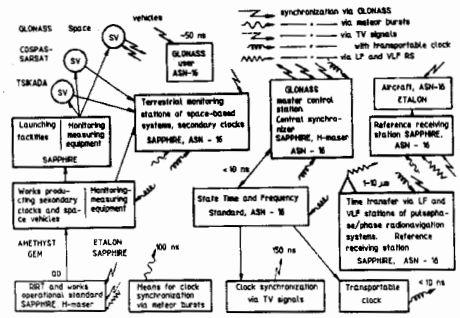


Fig. 2. Metrological monitoring of frequency/time standards

A metrological reliability of clock synchronization is realized by using various synchronization aids with relevant redundancy. The calibration of clock proper metrological parameters is effected in all the cases by comparing time data with the Russian State Reference in the end, giving an error from a few microseconds to nanoseconds depending on used synchronization means.

In order to provide an unification of measurements on the basis of unified hardware support with the metrological accuracy of order of 10^{-13} at the second hierarchic level of metrological support in Russia, SAPPHIRE cesium beam frequency standard is widely used as a main frequency keeper, among with other FS, in references, reference standards and users' operational aids of hardware complexes of radio systems.

The most stringent requirements on metrological parameters and time keeping reliability are imposed on secondary reference time/frequency complexes. The metrological reliability of time keeping is provided due to manyfold redundancy of references' main systems. Loss in time keeping is intolerable. An example is GLONASS central synchronizer (CS) as well as RIRT's operational reference with metrological parameters at the level of $10^{-13} - 10^{-14}$ using SAPPHIRE cesium beam FS and hydrogen maser (HM). The error of independent time keeping does not exceed 30 ns for 1 month. RIRT's operational reference is certified by the State

Metrological Service. The daily instability is $3 \cdot 10^{-14}$ (rms), the accuracy error compared to the State reference is $\pm 10^{-13}$.

The metrological monitoring of atomic frequency/time standards entering into the set of users' operational metrological aids may be performed using the reference frequency/time standard which includes two SAPPHIRE sets and measurement equipment with the instrument error of frequency measurement no more than $1 \cdot 10^{-13}$ for 1 s and $1 \cdot 10^{-15}$ for 1 day. Such a monitoring takes place in the course of production at works where hardware complexes (for example, space vehicles) are produced, as well as at technological complexes (an example is a launching complex).

In order to realize a stationary/mobile reference frequency/time standard for transfer of frequency/time units to onboard operational metrological aids (for example, to airborne/shipborne aids), the user time equipment is used. Such an equipment using an ensemble of SAPPHIRE sets comprises various reference receiving station including terrestrial stationary and mobile platforms as well as shipborne ones. These stations are integral parts of hardware complexes of radio systems and provide its frequency/time support. There are various ways to realize operational metrological aids entering into hardware sets of radio systems. For example, the spaceborne clock may be realized using AMETHYST or GEM FS, the similar way for the airborne clock is to use ETALON FS.

The sequential metrological monitoring of atomic FS/clocks, from the very instant of its release from works up to its operation in the hardware complexes and radio systems as a whole, may be clearly demonstrated using GEM FS as an example. This FS is a part of GLONASS spaceborne clock. While releasing from works, GEM is certified via measurement its metrological parameters with respect to RIRT's operational reference, being a part of spaceborne clock equipment. The next stage of metrological control of parameters of these time standards is performed at works where space vehicles are produced. Here these time standards as a part of space vehicle equipment are calibrated relating to monitoring/measurement equipment. The final stage of terrestrial monitoring of metrological parameters of time standards is its calibration, while being a part of spaceborne equipment at the position before launching. The further monitoring of metrological parameters of spaceborne clocks is realized within the framework of GLONASS synchronization system by comparing space vehicle's time with that of central synchronizer. A widely used means for time synchronization of time/frequency standards of various users, as well as for metrological monitoring of operational standards, reference standards and references, is an equipment for time comparison via GLONASS signals using single-channel one-frequency ASN-16 receiver interfaced with PC (Fig.3).

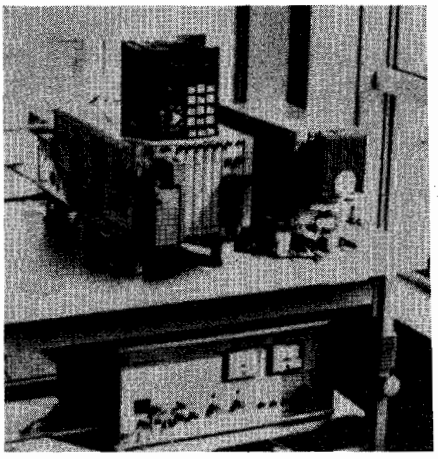


Fig.3. ASN-16 GLONASS receiver

The results of equipment tests conducted in RIRT from 1991 to 1992 have shown that, while operating in a mode of direct synchronization with respect to UTC (SU), the time comparison error is of 20 ns (1σ). In a mode of mutual synchronization this error is of 10 ns (1σ). This enables to realize a metrological monitoring of operational standards, reference standards and references with an error not exceeding $2,3 \cdot 10^{-13}$ and $1,16 \cdot 10^{-13}$ correspondingly. The measurement scheme is given in Fig.4.

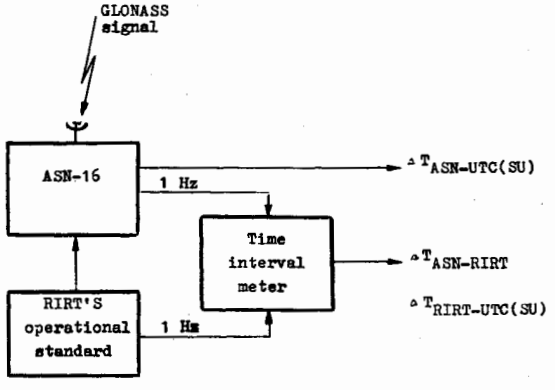


Fig.4. Time comparison between RIRT's Time and UTC (SU)

The similar results took place for tests of ASN-16 in Braunschweig in September of 1991. Error of UTC (PTB) - UTC (SU) comparisons via GLONASS signals was 25 ns in the mode of direct synchronization.

Conclusion

Owing to utilization of reference complexes, reference standards and operational metrological aids, using in the most of cases SAPPHIRE cesium FS and entering into the hardware sets of various radio systems with various means to transfer frequency/time data, the metrological monitoring at the level of 10^{-13} is provided through all the stages and in all the cases of its usage in radio systems. Provided also is the prescribed accuracy of frequency/time measurements at the second hierarchic level of metrological support for various radio systems and aids in Russia.

TIME TRANSFER BETWEEN USA AND FRANCE THROUGH LASSO

A preliminary analysis

C. Veillet, P. Fridelance
Observatoire de la Côte d'Azur
Av. Copernic, 06130 GRASSE, France

ABSTRACT

After one year of LASSO observations at McDonald Observatory (Texas) and Observatoire de la Côte d'Azur (France), an extensive set of data has been acquired. This paper will focus on the last sessions performed in January 1993, when good GPS measurements were also available. The precision of the clock comparison is found to be better than 100ps, even with few data, thanks to a processing using every information available.

Keywords : Laser, LASSO, Time transfer

1. INTRODUCTION

LASSO (LAser Synchronisation from Stationary Orbit), proposed in 1972 by M. Lefèvre and J. Gaignebet (Ref. 1), is an experiment for linking remote clocks by laser light. The LASSO flying package is made of a laser detector linked to an event timer able to record the arrival time of a laser pulse onboard on a time scale provided by a local Ultra Stable Oscillator. The host spacecraft carries also a retroreflector array which sends back part of the laser pulse to the ground. Transmitting/receiving stations on the Earth record the start and return time of the laser pulse sent to LASSO. As we will see later, such measurements allow a comparison between the station clocks. The first LASSO package, installed on a SIRIO satellite, ended its carrier in the Atlantic after the failure of the 5th Ariane launching. Thanks to the LASSO community, a second LASSO was built and mounted on Meteosat 3/P2, a meteorological satellite launched in mid-1988. The various moves of Met3 between 0 and 50°W on the geosynchronous orbit made the organization of the experiment quite uneasy. The first results were obtained in 1989 with the TUG laser station (Graz, Austria) and the Lunar Laser Ranging (LLR) facility at the Observatoire de la Côte d'Azur (OCA, Grasse, France) (Ref. 2).

With the second move of Met3 toward 50°W in 1992, a new campaign started between an LLR station (MLRS) located at McDonald Observatory (Texas, USA), and the OCA LLR equipment in France.

2. THE OBSERVATIONS

Successful common sessions were organized, and the first results were obtained in April 1992. Since that time, LASSO data were acquired on a regular basis, weather permitting. Ten days campaigns centered on New Moon were scheduled every lunar month, giving some common sessions. The need of a clear weather at both sites brought to a short number of common nights shown on Fig. 1.

The stations involved in this transatlantic phase are described in a previous paper (Ref. 3). Their main parameters are briefly given here :

	MLRS	OCA
Telescope diameter	75cm	150cm
Laser pulse length	200ps	350ps
Laser pulse energy	140mJ	350mJ
Met3 elevation	26°	14°

The laser pulse length is a very important parameter for both laser ranging and time transfer through laser light. When the target to be ranged is far (as for Met3), only a very small amount of the light transmitted comes back to the station, generally one photo-electron (a photon detected by the photomultiplier or the photodiode used as a receiver behind the telescope). As nobody knows where does this photon come from (beginning or end of the transmitted laser pulse), here is the fundamental uncertainty of the laser ranging technique. The shorter is the pulse, the better should be the measurement. As it will be seen in the following figures, the jitter of the clock offset determination will be lower with the MLRS data than with OCA where the pulse length is nearly twice as long as at MLRS.

3. THE DATA PROCESSING

Two stations are firing to Met3, receiving echoes and generating detection times on LASSO. The original processing of the LASSO data comes from the following equation:

$$(1) \quad \text{offset} = (s' + r') / 2 + c' - [\,(s + r) / 2 + c\,] - (l' - l)$$

where s is the start time of a laser pulse from the first station, r its return time, and l its detection time onboard LASSO. c is a correction term taking into account the fact the the arrival time of the light on the satellite is not exactly in the middle of the [s , r] interval. s', r', l' and c' are the same for a pulse transmitted and received by the second station.

Among these measurements, only l and l' are measured on the LASSO time scale, provided by the USO. If ($l' - l$) is short

enough, the uncertainty on the knowledge of the USO will not be important. The observation procedure is then established: insure that both the stations are firing so that the arrival times of their laser pulses are close enough to make ($l' - l$) of the order of a few milliseconds. The USO vanishes from the equation (1) and the clocks are linked !

Unfortunately, only one laser shot over 20 to 30 provides a return time, and the same proportion provides a detection time onboard of LASSO. Only a very low fraction of the shots provides what we call a *triplet*, i.e. a start time , a return time and a detection time by LASSO for the same laser pulse. The constraint of close arrival times on LASSO makes the probability very very small to get a triplet from each station with a small ($l' - l$). The use of equation (1) is no longer possible with the current statistics observed with the OCA and MLRS stations !

As it is not possible to rule out the USO from the data processing through a small ($l' - l$), we have to monitor its behaviour from the two stations involved. For that purpose, we can use all the triplets from a given station, and compute the quantity $(s + r) / 2 + c - l$ (2)
which is the offset between the ground clock and the flying USO. If observed over some time, we can determine the offset and rate of the USO as seen from the station clock. The result of such a determination is shown on Fig. 2. The larger the number of triplets used, the better the USO offset and rate determinations. Unfortunately, there are often only a very few triplets for a 10-minutes time span, and the monitoring of the USO becomes very weak. However, we have a good number of echoes, and also a good number of detection times on LASSO, but they do no match on the same laser pulse !

This information (return times and detection times) can be used in the following way : the echoes obtained over the considered time (generally ten minutes) allow us to determine with a quite good accuracy a model of the station-to-satellite round trip time.We use this model for computing a return time of the pulses transmitted and detected on LASSO, but without an echo received back at the station. Through this process, we obtain what we call now a *reconstructed triplet*, made of all the information provided by the pulses with returns, and of its own start and detection times. These reconstructed triplets, used as the real ones, provide the same information as previously, i.e. the offset and rate of the LASSO USO, as shown on Fig. 3.

3. THE JANUARY 1993 DATA

Three nights provided LASSO data simultaneaously with the acquisition of good GPS measurements. Unfortunately, these nights were not very clear at OCA. Due to the low elevation of Met3 as seen from Grasse (14° above the horizon), the amount of echoes and LASSO detection times has been lower than usual. However, the observations made at both sites are sufficient for a precise time transfer between Texas and Grasse.

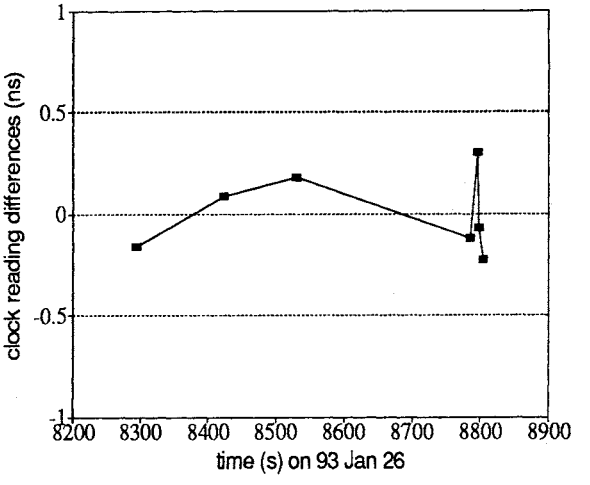


Figure 2 : An example of the LASSO clock monitoring from the ground. Here are used only real triplets obtained with the MLRS for ten minutes. Clock offset and rate have been determined with only 7 points. The uncertainty is however around 100ps on the offset.

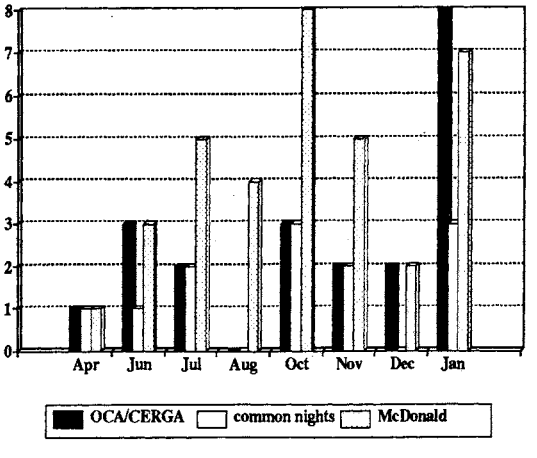


Figure 1 : The number of nights with LASSO data for the lunar months with at least one LASSO observation. Ten nights were scheduled each month. Due to poor weather conditions, only 12 nights provided common sessions, among which 3 in January with GPS data.

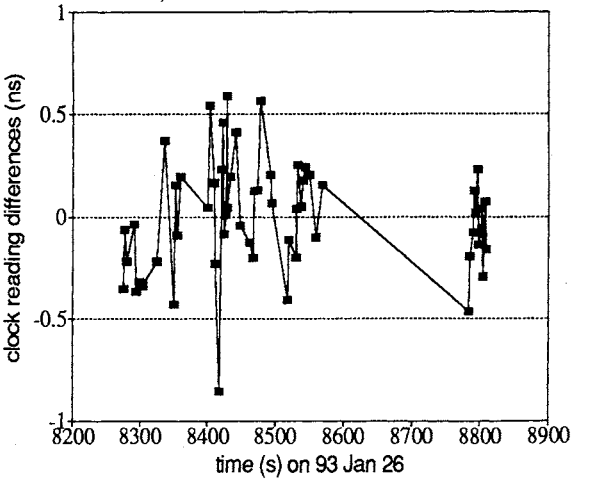


Figure 3 : For the same session as on Fig. 2, the LASSO USO is monitored from MLRS. But now, all the detection times recorded onboard are used, with the corresponding return times computed from all the echoes received for the ten minutes of laser ranging.

Table 1 and 2 summarize the results of the preliminary processing. The number of measurements are presented in the first columns in Table 2. The low number of echoes received at CERGA is clearly seen.The formal uncertainties given come from the fits made at various steps of the data analysis. They are confirmed at the end by the comparison of the determinations made on the same night, but independantly. In fact, every final clock offset is derived from a session without any relation to the other data.

The results presented in these tables are provisional for the following reasons :
- the corrections (the term *c* in the equation (1))have been ignored, and thus considered as a constant to be added later to the offsets found. They are not constant at the 10ps level.
- the variations of the various transit times in the electronics at OCA have been ignored. They have to be taken into account at the 50ps level.
- the data should be corrected for the rotation of Met3. Ignoring this spinning adds some jitter to the measurements, especially in the reconstruction process.
- the stations are not calibrated. The calibration will allow us to achieve an absolute time transfer, which could be compared to GPS, not only with an offset to be removed, but in an absolute way.

A comparison with GPS is possible however, at least at the 100ps level, as the errors implied by the preliminary processing can be considered as constant. The results of this comparison are presented in another paper (Ref. 4).

4. CONCLUSION

Waiting for these effects and corrections to be included in the data processing, we can conclude on very exciting statements:

- The LASSO concept, joined to the data processing using *all the information* through the reconstruction process, is able to monitor a flying clock with an uncertainty better than 50ps with a short observing time.
- Two stations working at the same time, but with no other constraint on their firing times, can link their clocks through LASSO with a 50ps precision. The accuracy of this link will depend on how well a station can measure (or stabilize) the transit times within its equipment (including its link to themaster clock).

Considering that the LASSO equipment is more thant ten years old, a new generation of laser link based on the same concept could reached a precision better than 10ps using the current data processing. A proposal for such a new experiment (TROLL, Time Relay in Orbit by Laser Light) is being

	date of offset determination of the LASSO clock (s)		LASSO clock offset (modulo 1ms, in ns)		LASSO clock offset uncertainty (ns)		LASSO clock rate (ps/s)		LASSO clock rate uncertainty (ps/s)	
day	OCA	MLRS	OCA	MLRS	OCA	MLRS	OCA	MLRS	OCA	MLRS
1993/01/21	8031.1	8197.7	140422.23	132092,09	0.055	0.050	8425.46	8425.72	0.11	0.38
1993/01/22	5965.1	6264.6	74303.55	66834,30	0.058	0.041	8416.03	8415.53	0.26	0.22
1993/01/22	6953.7	6990.5	82626.61	72946,17	0.043	0.049	8419.22	8419.68	0.24	0.22
1993/01/26	6860.3	6614.4	819215.88	810861,66	0.084	0.037	7320.97	7317.66	0.82	0.40
1993/01/26	8295.7	8267.4	831525.14	821161,38	0.5	0.046	7325.34	7324.65	2.3	0.17

Table 1 : The LASSO clock seen from MLRS and OCA, as resulting from the preliminary processing of the three nights (5 independant determinations) in January 1993, with also good GPS data. The uncertainties are formal, computed from the fit of the clock reading differences between LASSO and the laser stations.

day	number of triplets OCA true	OCA rec.	MLRS true	MLRS rec.	returns OCA	returns MLRS	epoch (s)	OCA-MLRS clock offset ns	uncertainty ns	OCA-MLRS clock rate ps/s	uncertainty ps/s
1993/01/21	2	67	2	35	12	66	8114.4	9933,56	0.08	0.26	0.40
1993/01/22	3	43	9	44	39	74	6114.9	9990,05	0.09	-0.50	0.34
1993/01/22	2	49	6	45	24	32	6972.1	9990,58	0.07	0.46	0.32
1993/01/26	0	11	16	84	4	82	6721.4	10154,26	0.13	-2.3	0.88
1993/01/26	0	2	18	108	19	98	8279.4	10155,96	0.50	-0.55	2.3

Table 2 : The MLRS and OCA clocks compared through LASSO, as resulting from the figures given in Table 1. The number of triplets is given in the first four columns after the date (*rec.* for reconstructed), together with the number of returns used for computing the return times in the reconstruction process. The uncertainties are formal. Each determination is completely independant on the others. The good consistency of the clock differences can be seen on the two nights where two determinations were possible.

prepared for an implementation on satellites carrying clocks (GLONASS, GPS, flying H-maser experiments could be equipped with) for both clock monitoring and scientific applications.

Acknowledgements : The authors are grateful to those who observed at the McDonald Laser Ranging Station (Texas) and at the OCA Lunar Laser Ranging facility (Grasse, France), and who made the laser data available to us. We thank also CNES, INSU and ESA/ESOC for their support.

REFERENCES

1. M. Lefèvre, J. Gaignebet. Utilisation scientifique possible de longues bases laser. *Space Research XIII* Akademie Verlag, Berlin 1973.

2. C. Veillet, D. Féraudy, J.M. Torre, J.F. Mangin, P. Grudler, F.Baumont, J. Gaignebet, J.L. Hatat, W. Hanson, A. Clements, J. Jespersen, M. Lombardi, D. Kirchner, G. Kirchner, H. Ressler, S. Fassl, W.J. Klepczynski, P.J. Wheeler, W.Powell, A. Davis, M. Granveaud, P. Uhrich. LASSO, Two-Way and GPS time comparisons : a (very) preliminary status report. *Proc. 22nd PTTI meeting,* pp. 575-582, 1990.

3. C. Veillet, P. Fridelance, D. Féraudy, Y. Boudon, P.J. Shelus, R.L. Ricklefs, J.R. Wiant. LASSO Observations at McDonald (Texas, USA) and OCA/CERGA (Grasse, France). *Proc 24th PTTI meeting,* 1992, in press.

4. F. Baumont, P. Fridelance, P. Grudler, C. Veillet, J.R. Wiant, W. Lewandowski, G. Petit. Preliminary report on the comparison of LASSO and GPS time transfers, *in these proceedings.*

PRELIMINARY REPORT ON THE COMPARISON OF LASSO AND GPS TIME TRANSFERS

F. Baumont, P. Fridelance, P. Grüdler, C. Veillet
Observatoire de la Côte d'Azur
Av. Copernic, 06130 GRASSE, France

J. Wiant
McDonald Observatory
Fort Davis, Texas, USA

W. Lewandowski, G. Petit
Bureau International des Poids et Mesures
Pavillon de Breteuil, 92312 Sèvres Cedex, France

ABSTRACT

The LASSO is a technique which should allow the comparison of remote atomic clocks with sub-nanosecond precision and accuracy. The first successful time transfer using LASSO has been carried out between the Observatoire de la Côte d'Azur in France and the McDonald Observatory in Texas, United States. This paper presents a preliminary comparison of LASSO time transfer with GPS common-view time transfer.

Keywords: Laser, LASSO, GPS, Time transfer.

1. INTRODUCTION

The idea of using satellite laser techniques for the comparison of remote atomic clocks, known as LASSO (Laser Synchronization from Stationary Orbit), was conceived by M. Lefebvre and J. Gaignebet in 1972. The first successful tests were accomplished in 1989 (Ref. 1). The first LASSO experimental link was established in 1992 between Observatoire de la Côte d'Azur (OCA) in Grasse, France, and McDonald Laser Ranging Station (MLRS) located at the McDonald Observatory in Fort Davis, Texas (Refs. 2,3). LASSO involves very simple physics, light travelling through space, and should easily allow sub-nanosecond time transfer.

At present, the use of the GPS common-view technique permits the comparison of remote atomic clocks at their full level of performance. We already know, however, that over the next few years new and better clocks will become available and that the GPS, as practiced now, will not then provide a system of adequate resolution. The estimated accuracy of intercontinental GPS time comparisons is of several nanoseconds (Refs. 4,5), but, up to now, this has not been verified by an independent method of comparable or better accuracy. In this context LASSO appears to be an outstanding tool for the evaluation and calibration of the GPS time transfer and other time transfer techniques.

During an experiment covering the period from 8 December 1992 to 28 January 28 1993, two remote atomic clocks, separated by about 8000 km, were compared by independent space links, LASSO and GPS common-view. On each site a laser and a GPS receiver were connected to a single clock, a commercial cesium standard (Fig. 1).

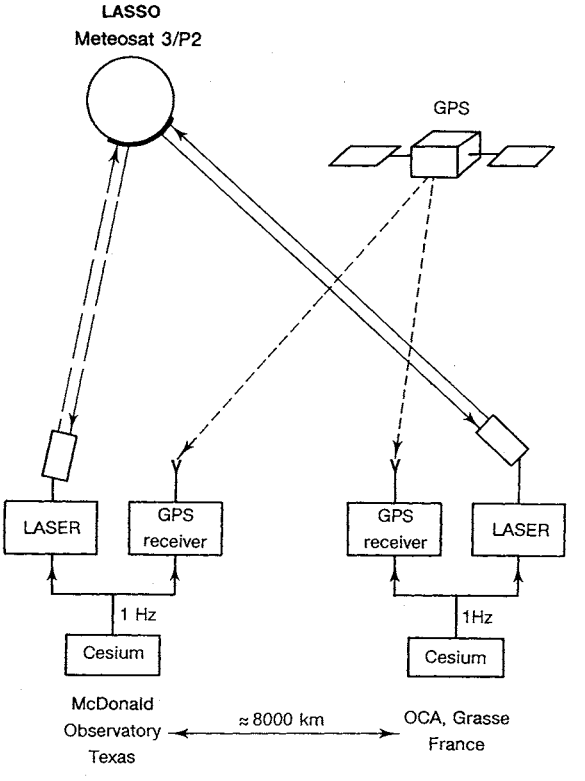


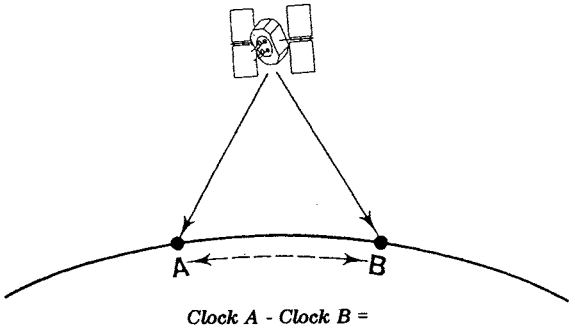
Figure 1. *The experiment configuration.*

The principal difficulty of this experiment was to obtain weather conditions under which LASSO measurements could be carried out. Unfortunately, throughout the experiment the sky was often covered and only in last week were five LASSO measurements accomplished. The experiment has now ended since the geostationary satellite Meteosat 3/P2, on which LASSO package is placed, has been moved to a location not visible from the OCA.

This paper, after giving a brief description of the GPS common-view and LASSO time transfers, reports the preliminary results of the comparison of these two methods.

2. GPS COMMON-VIEW LINK

In common-view time transfer, two remote stations receive signals from the same satellite at the same time and exchange the data to compare their clocks (Fig. 2). The main advantage of this method, introduced in 1980 for GPS (Ref. 6), is that satellite clock error contributes nothing (satellite time disappears in the difference). Also over distances of up to a few thousands of kilometres the impact of other errors, such as poor estimation of ionospheric delay or broadcast ephemerides, is diminished. During our experiment, time transfer were carried out over a distance of 8000 km. In this case the modelling of the ionospheric delay applied by GPS time receivers should be replaced by measurements of ionosphere and broadcast ephemerides should be replaced by post-processed precise ephemerides. Both improvements will be incorporated in the next report on this experiment.



Clock A - Clock B =
[Clock A - GPS time] - [Clock B - GPS time].

Figure 2. *GPS common-view time transfer.*

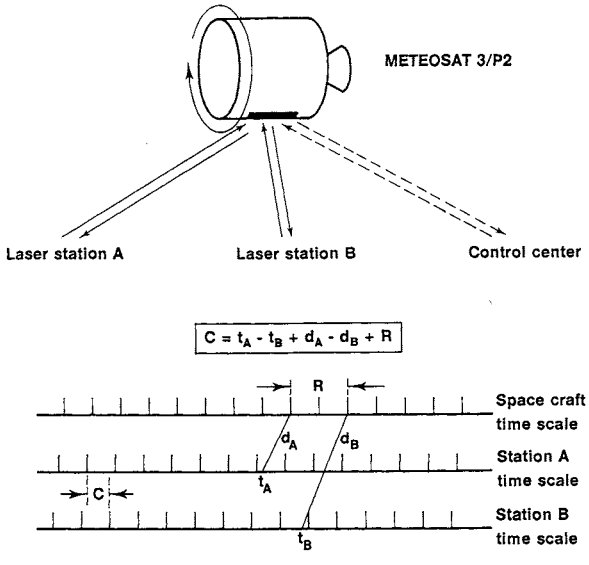


Figure 3. *LASSO time transfer.*

By using the same type of receiver at each site the consistency of the clock comparison is ameliorated as possible software errors are removed by the common-view approach. In the present study, GPS C/A Code time receivers were of the same type, AOA TTR5 at OCA and AOA TTR6 at McDonald.

The GPS TTR6 receiver, on loan from the BIPM, was sent back to France at the end of the experiment, where was compared with the OCA TTR5 receiver. This differential calibration has an estimated uncertainty of 2 ns.

In addition to single frequency GPS time receivers, double frequency GPS receivers providing ionospheric measurements were installed on both sites: a NIMS (NIST Ionospheric Measurement System) at OCA and a Trimble 4000 SST at McDonald.

The coordinates of the GPS antennas were expressed in the ITRF88 reference frame at OCA and the ITRF91 reference frame at McDonald, and were provided by the geodetic link with laser points at each site (Ref. 7). The estimated uncertainty of coordinates at each site is 5 cm. The uncertainties of the GPS ground antenna coordinates can have only a sub-nanosecond impact on the accuracy of this GPS common-view link.

The GPS common-view time transfer was realised from the about fifteen daily tracks of Block I and Block II satellites, following a special schedule (Fig. 4). During this experiment, the Block II satellites were subject to Selective Availability, so strict common views were required (Ref. 8).

A Vondrak smoothing (Ref. 9), which acts as a low-pass filter with a cut-off period of about five days, was performed on the raw GPS common-view values. The precision of this GPS link, estimated from the residuals of the smoothed values, is about 10 ns and should fall to about 4 ns after application of the measurements of ionosphere and post-processed precise ephemerides (Ref. 4).

3. LASSO LINK

The LASSO link was realized through quasi-simultaneous laser firing to the METEOSAT 3/P2 geostationary satellite which was located at longitude 50° West (Refs. 2,3). The on board LASSO payload comprises a laser pulse detector, an event timer monitored by quartz oscillator, a retroreflector and uses telemetry downlink transmitter of the meteorological partof the satellite. A laser pulse departure is recorded on the ground station clock, its arrival at the satellite is recorded on the on board-clock and the time of returning pulse is then recorded on ground station clock (Fig. 3). A set of these three epochs is called a triplet. The same scheme is repeated quasi-simultaneously on the second ground station. The recorded arrival times at the satellite are sent back to the Earth. Atomic clocks located in two laser stations are compared using pairs of triplets. A detailed description of the LASSO observation and data processing is given in Ref. 2 and in Ref. 3 in these proceedings.

Bad weather conditions at the OCA and McDonald prevented simultaneous laser observations for almost the entire period of this experiment. Only during the last week did three cloudless nights at both sites allow LASSO time transfer. The results are given in Table I and Figure 4. The estimated precision of each of the five LASSO comparisons is better than 100 picoseconds (Refs. 2,3). The laser equipment was not calibrated. The results of the calibration will be given in future reports on this experiment.

4. COMPARISON OF GPS COMMON-VIEW AND LASSO TIME TRANSFERS

As the GPS common-view link between the OCA and McDonald was calibrated, and its estimated uncertainty is about 10 nanoseconds, it serves as a good reference for a preliminary evaluation of the LASSO link.

For this experiment, the smoothed GPS values were interpolated for the times of occurrence of LASSO observations. Comparisons of the two techniques are given in Table I and Figure 4. Note that the GPS and LASSO results differ by a fairly constant bias with peak-to-peak discrepancy of about 15 ns. The mean of these differences is 192 ns. The root mean square of the residuals to the mean, which is taken as an estimation of the confidence of the mean, is 6 ns.

1993 Jan.	Epoch			OCA-MLRS by LASSO	OCA-MLRS by GPS	GPS - LASSO	GPS - LASSO -192ns
	h	m	s	ns	ns	ns	ns
21	2	15	14.4	9933.56	10133	199	7
22	1	41	54.9	9990.05	10175	185	-7
22	1	56	12.1	9990.58	10176	185	-7
26	1	52	01.4	10154.26	10349	195	3
26	2	17	59.4	10155.96	10350	194	2

TABLE I. *Comparison of GPS common-view and LASSO time transfers.*

The bias of about 192 ns between the GPS common views and the LASSO is certainly due to the lack of calibration of laser equipment. In the last column of Table I the 192 ns bias has been removed.

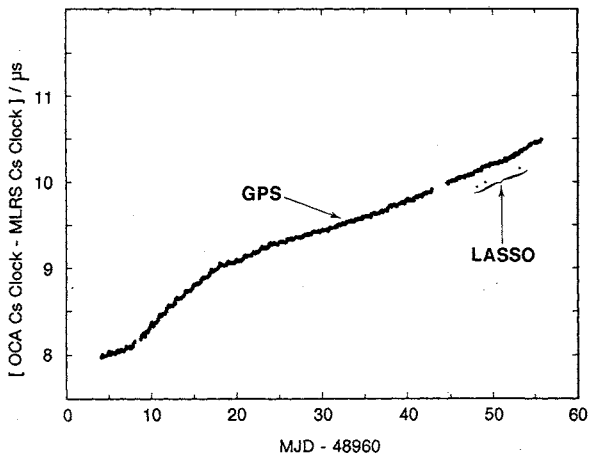


Figure 4. *[OCA Cs Clock - MLRS Cs Clock] by GPS common-view and by LASSO.*

5. CONCLUSION

This preliminary comparison of GPS common-view and LASSO time transfers between Western Europe and North America shows consistency within stated uncertainties and a bias of about 192 ns. The bias is certainly due to non-calibration of the laser equipment.

After ameliorating of GPS link by inserting measurements of ionosphere and post-processed precise ephemerides, and after differential calibration of the lasers, the agreement between two techniques should be improved.

Although LASSO, because of its sensitivity to weather conditions, is inherently unsuited for operational duties, it is certainly an excellent tool for the assessment of the accuracy of GPS, GLONASS and Two-Way time transfers.

Acknowledgements

The authors are grateful to Dr William Klepczynski of USNO for his help with this experiment. We also thank Frank Howard, PE, Director of Automated Surveys, the Texas Department of Transportation, for the loan of a double frequency GPS receiver at the McDonald Laser Ranging Station.

References

1. C. Veillet, D. Féraudy, J.M. Torre, J.F. Mangin, P. Grüdler, F. Baumont, J. Gaignebet, J.L. Hatat, W. Hanson, A. Clements, J. Jespersen, M. Lombardi, D. Kirchner, G. Kirchner, H. Ressler, S. Fassl, W.J. Klepczynski, P.J. Wheeler, W. Powell, A. Davis, M. Granveaud, P. Uhrich, LASSO, Two-Way and GPS time comparisons: a (very) preliminary status report, *Proc. 22nd PTTI meeting*, pp. 575-582, 1990.

2. C. Veillet, P. Fridelance, D. Féraudy, Y. Boudon, P.J. Shelus, R.L. Ricklefs, J.R. Wiant, LASSO observations at McDonald (Texas, USA) and OCA/CERGA (Grasse, France), *Proc. 24th PTTI meeting*, 1992, in press.

3. C. Veillet, P. Fridelance, Time transfer between USA and France through LASSO, *in these proceedings.*

4. W. Lewandowski, C. Thomas, GPS time transfer, *Proc. of the IEEE, Special Issue on Time and Frequency*, pp. 991-1000, July 1991.

5. W. Lewandowski, G. Petit, C. Thomas, Precision and Accuracy of GPS time Transfer, *IEEE Trans. I&M, Special issue for CPEM'92*, 1993, in press.

6. D.W. Allan and M.A. Weiss, Accurate time and frequency transfer during common-view of a GPS satellite, *Proc. 34th Ann. Symp. on Frequency Control*, pp. 334-346, May 1980.

7. W. Lewandowski, High Accuracy Ground-Antenna Coordinates for GPS Time Transfer, in *Proc. IAG Symposium G2-Permanent Satellite Tracking Networks for Geodesy and Geodynamics*, Vienna, August 1991, in press.

8. D.W. Allan, M. Granveaud, W.J. Klepczynski and W. Lewandowski, GPS time transfer with implementation of Selective Availability, *Proc. 22nd PTTI meeting*, pp. 145-156, 1990.

9. J. Vondrak, A Contribution to the Problem of Smoothing Observational Data, *Bull. Astron. Inst. Czechoslovakia*, 20, pp. 349-355, 1969.

MATERIALS

Chairman: J.P. Aubry

OBSERVATIONS OF DEFECTS AND SURFACE ACOUSTIC WAVES IN LITHIUM NIOBATE

A.ZARKA*, B.CAPELLE*, J.DETAINT**, J.SCHWARTZEL** and J.M.HODE***

* L.M.C.P. Université P. et M. CURIE. 4 Place Jussieu. 75252. PARIS CEDEX 05. FRANCE.
** C.N.E.T. Dept. PAB/BAG/MCT/CMM. 196 AV. H. Ravera 92220 BAGNEUX. FRANCE.
***THOMSON- SINTRA. DTAS. 399 Route des Crêtes.B.P.157.06903. VALBONNE CEDEX. FRANCE.

ABSTRACT/ RESUME

The extended defects in Lithium Niobate ($LiNbO_3$) have been studied by X-ray topography using the white radiation delivered by the D.C.I. synchrotron at L.U R.E. (Orsay, France). Very few low angle subgrains were observed in slices cut perpendicularly to the X, Y and Y-45° direction.The large intensity of the synchrotron source has permitted an experimental study of the acoustic waves (S.A.W.) in different filters using Rayleigh waves propagating in the Z direction in Y-cut plates. Furthermore, using the time structure of the radiation of the synchrotron and a synchronous excitation, stroboscopic observations were made. From that, it was possible to investigate the interactions of the S.A.W. with the crystal defects, plate edge and an acoustic absorbant. The observation of the displacement fields in high performance filters have revealed interesting properties of apodized transducers and M.S.C..This study shows that the stroboscopic topography is a good technique to observe and analyze the wave fields in surface waves devices.
S.A.W..Lithium Niobate.Stroboscopic X- ray Topography

1. INTRODUCTION

The S.A.W. devices are key components in modern analog signal processing particularly for filtering in television and in radiocommunication equipments. The rapidly growing interest for this latter application is an important factor of the new developments presently made to increase their performances. Among these developments several are relative to the obtention of large wafers of high crystalline perfection.

The first single crystals of significant size were synthetized at the Bell Laboratories (1). Like many other oxygen-octahedral ferroelectrics, the Lithium Niobate is obtained by the Czochralsky growth method at a very high temperature (1260°C). It can present a quite large non stoechiometry with an usual Li / Li + Nb ratio close to O.47. The crystal structure was determined by several workers (2-3) but very few studies were concerned with the crystal perfection.

Other developments regard the research of S.A.W. device structures which can lead to the better characteristics (reduction of the insertion losses, miniaturization and so on (4).
Among several piezoelectric materials which can be used for these applications, the Lithium Niobate has the larger coupling coefficients and hence permits to obtain filters with the largest bandwidths and other signal processing devices with a great efficiency.
Due to the extreme complexity of the three dimensional theory which describe enough accuratly the behaviour of the real S.A.W. devices, experiments are needed to understand their characteristics.On the other hand, the influence of the quality of the material on the properties of these devices is not until now very well understood.
The use of synchrotron radiation permits investigations of the quality of this material by X-ray topography. Such research is extremely difficult (or even impossible) with conventional sources, due to the very high X-ray absorption of these crystals.

In this study, we have characterized the crystalline defects of $LiNbO_3$ crystals by X-ray topography using the synchrotron radiation of the D.C.I. storage ring at the L.U R.E. (Orsay, France). The intensity of the time structure of the X-ray radiation of the synchrotron have permitted precise experimental study of surface acoustic waves.
In this paper, we present different results concerning the perfection of the presently available materials and the results of experiments made to observe the energy distribution of acoustic waves in different devices.

2. QUALITY OF THE CRYSTALS

2.1 Experimental settings

For crystals or filters too thick to be examined in a transmission Laue setting (5), reflection setting (5) with an oblique incidence was used.
In this case the film was adjusted parallel to the sample at a distance (10-20 cm) choosen to obtain in each case a good compromise between the resolution and the separation of the diffraction spots. For the thinner plates the conventional transmission Laue setting was used.

2.2 Experimental results concerning the crystals

In the following figures we represent typical examples of the different observations made.

In figure 1 is given an X-ray topograph of a X-cut slice. This topograph was obtained with the reflection setting. It may be observed that the quality of this crystal is good and that very few dislocations are present.

In Figure 2 is shown another reflection topograph of a Y-45° slice. This topograph as obtained with a diffraction vector having direction close to the normal to the surface. It may be observed a lattice of dislocations and a black and white contrast due to a subgrain boundary. The contrast corresponding to the subgrain boundary is a white line if the relative orientations of the two adjacent subgrains are such that the corresponding diffracted beam are divergent, which gives a white region with no image. In the case of relative orientations giving a convergence of the diffracted beam a darker contrast is

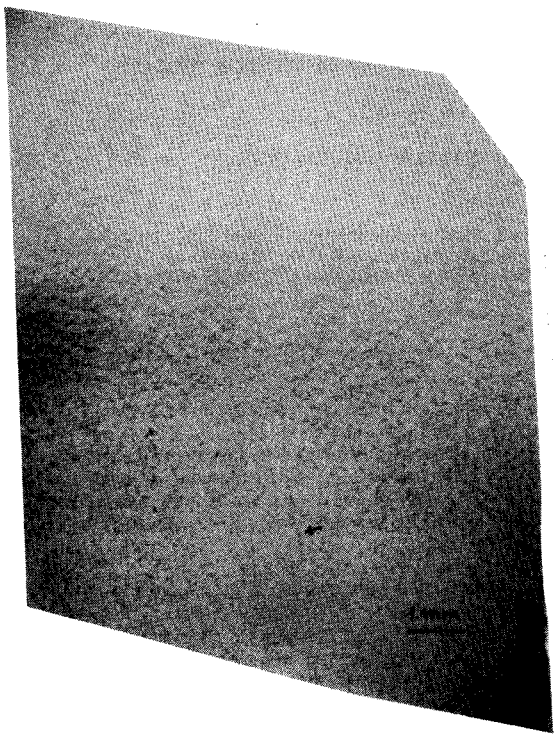


FIGURE 1



FIGURE 2

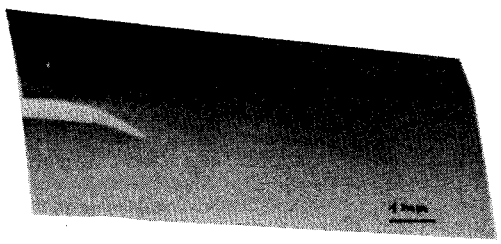
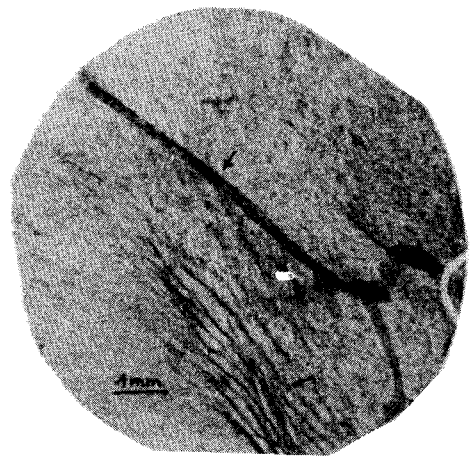
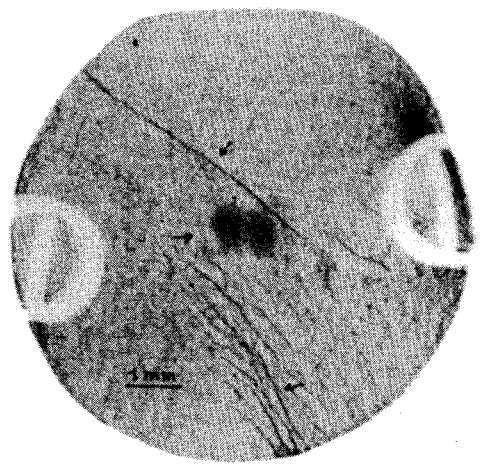


FIGURE 3



a



b

FIGURE 4

obtained owing to the overlapping of the images. This effect can be observed in Figure 2. It may be observed also that the dislocation density is quite large.

Another example is given in Figure 3. The quality of this crystal is enough good and this slice shows two kinds of defects: Straight lines dislocations and surface defects which are due to the imperfect polishing.

Figures 4a and 4b show an example of a slice studied by transmission. It corresponds to an Y+36° cut plate of about 100µm thick. Two types of defects are visible: a subgrain boundary and straight dislocations In Figure 4b the plate is vibrating in a overtone thickness stretch mode (bulk waves). The central dark part corresponds to the vibrating zone. The subgrain boundary presents another configuration because the corresponding diffraction vector is different than the previous one (Figure 4a).

3. STUDIES OF S.A.W. FILTERS

In Figure 5 is shown an X-ray topograph obtained on a simple transducer. It may be observed in the central part the dark contrast due to the vibrating zone. Due to the setting which correspond to a classical one the obtained image is an integrated one and the S.A.W. cannot be individualized. Nevertheless, the contrast is interesting and it can be observed a slight diffraction of the waves in this zone.

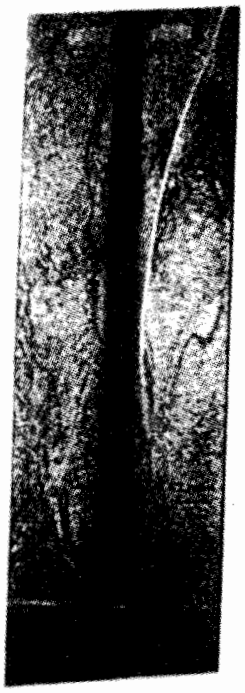


FIGURE 5

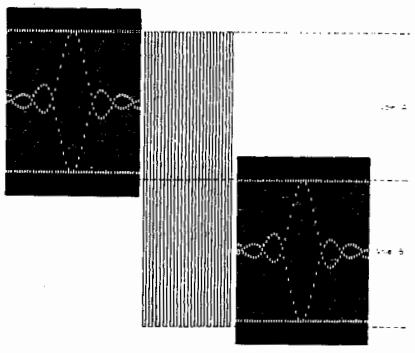
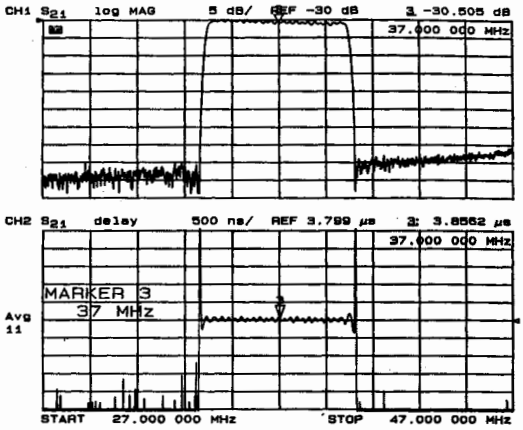


FIG. V.14. — *Mise en cascade de deux transducteurs apodisés.*

FROM " TRAITEMENT DU SIGNAL PAR ONDE ELASTIQUES DE SURFACE "
BY . M. FELDMAN AND J. HENAFF.
MASSON . C°. 1986.

a

FIGURE 6



b

FIGURE 6

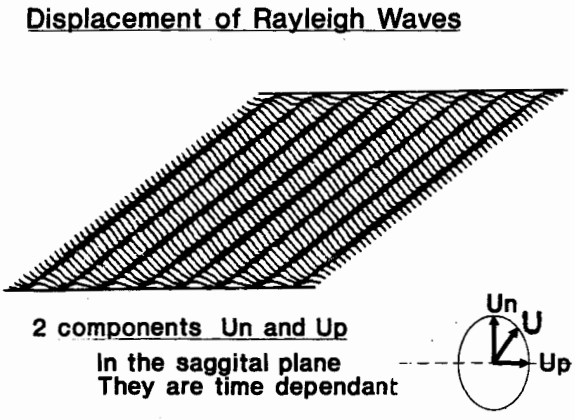


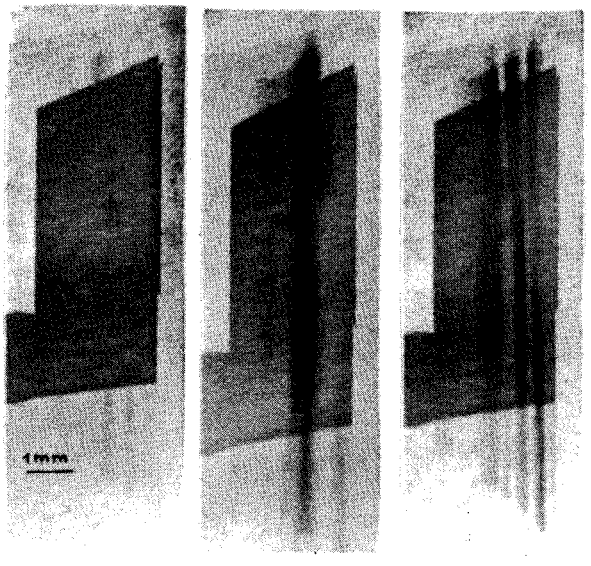
FIGURE 7

In Figure 6a is represented a drawing of a filter with apodized transducers and M.S.C.(Multi-Strip Couplers) (6-7). In Figure 6b it can be observed the response curve of the filters which have been studied by X-ray topography using the synchrotron radiation. In the upper part is shown the amplitude response and in the bottom the group delay.
The representation of the displacement of the Rayleigh waves is shown on the drawing of the Figure 7. The two principal components **Un** and **Up** of the displacement **U** are represented in the saggital plane. Naturally, they are time dependant and the properties of the corresponding Rayleigh waves are given in the following table 1.

VELOCITY (m/s)	3491.1
COUPLING COEFFICIENT	4.86
TEMPERATURE COEFFICIENT	94 ppm
ANGLE OF POWER FLOW	0.0

TABLE 1

This type of filter have been studied by X-ray topography by reflection using the white beam of the D.C.I. synchrotron. This setting is necessary because the whole device constitute a very absorbing mounting (substrate). An example of such topograph taken among the different spots obtained in a Laue pattern is shown in Figure 8. In this figure, it may be observed the homogeneous contrast due to the transducers and the M.S.C.(Figure 8a). In Figure 8b, the filter is in a asynchronous excitation and as in the general case (reflection) only the component perpendicular to the surface is visualised. Due to the fact that the excitation is not synchronous, the contrast of the S.A.W. is integrated and the fringes cannot be resolved individually. Nevertheless, some other fringes due to the displacement of the energy in the transducers are visible. In Figure 8c, the X-ray topograph was taken in the same experimental conditions, but in contrary to the previous case the frequency of the filter was adjusted to be an exact multiple of the frequency of the X-ray beam. This permits to obtain a stroboscopic image (5) (34.862 MHz) where the S.A.W. are individually resolved. Some observations can be done from this topograph: it can be observed that the S.A.W. are not very well confined in the transducer and in other hand that the vibrating energy is also confined in the perpendicular large fringes already observed in the previous topography (Figure 8b). It may be also observed that the contrast of these last fringes is very different for the same level of excitation.

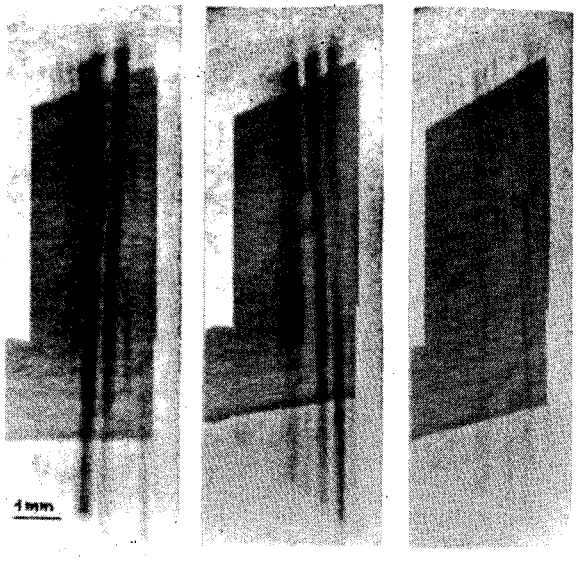


a b c

FIGURE 8

Influence of the frequency on the contrast.

The X- ray topographs presented in Figures 9a, 9b and 9c correspond to enlargement of the same zone of the filter. They have been taken with the same experimental conditions (same diffraction vector **g**, same θ Bragg angle, same incidence direction) Between these stroboscopic topographs, only the frequency has changed. They were taken respectively at 34.862 MHz (Figure 9a), 38.031 MHz (Figure 8b) and 41.200 MHz (Figure 9c). The principal features which can be observed are the fringes due to the S.A.W. and the very contrasted zones which reveal the flow of the energy in the transducer.



a b c

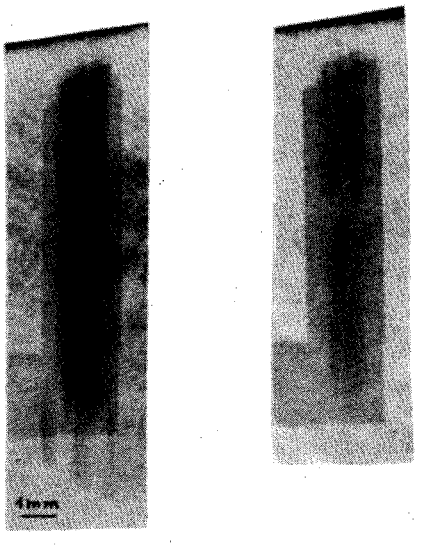
FIGURE 9

Influence of the θ Bragg angle.

Variing the θ, the contrasts obtained on the topographs change.

In particular, when the θ varies, the λ (X-ray wavelenght) changes and the depth of penetration of the X-ray in the filter is more or less important.

An example is given in Figures 10a and 10b. Obviously the contrast depends on the amplitude of the waves inside the filter and is less important when the angle θ is greater (Figure 10b).

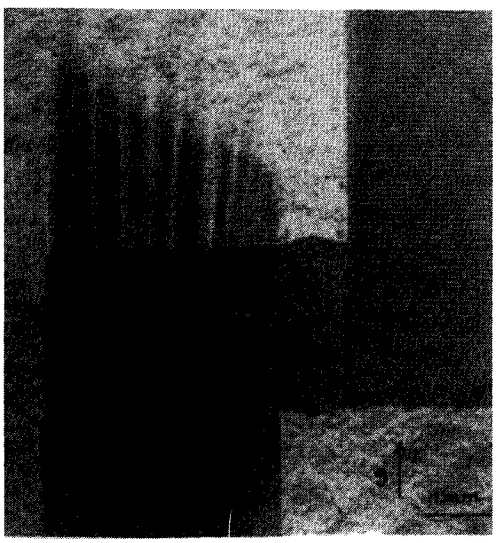


a b

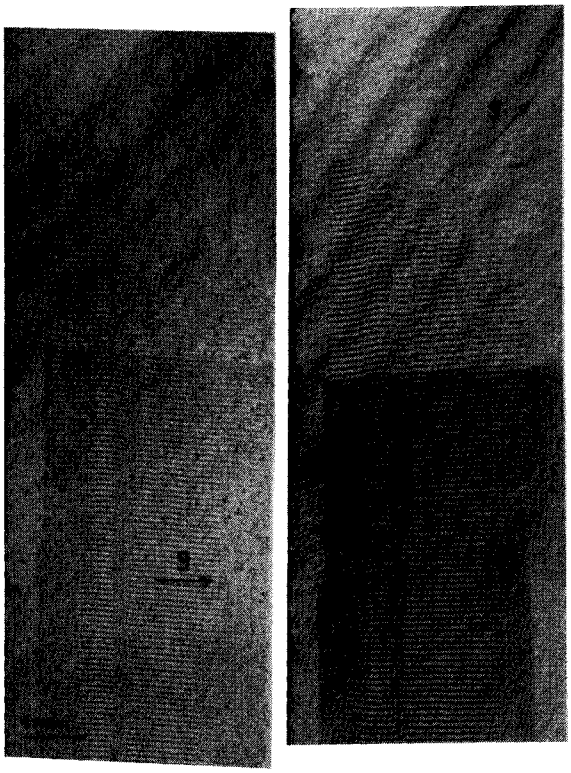
FIGURE 10

Influence of the g diffraction vector.

Depending on the direction of the **g** diffraction vector, different components of the displacement **U** are revealed. In Figure 11a the diffraction vector is parallel to the larger dimension of the transducer. In this case, it can be observed the wave front curvatures. In Figure 11b the visualisation of the modulation is obtained when the **g** diffraction vector is perpendicular. For the intermediary case the different components of the displacement are visible (Figure 11c,where **g** is at 45°).



a



b c

FIGURE 11

Influence of the defects.

In Figure 12a is presented a filter where two subgrains exist in the crystal. In the central part, an acoustic absorbent has be deposited. The role of this absorbent is to attenuate the losses by reducing the vibrating amplitude as it can be seen on the topograph. In contrary the two zones with the subgrain boundaries (enlarged in Figures 12b and 12c) show that the S.A.W. present phase shifts due to these defects.

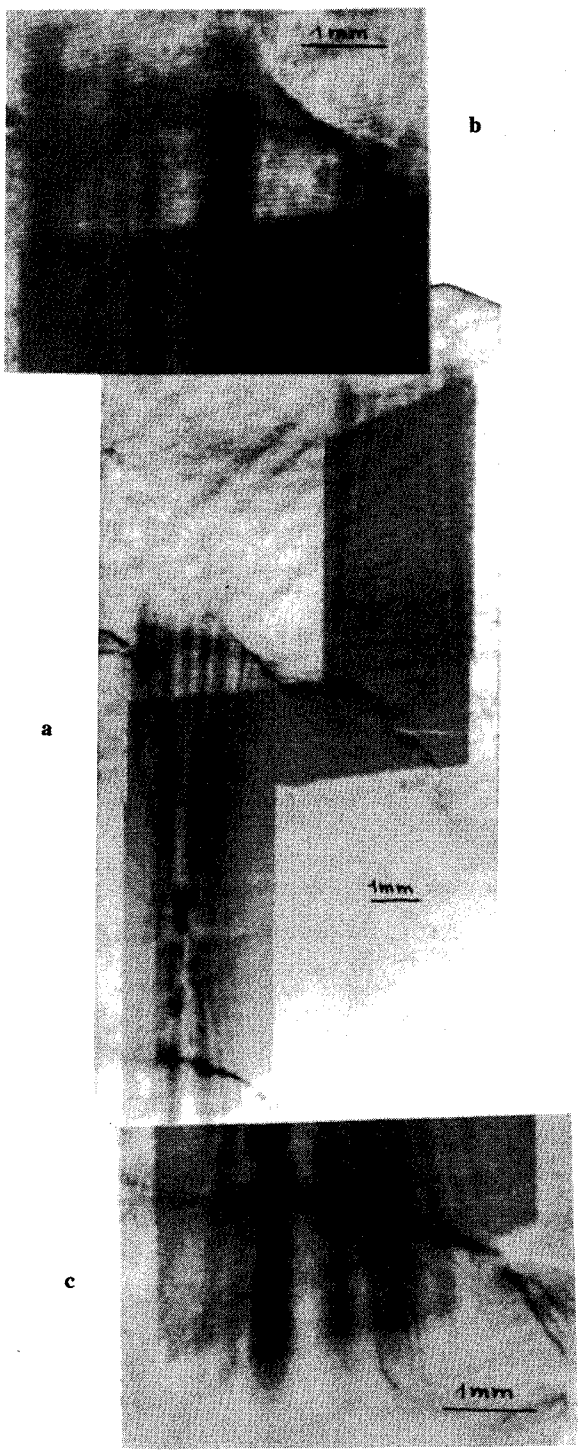


FIGURE 12

4. CONCLUSION

The large intensity of the synchrotron source has permitted an experimental study of Lithium Niobate crystals and of the surface acoustic waves (S.A.W.) in different filters by X-ray topography.
The studies of the apodized transducers by topography present different complicated contrasts which vary with the frequency. Generally these contrasts correspond to a modulation of the wave fronts and the intensity. It was possible to investigate the quality and the interactions of the S.A.W. with the crystal defects, plate edges, and an acoustic absorbent. The amplitude distribution in apodized transducers and in multistrip couplers (MSC) were also observed.

This study shows that the crystal quality of the lithium niobate has been greatly enhanced in the recent years in comparison with those produced some time ago

It was also shown that the stroboscopic topography is a good technique to observe and analyse the wave fields in surface waves devices. It is possible to obtain quantitative informations from densitometry and image processing.

These techniques can be used for devices using others materials.

5. REFERENCES

(1) BALLMAN. A.A. J. Am. Ceram. Soc. (1965) **48**. 112-114

(2) NASSAU. K., LEVINSTEIN. H.J. and LOIACONO. G.M. J. Phys. Chem. Sol. (1966). **27**. 983-996.

(3) ABRAHAMS.S.C., REDDY.J.M. and BERNSTEIN. A. J. Phys. Chem. Sol. (1966). **27**. 997-1012.

(4) ATEMAN. P. and DIEULESAINT E. Electronic Letters (1969) **5**; 25. 657-658.

(5) CAPELLE. B., DETAINT. J., SCHWARTZEL. J., ZHENG. Y. and ZARKA. A. Proc. IEEE Frequency Control Symposium (1992). 553-560.

(6) FELDMAN. M. and HENAFF. J. Traitement du signal par ondes élastiques de surface. Ed. Masson. (1986).

(7) MAERFELD. C. "Multistrip Couplers". Wave Electronic 2. (1976). 82-110.

TEMPERATURE DEPENDENCE OF ELASTIC CONSTANTS OF $GaPO_4$ AND ITS INFLUENCE ON BAW AND SAW DEVICES

W. WALLNÖFER, J. STADLER and P. KREMPL

AVL List Ges.M.B.H, Physics Division, A-8020 Graz, Austria

ABSTRACT/RESUME

Brillouin scattering in $GaPO_4$ was measured in the temperature range from -173 to 427°C. The scattering intensities at RT delivered the full elastooptic tensor. From the measured sound velocities, temperature coefficients of all the elastic stiffness constants were determined. These data and results from rotated Y-cut BAW resonators with Q values up to 60000 were used to study the BAW and SAW properties of $GaPO_4$. In both cases, a temperature compensated cut with parabolic frequency behaviour and higher coupling than in quartz was found, situated at $\vartheta = -16°$ for BAW and at $\mu = 57°$ for SAW.

Keywords: $GaPO_4$, elastic constants, temperature compensated cuts, BT cut, ST cut, SAW, elastooptic constants.

INTRODUCTION

The piezoelectric crystal gallium orthophosphate, homeotypic to quartz, has been presented to this forum in the last years[1,2,3,4]. The high stability of its α-quartz-phase up to 930°C and its strong electromechanical coupling make it a promising material for applications at high temperatures and large bandwidths.

An essential requirement for the use of a piezoelectric in acoustic applications is that it has orientations where the temperature dependence of the devices' frequency or delay time is small in the specific range. This is the case for quartz and berlinite. It has been shown[2,4] that $GaPO_4$ has a rotated Y-cut with temperature compensated resonance frequency near room temperature.

For the calculation of the temperature dependence of SAW devices, as well as for a more comprehensive study of the bulk resonator properties, the knowledge of the temperature coefficients of all 6 independent components of the elastic tensor is necessary. Up till now, only coefficients of selected components have been published[2, 5]. Consequently, a determination of the complete set was necessary.

BRILLOUIN SCATTERING MEASUREMENTS

Brillouin spectroscopy was chosen as method because it allows the complete determination of the elastic properties with few and small samples. We used five cubes with 1.5 - 2 mm side length. Details of the experimental apparatus and method were described at the 5th EFTF[4] where the determination of the elastic constants at room temperature[6] was reported. Due to the use of an Argon ion laser at 514.5 nm and a 90° scattering geometry, the frequencies of the thermal phonons involved in the scattering process were in the range of 15-20 GHz.

The room temperature values in Table 3 are the results of a more elaborate and exact evaluation procedure than those reported in Ref. 4 and thus supersede that preliminary data set. The sign of c_{14} has been changed in accordance to the IEEE standard 176-1978.

While the frequency shift of the Brillouin peaks yields information about the elastic properties, their intensities can be evaluated to determine the elastooptic tensor[7]. By using 20 measurements made with horizontally and vertically polarized input and output light in selected orientations, all 8 independent components of the elastooptic tensor of $GaPO_4$ could be determined. The intensities were calibrated with spectra taken from toluene. The results are shown in Table 1.

Table 1. Elastooptic constants of $GaPO_4$

p_{11}	p_{12}	p_{13}	p_{31}	p_{14}	p_{41}	p_{33}	p_{44}	p_{66}
.146	.333	.328	.417	-.009	.022	.067	-.165	-.187

Brillouin spectroscopy is advantageous for measurements at high temperatures, because there are no problems with electrodes, rising conductivity or electrical signal transmission. Thus, it was decided to extend the measurements to a large temperature range to determine the coeffiecients of the elastic constants.

For low temperatures a spectroscopic kryostat cooled with liquid nitrogen was used. Spectra were taken at temperatures down to 100 K. For the high temperature range a cell usable up to 700°C was constructed. However, at high temperatures the samples became increasingly opaque. This effect has been observed in berlinite[8] and attributed to the precipitation of water from

OH impurities. The Rayleigh scattering due to the formation of small water bubbles led to an increased background scattering above 250°C. This deteriorated the signal-to-noise-ratio and limited the possibility of exact determination of the sound velocities to below 430°C.

Series of spectra for seven scattering geometries were chosen for evaluation. Each series covered the temperature range from -173°C to 427°C (100 to 700 K) with 9 to 11 spectra at different temperatures, which made it possible to determine the temperature coefficient of the Brillouin shifts in this range. The temperature coefficients of the Brillouin shift frequencies determined by least squares fit are listed in Table 2.

Table 2. Temperature coefficients of the measured Brillouin shift frequencies for rotated X" direction.

Euler angles λ, μ, θ	mode (fast/middle /slow)	$T^{(1)}\Delta v$ [$10^{-6}K^{-1}$]	$T^{(2)}\Delta v$ [$10^{-9}K^{-2}$]	$T^{(3)}\Delta v$ [$10^{-12}K^{-3}$]
0,0,0	f	-16.95	-12.39	4.61
0,0,0	m	176.9	-353.	289.8
0,0,0	s	18.54	-37.8	50.4
0,90,90	f	-50.1	-17.05	9.34
0,90,90	s	49.3	-114.	55.9
90,90,45	f	-43.1	26.11	-21.58
90,90,135	f	-26.5	44.9	-41.

The sound velocities are calculated from the frequency shifts by

$$v = \frac{\omega}{k} = \Delta v \lambda \, (n_i^2 + n_s^2 - 2 n_i n_s \cos\theta)^{-1/2},$$

and depend on the elastic constants according to

$$(\rho\omega^2 \delta_{jk} - c_{ijkl} k_i k_l - \frac{e_{ij} k_i k_l \, e_{mkn} k_m k_n}{\varepsilon_{ij} k_i k_l}) u_k = 0.$$

To calculate the coefficients of the elastic constants from these results, some additional material data are necessary. The temperature dependences were determined in the following way: the temperature coefficients of the refractive indices were evaluated from the data of Ref.9 to Tn_{11} = -1.9 10^{-6}/K and Tn_{33} = -2.9 10^{-6}/K. The thermic expansion was determined from the x-Ray measurements in Ref.10 . The resulting first and second temperature coefficients of the elastic constants are listed in Table 3. The uncertainty of the data varies and is smallest for the longitudinal constants. The temperature behaviour of the constant c_{66} is the least reliable, since it has an effect mainly on the weakly scattering slow transverse mode in the xy-plane.

Table 3. Elastic constants of $GaPO_4$ and their first and second temperature coefficients. *) Values from Table 4 recommended.

	c [10^9 Pa]	$T^{(1)}c$ [$10^{-6}K^{-1}$]	$T^{(2)}c$ [$10^{-9}K^{-2}$]
c_{11}	66.58	-51.2	-47
c_{33}	102.13	-123.0	-51
c_{44}	37.66	-4.4	-40
c_{66}*	22.38	62.7	-9.5
c_{12}*	21.81	-277	-123
c_{13}	24.87	-27.8	-37
c_{14}	3.91	520	-196

A comparison with the respective values of quartz[11] and berlinite[12] shows that the longitudinal constants are similar, while c_{44}, c_{66}, c_{12} and c_{13} are clearly lower and c_{14} is much higher. Here again, these two materials are more similar to each other than to the homeotypic $GaPO_4$.

BULK RESONATOR MEASUREMENTS

In previous years[4] we reported at this forum results from measurements of rotated $GaPO_4$ Y-cut resonators. Since then, acoustic quality factors of the resonators were improved continuously, due to better crystal purity and processing technique. Fig.1 shows as example the frequency response of a bevel-edged Resonator with a Q factor up to 60130.

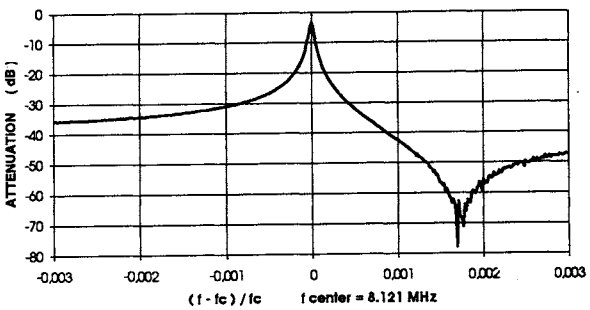


Fig. 1: Frequency response of a $GaPO_4$ Y+53° cut with high Q factor

The resonance of an unrotated GaPO4 Y-cut resonator was measured up to 670°C. The frequency/temperature behaviour is shown in Fig. 2. The elastic properties of this cut are determined solely by the component c_{66}. Since the temperature range of this measurement coincides with the main part of the Brillouin measurements range and c_{66} is the most uncertain component to be determined by Brillouin scattering, we replaced the values of Tc_{66} in the Brillouin data set used for further calculations with those evaluated from this resonator, shown in Table 4.

Table 4: Temperature coefficients evaluated from Y cut resonator; Tc_{12} calculated with Tc_{11} from Table 3.

	$T^{(1)}c$ [$10^{-6}K^{-1}$]	$T^{(2)}c$ [$10^{-9}K^{-2}$]
c_{66}	47.2	27.1
c_{12}	-242	-198

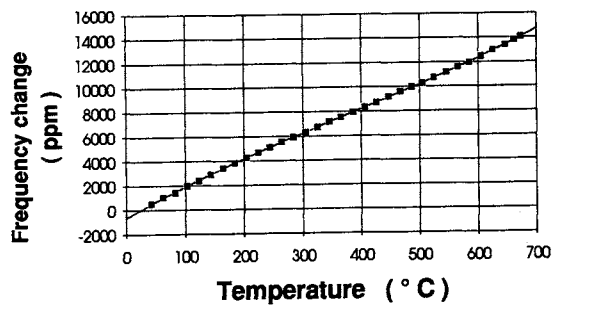


Fig. 2: Thermal behaviour of a $GaPO_4$ Y-cut resonator up to about 670°C

The experimental work has advanced to the point where a reliable fit of the temperature coefficients of the three involved elastic constants c_{14}, c_{44} and c_{66} of $GaPO_4$ could be done, allowing an independent test of the Brillouin results.

The resonance frequency data of the five resonators listed in Table 5 in the temperature range from 0°C to 200°C were used. Temperature dependences of thickness and density were calculated using the thermal expansion data from Ref. 13. The mass loading due to the electrodes was corrected numerically.

Table 5: Data of selected $GaPO_4$ resonators

Angle ϑ, °	Frequency constant, m/s	Coupling factor, %
-18.2	1251.2	9.8
0.0	1270.6	15.1
15.8	1372.7	15.8
35.2	1546.7	12.0
53.0	1607.7	6.8

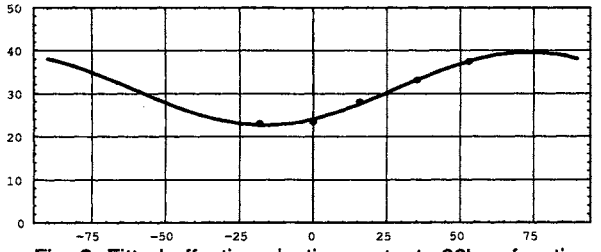


Fig. 3: Fitted effective elastic constant c66' as function of the angle and values from five $GaPO_4$ resonators

Using the formula for $GaPO_4$ thickness shear resonators rotated by ϑ from the Y cut,

$$f_R = \frac{1}{2t(T)}\sqrt{\frac{c'_{66}(T)}{\rho(T)}};$$

$$c'_{66}(T) = c_{66}(T)\cos^2\vartheta + c_{44}(T)\sin^2\vartheta + 2c_{14}(T)\sin\vartheta\cos\vartheta$$

the values of the relevant components of c^E were determined by least square fit at different temperatures. The resulting room temperature values of the effective elastic constant $c'_{66}(\vartheta)$ and the measured coupling factors are plotted versus the rotation angle in Fig. 3. The temperature coefficients of the c^E components are listed in Table 6, and their temperature dependence is plotted in Fig.4. The differences to the Brillouin values can be attributed partly to the different temperature ranges over which the determination was done, partly to dispersion effects related to impurities, as described at the end of this section, and to different material qualities.

Table 6: Temperature coefficients of elastic constants c^E from resonator data

T/°C	$T^{(1)}c_{14}/10^{-6}\,K^{-1}$	$T^{(1)}c_{44}/10^{-6}\,K^{-1}$	$T^{(1)}c_{66}/10^{-6}\,K^{-1}$
25	350.4	9.4	34.7
100	540	-22.2	46.5

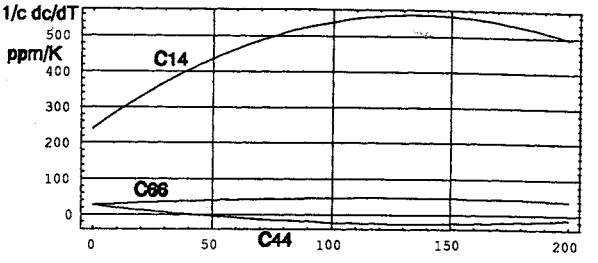


Fig. 4: Temperature dependence of the linear temperature coefficient for the elastic constants c_{14}, c_{44} and c_{66}.

With these results, the temperature coefficients of frequencies at room temperature plotted in Fig. 5 were calculated. The two resulting temperature compensated cuts are located at ϑ=-16° and -73,3°. The second one has not yet been verified experimentally. However, it is positioned in a range of low electromechanical coupling and thus probably not interesting for practical purposes. The temperature compensated cut near -16° was experimentally verified[4]. It follows a parabolic curve with a second temperature coefficient of -18 $10^{-9}/K^2$. This is about half the value of the quartz BT cut, but the behaviour is not as good as the quartz or berlinite AT cut. The electromechanical coupling factor of the $GaPO_4$ BT cut is 0.124.

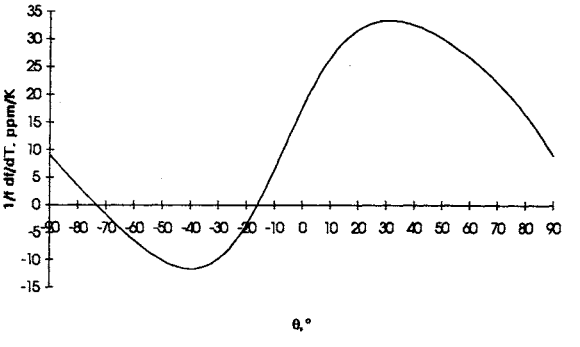


Fig. 5: Linear temperature coefficient of the resonance frequency as function of cut angle at 25°C

At the third EFTF[2] a cut with quartz AT cut like frequency/temperature behaviour was presented. This behaviour can be attributed to the high OH impurity content of the crystal material in question, which, according to infrared absorption measurements, was 1240 ppm. As observed in berlinite[14,15] this leads to a peak in the temperature dependence of the acoustic damping and a corresponding flattening or step in the frequency curve. Similar loss and frequency effects of high OH concentration in $GaPO_4$ were observed by Philippot et al.[16]. The effect is attributed to transitions between different impurity configurations in the crystal lattice.

SURFACE ACOUSTIC WAVE PROPERTIES

With complete elastic temperature behaviour, the relevant surface acoustic wave properties can be calculated. Particularly interesting for acoustoelectric applications are the electromechanical coupling constant and the temperature dependence of the delay time.

For the calculations, a program was developed using the algorithm described in Ref. 17, by which the determinant root search for the calculation of the partial waves is changed to an eigenvalue equation in the decay constants.

The properties determined by this program comprise bulk and surface wave velocity for free and metallized surfaces, temperature coefficients of velocity, frequency or delay, coupling factor, beam steering angles, diffraction coefficient, reflection at grooves and fingers and the amplitude distribution of displacements and potential of the wave. These properties can be calculated for surface waves on crystals of any piezoelectric symmetry class with any cut orientation and temperature, including Bleustein-Gulyaev waves and leaky waves.

There are 3 degrees of freedom in the orientation of SAWs. However, in crystal class 32, one mainly uses singly rotated cuts with the propagation direction along the X-axis, because these X-boule-cuts offer easy orientation with low misadjustment sensitivity in two of the 3 angles, no beam steering effects and rather high coupling factors.

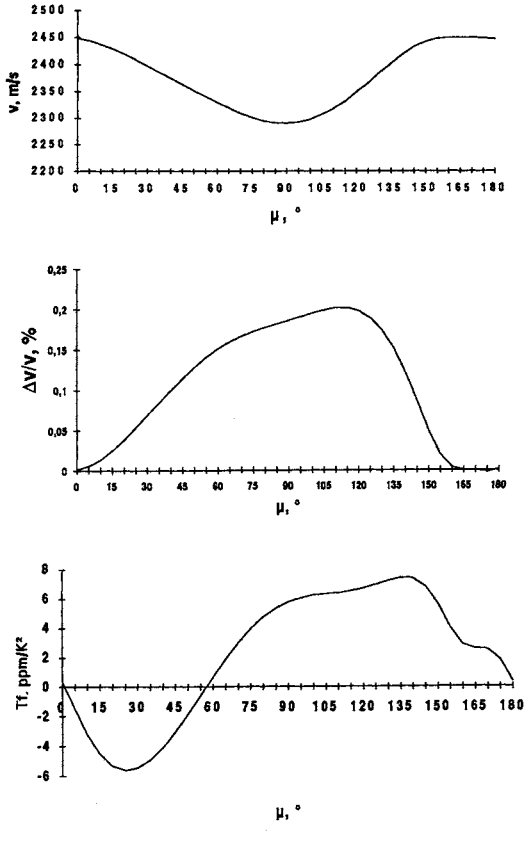


Fig. 6: Properties of SAWs in $GaPO_4$ X-boule orientations: a) Phase velocity, b) coupling factor, c) temperature coefficient of frequency.

Figure 6 show results for the $GaPO_4$ X-boule orientations. It can be seen in Figure 6 that temperature compensated cuts do indeed exist. While one of them has a low coupling, the cut near $\mu = 57°$ has a coupling $\Delta v/v$ of .146%, which is more than twice the value of the Quartz ST cut, although not quite as good as in berlinite. Using the resonator data set from Table 6 instead of the corresponding Brillouin values, the temperature compensation at room temperature occcurs at $\mu = 54°$. We propose to name this orientation, in analogy to quartz, the GaPO4 ST cut. Properties of the ST cuts of quartz homeotypes are compared in Table 7.

Table 7. Properties of ST cuts of quartz homeotypes.

	$GaPO_4$	$AlPO_4$	SiO_2
Cut angle μ [°]	57	80.4	47.25
Phase velocity v [m/s]	2335	2741	3158
Coupling $\Delta v/v$[%]	.146	.245	.058
$T^{(2)}f$ [10^{-9} K^{-2}]	-50	-220	-32

A special feature of $GaPO_4$ is its high stability up to the transition to the cristobalite phase at 930°C. This can also be seen from the SAW behaviour. The temperature coefficient of delay for cuts near the ST cut stays rather low up to the measurements limit at 430°C. Since the crystal does not come into the vicinity of a displacive transition, one can extrapolate that temperature compensated SAW cuts exist also at higher temperatures. This might be possible right up to the cristobalite transition, which is restructuring and thus not preceded by a deterioration of stability.

CONCLUSION

The temperature dependence of the full set of elastic constants of $GaPO_4$ was determined for the first time in the range of -173°C to 427°C by Brillouin scattering. The temperature coefficients of c_{66}, c_{14} and c_{44} near room temperature were obtained independently from resonator measurements with high quality factors.

With these data, temperature compensated cuts for BAW and SAW devices were calculated. For both wave types, a cut exists with higher coupling than quartz and with a parabolic temperature dependence whose second coefficient is 18 and 50 $10^{-9}/K^2$ for the BT and ST cut, respectively. Temperature compensation can be achieved up to very high temperatures.

As number, size and quality of the available $GaPO_4$ crystals are continually increasing, research on the elastic and piezoelectric properties up to 930°C is an important task for the development of acoustoelectric applications of this promising new material.

ACKNOWLEDGEMENTS

The work on BAW resonators was supported by the Austrian Forschungsförderungsfonds für die gewerbliche Wirtschaft under project Nr, 3/7992, which is gratefully acknowledged.

REFERENCES

1) G.F. Engel and P.W. Krempl, Proceedings of the Second European Frequency and Time Forum, Neuchatel, 1988, pp.827-832
2) G.F. Engel, P.W. Krempl and J. Stadler, Proceedings of the Third European Frequency and Time Forum, Besancon, 1989, p.50
3) J. Stadler, P.W. Krempl and G.F. Engel, Proceedings of the Fourth European Frequency and Time Forum, Neuchatel 1990, p.567-575
4) P.W. Krempl, J. Stadler, W. Wallnöfer, W. Ellmeyer and R. Selic, Proceedings of the Fifth European Frequency and Time Forum, Besancon 1991, p.143-147
5) I.A. Danikov, O.V. Zvereva, V.I. Ivannikov, E.F. Tokarev and F.Sh. Khatamov, Kristallografiya 36, 504-507 (1991) [Sov. Phys. Crystallogr. 36, 277-279 (1991)]
6) W. Wallnöfer, P.W. Krempl and A. Asenbaum, to be published
7) R. Vacher and L. Boyer, Phys.Rev.B. 6, 639-673 (1972)
8) J.C. Doukhan, B. Boulogne, P. Cordier, E. Philippot, J.C. Jumas & Y. Toudic, J. Crystal Growth 84 (1987) 167-179
9) S. Defregger, G.F. Engel and P.W. Krempl, Phys. Rev. B43, 6733-6738 (1991)
10) O. Baumgartner, A. Preisinger, P.W. Krempl and H. Mang, Z. Kristallogr. 168, 83-91 (1984)
11) B.J. James, Proceedings of the 42nd Annual Frequency Control Symposium (1988), 146-154
12) D.S. Bailey, J.C. Andle D.L. Lee, W. Soluch, J.F. Vetelino and B.H.T. Chai, Proc. of the 1983 IEEE Ultrasonics Symposium, 335-340
13) K. Kosten and H. Arnold, Z. Krist. 152 (1980) 119
14) R.F. Steinberg, M.K. Roy & A.K. Estes, Proc. IEEE 1984 Ultrasonics Symposium 279-284
15) E. Philippot, A. Goiffon, M. Maurin, J. Détaint, J. Schwartzel, Y. Toudic, B. Capelle and A. Zarka, J. Crystal Growth 104 (1990) 713
16) E. Philippot, A. Ibanez, A. Goiffon, B. Capelle, A. Zarka, J. Schwartzel and J. Détaint, Proceedings of the 6th European Frequency and Time Forum, Noordwijk 1992, p.383-388
17) E. Langer, S. Selberherr and P.A. Markowich, Proc. IEEE 1983 Ultrasonics Symposium 1157-60

X RAY TOPOGRAPHY STUDY OF GALLIUM PHOSPHATE CRYSTALS AND RESONATORS

B.Capelle*, A.Zarka*, J.Schwartzel**, J.Detaint**, Y.Zheng*, A.Ibanez***, E.Philippot***

*L.M.C.P. Universités P VI et P VII, CNRS, 4 Place Jussieu 75252 Paris France
**CNET Laboratoire de Bagneux, 196 Av.H.Ravera 92225 Bagneux France
***L.P.C.M.S. Université de Montpellier, CNRS-URA D0407, case 03, 34095 Montpellier France

Abstract.

Using the synchrotron radiation delivered by the DCI storage ring at the LURE (Orsay, France), we have studied by the X-ray topography technique the crystalline perfection of $GaPO_4$ samples grown in different conditions. The growth of the gallium phosphate on a berlinite seed is analysed by X-ray section topography. It was observed different vibration modes of an approximately rhombohedral face cut and a near AT cut thickness shear resonators.

Introduction.

The gallium phosphate ($GaPO_4$) is, as berlinite ($AlPO_4$), a material which is a close piezoelectric analog of quartz and possesses similar cuts. It was shown that this material has very attractive properties for the applications to piezoelectric devices, due to its large coupling coefficients and to the existence of compensated cuts [1,2]. Moreover, for the AT cut, the first results obtained seem to indicate a value of coupling coefficient higher than those of quartz and berlinite. This induces the possibility to obtain larger band-pass of filters. On the other hand, there is no α-β transition for this material and, then, it can be used in a large range of temperature (as much 900°C), for sensor devices for instance.

We present here an X-ray topography study of crystals obtained with different growth approachs, and some vibration modes of gallium phosphate resonators. The plates of GaP04 were studied by X-ray transmission topography using the white beam delivered by the DCI storage ring at LURE. We have used conventional X-ray topography to characterize this material and stroboscopic X-ray topography [3] to analyse the vibration modes.

All the gallium phosphate crystals presented in this paper have been made by the group of Etienne Philippot at Montpellier.

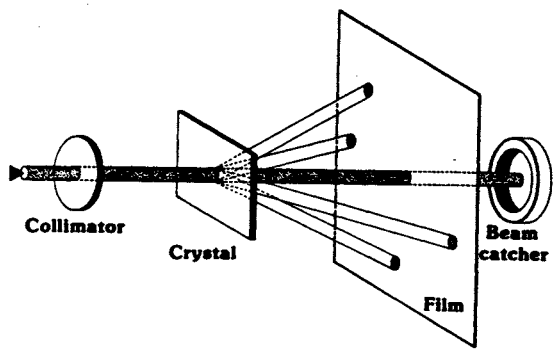


Figure 1- The transmission Laue set-up.

The experimental set-up.

The figure 1 presents the principle of the experimentl set-up which is a Laue set-up. The sample is placed in the white X-ray incident beam delivered by the synchrotron source. On the film, we obtain a Laue pattern on which the crystal is simultaneously imaged with different diffraction vectors. Two types of topographs can be obtained. The first one is a traverse topograph which is obtained when a large incident X-ray beam is used, the second one is a section topograph when the incident beam is limited by a fine slit of about fifteen microns wide. With this second type of topograph we can see the section of the crystal. Let us see a pattern made in the vertical plane and shown on the figure 2.

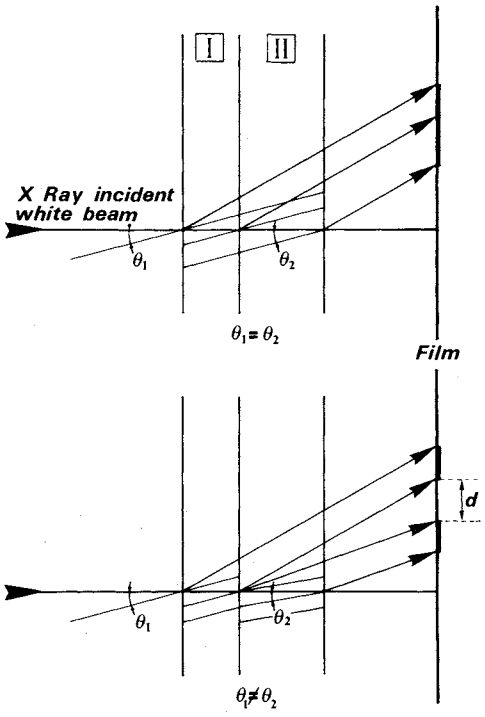


Figure 2- Section topograph of a crystal made with two parts without distortions in a and with distortions in b.

We assume that the crystal is made with two parts which are in our case an epitaxial growth in I and a substrate in II, this crystal is a slice perpendicular to the plan of this pattern .

In a this is the case where the two parts are the same structure and where there are no distortions. The diffracting planes are the same for the two parts. In this case, the angles θ_1 and θ_2 between the incident beam and the diffracting planes of the two parts respectively are equal, and the diffracted beams by the two parts are parallel. Then, we obtained on the film only one image for the two parts.

In b this is the case where the diffracting planes are not identical for the two parts because of different structures or parameters, distortions or a bad epitaxial growth. In this case the two angles θ_1 and θ_2 are different and the diffracted beams are not parallel because they make an angle equal to $2\theta_1$, or $2\theta_2$ with the incident beam. Then, we obtained, on the film, two separated images for the two parts of the crystal. If we know the distance between the crystal and the film, we can measure d on the film and calculate the disorientation $\Delta\theta$ between the two parts.

We will see an example of this effect farther.

$GaPO_4$ crystals growth.

The $GaPO_4$ crystals growth has been carried out through the three methods already used for berlinite [4,5]:

-the slow heating method (SHT);

-the horizontal composite gradient method (HTG);

-the vertical reverse temperature gradient method (VTG).

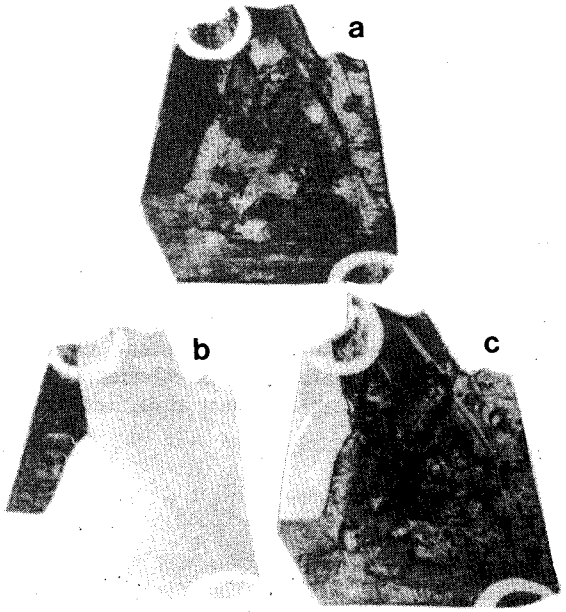


Figure 3- Topographs of a gallium phosphate crystal obtained by successive growths from a small spontaneous nucleation by SHT method in phosphoric acid.

The figure 3, in a, shows a topograph of a gallium phosphate crystal obtained by successive growths from a small spontaneous nucleation by slow heating method in phosphoric acid. Many defects appear, dislocations around the crystal and several twins mixed in the central part. Due to the different values of the structure factor for the two electrical twin domains for fixed reflections it is possible to see on two different topographs each twin domain.

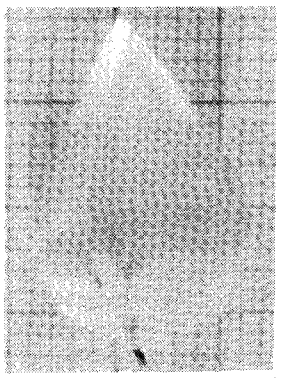
This is the case for the two topographs on the figures 3c and d, where one isolated electrical twin can be seen. The other twins in the central part of this crystal are optical twins.

An another example of this type of crystal is shown on the figure 4. Note that, like in the precedent case, the constrast in the central part of this crystal are very complicated to analyse because the superposition of many defects.

The first step to improve the crystalline quality of this material is to select seeds among plates cut in the crystals obtained from the spontaneous nucleation. That way the crystal on the figure 5a has grown, on the bottom we can see the fixation of the seed. The SHT method in



Figure 4- Topograph of a gallium phosphate crystal obtained by successive growths from a small spontaneous nucleation by SHT method in phosphoric acid.



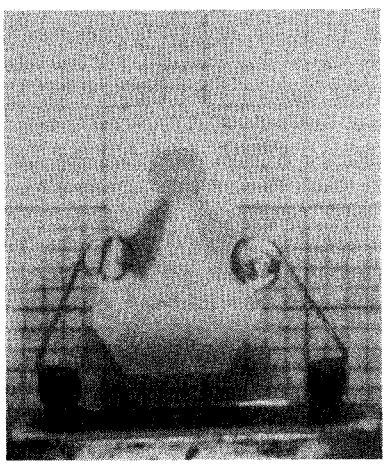


Figure 5- Gallium phosphate crystal obtain from a seed selected among plates cut in the crystals obtained from the spontaneous nucleation in a and in b a slice of this crystal.

phosphoric acid has been used. This method where the temperature is slowly increased from about 150 to 180°C with a speed equal to 0.5 to 2°C/day, is now well known. One's major drawback is the impossibility to obtain large crystals in one operation because of the limited quantity of gallium phosphate available in the solution. In b is presented a slice parallel to a face of the rhombohedron of this crystal that we have study.

The optical observation of this slice shows (Fig.6b) an electrical twin, probably due to the fixation of the seed, and a strong black and white contrast due to the fixation of the first growth from a small spontaneous nucleation. On the topograph (Fig.6a) strong contrasts around the fixation of the first growth and, in the same area of the fixation of the seed that the twin, some cracks appear in white contrast .

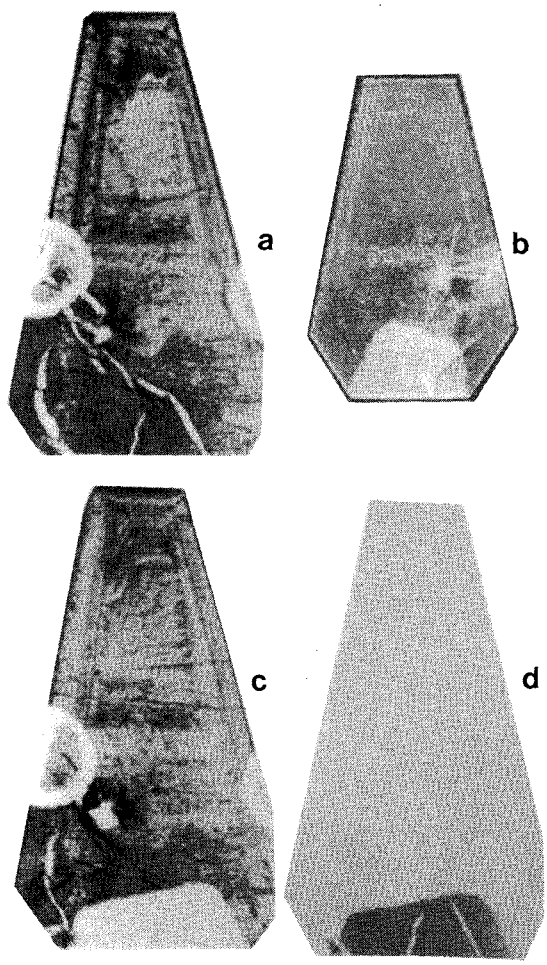


Figure 6- Topographs in a, c, d and optical observation in b, of the slice of the figure 5b.

In c and d two topographs obtained with two different diffraction vectors show the two electrical twin domains. The quality of this crystal is better, in particular in the top part, than one of the crystals of the first generation.

Anothers slices of the same crystal are shown on the figure 7. We see again the electical twin and a strong contrast due to the strains induced by the fixation of the first growth which exist in all the crystal . These defects make unusable the half of the crystal. But the top part of the sample in a seems better. The quality of

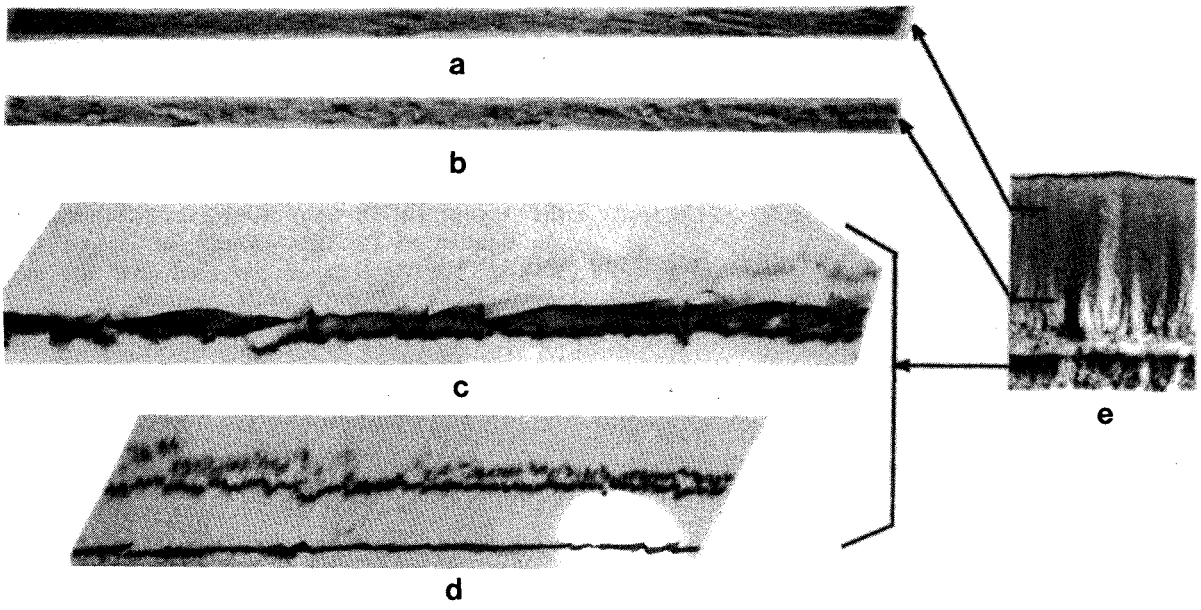


Figure 10- Section topographs of an epitaxial growth of gallium phosphate from berlinite seed obtained with an horizontal slit.

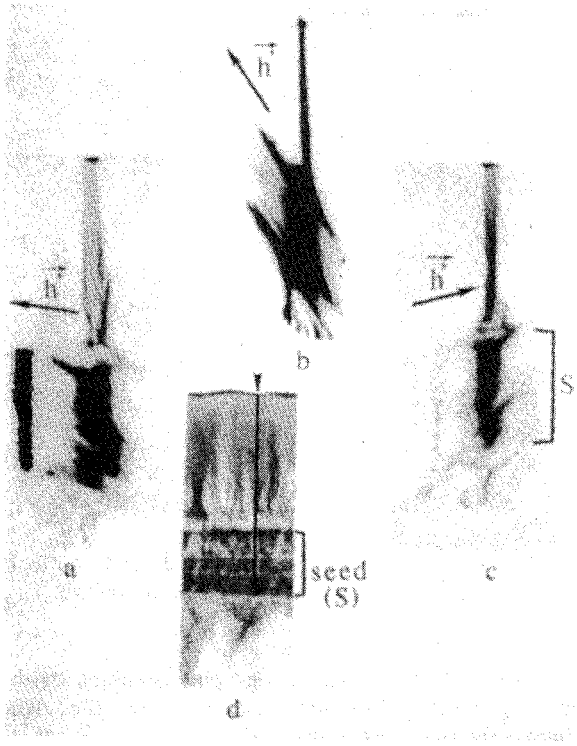


Figure 11- Section topographs of an epitaxial growth of gallium phosphate from berlinite seed obtained with an vertical slit.

images of the epitaxial gallium phosphate. The aspect of these images is due to the different values of the strain at each point of the interface. In c we have the same thing, but the seed image is like a ribbon. The two other section topographs show that the quality increase far from the seed. In b we have many defects and in a the quality becomes better.

If we put a fine vertical slit we obtain one another set of section topographs (Fig.11).

In d a traverse topograph with the X zone, the seed and the $\bar{X}$ zone indicates the position of the vertical slit. The section topographs, in a, b and c, are obtained simultaneously with different diffraction vectors; h' is the projection of the diffraction vector on the film. In a, the image of the seed appears very well on the left, the images of the two gallium phosphate layers are not separated because the strong strains give over large images. On the top the gallium phosphate of the X zone becomes better progressively far from the seed. At the interface between the seed and the X zone no image appears showing that there is an amorphous zone, alone the epitaxial gallium phosphate seems to be unbroken. This amorphous zone is also visible on the traverse topograph. In b and c two anothers section topographs. In c, the image of the seed, on the left, is very pale and the unbroken gallium phosphate between the seed and the X zone can be seen again.

This result is very interesting and shows how to obtain large good seeds of gallium phosphate.

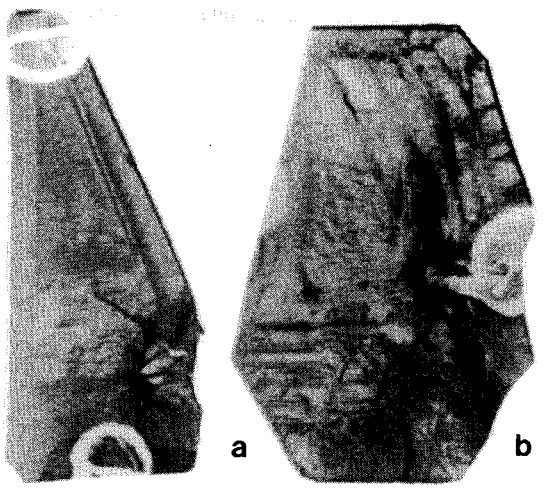


Figure 7- Topographs of two others slices of the crystal of the figure 5a.

the sample in b is worse because this slice is cut in the middle of the crystal and therefore is near the seed. The quality becomes better when the slice is cut far from the seed.

This way does not seem to be the good way to obtain crystals with a good quality and large sizes. Also another approach is necessary.

This new approach consists of the use of a berlinite seed for the growth of the gallium phosphate. Indeed, large berlinite seeds (four to eight centimeters long) are now available.

We become by a growth with a Z berlinite seed in sulfuric acid and VTG method. V_X being the only significant growth rate, flat crystals are obtained with a very thin epitaxy of gallium phosphate in the Z direction. An example is given in figure 8. We continue the growth from this crystal in phosphoric acid and composite gradient method to have a significant V_Z growth rate. This method, with nutrient in the colder side of the glass autoclaves, allows to obtain larger crystals than the SHT method. The fluid circulation, from cold to hot zone due to the horizontal temperature gradient, is done at constant temperature or coupled to a slow heating rate of 0.25 to 1.0°C/day. In both cases, quite good results are obtained for the crystal growth experiments.

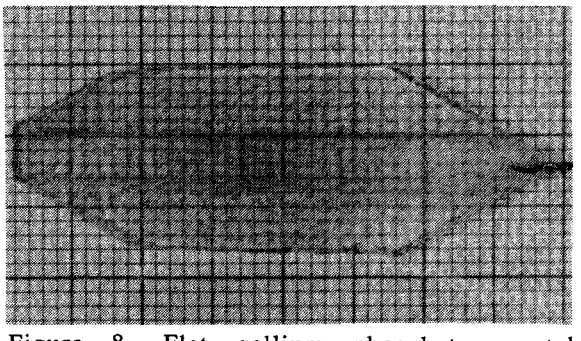


Figure 8- Flat gallium phosphate crystal obtained from Z berlinite seed in sulfuric acid and VTG method.

Then we can cut gallium phosphate X seeds for a growth in sulfuric acid and the VTG method or Z seeds for a growth in phosphoric acid and the SHT method or the composite gradient method.

Topographs of the X zone and the top part of the berlinite seed and of the berlinite seed and a part of the $\bar{X}$ zone of the crystal shown in the figure 8 are in the figure 9 a and b respectively. We can see, at the interface between the berlinite seed and the gallium phosphate X zone, strains which decrease when the distance to the seed increases. To analyse the epitaxial growth of the gallium phosphate on the berlinite seed the X-ray section topography is a very good method.

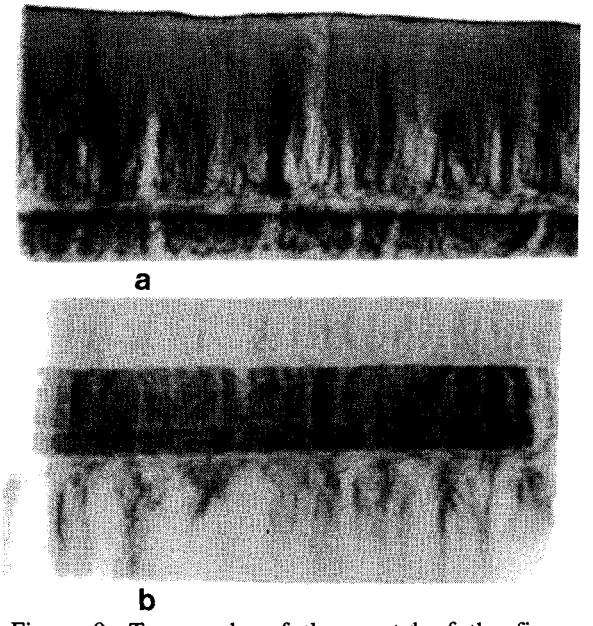


Figure 9- Topographs of the crystal of the figure 8 showing the berlinite seed.

The results are shown in the figure 10. In e the same traverse topograph that in the figure 9a and in a, b, c and d different section topographs obtained with a fine horizontal slit.

In c and d two section topographs made at the seed-level. We have here the case where the crystal is made with 3 parts, a berlinite substrate, the seed, and two epitaxial gallium phosphate growth on both sides. These two section topographs are obtained with two different diffraction vectors. Due to the strains, separated images for the different parts are obtained. In d we can see, at the bottom, the image of the berlinite seed and above the two

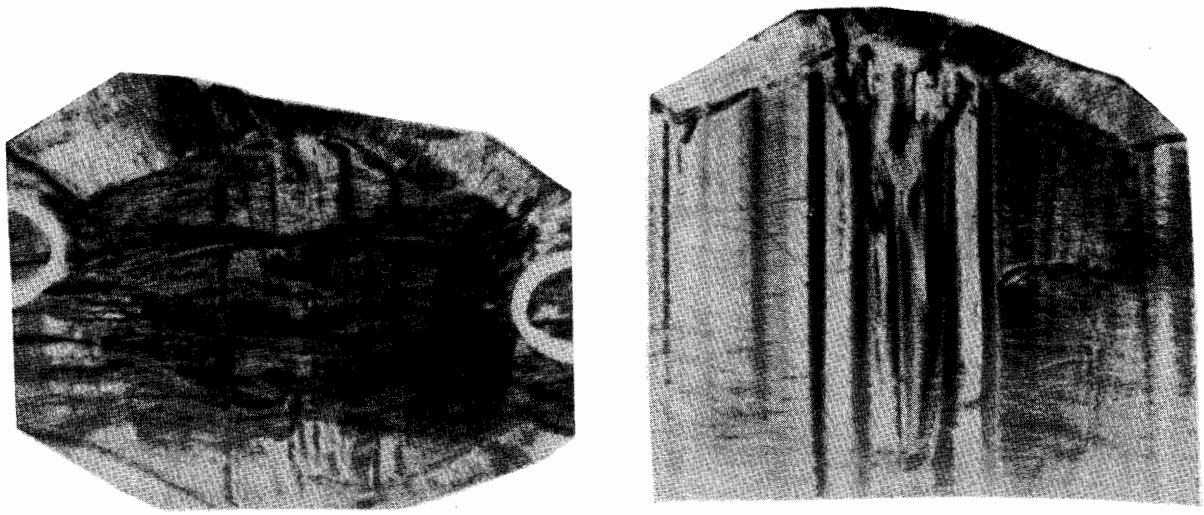


Figure 12- Topographs of two crystals growth from gallium phosphate seeds obtained from a growth on a berlinite seed.

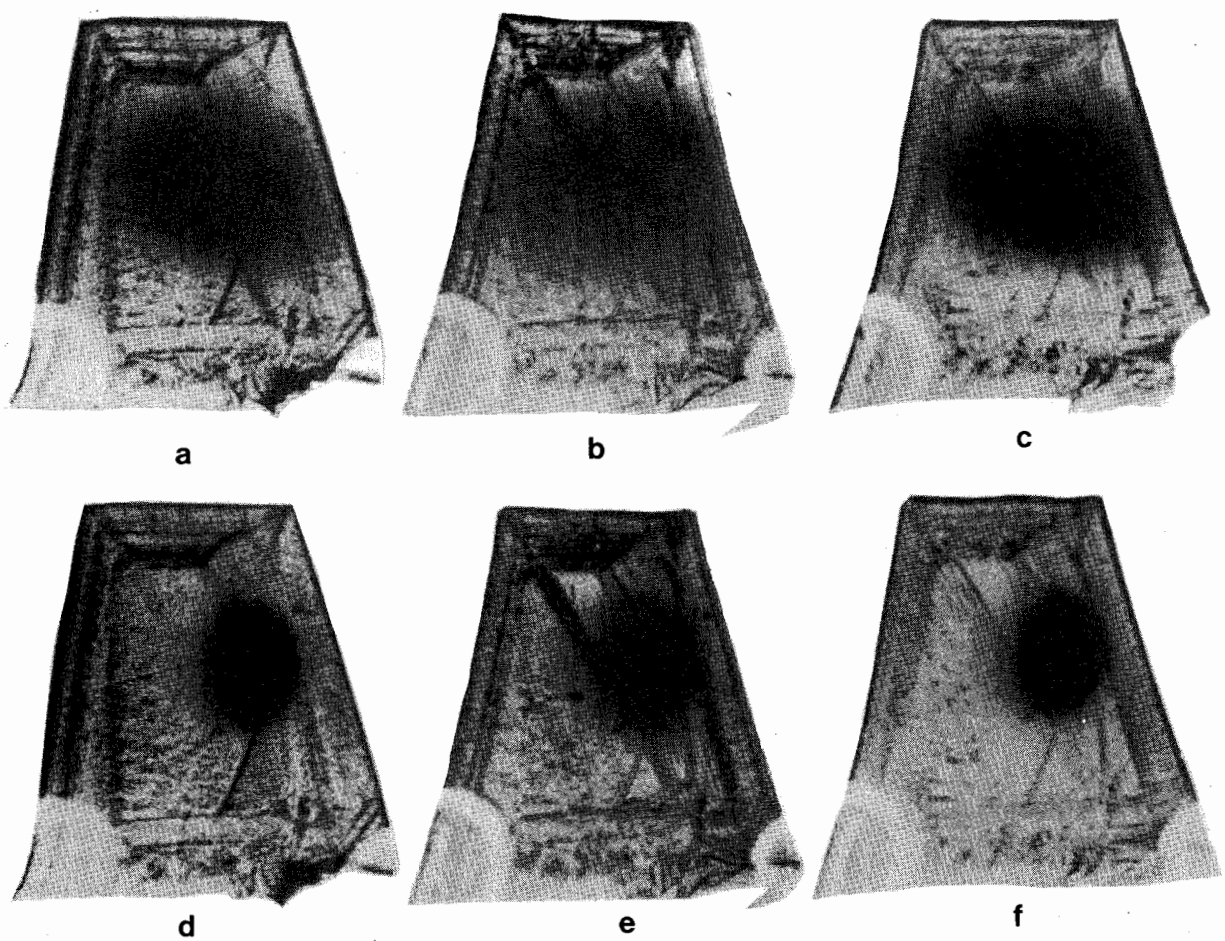


Figure 13- In a, b and c three topographs of the fundamental mode (about 6 MHz) of a gallium phosphate resonator recorded simultaneously in the same Laue pattern with three different diffraction vectors. In d, e, f three topographs of the third overtone.

First examples of growth from this type of seed are shown in the figure 12. The better crystal of these two crystals, in b, has been obtained in phosphoric acid and SHT method with successive growths. The conditions are not yet optimized but the quality is good. We have no twins and we can see individual dislocations. We have used the same method to obtain the crystal in a, but it presents more dislocations.

Resonators

We have made some observations of vibration modes of an approximately rhombohedral face cut and a near AT thickness shear resonators.

The crystal shown in the figure 13, a slice approximately parallel to a face of the rhombohedron, has been obtained from a seed cut in a crystal obtained from a spontaneous nucleation by the SHT method in phosphoric acid. The three topographs of the fundamental mode (about 6 MHz) in a, b and c have been recorded simultaneously in the same Laue pattern with three different diffraction vectors. We can see in b the apparition of a flexion mode coupled with the principal one. This is the third overtone which is on the three topographs d, e, f recorded also in the same Laue pattern.

The modes interact with the defects, mainly growth sector boundaries, which shift the modes with regard to the electrodes. The electrodes are visible in c.

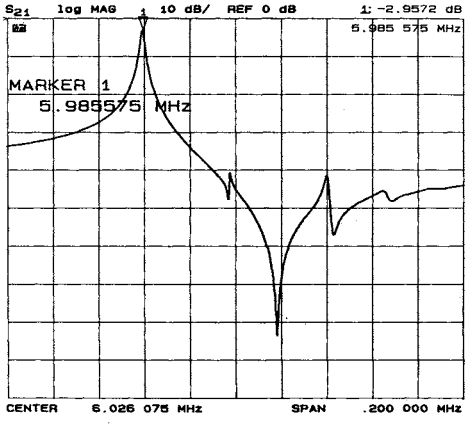


Figure 14- The response curve of a Y - 15° cut gallium phosphate resonator. The frequency of the fondamental mode is about 6 MHz and the Q factor is equal to 11.000.

We have also observed a crystal growth from a gallium phosphate seed obtained from a berlinite seed. The figure 14 shows the response curve of a resonator made with this crystal. It is a Y - 15° cut, the frequency of the fondamental mode is about 6 MHz and the Q factor is equal to 11.000. The principal mode is weakly attenuated and some anharmonic modes including one asymmetrical can be seen.

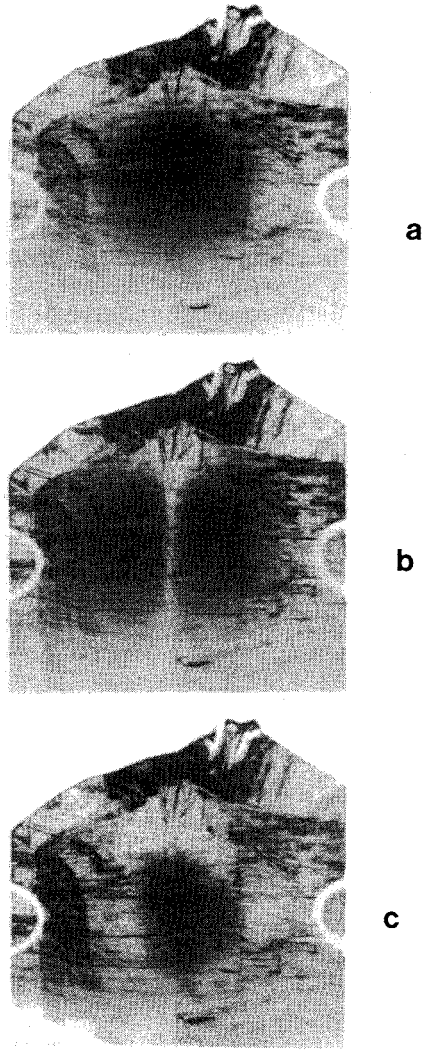


Figure 15- Topographs of modes corresponding to the figure 14. In a the fondamental mode appears with coupled modes on the top of the topograph. In b the asymmetrical anharmonic mode and in c the fifth overtone mode.

The topographs of these modes are shown in the figure 15. In a the fondamental mode not very confined but more than in the quartz, appears with coupled modes on the top of the topograph. In b we can observe the asymmetrical anharmonic mode and in c the fifth overtone mode. All these modes interact with the dislocations like you can see it in particular in the middle of the topograph in c, moreover the boundaries of the vibrating parts are diffuse.

CONCLUSION

In this investigation, it has been demonstrated the feasability of gallium phosphate epitaxy on large berlinite seeds in sulphuric acid medium. From these new seeds, the crystal growth of large gallium phosphate crystal in phosphoric acid medium can be undertaken.

From these first characterizations, the good quality of the gallium phosphate epitaxy on berlinite seeds appears. This method is a good solution to obtain gallium phosphate crystals with a good quality and large sizes.The first results of the resonators are very promising.

[1] P.Krempl, J.Stadler, W.Wallnöfer, W.Ellmeyer and R.Sclic; Proc. 5th European Frequency and Time Forum, Besançon (1991) p.143.

[2] J.Detaint, J.Schwartzel, C.Joly, B.Capelle, A.Zarka, Y.Zheng, Y.Toudic, E.Philippot; Proc. 6th European Frequency and Time Forum, Noordwijk (1992) p.383-388.

[3] A.Zarka, B.Capelle, J.Detaint , J.Schwartzel; Proc. 41st Annual Frequency Control Symposium, Philadelphia (1987)

[4] J.C.Jumas, A.Goiffon, B.Capelle, A.Zarka, J.C.Doukhan, J.Schwartzel, J.Detaint, E.Philippot; J. of Cryst. Growth,80 (1987) 133.

[5] E.Philippot, A.Goiffon, M.Maurin, J.Détaint, J.Schwartzel, Y.Toudic, B.Capelle, A.Zarka; J. of Cryst. Growth, 80 (1990) 713-726.

ELECTRIC RELAXATION BY MOVEMENT OF IMPURITIES IN QUARTZ.

C. POIGNON,G. JEANDEL and G. MORLOT.

Laboratoire Infrarouge Lointain.
Faculté des Sciences B.P.239.
54506 VANDOEUVRE-LES-NANCY.FRANCE.

ABSTRACT.

We have made measurements of thermoionic current (I.T.C.) in quartz samples in order to study the mobility of ionic impurities in this material. The diversity of studied samples,donated by the S.I.C.N. will permit a better characterisation of the purification process known as 'sweeping'. Plates extracted from crystals of different qualities, cut perpendicularly to the Y or Z axes have been studied. Certain plates have been cut both near to and far from the germ, whereas others were treated before analysis. The thermoionic currents of the different plates were compared and the effect of the sweeping on the I.T.C. current was observed.

Keywords: quartz , impurities , sweeping , Ionic Thermal Current.

1. INTRODUCTION.

Quartz is a piezoelectric material which is used in many electronic systems because of its favourable electrical properties, and is consequently the object of many studies. A very high temporal stability of frequency is often necessary for some sophisticated systems.However,defects such as impurities and dislocations can appear during the growth and the purification of a quartz crystal which can lead to changes in the quartz resonator's performance level . In addition to hydrogen which is systematically present in such crystals, the major defect seems to be an Al^{3+} ion which substitutes Si^{4+}. In order to conserve electroneutrality of the crystal lattice, a positive compensating charge is necessary. Such compensating charges can be provided by an alkaline ion $M^{+}(Li^{+},Na^{+},K^{+})$ or by a proton in an interstitial position, or hole, located in an oxygen ion adjacent to an Al^{3+} ion. Such a charge located near to each substitutional ion produces centres $Al\text{-}M^{+}$, $Al\text{-}H^{+}$, Al-hole that are detectable with different techniques. Dislocation density is studied by X ray topography (ref.1). The proton forms an OH^{-} radical which is infrared active (refs.2,3), the $Al\text{-}M^{+}$ centres can be studied by dielectric relaxation measurements (ref.4) or by acoustic loss measurements (refs.5,6), and the Al-hole group can be detected by E.S.R. techniques. These centres can be destabilized by electrodiffusion or by irradiation.

The sweeping technique is used to restrict the number of impurities in the material. An electric field is applied to a quartz sample (parallel to **c**) which is heated to approximately 500°C. Thermal agitation can liberate charged particles which have the least tendancy to bind onto the crystalline structure. These ions are then drawn towards the surface of the crystal along channels parallel to **c** under the action of the electric field. Although employed in industry for a long time, the sweeping technique is still poorly understood from a theoretical point of view.

Since quartz is an insulating material (Eg≈9eV) (ref.7), and its electrical conductivity is ionic, the thermoionic currents for several high purity samples were measured using the I.T.C. (Ionic Thermal Current) method, in order to determine the best conditions for sweeping.

2. THE I.T.C. METHOD.

The first measurements of this type were made by Bucci (ref.8).

2.1 PRINCIPLE.

An electric field is applied to a sample,then either a displacement of charge around the equilibrium position (rotation of dipoles) or a displacement of charge inside the material from one equilibrium position to another (movement of space charge) (ref.9) is created. The polarization caused by these effects is then measured using either insulating or blocking electrodes which are applied to the sample. Two kinds of I.T.C. measurements exist (refs.10,11):

-T.S.D.(Thermally Stimulated Depolarization)
-T.S.P.(Thermally Stimalated Polarization)

The T.S.D. measurement consists of applying electric tension to the terminals of a sample at a fixed temperature To, so as to move the ions that are least binded to the crystalline structure from their equilibrium position. In general,such easily moved ions are impurities.After the sample is cooled, the electric tension is removed and the depolarization current is measured while the temperature is raised.

For the T.S.P. measurement, the electric field is applied as soon as the temperature To is reached and the current measurements are carried out under the electric field.

The T.S.D. method was chosen because of the inherently higher precision offered by this method.

Blocking electrodes of chromium (the metallization is performed by evaporation directly onto the sample in a vacuum) were used since this material is employed for the manufacture of resonators in the same conditions.

2.2 INDUSTRIAL TREATMENT: SWEEPING.

Sweeping is an industrial process used for the purification of crystals, and is based on the principle that the α-quartz structure contains large channels parallel to **c** where the interstitial impurities can circulate. King (ref.12) was the first to develop this process in order to improve the quality of the material by changing the concentration of interstitial cations.An intense electric field (2500 V/cm) is applied to terminals of the crystal at a temperature of approximately 500°C for several days. The impurities (M^+) which are susceptible to the change in electrical conditions move to the anode under the effect of the temperature and the electric field, and form an alkaline deposit. After a period on the order of 10 to 12 days, the crystal is naturally left to cool to ambient temperature, at which point the tension is removed.and the extermities of the sample are then either cut off or removed by polishing.

The I.T.C. method was chosen for the study of the ionic impurities in the quartz samples because of the physical similarities between the process of electrodiffusion and this method. The measurement of thermoionic currents can be used to better characterize the mobility of impurities as a function of the temperature. The effect of the intensity and the duration of the application of the electric field on the purification can be measured.It should thus be possible to improve the conditions under which sweeping is performed through better understanding of the influence of different parameters involved in the process.

2.3 THEORETICAL STUDY.

Before examining the experimental results,the relaxation by rotation of dipoles and displacements of space charges are studied since these are the phenonema which allow the electrical current to be measured.

2.3.1 Relaxation by rotation of dipoles.

If it is supposed that relaxations are not interactive and a single relaxation time τ can therefore be considered for a given phenomenon:(ref.11)

$$\tau(T) = \tau_o \exp(E/kT)$$

where τ_o is a constant and E is an activation energy.

If it is furthermore supposed that the orientation of dipoles is alone responsible for the creation of a depolarizing field when the medium is subjected to an electric field. The contribution of ionic and electronic currents to the depolarization is considered negligible.In the absence on an electric field, a rize in temperature (linear in the case: T=bt+To) causes a current to appear.
The current density is given by:

$$j(t) = -\frac{dP}{dt}$$

A peak of the current at the temperature Tm corresponds to a maximum in j(t). If P=Po at To,then:

$$\left(\frac{dj}{dT}\right)_{T=Tm} = 0$$

and

$$\frac{E\ b\ \tau_o}{k\ Tm^2} \exp\left(\frac{E}{kTm}\right) = 1$$

2.3.2 Movement of space charges.

However if it can be supposed that the depolarizing field is due to solely space charges, the density of current can be written:

$$j(t) = -\frac{dP}{dt}$$

and

$$j(t) = [\ \sigma\ ]\ E$$

where $[\ \sigma\ ]$ is the tensor of conductivity.

Quartz is an anisotropic material, which means that the conductivity is not the same in all directions. As quartz is not pyroelectric,its polarization is due uniquely to the effect of the electric field:

$$P_i = \varepsilon_o\ \chi_{ij}\ E_j$$

where $(i,j) \in \{x,y,z\}^2$

and $[\ \chi\ ]$ is the tensor of susceptibility.

P and **E** are not colinear, $[\ \chi\ ]$ can not reduced to a simple constant. If the quartz crystal is cut perpendicularly to **c** , the field **E** is applied in the direction parallel to Z (**c**). This can be expressed by:

$$P_i = \varepsilon_o\ \chi_{iz}\ E \qquad \text{with } \mathbf{E}(0,0,E)$$

and

$$j_i = \sigma_{iz}\ E$$

The conductivity conforms to the Arrhenius law:

$$\sigma_{iz} = \frac{\sigma_{io}}{T} \exp\left(-\frac{E_i}{kT}\right)$$

where E_i is an activation energy in the direction i (i=x,y,z) and σ_{io} is a pre-exponential factor.

The current density can then be written:

$$j_i = -\frac{dP_i}{dt} = \sigma_{iz}\ E$$

The measurement is carried out in the direction Z (**c**),

$$j_z = -\frac{dP_z}{dt} = \sigma_{zz}\ E$$

$$j_z = \frac{P_{oz}\sigma_{zo}}{\varepsilon_o\chi_{zz}T} \exp\left[-\frac{E_z}{kT} - \frac{\sigma_{zo}}{b\varepsilon_o\chi_{zz}} \int_{To}^{T} \exp\left(-\frac{E_z}{kT'}\right)\frac{dT'}{T'}\right]$$

The peaks of current correspond to the maxima in j_z at Tm:

$$\frac{\sigma_{zo}kTm}{b\varepsilon_o\chi_{zz}(E_z-kTm)}\exp\left[-\frac{E_z}{kTm}\right]=1$$

This relation allows one to determine the activation energy of the impurity corresponding to a given peak.It is necessary to know the values of σ_{zo} and χ_{zz} in order to use this formula.

If the quartz crystal is cut perpendicularly to Y,the electric field is applied in the Y-direction and the measurement of the current density is carried out in the Y-direction . In this case,the index z is replaced by the index y in the preceding formulas.

3. RESULTS AND DISCUSSIONS.

3.1 CHARACTERIZATION OF SAMPLES.

Two sorts of crystals were studied: one of high quality (n°272) and the other of standard quality (n°511). Several blocks were extracted from a single quartz crystal. Plates were cut from each of these blocks perpendicular to the Y or Z axes. The following plate configurations were studied:
-Untreated plates, cut both near to and far from the germ;
-Plates purified by sweeping either in air or in a vacuum. These plates were cut at both the cathode and the anode.
This diversity of conditions will permit us to determine the influence of different parameters.

3.2 EXPERIMENTAL CONDITIONS.

The blocking electrodes used were made of chromium.A tension of 1600 V was applied to each sample for 14 hours at 350 K. After cooling, the temperature was raised linearly at rate of 0.05 K/s. All plates cut perpendicular to the Z axis were 0.75 mm thick. Since the electrode surfaces differred from one sample to another, it is preferable to present the evolution of current densities as a function of the temperature and not of the intensity itself.

3.3 PRESENTATION AND INTERPRETATION OF CURVES.

Several curves have been already published and commented on.(ref.13)

3.3.1 Standard quality samples:n°511.

The current density of a 511 sample (cut Z),purified in air, from which plates were drawn near to the cathode is presented in the Figure 1.That of a sample drawn near to the anode obtained in Figure 5 of reference 13 is shown for the sake of comparaison.It can be seen that the two curves are practically identical,with a large peak appearing in both. It appears that the measurement of current density as a function of the temperature does not depend on the spot from which the sample plates were obtained from the original crystal.

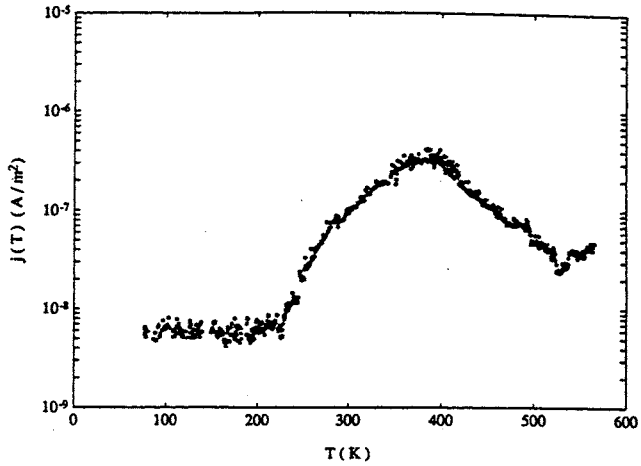


Figure 1. Current density T.S.D. for a plate of a 511 sample which underwent sweeping in air. Plate cut near the cathode.

Figure 2 shows the evolution of the current density in a plate cut perpendicular to the Y axis from a 511 sample, purified in a vacuum. The curve is similar to that of a sample with the same characteristics, but purified in air. It thus appears that sweeping in air and in a vacuum leads to the same current density in the Y-direction. Furthermore, it can be seen that the current density in the plate swept in air and cut perpendicularly to Y is inferior to that of the plate swept in air, but cut perpendicularly to Z.This can be explained by the anisotropy of the quartz structure, which favours conductivity in the Z-direction.

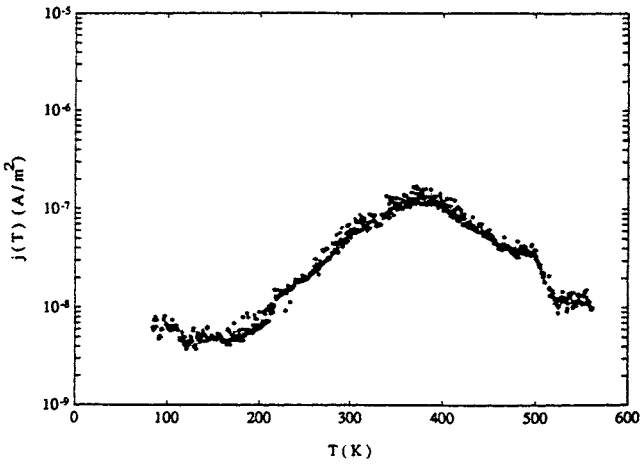


Figure 2 . Current density T.S.D. for a plate of 511 sample, cut perpendicularly to Y, swept in a vacuum.

3.3.2 High quality samples:n°272.

Figure 3 presents current densities measured in the high quality quartz samples , cut perpendicularly to the Z axis, and taken near the germ (n) and far from it (f). The curves are qualitatively similar over the range of 100 K to 400 K.Over this range,they both have peaks at 102 K, 144.5 K, 213 K, 327 K. Nevertheless, the peaks of curve (n) are higher than those of curve (f). However, the fifth peak appears at a temperature of 475 K in curve (f),but at a lower temperature of 425 K in the curve (n). Additionally,

there appears to be two additional peaks at approximately 280 K and 355 K in curve (n).

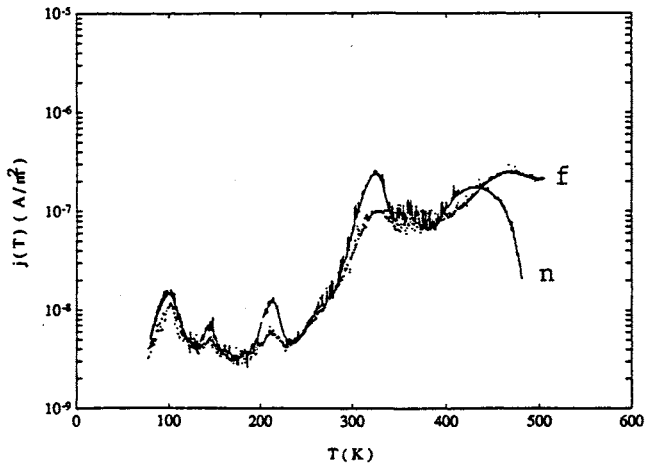


Figure 3 . Current density T.S.D. in a plate drawn near to the germ (n) and a plate drawn far from the germ (f).

The current density in a plate swept in a vacuum and cut from near the anode is shown in Figure 4. There are two peaks,one near 212 K and the other near 470 K. The current density in this sample is obviously lower than the one in the untreated sample.It can then be concluded that sweeping effectively reduced the impurity concentration(s). However,the two peaks which are seen in Figure 4, are also present in the curves corresponding to the unpurified samples shown in Figure 3. The impurities which correspond to these peaks are thus not eliminated by sweeping carried out in a vacuum.

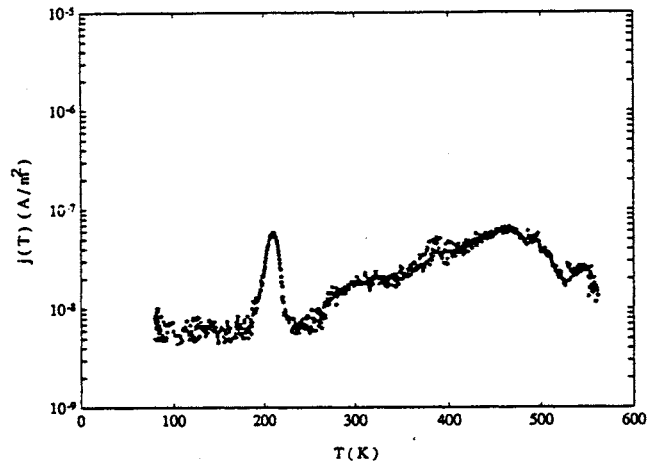


Figure 4 . Current density T.S.D. for a plate swept in a vacuum, cut near the anode.

4. CONCLUSION.

The results of this study of a 272 quartz sample confirm those found in the earlier examination of the 511 sample (ref.13),namely: more impurities are found near the germ.

The current density in the sample cut perpendicularly to the Y axis is lower than that in the sample cut perpendicularly to the Z axis. This can be explained by the anisotropy of the quartz structure which favours conductivity in the Z direction.

It has been demonstrated that sweeping has an effect on the thermoionic current. The current density is lower in a sample which has been purified by sweeping than in an untreated sample For the sample 272, sweeping eliminated certain,but not all,impurities in sample number 272 since two peaks remain in the curve of the purified sample. Sweeping in a vacuum does not remove these impurities.

Further studies will be carried out in order to determine more precisely the influence of both the intensity of the applied tension and the polarization time on the measurement of the thermoionic current. The identification of impurities which correspond to the peaks observed in the current-temperature curves will also be examined.

BIBLIOGRAPHY.

(ref.1) DETAINT J.,PHILIPPOT E.,JUMAS J.C., SCHWARTZEL J.,ZARKA A.,CAPELLE B and DOUKHAN J.C, "Crystal growth, physical characterization and baw devices applications of berlinite" , PROC. 39th FREQ. CONT. SYMP., 234-246,(1975).

(ref.2) KATS, PHILIPS RES. REP., 17, 133 (1962).

(ref.3) PARK AND NOWICK, "Dielectric relaxation of point defects in α-quartz" , PHYS. STAT. SOL. (a) 26, 617 (1974).

(ref.4) BACHHEIMER J.P. and DOLINO G., "Study of quartz sweeping in α and β phases", PROC. 43th FREQ. CONT. SYMP.,(1989).

(ref.5) FRASER D., PHYSICAL ACOUSTICS V . ed.MASSON , p.59-108 (1968).

(ref.6) MARTIN J.J., "Aluminum - related acoustic loss in AT-cut quartz crystals", J. APPL. PHYS. Vol. 56 n°9 (1984).

(ref.7) JAIN AND NOWICK, "Electrical conductivity of synthetic and natural quartz crystals", J. APPL. PHYS. 53(1), 477 (1982).

(ref.8) BUCCI C., "Ionic thermocurrents in dielectrics", PHYSICAL REVIEW Vol.148 n°2 816 (1966).

(ref.9) JEANDEL G.,WARIN A.,MORLOT G., "Dielectric constant and I.T.C. measurements of $BaF_2:Dy^{3+}$ " , PHYS. STAT. SOL. (a) 108, 185-187 (1988).

(ref.10) MULLER P., "Relationship between thermally stimulated depolarization and conductivity",PHYS. STAT. SOL. (a) 23,165 (1974)

(ref.11) MULLER P., "Electric relaxation currents in high-resistivity solids", PHYS. STAT. SOL. (a) 67,1,11-65 (1981).

(ref.12) KING, BELL SYSTEM TECH. J. 38, 573 (1959).

(ref.13) JEANDEL G., WARIN A., POIGNON C., AUBRY J.P., MORLOT G.,"Mesure de courants thermoioniques dans le quartz", PROC. OF 6th EUROPEAN AND TIME FORUM , Noordwijk,NL, 389-392.(1992).

OPTICAL STUDY (I.R. AND VISIBLE) OF THE X-RAYS IRRADIATION EFFECTS ON SWEPT QUARTZ (AT AND SC CUTS).

J-P. BACHHEIMER

Université Joseph Fourier (Grenoble I)
Laboratoire de Spectrométrie Physique (associé au C.N.R.S.)
B.P. 87 - 38402 Saint-Martin-d'Hères Cedex (France)

ABSTRACT

X-ray irradiation of quartz crystal can induce variations of the infrared OH absorptions and darkening (Al-hole color center) depending on the relative Al and OH impurity concentrations. These effects are correlated with resonator frequency variations which can be reduced using a preventive electrolysis process (sweeping).
We present both variations of infrared and visible spectra induced by X-irradiation of thin AT (medium quality) and SC (good quality) cuts which were previously swept in different conditions under vacuum and air atmosphere (4th E.F.T.F). It appears particularly on AT cuts that sweeping under extreme conditions such as higher temperatures and electric fields increases the radiation induced coloration and therefore could lower the radiation hardness of the resonators.

Key words : swept quartz, radiation effects, VIS and IR spectra.

INTRODUCTION

It is well known (ref. 1, 2) that radiation hardness of quartz crystal resonators is improved by a high temperature electrolysis treatment ("sweeping"). This occurs because sweeping gives preventively effects similar, but more stable, to radiation effects on the Al-M^+ ($M^+ = Na^+, Li^+$) and OH centers which are introduced during hydrothermal growth of cultured quartz. Under ionising radiation,up to saturation, the Al-M^+ and OH are dissociated and form Al-hole and Al-OH centers , in quantities depending on the relative concentration of as grown H and substitutional Al center, while the M dwell in the vicinity of the Al. Under electric field, electrolysis removes the M^+ towards the cathode, and the Al-M centers are converted into Al-OH and Al-hole centers, the proportion of which depends further on the availability of hydrogen which can be supplied from the surrounding gas. Sweeping in an air (wet) atmosphere replaces the alkali ions by hydrogen to form Al-OH, while sweeping in a vacuum atmosphere should form, at the completion, Al-hole centers after the depletion of as grown OH which supply the Al-OH centers which are temporarily observed before the completion.
It is optically possible to monitor radiation as well as sweeping effects because, (at room temperature) : i) as grown OH displays infrared peaks labeled S_4 (largest 3585 cm^{-1} or 2789 nm), S_3, S_2 , S_1 on a large background absorption (usually measured at 3500 cm^{-1} and 3410 cm^{-1}), ii) Al-OH displays two I.R. peaks at 3380 cm^{-1} ("e_2"-large) and 3306 cm^{-1} ("e_1" - weak), iii) Al-hole is involved to the A-bands smoky coloration (ref. 3, 4) which absorption peaks at 460 nm.
Radiation effects on swept sample could be in principle deduced from the above models : when the sample is insufficiently swept,we do observe an increase of Al-OH and/or Al-hole absorptions, when the sample is completely air swept,Al-OH and S_4 are not affected (apart a temporary Al-hole conversion), and when the sample is fully vacuum swept it should be also not affected as it should remain only Al-hole centers. The fact appears more complicated because some authors (ref. 5, 6, 7) and ourselves have observed sometimes (but not always) after complete air sweeping , a slight radiation-induced reduction of Al-OH and S4 strength and after vacuum sweeping a radiation induced increase of Al-hole concentration ; we have also shown (ref. 8) that this increase takes place even when Al-OH still exists before the final

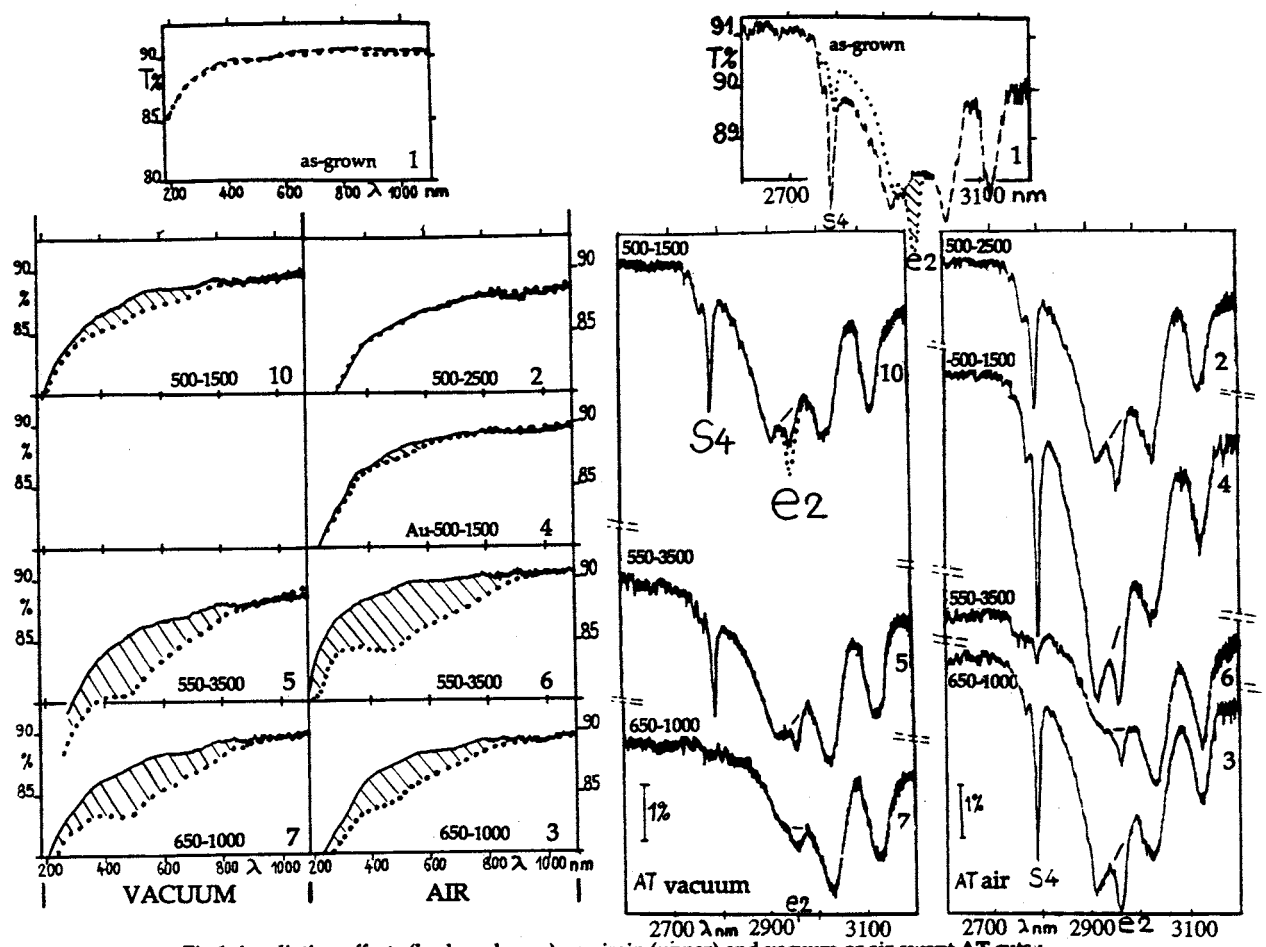


Fig.1: irradiation effects (hachured zone) on virgin (upper) and vacuum or air swept AT cuts : VIS(left) and IR (right) transmission spectra. Virgin : - - - ; swept : — ; irradiated: .....

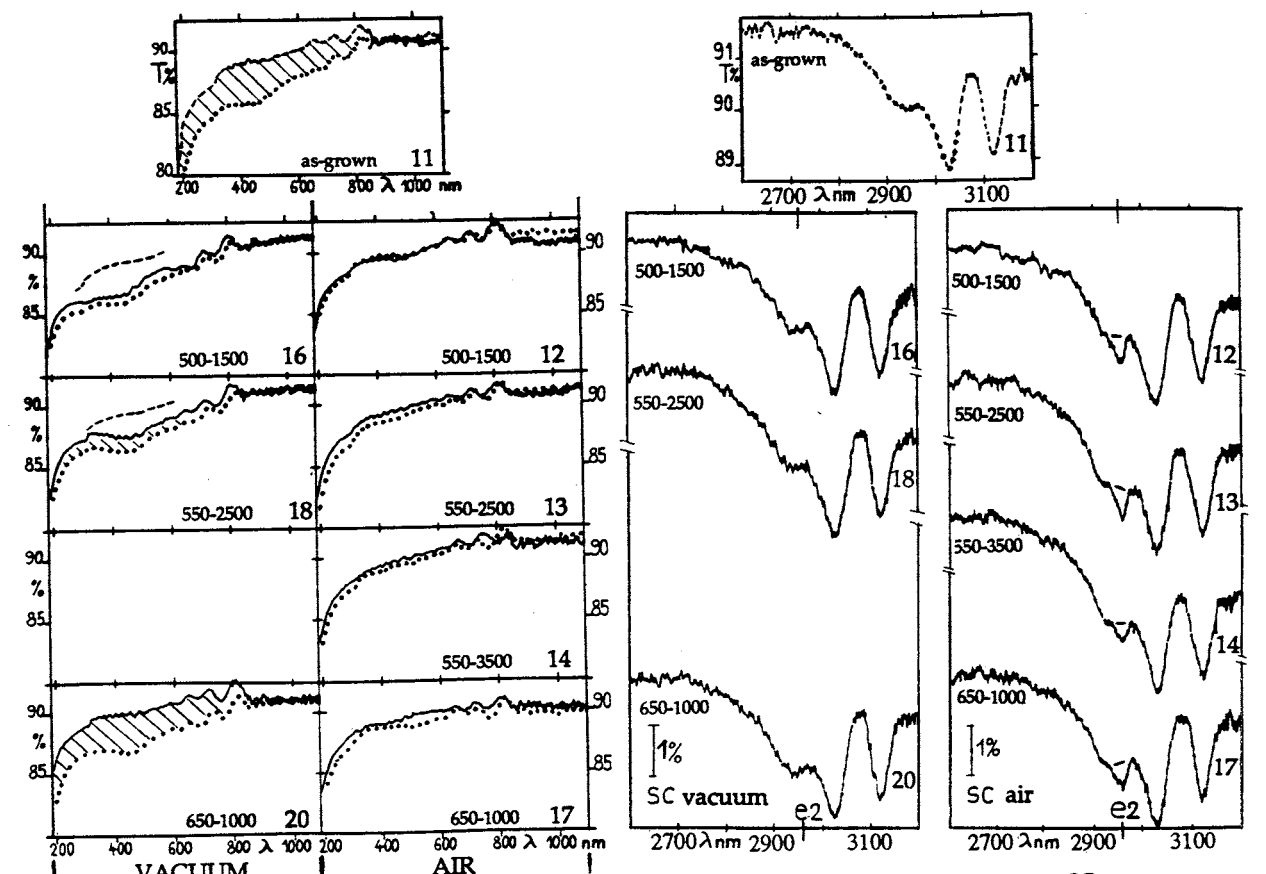


Fig. 2: irradiation effects (hachured zone) on virgin (upper) and vacuum or air swept SC cuts : VIS(left) and IR (right) transmission spectra. Virgin :- - - ; swept : —; irradiated :.....

stage of full vacuum sweeping. To summarize the effects of radiation on swept samples, one can say that when the sample is fully air swept Al-OH strength is not affected or decreases slightly, and when the sample is partly or fully vacuum swept, Al-hole increases while remaining Al-OH is constant or decreases slightly. It is also worthwhile to keep in mind that the Al-OH and Al-h centers created by irradiation are annealable to 400°C (maximum) unlike of those created by sweeping.
Frequency variations of the quartz resonators are directly connected to the variations of the strength of these centers because both are implicated in acoustical absorptions (Al-h directly, Al-OH forms at the expense of Al-Na which shows an acoustic peak and further Al-OH converts temporarily into Al-h under irradiation (ref. 9)). Radiation induced coloration is an usual test to select quartz and test sweeping effectiveness (ref. 9, 10). The value of 633 nm He-Ne laser light absorption measurement has also recently (ref. 11,12) been pointed out to scan Al-hole formation as an alternative no destructive technique to E.S.R. measurements.
In this paper we compare results of I.R. and visible spectra obtained before and after irradiation on the fourteen thin AT and SC cuts we have swept under different conditions (air, vacuum, temperature, electric field) as reported shortly in the 4th EFTF (1990) (ref 13). It has been shown (ref. 14) that there are no significant differences between Co^{60} and X-ray radiation resonator response.

EXPERIMENTAL

Irradiations were performed at room temperature (r.t.) for about 12 hours (close to saturation) with a 40 kV - 30 mA copper target X-ray generator (distance sample target about 10 cm).
All the optical spectra were also obtained at r.t. using a Lambda 9 Perkin Elmer double beam spectrometer with a beam size area limited to about 7 mm^2. Care was taken to always explore the same position in respect of the seed for each series of samples. In the ranges of interest, the light beam reveals a certain polarization so that the results are close to those which could be obtained from polarized light. To maximize the absorption $\vec{E} // Z$ = "Π" polarization for the visible spectra and $\vec{E} \perp Z$ polarization for the I.R. spectra were adapted.
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 sweeping and an estimation of the Al content deduced from the maximum Al-OH peak strength (according to our calibration)

	Samples	α_{3500} cm-1	α_{3410} cm-1	Al ppm weight
Table I	AT(n°1...10)	0.08	0.145	7 ± 2
	SC (n°11..20)	0.02	0.080	3 ± 1

Electric fields were applied by means of Pt pressed foil electrodes (gold evaporated electrodes for n° 4) on the 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. Sweeping currents, activation energies and ohmicity tests have already been reported (ref. 13)

RESULTS

Fig. 1 (AT cuts) and Fig. 2 (SC cuts) show the I.R. (right part) and UV-VIS spectra (left part) obtained after sweeping (full line) and after subsequent irradiation (dot line). Due to bad surface state particularly for AT cut, an anomalous extra absorption sometimes appears in the UV-VIS range which does not affect the conclusions. We see from the virgin samples -upper part of these figures- that besides their marked difference between their as grown OH content, as shown by I.R. spectra, the two kinds of quartz have very different [OH]/[Al] content ratios : in AT cut the e_2 (Al-OH) peak appears under irradiation,while there is no effect on the VIS spectra (therefore there is enough OH to saturate all the Al centers), on the contrary for SC cut, the radiation induced e_2 peak is negligible and there is a pronounced visible (Al-hole) absorption showing that the Al content is not negligible (this is confirmed from the air swept sample (n° 12) where Al-OH appears quite clearly with an amplitude about one half of the Al-OH obtained from AT cut).
Irradiation effects are summarized in table II in which the upper part gives the AT cuts results and the lower part the SC cuts results. We have used arrows to roughly indicate the effects of irradiation on the e_2 peak (Al-OH) and on the maximum visible (at 460 nm) absorption related to Al-hole centers (any S_4 variations are reported in the "comments" case). Thus horizontal arrows in both Al-OH and Al-hole cases mean no measurable effect of irradiation after treatment, which we can infere therefore minimum radiation induced frequency shift (due to these centers of defects).

	Sweeping conditions temp. (°C), E (V/cm), time (days) : vacu./air	VACUUM						AIR					
			Al-OH		460				Al-OH		460		
		n°	bef.	aft.	bef.	aft.	comments	n°	bef.	aft.	bef.	aft.	comments
AT	none	1	0	↗	0	→	virgin ; S4↘ [Al] < [OH]						
	500 - 1500 -15/6	10	∿X/2	↗	∿0	↗	insuf. sweep. S4↘ ; perhaps grad. defects //Z	4	X	→	∿0	∿→	sweep. slightly too long with Au évapor. électr.
	500 - 2500 - /6							2	X	→	0	→	good sweeping
	550 -2500 - 7/15 +3500	5	∿X/2	→	∿0	↗↗	no complete vacuum sweep.	6	∿X/2	→	∿0	↗↗	too energetic sweep. effects similar to vacu.
	650 - 1000 - 4/4	7	∿X/3	→	∿0	↗↗	no complete vacuum sweep.	3	X	→	∿0	↗↗	effects similar to no complete vacu. sweep.
SC	none	11	0	→	0	↗↗	virgin ; [AL]>>[OH]						
	500 -1500 - 13/7.7	16	0	→	X	→	good vacuum sweeping	12	X	→	0	→	good air - sweeping
	550 - 2500 - 10/5	18	0	→	∿X	→	slightly more radiation sensit.	13	X	→	0	→	"
	550 - 3500 - /7							14	∿X/2	→	0	→	"
	650 - 1000 - 5/11	20	∿0	→	0	↗↗	very insuf. sweep.	17	X	→	∿0	→	"

Table II : schematic representation of the effects (arrows) of irradiation on the Al-OH and 460 nm (Al- h) absorptions of the AT and SC cuts after the treatment described in the first column (bef.=before irradiation; aft.=after irradiation).
symbols : X = maximum observable intensity ; → no effect ; ↗ slight increase ; ↗↗ strong increase

The first row of the data from the investigated AT and SC cuts show the radiation response of the virgin sample.
Results from table II suggest the following comments concerning radiation hardness of the swept samples :

- AT cuts : the best air sweeping is 500°-2500V/cm-6 jours (n° 2), the gold evaporated sample (n° 4) seems to be a little less good ; increasing of temperature and electric field spoils the radiation hardness as coloration appears after irradiation; the vacuum sweepings will probably give about the same radiation response as the air swept material (apart n° 10 which appears very insufficiently swept for an indeterminate cause perhaps in relation with anomalous sweeping currents).

- SC cuts : apart from n° 20 which is practically not swept (corresponding to anomalous low sweeping currents), we obtained two typical sets of good air swept ($e_2 \neq 0$, no irradiation effect) and full vacuum swept (S4, $e_2 \equiv 0$, Al-h $\neq 0$, no or little affected by irradiation). In principle under static irradiation air-swept samples should be as good as vacuum swept sample n° 16 but for pulse irradiation the vacuum swept samples could be better (probably improved by prophylactic irradiation for n° 18). We also note here that it is not necessary to increase (at practically equal duration) temperature and electric field to improve radiation hardness of the swept samples.

CONCLUSION

From the observation of visible spectra in addition to I.R. spectra we have been able to define, with the help of irradiation tests, the achievement of the various air and vacuum sweepings we had performed on AT and SC thin cut crystals. In addition to radiation sensibility tests, both spectra are particularly useful for the SC cuts which have here a poor as grown OH content with relatively high Al content and which therefore are able to give only visible absorption after vacuum sweeping. Performing the whole extended VIS spectra instead of only 633 nm absorption allows an easier distinction for small absorptions and background corrections due to imperfections of suface polishing.
Optical radiation induced sensitivity has also shown that sometimes (here particularly for the AT air swept cuts) increasing of temperature and electric field of the electrolysis can decrease the radiation hardness because that could increase the amount of radiation induced Al-hole centers (however it is possible that this could be overcome by a reduction in the sweeping times). In this respect the role of the quartz quality has been shown.

ACKNOWLEDGEMENTS

The author thanks J.P. AUBRY (C.E.P.E. Thomson) who encourage him to these studies by furnishing the samples.

REFERENCES

[1] J.C. King and D.R. Koehler, in Precision Frequency control ch. 3 pp. 147-159 (1985) ; E.A. Gerber and A. Ballato, Eds, Academic Press.

[2] J.J. Martin, IEEE Trans. Ultrasonics, Ferro and Freq. Cont. 35, 288 (1988).

[3] N.J. Koumvakalis, J. Appl. Phys. 51, 5528 (1980).

[4] E. W.J. Mitchell and E.G.S. Paige, Philosophical Magazine 46, 1353 (1955).

[5] F.K. Euler, H.G. Lipson and P.A. Ligor, Proc. 34th An. Freq. Cont. Symp. IEEE p. 72 (1980).

[6] A. Kahan, H.G. Lipson, Proc. 39th An. Freq. Cont. Symp. IEEE p. 255 (1985).

[7] A. Kahan, F. K. Euler, H.G. Lipson and C.Y. Chen, L.E. Halliburton, Proc. 41st. An. Freq. Cont. Symp. IEEE p. 216 (1987).

[8] J. P. Bachheimer , to be published.

[9] T.M. Flanagran , IEEE Trans. on Nuclear Science NS 21, p. 390 (1974).

[10] G. Kaitz Proc. 38th An. Freq. Cont. Symp. IEEE, p. 245 (1984).

[11] J. P. Anhes, P. Garcia, and D.R. Koehler, OSA Techn. Dig. Series 7, p. 20 (1987).

[12] H.G. Lipson, IEEE Trans. Ultra. Ferr. and Freq. Cont. 37, p. 524 (1990).

[13] J.P. Bachheimer and P. Palleau, 4th Europ. Freq. and Time Forum,p.723 (1990) Neuchâtel University (Switzerland).

[14] L.J. Palkuti and Q.T. Truong, Proc. 38th An.. Freq. Cont. Symp. IEEE, p. 55, (1984).

SPACE 2

Chairman: S. Feltham

Ion Standards for Space

L. Maleki
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

A number of important and fundamental tests of relativity and astrophysics require the environment of space and the service of a space-borne ultra-stable frequency standard. Limitations of size, mass, and power available on spacecraft imply that the ideal frequency standard for these tests must have the stability of a hydrogen maser in a package similar to typical field and particle instruments.

Based on our recent success with the development of a mercury ion standard at JPL, we have determined that a trapped ion standard offers the most promise for spacecraft applications. While the stability achieved with the JPL mercury ion standard meets the requirements stated above, further work is required to meet the requirements of size , mass and power consumption. We have performed a study that has pointed to the possibility of developing ytterbium ion standard with a mass of about 15 Kg and power requirement of less than 30 W. The reduced mass and power consumption can be achieved since ytterbium ion can readily be excited with semiconductor lasers, and is also suitable for applications of optical fibers and solid state detectors. Based on this study we are developing a ytterbium ion standard and various associated subsystems to demonstrate the feasibility of this concept.

In this paper details of the study pointing to the use of the trapped ytterbium ion standard for space applications will be discussed. The progress in the development of the standard currently underway at JPL will also be presented and future areas of work will be identified.

THE ARGOS DATA COLLECTION AND LOCATION SYSTEM
System description, performances and applications

M. Brunet, G. Portas*, A. Goasguen***

*CNES, Centre National d'Etudes Spatiales, Toulouse, France
**CLS, Collecte Localisation Satellites, Toulouse, France

ABSTRACT

The Argos system offers world-wide capabilities for location and environmental data collection for fixed and moving platforms using low earth orbit satellites.

Argos is the result of a cooperative programme between the Centre National d'Etudes Spatiales (CNES, France), the National Aeronautics and Space Administration (NASA, USA) and the National Oceanic and Atmospheric Administration (NOAA, USA). The system has been operational since 1978 and uses the two Tiros-N/NOAA meteorological satellites (sun-synchronous, 850 km polar orbit).

The Argos system is managed by the two organizations, CNES and NOAA, which have cooperated in its accomplishment. The CNES provides the Argos on-board processor manufactured by Dassault Electronique. NOAA provides funding for the spacecraft, its in-orbit control and the specific ground segment. The satellites are built by General Electric under a NASA contract. A subsidiary of CNES, the CLS (Collecte, Localisation, Satellites) is in charge of the Argos ground segment operations, the relations with the customers and the promotion of the system.

What is unique about the Argos system is the availability of a location together with collection of environmental data. Several hundred thousands of bits are collected daily through the system and the location accuracy provided varies from one hundred meters to one kilometer according to the type of transmitter used. These performances correspond to the needs of the user' communities : meteorologists, hydrologists, biologists, oceanographers, physicists and others.

The location is based on a one-way random access up-link from platforms to satellite using the Doppler effect. The location accuracy depends directly on the quality of the ultra-stable oscillators (USO) used by the platforms and by the one, on-board satellite processor, which measures the received frequency.

This paper describes the general characteristics of the system, the performances of the Argos USOs, in-orbit and on the ground, and presents some typical applications.

1. THE ARGOS PROGRAMME

1.1. The Argos Mission

Argos is a satellite-based data collection and location system. Its mission is to locate fixed and mobile platforms and to collect the environmental data they are transmitting.

This system is the result of a cooperative programme between the Centre National d'Etudes Spatiales (CNES, France), the National Aeronautics and Space Administration (NASA, USA) and the National Oceanographic and Atmospheric Administration (NOAA, USA). The agreement signed in 1974 was renewed in 1986 and offers a continuous service beyond the year 2000.

The Argos system is managed by the two organisations which have concurred in its realisation : CNES and NOAA. The CNES provides the receiver-processor which is manufactured by the Dassault Electronique company. The NOAA supplies the funds for the satellites, their in-orbit control and the raw data distribution. The spacecraft are built by the General Electric Aerospace Company under a NASA contract. A CNES subsidiary, CLS (Collecte, Localisation, Satellites) is in charge of the Argos ground segment, the relations with the users and the promotion of the system.

The first launch happened in October 1978. This year, 1993, will see the launch of the ninth Argos receiver-processor.

The performances required by the Argos mission are as follows :

- In location : precision between near one-hundred meters to one kilometer, from one to a few number of locations per day.
- In data collection : processing of thousands of platforms, world-wide, with data requiring one or few collections per day. Each transmission contains up to 256 bits of data.
- A probability of location and data collection better than 95 % in a 12-hours period of time.
- A delay between data collection from platforms and the distribution to the users : between one and three hours.

1.2. The Argos system

1.2.1. The Argos system components (figure 1)

The Argos system comprises a "space segment" composed of two satellites each one with a receiver-processor, a "ground segment" which includes the telemetry acquisition stations, the processing and distribution centers and a network of orbitography beacons and a third component composed by all the user's platforms.

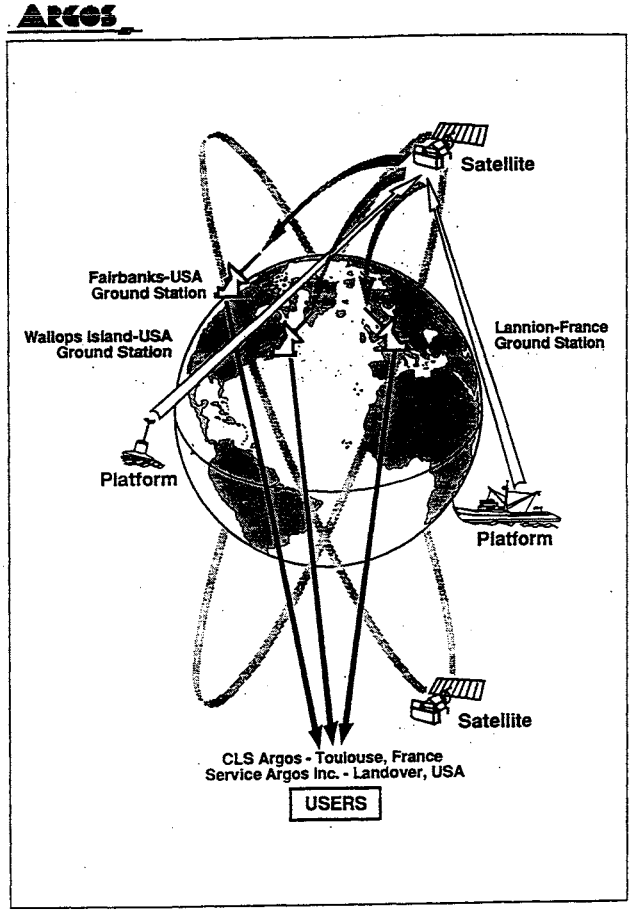


Figure 1. The ARGOS SYSTEM

1.2.2. The space segment

The receiver-processor receives on a random basis the messages transmitted by the platforms, processes them and includes them in the general telemetry of the satellite.The satellites carrying the Argos instrument are the NOAA quasi polar-orbiting spacecraft of the Advanced-Tiros-N series. The two simultaneous operational satellites are on circular orbits at an altitude of 850 km. These orbits are sunsynchronous : each satellite passes over the North and South poles at each revolution and crosses the equatorial plane at fixed local solar times. At every moment, a satellite sees all the platforms within a 5000 km diameter circle. As the satellites orbits, this circle sweeps a swath 5000 km in width encompassing the earth. As a result of the earth's rotation, two successive swaths are separated by 2800 km at the equator. The number of orbits per day is approximately 14 for each satellite as the orbital period is around 100 minutes. This choice of orbit guarantees complete coverage of the earth surface.

The number of daily passes over a given platform depends on latitude. At the poles, the satellites see all platform, each at every pass, that is 28 times a day, while at the equator this number is around seven. Pass duration however is independent of site latitude and has a mean value of ten minutes.

1.2.3. The ground segment

Three telemetry acquisition ground stations are used. Two are operated by NOAA, one in Alaska (USA) the other in Virginia (USA) and the third is at Lannion (France). They receive the raw data and distributes them to the Argos data processing centers, one in Toulouse and the other in Washington (USA) by dedicated permanent links.

The Argos data processing centers perform real-time processing of data sensors, location of platforms, creation and updating of result files, dissemination and dialogue with users.

Operating 24 hours a day, 365 days a year, the Argos centers aim at maximum availability. The data are distributed to the users via many different data communication systems.

As part of the Argos programme, CLS company, situated at Toulouse, is the focal point for all the users except the North American ones which are served by an American subsidiary, Service Argos Inc., located near Washington DC.

The ground segment includes also a network of ten orbitography beacons installed in accurately known places (geodesic points). They are used to calculate continuously the orbit parameters and are the reference system for the location determination.

1.2.4. The location principle

The Argos platform location is based on the Doppler-Fizeau effect : the frequency received by the on-board processor is a function of the radial velocity between the platform and the satellite. The platform transmitted frequency being known (all are at 401.650 MHz) and supposed constant, the received frequency at a moment "t" allows to define a locus for the platform : at the intersection between the Earth and a cone centred on the satellite, with its velocity vector as axis and a half apex angle function of the Doppler frequency. Theoretically, the reception of at least two messages at t1 and t2 allows to define two possible locations for the platform symmetrical against the satellite track on Earth. In practice, four messages are required. It is possible to resolve the ambiguity and find the true location by various means. For instance if there are more than three messages, the calculated locations for the true point are less scattered than for the image one.

Thè location precision and accuracy depends on the knowledge of the orbit, the time of the frequency measurement, on the earth-platform-satellite geometry for each message, on the number of messages available and for the most important part on the stability of the platform oscillator during the satellite pass. Depending upon these parameters, the location accuracy ranges from 150 m for very good oscillators to around 1 km. The on-board receiver - processor ultra stable oscillator (USO) is specified to introduce a negligible error on the location compared to the other factors.

1.2.5. The random access data collection

Each platform transmits independently from the others a short message (0.36 s to 0.92 s) with a repetition period from one to some minutes. The processor receives messages at random versus time but also with random frequencies : the intensity of the Doppler effect depending on the satellite-platform geometry, spreads the received frequencies in a ± 9 kHz range. If two messages are received simultaneously but with enough frequency separation they are correctly processed.The current receiver-processor can process four messages at a time. If simultaneous messages are too close in frequency, they interfere and are lost. The random access data collection is characterised for a given message, by its probability to be correctly processed. The mean duration of a satellite pass being around ten minutes, for a platform transmitting every minute to be located requires four messages correctly processed out of the ten transmitted during the pass. Only one out-of-ten would be required for a data collection without location.

We have already seen that the number of passes per day varies from a minimum of six at the equator to a maximum of 28 at the poles. For the mean latitudes, this number is around twelve.

1.2.6. The platforms and the Argos's users

Generally speaking a platform is any set of sensors, fixed or mobile, maritime, terrestrial or aerial, equipped with an Argos transmitter periodically transmitting at 401.650 MHz.

The Argos system offers the possibility to automatically collect and transmit data from hardly accessible or scarcely inhabited regions as well as from dangerous environment. Maritime platforms are used as fixed or drifting buoys as well as on board various ships for the study of meteorology, oceanology, fishing. Land platforms are used in as diverse fields as hydrology studies of snow and ice, drifting icebergs, volcanology, seismology, animal tracking or balloons borne instrumentation and again meteorology. More than 2000 platforms are currently in activity.

The Argos system is managed by an Operations Committee, a bilateral (France USA) organisation made up of representatives of CNES and NOAA agencies with the participation of NASA.

2. SPACE SEGMENT : THE RECEIVER-PROCESSOR

2.1. Electrical Schematic diagram (figure 2)

The receiver-processor electrical schematic diagram is shown in figure 2. It comprises :

- a UHF receiver centered on 401.650 MHz, an ultra stable oscillator (USO) which is the reference for the time counter, the local oscillators of the receiver, the Doppler frequency counter and creates the clocks for various electronic circuits,
- four Data Recovery Units (DRU), each including a phase lock loop (PLL) to process simultaneously four messages,
- a spectrum analyzer for the detection of messages and pre-tuning of PLLs,
- a control unit which monitors the processor, assigns the detected messages to the DRU available, according to an algorithm which optimises the message probability access,
- a telemetry formatter which creates the down link messages. These telemetered messages contain the sensor data and the two data needed to locate the platform : the Doppler frequency measurement and the time of this measurement.

The main characteristics and performances of the receiver-processor are summarised as follows :

- total mass of the units composing the instrument : 19 kg,
- power consumption : 19 W,
- receiver frequency band : 24 kHz centered on 401.650 MHz,
- power level of the received messages: from -161 dBW to -138 dBW,
- correct processing probability : > 99%,
- bit error rate: $< 2.10^{-5}$,
- Doppler frequency measurement precision: standard deviation: ≤ 350 mHz,
- USO stability: 1.10^{-10}/day and 1.10^{-10}/100 ms (see thereunder chapter 4),
- theoretical reliability after two years in orbit: > 90%.

2.2. From Argos-1 to Argos-2

The Argos programme being a recognized success, CNES decided in 1986 to develop a second generation of receiver-processors for a three fold increase in the system capacity while keeping the same range of location precision. The new receiver-processor, Argos-2, includes eight DRUs instead of four and the received band is widened from 24 to 80 kHz. Due to a large use of digital technology (FFT spectrum analyser, digital PLL...) these upgraded performances have been reached within an acceptable increase of mass and power consumption.

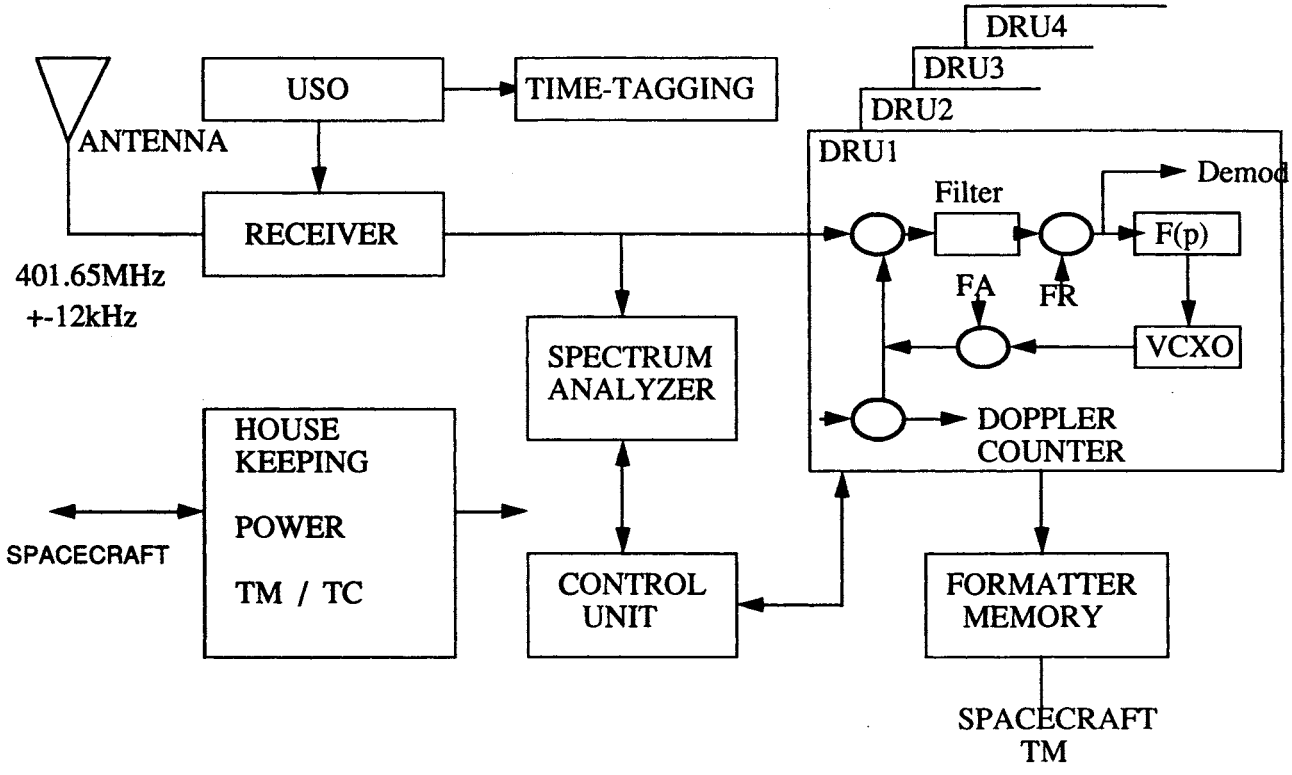


Figure 2. On board ARGOS Electrical diagram

2.3. The Argos Launch List

The following launch list is an illustration of the continuity of the Argos programme from 1978 up to 2000 and beyond... The TIROS-N and NOAA-A through J spacecraft are equipped with an Argos-1 generation receiver-proccessor, while the new spacecraft, NOAA-K through N, currently under development will be provided with a second generation one.

Spacecraft name on ground	Spacecraft name in orbit	Launch date	Operational life time
TIROS-N	TN	13 October 1978	2 years 4 months
NOAA-A	NOAA-6	27 June 1979	6 years 1 month
NOAA-B	-	29 May 1980	Launch failure
NOAA-C	NOAA-7	23 June 1981	3 years 7 months
NOAA-E	NOAA-8	28 March 1983	2 years 6 months
NOAA-F	NOAA-9	12 December 1984	3 years 11 months
NOAA-G	NOAA-10	17 September 1986	4 years 11 months
NOAA-H	NOAA-11	24 September 1988	4 years 5 months in operations (*)
NOAA-D	NOAA-12	14 May 1991	1 year 9 months in operations (*)
NOAA-I	-	planned in June 1993	-
NOAA-J	-	in 94	-
NOAA-K	-	in 95	-
NOAA-L through N	–	in 97 and beyond	-

(*) as of February 1993

3. GROUND SEGMENT: (FIGURE 3)

The recorded data are dumped each time a satellite is in visibility of one of the three ground receiving stations (Fairbanks, Alaska and Wallops Island, Virginia in the USA, Lannion in France).

They are then forwarded to the Toulouse and Washington Global Processing Centers (GPC) by permanent dedicated links.

Sensor data processing (conversion of received data into digital form or user-defined physical units) and location determination are performed at both GPCs. Location accuracy is basically a function of Platform Transmitter Terminal (PTT) oscillator stability and figures around 350 m are typical, irrespective of platform position. Value-added services extend the range of processing capabilities. Under certain conditions location accuracies to 50 m can be derived.

Platform location is achieved solely by measuring the Doppler shift in the carrier frequency of received messages. The transmit frequency must therefore remain stable throughout the pass. Each frequency measurement yields a set of possible PTT positions, the locus of which is a cone with the spacecraft at the apex and its velocity vector as the axis. Assuming the platform elevation is known, the intersection of these cones with the elevation sphere yields the solution. The location is valid if the spacecraft receives at least four messages during a single pass.

The primary sources of error are:

- accuracy of PTT elevation information. While this is clear for maritime applications (drifting buoys, shipping, marine animals, etc) there can be a problem with programs involving balloons or terrestrial animals, for example.
- stability of PTT oscillator frequency during the 10-minute (mean) satellite overpass. Although there is no technical problem, users may encounter some difficulty in designing their terminals (cost, weight, consumption,...) depending on their application.

The location reference system comprises ten special orbitography beacons installed around the world at precise geodesic positions. Once a day, Doppler measurements from these reference beacons are processed to compute the orbit of the spacecraft. As part of the orbit determination process, the frequency of each beacon is computed. The Toulouse reference beacon is synchronized on the CNES frequency and time reference. This allows to compute the frequency of the onboard Ultra Stable Oscillator (USO) and to synchronize its clock on UTC.

The GPCs are designed for maximum availability and minimum data loss to the user. They operate around the clock, every day of the year.

User access to data is through any of four channels:

- The Global Telecommunications System (GTS) a system for data exchange between the weather services of all countries. The GTS is operated as part of the World Weather Watch (WWW);
- international telex network;
- public switched telephone network (PSTN);
- national data transmission networks in each country, connected to the French Transpac or US Tymnet networks.

Throughput time - the interval between data reception by the spacecraft and results being made available to users -varies according to the path taken.

World level: data pass from the onboard receiver to the onboard tape recorder, then to the GPC via the ground station. Global coverage is thus assured. Around two-thirds of the results are on users' desks and terminals less than three hours after message collection aboard the spacecraft.

Regional level: The GPCs are connected to several direct readout stations, including Fairbanks, Wallops Island and Toulouse. Taking this route, PTT data from the Arctic regions, the North Atlantic, Europe and West Africa are received and processed in less than 30 minutes.

The GPCs keep data on file for three months which is particularly useful for long-term programs. Data are available to the user on either tape, floppy diskette or printout.

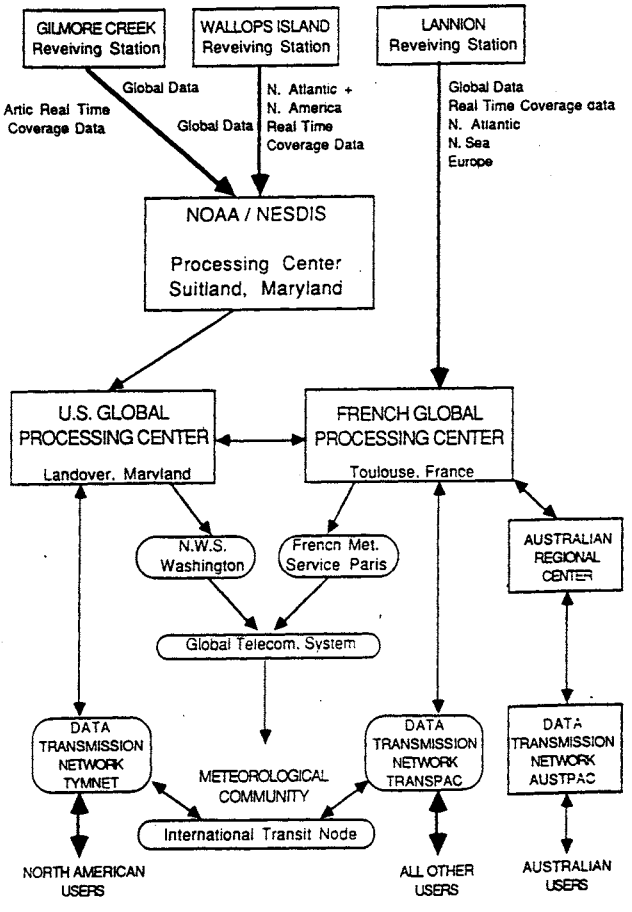


Figure 3. Flow of Argos data

User platforms (figure 4): Each platform carries an Argos PTT that assures the satellite uplink. The PTT is powered by rechargeable or disposable batteries, or solar panels or other external sources, and equipped with an antenna. Size and weight are reduced to a minimum. The PTT connects directly to the sensors, and transmits up to 32 measurements, encoded in words of up to 32 bits; the maximum message length is 256 bits.

Argos PTTs are simple to operate, light (under 1 kg), inexpensive, and low on power (200 mW on average). The result: Argos is easily implemented and suitable for a very broad range of conditions.

All PTTs transmit on the same frequency (401.650 MHz) at intervals of 90-120 s for location PTTs and 200-300 s for data collection-only types. Each transmitted message contains the PTT ID # and the measured values currently in memory. Message length, a function of the number of transmitted measurements, is always under one second (360-920 ms).

PTTs are type-certified. About 15 manufacturers, mainly in USA, Japan and France are manufacturing more than 3200 PTTs a year.

4. ARGOS USO SPECIFICATIONS AND PERFORMANCES

4.1. Argos USO specifications

The main characteristics required for the on-board (OB), orbitography determination beacon (ODB), and user PTT USOs are given in the following table 1.

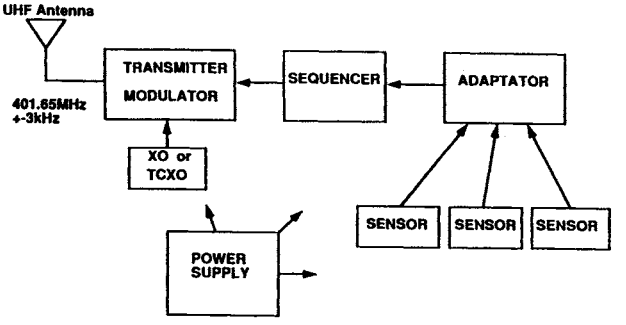


Figure 4. ARGOS PTT bloc diagram

Parameters	OB USO	ORBITO. USO	PTT USO
• Frequency	~ 5MHz	~ 5MHz	5 to 25 MHz
• DC supply variation	$\pm 5.10^{-11}$ (1%)	$\pm 5.10^{-11}$ (1%)	$\pm 5.10^{-11}$ (1%)
• Load variation	$\pm 5.10^{-11}$ (20%)	$\pm 5.10^{-11}$ (20%)	$\pm 5.10^{-11}$ (20%)
• Pressure variation	$\pm 1.5.10^{-8}$/bar	-	-
• Vibrations (sine, random)	$\pm 5.10^{-9}$	-	-
• Magnetic variation	$\pm 5.10^{-11}$/gauss	-	-
• Temperature variation	$\pm 2.10^{-10}$ (air, vacuum, 5 to 35°C)	$\pm 2.10^{-10}$ (air, 5 to 35°C)	$\pm 1.10^{-6}$ (air, -25 to 60°C)
• Short term stability τ = 0.1 s	1.10^{-10}	1.10^{-10}	1.10^{-9}
• ARGOS mid term stability with 5 °C/hour Max Slope Residual (1 σ)	 $\pm 5.10^{-12}$/min 5.10^{-10}	 $\pm 2.5.10^{-12}$/min 2.10^{-10}	 $\pm 5.10^{-10}$/min 1.10^{-9}
• Long term Aging 1 day after 1 month 10 years	 $\pm 1.10^{-10}$/day $\pm 2.10^{-7}$	 $\pm 1.10^{-10}$/day $\pm 2.10^{-7}$	 - $\pm 5.10^{-6}$ over 5 years
• Phase Noise in dBc/Hz 10 Hz 1 kHz	 - 120 - 150	 - 120 - 150	 - -
• DC Power air, -10 °C Vacuum, -10 °C	 8 W 2.8 W	 8 W 2,8 W	 < 20 mW -
• Power supply voltage	12 VDC	12 VDC	suppliers defined
• Random vibration	16.6 g rms	-	-
• Sine vibration	20 to 500 Hz: 10g 500 to 1000 Hz: 10 to 20g 1000 to 2000 Hz: 20g	-	-
• Radiations (3 years)	7.5 krads	-	-

Table 1. ARGOS USO SPECIFICATIONS

4.1.1. OB and ODB USO's frequency are about 5 MHz (5.146 795 MHz and 5.020625 MHz respectively). Frequency stabilities versus DC supply, load, pressure, vibrations are very classical. Frequency stability versus magnetic field is necessary for space USO, due to two components:

- the magnetic field generated by the Earth, at the altitude of the NOAA spacecraft,
- the magnetic field generated by magneto-torquers of the spacecraft.

The very important characterisations of ARGOS USO s for a precise orbit determination (± 100 m) are :

- the short term stability, expressed in ALLAN variance: $\sigma y\ (\tau) \leq 1.10^{-10}$ over 100 ms, duration of the on board Doppler count.
- the ARGOS medium term stability, defined by slopes (min^{-1}) and residuals at 1 sigma, of 40 samples of frequency measurements made over 100 ms every 30 seconds. Slopes (S) and Residuals (R) are calculated by the least squares method. They must be compliant over all environmental cconditions: temperature, magnetic field and radiation variations in orbit, and temperature, pressure variations for ODB USOs.

4.1.2. Users PTT frequency must satisfy several specifications. The nominal frequency of OB USOs depend on the architecture of transmitters which transmit at 401.65 MHz ± 2 kHz. For example, the use of a VCO and a divider by 81 leads to a USO nominal frequency value of 4.958642 MHz. In order to locate user PTT within ± 1 kilometer, the mid term stability of USOs, also expressed by S and R, should be less than ± 5.10^{-10}/min and 1.10^{-9}. They must be compliant over environmental conditions (temperature, vibrations, ...). DC power supply must be as low as possible, less than 10 to 20 milliwatts, in order to assure several years of lifetime.

4.2. Argos USOs procurement

4.2.1. The ARGOS OB Flight Model (FM) USOs were procured from FEI (USA) since 1974: 11 redundant units for the 8 first instruments (TIROS-N, and NOAA-A to NOAA-G), and 8 single units for NOAA-H to NOAA-M, with a total of 30 single FM including spare models. From NOAA-N, FM USOs will be procured from THOMSON-CEPE in France. ARGOS FEI USOs use AT PQ swept crystals (SC cut crystals for NOAA-K to NOAA-M USOs), two ovens for thermal control and glass Dewar technique. For the first 11 instruments, it was necessary to suspend USO inside receiver box to reduce mechanical vibration levels. The new models, more robust, are placed directly on the baseplate of the receiver.

4.2.2. ODB USOs (15 units, including spare models) were provided in 1976 from OSCILLOQUARTZ S.A. (Switzerland). Models used are B5400 type, using AT cut crystals, one oven and glass Dewar technique. In 1986, the ODBs were replaced by new electronics, and OSA USOs were replaced by PMT-P5 type USOs from THOMSON-CEPE in France.

4.2.3. Users PTT USOs were, in a first step, provided by CNES to several companies. We have chosen TCXO. Since 1980, TCXO are supplied directly by the manufacturers of PTTs. Performances of PTT USOs are verified during PTT type certification.

4.3. Argos USO performances

4.3.1. OB and ODB USOs are tested in detail, first at the suppliers, then by CNES. Main characteristics (short term, mid term stabilities, consumption, thermal vacuum and vibrations tests) are measured at equipment level in France, and at satellite level in USA before launch. Each ODB USO is also tested in detail at the supplier, then after integration in orbit Determination Beacon, and during the installation in each site.

The Table 2 give typical performances of the OB and ODB USOs, measured by suppliers and CNES.

Parameters	FEI OB USO	OSA ODB USO	CEPE ODB USO
• Short term stability ALLAN variance $\tau = 0.1s$	3.10^{-12}	$1.5.10^{-11}$	7.10^{-12}
• ARGOS mid term stability with $\theta = 0.1$ °C/min			
Slopes	$\pm 1.10^{-12}$/min	$\pm 2.10^{-12}$/min	$\pm 3.10^{-12}$/min
Residuals (1 σ)	1.10^{-11}	2.10^{-11}	4.10^{-11}
• Long term aging after 1 month	5.10^{-11}/day	5.10^{-11}/day	1.10^{-10}/day
• Temperature sensitivity			
+ 5 to 35 °C	-	2.10^{-10}	$1.5.10^{-10}$
in air	1.10^{-10}	-	-
in vacuum	3.10^{-10}/bar	-	-
• Pressure sensitivity	$< 1.10^{-11}$/gauss	-	-
• Magnetic sensitivity			
• Phase Noise in dBc/Hz			
10 Hz	-125	-125	-125
1 kHz	-150	-150	150
• DC power at 20 °C			
in air	1.6 W	3.2 W	1.6 W
in vacuum	1.3 W	-	-

Table 2. Typical ARGOS OB and ODB USO performances

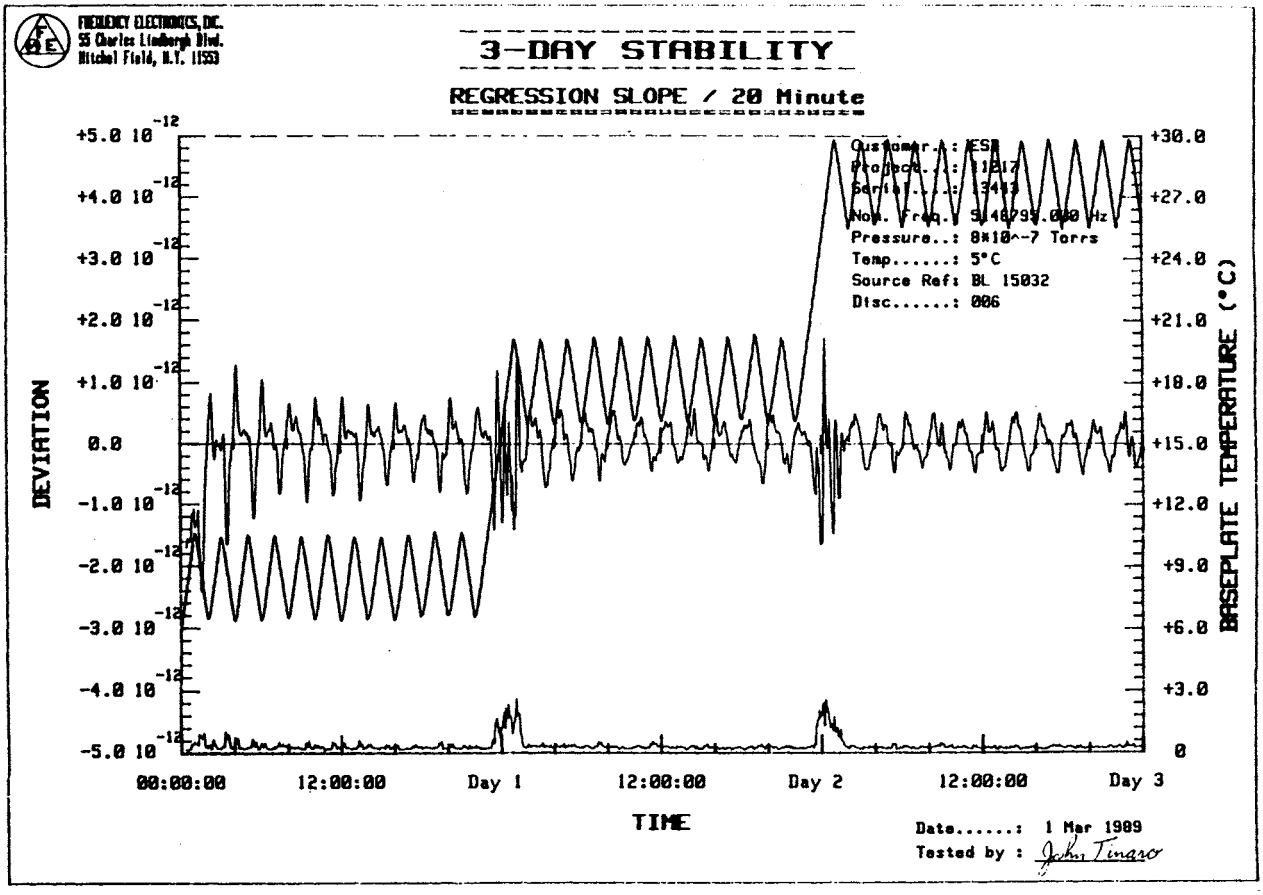


Figure 5. ARGOS Medium term stability of FM USO

Figures 5 gives the ARGOS medium term stability (Slopes S and Residuals R) of FM unit S/N 8902-13443, measured by FEI during thermal vacuum test.

4.3.2. Argos USO Performances via Orbit Measurements

ODBs were installed during 1977-1978 period. The first satellite (TIROS-N) was launched in October 13, 1978. The CNES Orbit Determination Department processed on-board Doppler data using specific software. Several parameters were adjusted:

- classical orbit elements,
- dynamic parameters, such as ionosphere,...
- frequency deviation between OB and ODB USO, for each pass when the satellite is within a range of a beacon. It is possible to calculate absolute frequency deviation of the OB USO, when the satellite is over the CNES at Toulouse (France), because we have installed a Master Beacon slaved to a cesium clock with a well known frequency measured by GPS ($+7.10^{-13} \pm 1.10^{-13}$). The accuracy of frequency determination by this method is estimated to 2 to 3.10^{-9}.

Several passes per day of NOAA-satellites over Toulouse permit to know the drift of each OB USO in orbit. After that, it is possible to calculate, in absolute, the frequency of each ODB USO.

Since 1982, CLS computes monthly the frequency of each OB and ODB USO (one value per day).

Figures 6 and 7 give two examples of drift in orbit (NOAA-9 and NOAA-10 ARGOS USO). We can see a negative slope versus time, probably due to irradiation on crystals. Figure 8 gives, for 7 flight models, the mean drift versus time, taken at the instant of their power-on in orbit.

Figures 9 and 10 give two examples of ODB USO drift (WALLOPS in USA and CAMBERRA, Australia). From 1980 to 1987, we can see positive drift given by B5400 OSA USOs; then, since 1987, the negative drift of PMT-P5 CEPE USOs. Figure 11 gives, for the last USOs, the mean drift versus time. After 2 years the drift is between 0 to -3.10^{-10} per month.

Some frequency variations of 3.10^{-9} peak to peak over 2 or 3 months to CAMBERRA beacon in mid 1983 can be observed. Theses effects are not due to USO itself, but due to seasonal fluctuations which are not well modelled, and which are arbitrarily allocated to frequency deviations.

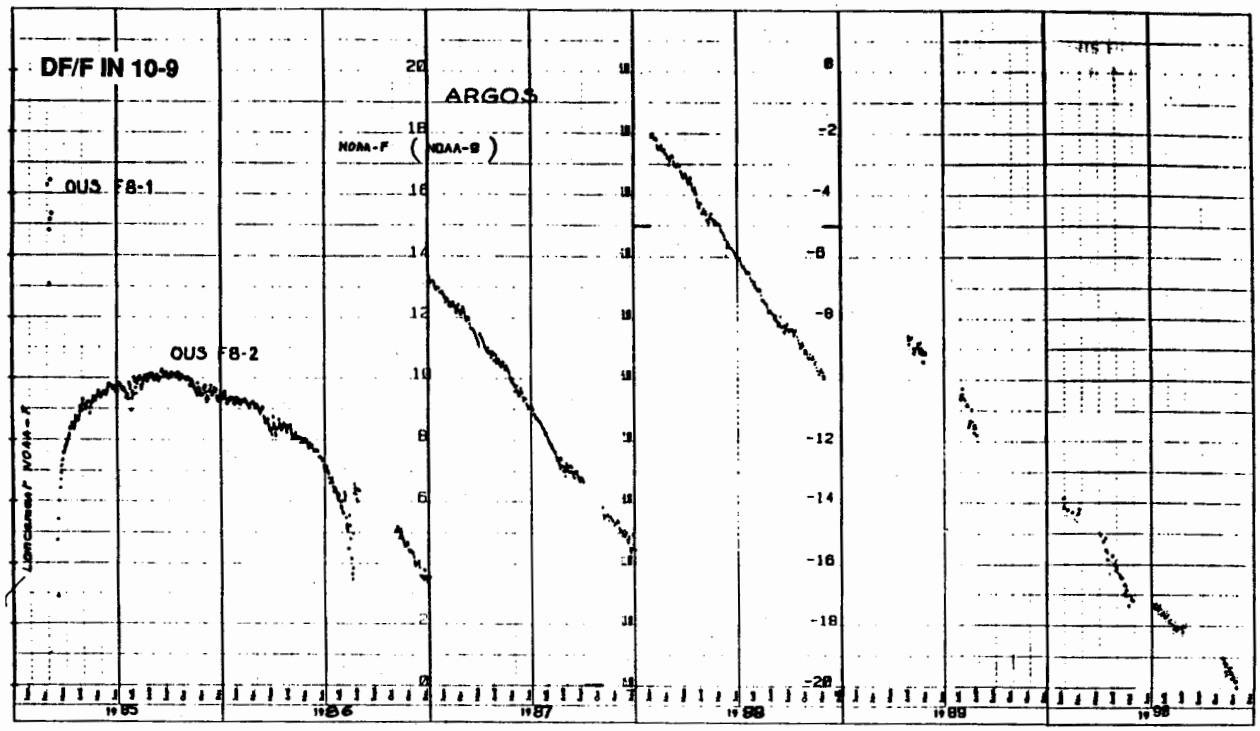


Figure 6

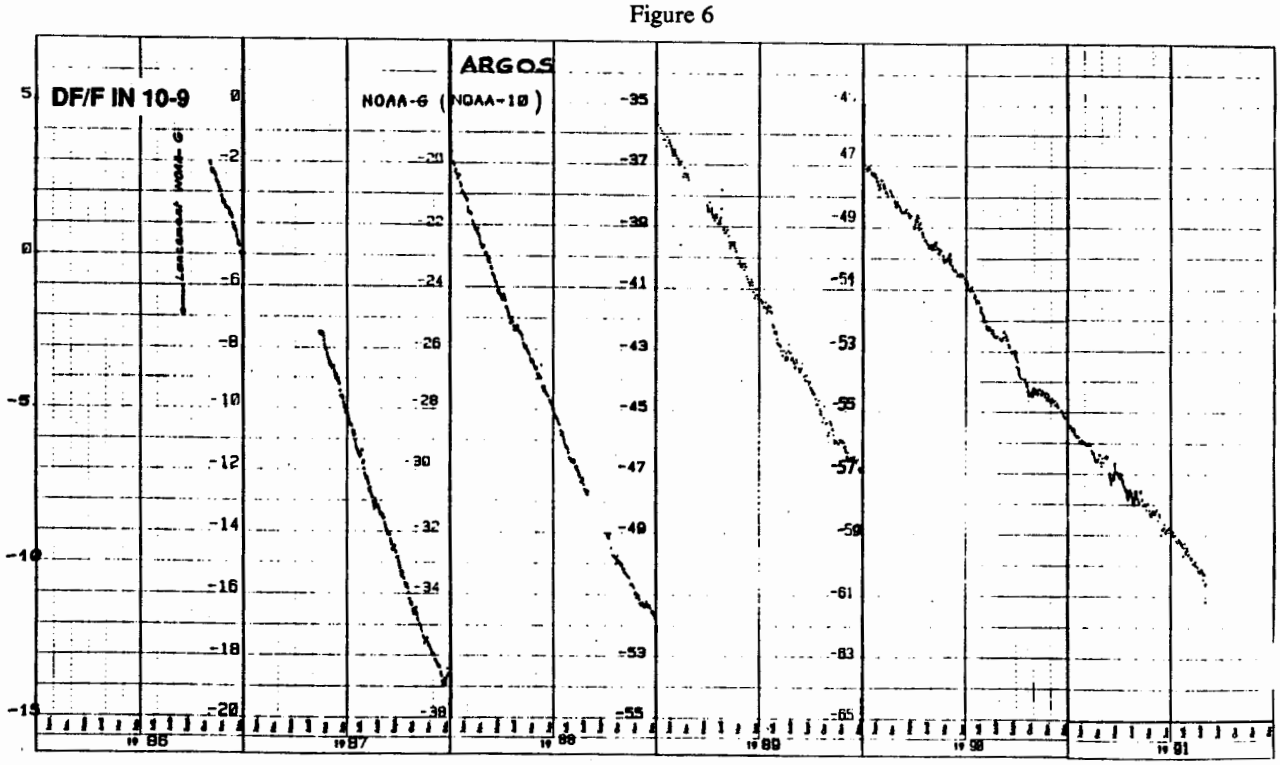


Figure 7

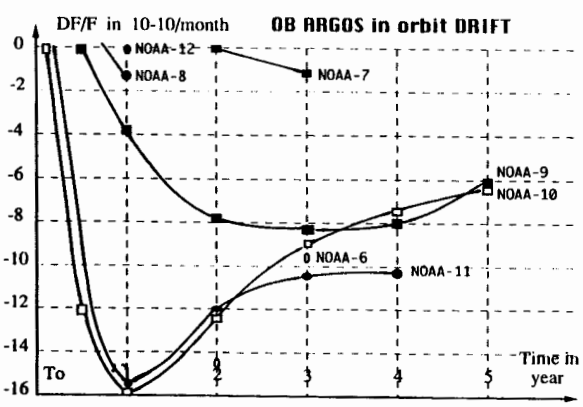


Figure 8

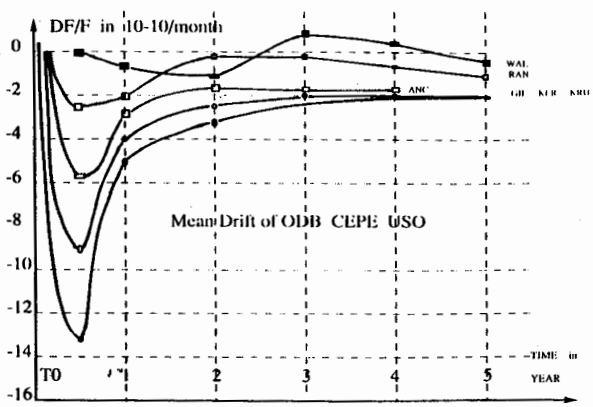


Figure 11

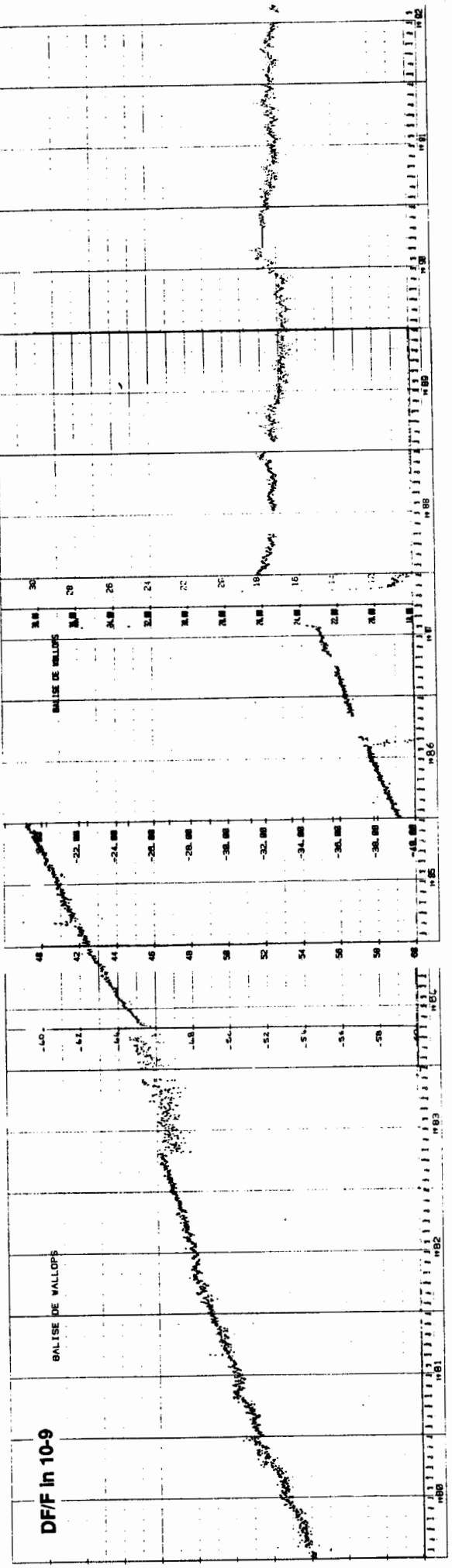


Figure 9

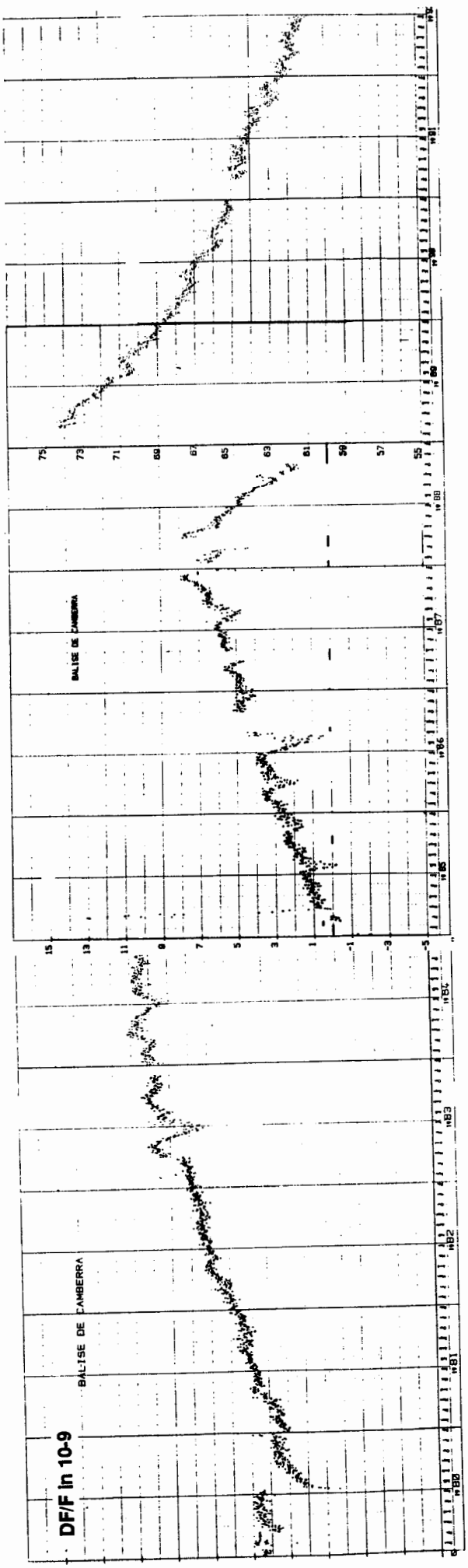


Figure 10

5. ARGOS SYSTEM APPLICATIONS

The Argos system permits the transmission of data recorded in locations too isolated to be integrated into telephone networks and too restricted for satisfactory surface VHF links. Applications cover a wide range of environmental studies, that is, for studies of the sea, land and air, as well as for biological research (birds, animals, mammals).

Drifting buoys

Of the existing suite of Argos applications, oceanography is by far the largest. The use of buoys to measure surface and sub-surface environmental parameters is now common practice. Beginning with the First Global Garp Experiment (FGGE), in the late 1970's, today with the Tropical Ocean Global Atmospheric (TOGA) Program, and in the future, the World Ocean Circulation Experiment (WOCE), Argos has been, and will be, a direct participant and contributor to information required for research involved with the air/sea boundary.

It should be noted that ocean observations are of interest not only in oceanography, but also in meteorology, climatology, glaciology and marine biology, as well as in industries such as offshore oil, fishing and merchant shipping.

Ice buoys

The use of PTTs on ice in polar regions is becoming increasingly common. The purpose is to track ice movement and thereby study the risks of collision with oil drilling rigs, ships, or towed vessels. Ice buoys are sometimes equipped with satellite navigation receivers that compute iceberg positions extremely accurately. In these cases, the Argos PTT is used to transmit another position as a back-up, however, future Argos system enhancements will result in the computation of very accurate location calculations.

Ice buoys can be deployed by a variety of means, including helicopter, light aircraft and parachute. They are designed to withstand extremely low temperatures. Powered by lithium batteries, the PTTs have a life of more than two years.

Moored buoys

The need for environmental data in specific locations over long periods of time has led to buoys being moored in the open seas. These buoys, by design, are larger and heavier than the drifting variety, since they must support an anchor chain and be able to withstand buffeting by the wind and sea. Their loss involves not only a high financial cost, but can also create a serious hazard to navigation. Used in conjunction with Argos PTTs, such buoys can be readily located and their data automatically transmitted.

Land observations

In addition to conventional meteorological stations measuring pressure, temperature, and wind speed and direction, several new specialized stations have been developed for sensing areas such as snow data, hydrology, volcanology, seismology, and others. These stations can be powered by connection to a local electrical source, or by battery or solar supply, and can be fully automatic for use in remote or inaccessible areas.

Meteorological stations

Sensors used with Argos equipped meteorological stations normally need to meet rather stringent specifications. Sensed parameters tend to be the same as those for conventional (human operated) meteorological stations, and include atmospheric pressure, temperature, humidity, dew point, precipitation data and wind information.

Snow data stations

Avalanche predictions are performed through networks of automatic measurement stations designed specifically to meet the rigors of remote high-mountain locations.

Measurements made at these stations are typically: depth of snow cover, measured by an ultrasonic sensor, by which the precipitation equivalent and progressive packing of the snow cover can be estimated; wind speed, measured by an anemometer; air temperature, measured under a shelter; and, the vertical temperature profile of the snow cover. Stations often house a microprocessor which manages station operation and can, for example, also compute average values (compress the data), prior to transmission to the satellite(s).

Hydrological stations

These stations are normally equipped with limnographs and rain gauges, but other parameters such as temperature can also be measured. Since the number of daily satellite passes is insufficient to enable river level variations to be monitored, hydrological stations are often equipped with electronic memories.

Balloons

Complex physico-chemical phenomena in the various layers of the atmosphere can only be measured and studied "in situ". Very long flight time balloons capable of staying aloft for several weeks or even months, and carrying instruments weighing around 50 kg, open up the possibility of collection of temperature, atmospheric pressure, humidity, and balloon internal pressure on a global scale. The Argos system has also been used experimentally for the tracking of long-duration manned balloon flights, for example the flight Tokyo-San Francisco in Double Eagle V, and others.

Animal tracking

Efforts to study and protect certain animal species call for information as to their distribution and migration. A number of satellite-based programs have been undertaken to track whales, dolphins, sharks, turtles, birds, wild pigs, caribou, muskox, bear, and so on. Perfecting the use of Argos in animal tracking is an on-going activity.

6. CONCLUSIONS

Since the 60s, several space systems have been used for needs in navigation and radiolocation.

In navigation, the US Transit system was declared operational in 1964, and is being replaced by GPS. In these systems, each user receives data from satellites, and determines his position himself. He is the only one to know his position.

Radiolocation systems, such as ARGOS, offer wider possibilities: location of fixed or mobile platforms (PPTs), collection of data they are transmitting, processing and distribution to the users.

The ARGOS system, a cooperative program between NASA, NOAA and CNES, has been proposing a continuous world-wide service since 1978, and will continue beyond 2000. CNES provives ARGOS payloads, which are embarked on NOAA spacecraft. CLS, a subsidiary of CNES, is in charge of ground segment operations, relations with the customers, and the promotion of the system. The ARGOS location is based on one-way random access up-link, from automatic PTTs, using the Doppler effect. Location accuracy (150 m to 1 km) depends strongly on the stability of USOs. The on-board USOs were procured from USA. In 1992, CNES decided to procure them in France. The ARGOS PTTs are very simple to operate, light (less than 1 kilogram), inexpensive, and low on power (200 milliwatts on average) which assures several years of lifetime. PTTs connect directly to the sensors, and transmit up to 32 measurements, or 256 bits. Data are distributed to users within one to three hours.

The ARGOS system being a recognized success, CNES decided in 1986 to develop a second generation of instruments, with a three fold increase in the system capacity (several thousands of PTTs).

ARGOS covers a wide range of terrestrial, aerial and maritime environmental studies, as well as biological researchs on birds and mammals, and is well adapted to ecological survey. ARGOS has been, and will be a direct participant and contributor to global meteorological programs (FGGE, TOGA, WOCE).

REFERENCES

[1] CNES, Cours de Technologie Spatiale, "Location and Navigation Systems", 1989 - CEPADUES-EDITIONS, 111 rue N. Vauquelin - 31100 Toulouse, France.

[2] D. Hernandez, C. Derieux, "Radionavigation, radiolocalisation et collecte de données par satellites au CNES" ; revue "L'Aéronautique et l'Astronautique", n° 154-155, 1992-3/4 - pp 66-74.

[3] Argos's News Letter, periodic publication, CLS, 18 avenue E. Belin, 31055 Toulouse, France.

[4] CLS/Service Argos - User's Manual.

[5] CNES/CLS/Service Argos, Argos platform transmitter terminal : general specifications and certification.

MICROWAVE LINKS FOR PRECISE TIME- AND FREQUENCY TRANSFER BETWEEN GROUND- AND SPACE-BASED CLOCKS

S. Starker*), E. Sappl*), W. Schäfer**)

*) DLR, Institut für Hochfrequenztechnik, Oberpfaffenhofen
**) Universität Stuttgart, Institut für Navigation, Stuttgart

The principles of precise T/F-comparison between clocks in space and on the ground are summarized. The influence of asymmetries in up- and down-links on the accuracy are considered and measures for their determination and correction are presented. After pointing out two existing solutions, the concept of an improved microwave link is outlined, suited for the comparison of H-maser clocks in space and on the ground.

T/F-Transfer, microwaves, clocks in space;

INTRODUCTION

Despite remarkable improvements in one-way time distribution by satellite-navigation-systems as well as by VLBI, there will remain a lot of applications in the future, which will require two-way or multi-way time transfer with utmost accuracy. The need for such links will rise with the growing frequency stability of oscillators just under development. The use of microwave frequencies for this purpose is desirable because of their relative low weather dependence. In the following one of the most demanding-applications of a multi-way time transfer links will be considered: the comparison of clocks in space and on the ground. Until now, two experiments have been performed, where atomic clocks in space were controlled from the ground by multi-way methods: the well known redshift experiment (1976) with a H-maser on board of a rocket, and the NAVEX-experiment (1985) with a Cs- and a Rb-clock on-board of the Spaceshuttle. Both experiments used very different techniques and methods in persuing different goals. In the following a general overview on the principles of multi-way T/F-transfer and its special problems will be given.

Because the performance of such links will be limited in the end by systematic errors, the following considerations shall be concentrated on these. Random errors and their reduction by transmitting power, noise figures and bandwidth is a different problem and will not be regarded in this paper.

After a short insight in the methods, which were used in the above mentioned experiments, the concept for a microwave link with utmost accuracy is sketched, which should be suited for T/F-comparisons in a H-maser clock experiment proposed in (Refs. 1, 2).

TWO-WAY TIME-COMPARISON

Time-comparison denotes the comparison of simultaneous readings at two different time-scales, which may be produced by two distant clocks A and B.

To do this, pairs of simultaneous events have to be indicated on the two scales t_A and t_B. These pairs of simultaneous distant events can be determined following the definition of simultaneousnes by A. Einstein, visualized in fig. 1a:

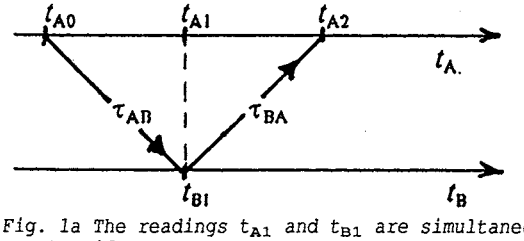


Fig. 1a The readings t_{A1} and t_{B1} are simultaneous events, if $\tau_{AB} = \tau_{BA}$

An electromagnetic signal starts from clock A, when the reading at that clock is t_{A0}; it arrives at clock B just when this clock shows the reading t_{B1}; it is sent back immediately and arrives at clock A, when this clock shows the reading t_{A2}.

In the case of symmetrical propagation delays ($\tau_{AB} = \tau_{BA}$) it is quite evident that

$$t_{A1} = \frac{1}{2}\,(t_{A0} + t_{A2}) \qquad (1)$$

and that the reading of t_{A1} at clock A and the reading of t_{B1} at clock B are simultaneous events. Thus the time difference of the two timescales in this moment is exactly

$$\Delta t_{AB}(t_{A1}) = t_{A1} - t_{B1} \qquad (2)$$
$$= \frac{1}{2}\,(t_{A0} + t_{A2}) - t_{B1}.$$

The calculation of Δt_{AB} can be carried out with equ. 2 as soon as the readings of clock A have been transmitted to B (or vice versa) by a data link.

Technical realizations of the principle sketched above will always have some asymmetry in propagation delays (fig 1b)

$$\tau_{AB} - \tau_{BA} = 2\Delta\tau \neq 0. \qquad (3)$$

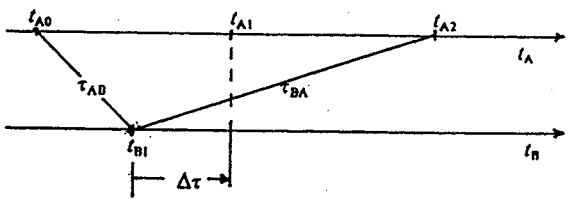


Fig. 1b In the case of asymmetric delays $\tau_{AB} \neq \tau_{BA}$ the readings t_{A1} and $(t_{B1} + \Delta\tau)$ are simultaneous events. The determination of $\Delta\tau$ requires additional measurements.

Considering this asymmetry, equ. 2 gets a more general form,

$$\Delta t_{AB}(t_{A1}) = \frac{1}{2}\,(t_{A0} + t_{A2}) - (t_{B1} + \Delta\tau), \qquad (2a)$$

which may be used as starting point for the following considerations.

In stationary configurations, $\Delta\tau$ is generally very small and may be neglected in most cases. But in time comparisons between clocks in space and on the ground this term can be relatively large and the accurate determination of its numerical value is one of the main problems.

The main reasons for this asymmetry are:
- the relative velocity between clocks
- asymmetries in propagation velocities.
- asymmetric delays in transmitters, receivers and cables

Each of these influences requires special additional measures, and has to be considered in more detail:

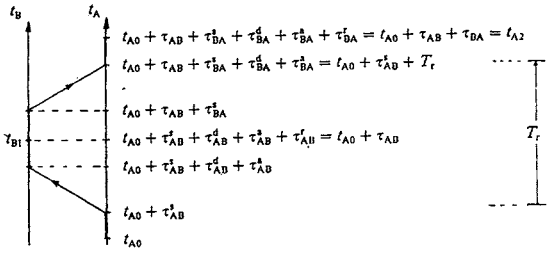


Fig. 2: Composition of the Two-Way Transit Time $(\tau_{AB} + \tau_{BA})$.

Fig. 2, a detailed version of fig. 1, shows the composition of the transit times τ_{AB} and τ_{BA}. The term τ_{AB} is the sum of transit time τ^d_{AB} of the signal in vacuum from antenna at A to antenna at B, small delay times τ^s_{AB} within the transmitter at A, τ^r_{AB} through the receiver at B and a very small delay time τ^a_{AB} caused by the atmosphere (mainly by the ionosphere). The term τ_{BA} is composed by the signal vacuum transit time τ^d_{BA} from antenna at B to antenna at A, the transmitter delay time τ^s_{BA} at B, the receiver delay time τ^r_{BA} at A and the atmospheric delay time τ^a_{BA} for the path from B to A. Defining the asymmetries

$$\Delta\tau_d = \frac{1}{2}\,(\tau^d_{AB} - \tau^d_{BA}), \qquad (4a)$$

$$\Delta\tau_a = \frac{1}{2}\,(\tau^a_{AB} - \tau^a_{BA}) \qquad (4b)$$

$$\Delta\tau_e = \frac{1}{2}\,((\tau^s_{AB} - \tau^s_{BA}) + (\tau^r_{AB} - \tau^r_{BA})), \qquad (4c)$$

in distance, atmospheric propagation and equipment delay times we get

$$\Delta\tau = \Delta\tau_d + \Delta\tau_a + \Delta\tau_e.$$

The term $\Delta\tau_d$, which depends on the relative velocity $\dot r$ and on the distance r between A and B, can be determined by an additional measurement of the two-way Doppler-shift f_{DA2} of the frequency f_A and the travelling time T_r

$$\Delta\tau_d = -\frac{1}{2}\,((\tau_{AB} + \tau_{BA}) - (\tau^s_{AB} + \tau^r_{BA}))\cdot\frac{\dot r}{c} = -\frac{T_r}{2}\cdot\frac{\dot r}{c} \approx \frac{T_r}{2}\cdot\frac{f_{DA2}}{f_A} \qquad (5)$$

The determination of $\dot r$ can also be performed by a two-way Doppler-measurement with the frequency f_B, as will be discussed later.

The atmospheric term $\Delta\tau_a$ is mainly caused by the frequency dispersion of the ionosphere in connection with the use of different signal frequencies f_{BA} and f_{AB} in the up- and down-link.

$$\Delta\tau_a = \frac{1}{2}\cdot 1{,}34\cdot 10^{-7}\cdot \mathrm{TEC}\,(f_{AB}^{-2} - f_{BA}^{-2})\ [\mathrm{s}] \qquad (6)$$

To determine the total electron content TEC, which varies with time and signal path, an additional signal with a different frequency f_c has to be transmitted parallel to the main signal. Assuming another frequency f_c in the downlink, the difference $\Delta\tau_{AC}$ between the propagation delays of f_A and f_c can be used to calculate the value of TEC

$$\mathrm{TEC} = \frac{\Delta\tau_{AC}\cdot f_c^2\cdot f_A^2}{1{,}34\cdot 10^{-7}\,(f_c^2 - f_A^2)} \qquad (7)$$

In addition a frequency dependence of the receiving equipment will have to be taken into account. This requires a correction of the measured difference $\Delta\tau_{AC}$ by a term $\partial\tau(f_c + f_{DC},\ f_A + f_{DA})$ to get the true value of $\Delta\tau_{AC}$.

Describing the equipment delay in each direction AB and BA by a value $\tau^e = \tau^s + \tau^r$ for the mean frequency and a value $\rho\cdot\Delta f/f$ for a frequency deviation Δf, the frequency dependence of the equipment asymmetry can be described as a function of the Doppler shift

$$\Delta\tau_e = \frac{1}{2}\,(\tau^e_{AB} - \tau^e_{BA}) + \frac{1}{2}\,\rho_{AB}\cdot\frac{f_{D1}}{f} - \frac{1}{2}\,\rho_{BA}\cdot\frac{f_{D2}}{f} \qquad (8)$$

with $\tau^e + \rho\cdot\frac{f_D}{f} = \tau^s + \tau^r;\ (\rho = \mathrm{const.})$

τ^e_{AB}, τ^e_{BA} and ρ have to be determined as accurate as possible by calibration measurements.

FREQUENCY COMPARISON

During the transfer of the frequency f_A from clock A to clock B, it suffers a Doppler shift f_{DA1}, which depends on the relative velocity $v_r = dr/dt = \dot r$ according the equation

$$f_{DA1} = -f_A\cdot\frac{v_r}{c} \qquad (9)$$

The amount of the frequency shift f_{DA1} can be determined from the measurement of the two-way Doppler shift, if the signal is sent back immediately from B to A (Fig. 3a). The necessary phase coherent frequency transposition at B by a factor p has to be removed at A before the measurement of the two-way Doppler shift f_{DA2} can be performed.

After transmitting the measurement results from A to B by a data link, the one-way Doppler shift f_{DA1} can be determined and corrected at B using the relation

$$\frac{f_{DA1}}{f_A} = \sqrt{1+\frac{f_{DA2}}{f_A}} - 1 \qquad (10)$$

$$f_{DA1} \approx \frac{1}{2} f_{DA2} \left(1 - \frac{f_{DA2}}{4 f_A}\right) + \ldots$$

The ratio f_{DA1}/f_A is $< 10^{-5}$ and can be neglected in many cases.

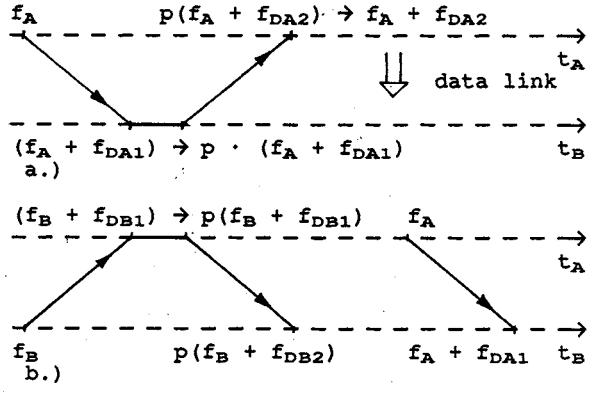


Fig. 3 Alternatives a.) and b.) for a direct frequency comparison under the assumption of symmetric atmospheric velocities $v_{down} = v_{up}$.

An alternative possibility for the determination of v_r/c and f_{DA1} is a 2-way Doppler shift measurement using the frequency f_B (Fig. 3b).

In this case a very similar relation is valid

$$\frac{f_{DA1}}{f_A} = \sqrt{1 + \frac{f_{DB2}}{f_B}} - 1 \qquad (11)$$

because of

$$\frac{f_{DA1}}{f_A} = \frac{f_{DB1}}{f_B} \; ; \qquad \text{and} \qquad \frac{f_{DA2}}{f_A} = \frac{f_{DB2}}{f_B} \; ;$$

but the absolute value of f_{DA1}depends now also on the ratio f_A/f_B, which requires an iterative determination of Doppler shift f_{DA1} and the frequency difference Δf_{AB}, or on additional frequency comparison at A.

$$f_{DA1} \approx \frac{1}{2} f_{DB2} \cdot \frac{f_A}{f_B}\left[1 - \frac{f_{DB2}}{4 f_B}\right] + \ldots$$
$$\approx \frac{1}{2} f_{DB2}\left[1 - \frac{f_{DB2}}{4 f_B}\right](1 + \frac{\Delta f_{AB}}{f_B}) \qquad (12)$$

In the case of asymmetric atmospheric delays ($v_{down} \neq v_{up}$), caused by the frequency dispersive behaviour of the ionosphere, the situation is more complicated. The one-way Doppler shift of a transposed frequency $q \cdot f_A$ is not longer proportional to q and f_{DA1}, but has to be described by the new term $q \cdot f'_{DA1}$. To achieve the necessary information on the frequency dispersion, additional frequency comparisons have to be performed at B between $(f_A + f_{DA1})$, $(f_A + f'_{DA1})$ and f_B.
The general model $f_D = x \cdot f + y/f$ for a frequency shift caused by relative motion ($x = -\dot{r}/c$) and ionosphere ($y \sim d(TEC)/dt$) (Ref. 4) results in three equations for x, y, f_{DA1}

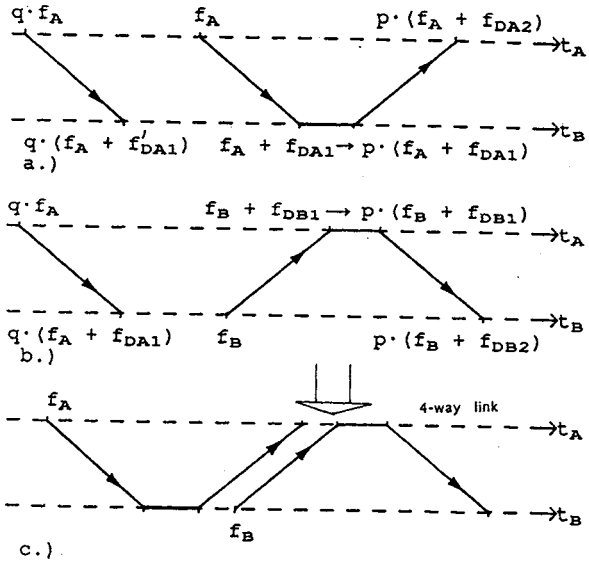


Fig. 4 Alternatives for a direct frequency comparison in the case of asymmetric atmospheric delays ($v_{down} \neq v_{up}$).

$$f_{DA1} = x \cdot f_A + \frac{y}{f_A} \,,$$

$$f'_{DA1} = x \cdot f_A + \frac{1}{q^2}\,\frac{y}{f_A} \,,$$

$$f_{DA1} - f_{DA1} = x \cdot f_A + \frac{1}{p^2}\,\frac{y}{f_A} \; ; \; (\frac{f_{DA1}}{f_A} \ll 1)$$

which yield the needed frequency shift

$$f_{DA1} = \frac{1}{2}\left\{f_{DA2} - \frac{1-p^2}{1-q^2}\,(\frac{q}{p})^2\left[f'_{DA1} - f_{DA1}\right]\right\} \qquad (13)$$

as a function of measured quantities f_{DA2} and $f'_{DA1} - f_{DA1}$, under the assumption that $f_{DA1}/f_A \ll 1$.

An alternative possibility for the determination of f_{DA1} in the case of asymmetric atmospheric velocities is a two-way Doppler shift measurement using frequency f_B (Fig. 4b) similar the case shown in Fig. 3b. From the measured quantities f_{DB2}, $f_B - f_A + f_{DB1}$ and $f_A - f_B + f_{DA1}$ we get the values of f_{DB2}, $f_{DA1} + f_{DB1}$ and $f_A - f_B + f_{DA1}$. The same procedure as above yields the one-way Doppler shift

$$f_{DA1} = \frac{\left(\frac{f_B}{f_A}\frac{1}{q^2} - \frac{f_A}{f_B}\right) f_{DB2}}{\left(1 + \frac{f_A}{f_B}\right)\left(1 + \frac{1}{p^2}\right) - 2\left(1 + \frac{f_B}{f_A}\frac{1}{q^2}\right)} \qquad (14)$$
$$- \frac{\left(\frac{f_B}{f_A}\frac{2}{q^2} - \frac{f_A}{f_B}\left(1 + \frac{1}{p^2}\right)\right)\left(f_{DA1} + f_{DB1}\right)}{\left(1 + \frac{f_A}{f_B}\right)\left(1 + \frac{1}{p^2}\right) - 2\left(1 + \frac{f_B}{f_A}\frac{1}{q^2}\right)}$$

which requires an iterative determination of f_{DA1}, because it depends also on f_A/f_B (see remark to equ.(12)).

The first approximation (valid for $(f_A - f_B)/f_B \ll 1$) is

$$f_{DA1} = \frac{1}{2} \frac{(1-\frac{1}{q^2})\ f_{DB2} - (1+\frac{1}{p^2}-\frac{2}{q^2})\ (f_{DA1}+f_{DB1})}{\frac{1}{q^2}-\frac{1}{p^2}} \qquad (15)$$

Chosing $1 + \frac{1}{p^2} = \frac{2}{q^2}$ the measurement $f_{DA1} + f_{DB1}$ does not appear in f_{DA1}. The measurement $f_B - f_A + f_{DB1}$ and its transfer to clock B by a data link is not necessary in this case. This condition is fulfilled in the well known Doppler cancelling system in the red shift experiment (Ref. 4).

The methods described above can be used for continuous frequency comparisons as long as a direct signal contact exists between A and B. A combination of both, that is simultaneous two-way Doppler measurements at A and at B leads to a four-way link, proposed by R. Vessot (Ref. 3), with several advantages, as for instance the availability of the Doppler information at both clocks (Fig. 4c).

For frequency comparisons over the duration of several satellite revolution periods the signal contact to the ground station will be interrupted after each satellite pass for more than an hour.

In this case the relative frequency deviation between the clocks A and B has to be determined by measuring the rate R_{AB} between the two clocks, which is defined as the change of Δt_{AB} with time (Fig. 5).

$$\Delta t_{AB}\ (t_A) = \Delta t_{AB}\ (t_{A1}) + R_{AB}\ (t_{A1})\ (t_A - t_{A1}); \quad (t_{A1} < t_A < t_{A4}) \qquad (16)$$

$$R_{AB}\ (t_{A1}) = \frac{d\ \Delta t_{AB}}{dt_A}\ .$$

To achieve the necessary data, repeated time comparisons have to be performed during different satellite passes.

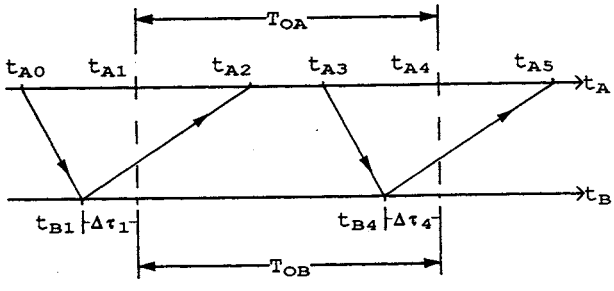


Fig. 5 Frequency comparison over several satellite passes by rate measurments.

With the notations of fig. 5, the rate R_{AB} can be achieved from the values T_{OA} and T_{OB}, which describe a time interval of equal length at the different time scales t_A and t_B

$$R_{AB}(t_A) = \frac{T_{OA} - T_{OB}}{T_{OA}} = 1 - \frac{T_{OB}}{T_{OA}} = 1 - \frac{f_B}{f_A}$$

$$R_{AB}(t_B) = \frac{T_{OA} - T_{OB}}{T_{OB}} = \frac{T_{OA}}{T_{OB}} - 1 = \frac{f_A}{f_B} - 1\ . \qquad (17)$$

As Fig. 5 indicates the measured value T_{OB} and therefore also R_{AB} is influenced by the change of $\Delta\tau$ with time

$$\Delta R_{AB}(t_A) = \frac{d\ \Delta\tau(t_A)}{dt_A} = \Delta\dot{t}_d(t_A) + \Delta\dot{t}_a\ (t_A) + \Delta\dot{t}_c\ (t_A) \qquad (18)$$

These terms have to be determined by measurements or by estimations and have to be introduced as correction-terms in the final result.

REALIZATION OF A FREQUENCY COMPARISON LINK IN THE RED SHIFT ROCKET EXPERIMENT

In this experiment the output frequency of a H-maser clock on board of a rocket was compared with the H-maser clock on the ground continuously during a time interval of nearly two hours (Ref. 4). The method applied is in essence a 3-way link as shown in Fig. 4. The signal frequencies are chosen in such away that the influence of the ionosphere on frequency f_A and f_B is exactly in the ratio 1 : 2 and can be eliminated by subtraction on the ground as shown in fig 6. The relative accuracy achieved in this experiment was better than 10^{-14} during the measurement interval of about 2 hours (Ref. 4).

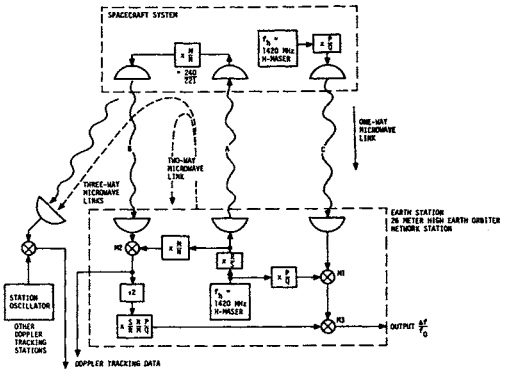


Fig. 6: A schematic diagram of the Doppler cancelling system used in the red shift rocket experiment (Ref. 3). (R/S = 82/55; P/Q = 76/49; N/M = 221/240).

REALIZATION OF A TIME COMPARISON LINK IN THE NAVEX SHUTTLE EXPERIMENT

In the navigation experiment NAVEX a navigation payload including a Cs- and a Rb-clock was tested on board of the Space Shuttle (Ref. 5). The one-way navigation signal from the Shuttle was completed by an upward signal in such a way that a two-way time comparison with pseudo noise code signals could be performed. The on-board measurement data were transferred to the ground by a 250 b/s data link, which used the PN-code signal as a carrier according the spread spectrum principle. For the determination of an asymmetric influence of the ionosphere a comparison of the L-band propagation delay with simultaneous radar measurements in the C-band was provided (fig. 7). The residual error in the two-way time comparison was about 5 ns, the relative error in frequency comparison $2 \cdot 10^{-14}$ during an observation time of 72 hours.

The visibility conditions of about 300 - 500 s contact time followed by about 90 min loss of signal between two passes were very similar to the conditions of a future H-maser clock experiment. Under these constraints the frequency comparison had to be performed by a rate measurement according to fig. 5. For minimizing the influence of asymmetries it was effective, to concentrate the time comparison measurements on the culmination area and to take advantage of existing symmetries:

The culmination time t_{AK}, the time of minimum distance between the two clocks during a pass is defined by

$$\dot{r}(t_{AK}) = 0.$$

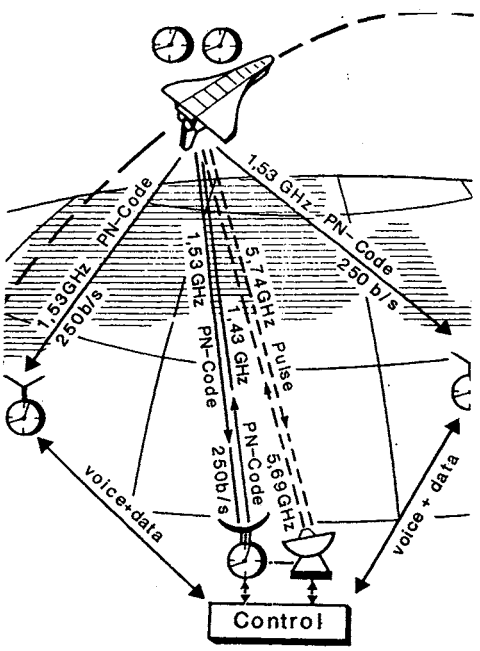


Fig. 7: Microwave Links for one- and two-way T/F-comparisons in the NAVEX experiment during Spacelab Mission D1 (Ref. 5).

This culmination time may be determined as the moment of vanishing Dopplershift or in a good approximation as the moment of minimum two-way transit time $\tau_{AB} + \tau_{BA}$. Equation (5) shows that the asymmetry $\Delta\tau_d$ vanishes at the culmination time. Equation (8) shows that the Doppler frequency dependent part of equipment asymmetry vanishes and we have at culmination time $\Delta\tau_e = (\tau^e_{AB} - \tau^e_{BA})/2$. The interpolation of the observation $(t_{A0} + t_{A2})/2 - t_{B1}$ within a neighbourhood of culmination time yields the value of this quantity at the culmination time and consequently the value of the time difference of the two time scales at this moment

$$\Delta t_{AB}(t_{AK}) = [(t_{A0}+t_{A2})/2-t_{B1}]_{t_{AK}} + (\tau^e_{AB}-\tau^e_{BA})/2 + \Delta\tau_a\ (t_{AK});$$ (from Equ. (2a)).

The atmospheric asymmetry is described by equation (6).

A PROPOSED T/F-COMPARISON LINK FOR FUTURE APPLICATIONS OF H-MASER-CLOCK IN SPACE

The future operation of H-maser clocks in space will require microwave links of utmost precision and accuracy. For economical reasons a solution was proposed (Refs. 1, 2, 7), using a recently developed equipment for precise range and range rate measurements (PRARE) (Ref. 8) as a two-way time comparison link.

Fig. 8 shows one of four identical channels, which are used by PRARE for measurements between the satellite and a fully coherent ground receiving/transmitting station. Although designed primarily for satellite ranging purposes, the main functions are very similar to the NAVEX link mentioned before with some improvements: The two-way PN-code link for time comparison will have larger bandwidth (10 MChip/s). A higher RF carrier frequency (X- or Ku-band) is used, which will be less influenced by the ionosphere. An additional S-band link to the ground will allow a TEC-determination according equ. 7. The signal links carry data information as visualized in Fig. 3. The most severe problem will be the determination of the value and the stability of equipment asymmetries. For this purpose, RF-carrier phase measurements are to be performed in addition to time delay measurements using PN codes. Prelaunch calibration on the space segment and measurements on 30 ground stations, which have been completed recently, show for the X-band link residual bias errors in the order of ± 30 ps. From this a relative uncertainty of $1.5 \cdot 10^{-15}$ in frequency comparison can be expected in a measurement interval of about 12 hours. If these numbers can be verified during flights on-board the Russian spacecraft Meteor-3 (to be launched Aug. 1993) and on ERS-2 (1994/1995) a time transfer with an accuracy of < 100 ps and a frequency comparison of 10^{-15} within 24 hours can be expected.

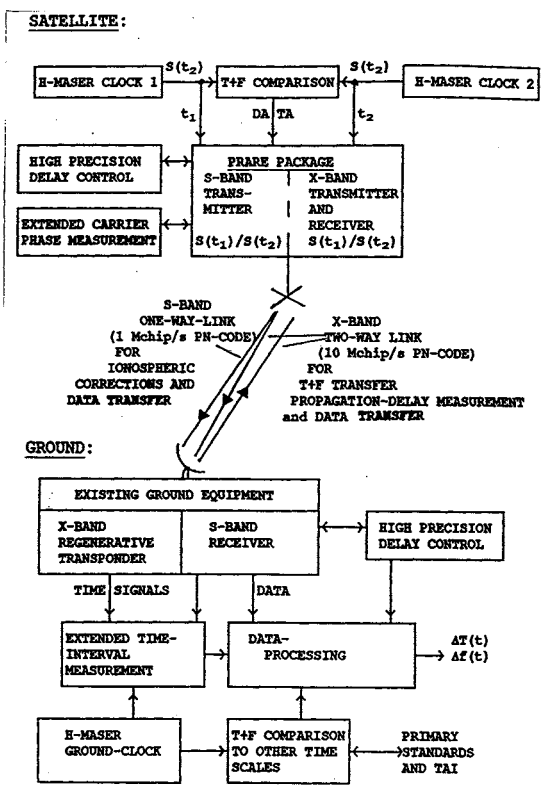


Fig. 8 A modified and extended PRARE-Link as proposed for precise T/F-Transfer.

(1994/95) a time transfer with an accuracy of < 100 ps and a frequency comparison of 10^{-15} within 24 hours can be expected.

SUMMARY

For precise T/F-comparisons between atomic clocks in space and on the ground multi-way microwave links are unavoidable. The accuracy of these transfer links is mainly limited by asymmetric delays in the up- and down-ward link. Additional measurements and estimations are necessary for the determination and correction of these asymmetric influences. In the case of time comparison the correction of asymmetries requires the determination of the relative velocity between the clocks, the determination of the TEC of the ionosphere, and the calibration and control of equipment delay times. For precise frequency comparisons the acceleration between both clocks, the change of TEC with time and the stability of equipment delays have to be determined or controlled.

Up to now, two such links have been developed and applied: an S-band frequency comparison link in the red shift rocket experiment, and an L-band time comparison link in the NAVEX Shuttle experiment. In the first case sinusoidal signals were used and the asymmetries caused by relative velocity and ionosphere were eliminated by a 3-way Doppler cancelling system. In the second case spread spectrum signals were used in a two-way mode with a simultaneous data modulation on the time signal. A coarse determination of the ionospheric influence was provided by parallel radar measurements in the C-band; because of the longer experiment duration of several days the equipment delays had to be controlled before and after each ground station contact.

For future experiments and applications with H-maser clocks in space also spread spectrum signals and carrier phase measurements shall be used for time and data transfer. An existing equipment for precise ranging (PRARE) once properly modified, appears to be suited, for this purpose.

REFERENCES

[1] Busca, G.; Starker, S.; Feltham, S.;
In Orbit Demonstration of a H-Maser Clock System.
Proc.: 23rd Anual Precise Time and Time Interval (PTTI) Applications and Planning Meeting, NASA-CP 3159, 1992, 11 p., pp. 385-399, 4 Figs., 11 refs. (Vortragender: G. Busca).

[2] R.F.C. Vessot, E.M. Mattision, G.U. Nystrom, R. Decher, S.J. Feltham, G. Busca, S. Starker and S. Leschiutta;
Hydrogen Masers for Space.
Proc. of the 6th European Frequency and Time Forum, ESA-SP 340 (1992), 7 p., pp. 19-25, 8 figs., 1 tab., 17 refs.

[3] R.F.C. Vessot
Applications of Highly Stable Oscillators to Scientific Measurements.
Proc. IEEE, July 1991, Special Issue on Time and Frequency, pp. 1040-1053.

[4] R.F.C. Vessot, M.W. Levine;
NASA Experimental Final Redshift Report.
GPA Project Report, 1980, Contract NAS8-27969.

[5] Starker, S.; Arko, D.; Hammesfahr, J.; Nau, N.; Sappl, E.; Schild, H.; Schimmel, R.; Stelzel, F.;
Navigation Experiment NAVEX on Spacelab Mission D1. Final Report, DLR, Institute of Radio Frequency Technology, D-8031, Wessling, 1990.

[6] E. Sappl
The Comparison of a Satellite Clock and a Ground Clock by Two-Way Time Transfer.
Z. Flugwiss. Weltraumforsch. 16 (1992), pp. 370-374.

[7] W. Schäfer
The Application of the PRARE Principle for Precise Two-way Time Transfer to Space-based Atomic Clocks,
6th European Frequency and Time Forum, Noordwijk/The Netherlands, 17-19 March, 1992, ESA-SP 340 (1992), pp 41-47, 8 figs., 3 refs.

[8] W. Schäfer, H. Wilmes
Precise Range and Range Rate Equipment (PRARE)-System Design and Status of Development,
Workshop on Advances in Satellite Radio Tracking, Austin, Texas, U.S.A., 5-6 Mai 1986, 14 p., 6 figs., 6 refs.

GPS PRECISE TIME, FREQUENCY, & EPHEMERIS DETERMINATION FOR MANNED SPACE FLIGHT

Alfred Anderman

Rockwell Space Systems Division,
12214 Lakewood Blvd., Downey, CA, 90241

ABSTRACT

Complexity of future manned spaceflight mission control can be significantly reduced by integrating GPS, the source of precise time and frequency, into telemetry, tracking and command (TT&C). Future telecommunications, space tracking, electronic intelligence, metrology, navigation, and data acquisition will thereby be served, including: 1) Onboard ephemeris determination, 2) Reduced synchronization time for time division multiple access (TDMA) links, and 3) In-flight clock calibration. Precise ephemeris aids mission success, status monitoring, also rendezvous and docking, and by eliminating the need for groundbased tracking, it can enhance autonomy and reduce communications bandwidth and ground support costs.

KEYWORDS: Autonomy, Ephemeris, GPS, Precise Frequency & Time, Synchronization, TDMA, Tracking

1. INTRODUCTION

Future space vehicles will require a very precise time reference for communication and navigation functions, including: 1) Timing signals to maintain user synchronization within various digital spread spectrum communications networks, 2) Accurate clocking to encryption devices for military missions, to maintain security, and 3) Generation of a "real time" time base to support integrated navigation functions, network acquisition, sensor data processing, correlation and fusion. Some speculation concerning potential additional GPS-aided space applications is included at the end of this paper.

Independent redundancy, to assure reliability, also implies synchronization among the independent data sources, either 1) Loosely, by buffering, comparing or voting, signalling concensus and/or marking completed intervals, or 2) Tightly, by hardware comparison or voting, and a common time reference, whereas loose synchronization can employ separate time references.

2. SPACE SHUTTLE ORBITER (SSO) TIMING & CONTROL

2.1 MASTER TIMING UNIT

The Space Shuttle stable crystal-controlled master timing unit provides serial time reference signals to the onboard computers, pulse code modulation master units (PCMMUs), various time display panels, and synchronization to instrumentation and other subsystems, using separate GMT and Mission Elapsed Time (MET) time accumulators. Time outputs are a) Serial time code continually updated at 100 pps rate (IRIG-B), and b) Self-clocking Manchester II biphase demand outputs read out upon receipt of externally supplied enable signals.

Frequency stability is better than 1 part in 10 E9 per day long term, and 1 part in 10 E10 standard deviation short term, using the Allan variance formula for averaging times of 0.5, 1, 2, and 10 seconds, but requires 72 hr. stabilization time after a power-off state at 35 deg.F. Master Timer derived time/clock/sync inputs to Shuttle operational instrumentation (OI) are shown in Figure 1.

2.2 AVIONICS PERFORMANCE

The estimated three-sigma position and velocity errors at main engine cutoff (MECO) of about 4600 ft and 20 ft/sec could be updated in the post-MECO state to about 100 ft and 0.6 ft/sec in about 2 minutes, using GPS.

On-orbit TDRSS provides adequate state vectors using two-way range rate (Doppler) data through both satellites within one revolution, in about 60-90 minutes. With only one relay satellite available, the 1.5 pass minimum tracking interval is about 115 minutes. An integrated time line suggests that 120-195 min. are required from the start of tracking to maneuver execution. With a full GPS constellation the state could be determined within about 2 minutes, as for the post-MECO state. Updating and resetting externally requires use of the data processing system.

For ground entry navigation, using C-Band radar, the mission control center (MCC) updates onboard knowledge of position after blackout at about 150 kft. altitude. GPS could provide accurate post-blackout navigation capability to about 200 ft and 1.2 ft/sec, or much better in differential mode. An orbiter GPS Development Flight Test (DFT) demonstration is planned for 1993 (Figure 2).

3. FUTURE MANNED SPACE TRANSPORTATION

Future manned space transportation will probably depend on a mix of vehicles. On-going NASA studies have resulted

in preliminary designs of Advanced Manned Launch Vehicle (AMLS) and Personnel Launch System (PLS), while anticipating continued use of the Shuttle Fleet through the first quarter of the 21st century (Ref. 8). A comparison of vehicle sizes is shown in Figure 3.

4. PRECISE TIME, FREQUENCY AND EPHEMERIS DETERMINATION

Of the various available systems, only Loran and GPS are capable of time synchronization to better than 100 nanoseconds, as shown in Figure 4, while differential GPS is capable of another order of magnitude improvement.

4.1 ON-BOARD TIME AND FREQUENCY

A spaceborne GPS system can provide global precise time accuracy of 103 ns (Ref. 3), without using DGPS. Ten ns accuracies can be obtained by correcting for ionospheric delays with simultaneous view for coordinated users.

The Russian Glonass, of similar accuracies relative to Glonass UTC, has a known relationship to UTC(USNO) via the International UTC maintained by the Bureau Internationale des Poids et Mesures (BIPM), Sevres, France, and could supplement GPS data to achieve FAA/ICAO required reliability/availability levels for commercial aviation.

GPS time is determined by the USNO Master Clock (MC) of the USNO Time Service Department (TS). GPS can provide UTC (USNO) time, after correcting for the difference from UTC by an integral number of seconds. The GPS system continuous time scale is referenced to midnight 5 Jan 1980, whereas UTC and Russian GLONASS time contain leap seconds to allow for the deceleration of the earth rotation. This number is provided in the navigation message, because it changes every time there is a leap second.

Upon network synchronization, the clock, timing signals, and real time base, derived from accurate crystal or rubidium frequency standards, of performance as summarized in Figure 5, will be slaved by software control loops to the network time base, controlled by an assigned network controller, referenced using GPS, UTC or GMT.

Treating GPS as a Precise frequency and time distribution system, semiconductor laser optical clock distribution can provide a virtually jitter-less timing source. Accumulated dynamic clock jitter between ports of both correlated and uncorrelated sources is of the order of picoseconds (Ref. 5).

4.2 PRECISE EPHEMERIS DETERMINATION

Precise ephemeris data is important onboard for mission success, status monitoring, also for rendezvous and docking. TDRS measures to better than 100 m accuracy, but does not give instantaneous position fixes, as does GPS. Presently 60 to 195 minutes elapses between the start of tracking and the start of executing maneuvers. In low earth orbit, unpredictable atmospheric drag is the largest limit on prediction.

5. COMMUNICATIONS AND DATA SOURCE SYNCHRONIZATION

Using a GPS timing reference, communications with other GPS users can interoperate immediately without first exchanging timing signals. Sub-microsecond accuracy timing derived from GPS brings new meaning to "real time" in computer and communications synchronization.

5.1 END-TO-END LINK SYNCHRONIZATION

Synchronization needs include carrier, clock, code, and network. A representative end-to-end communications block diagram, are provided in Figure 6 (Ref. 1).

TDRSS provides short code PN lock acquisition with > 90% probability within 20 seconds at S/No values between 34 and 73 dB-Hz, and within 7 seconds for S/No > 37 dB-Hz, and carrier within 10 seconds. TDRSS user transponders provide doppler compensation within 1500 Hz of actual transponder center frequency, allowing for about 1% velocity uncertainty. GPS-derived velocity error is closer to 0.01%. Code acquisition strategies are matched to TDRSS-derived predicted position uncertainties.

With reduced GPS position and velocity uncertainties, acquisition windows could be considerably reduced in both time and frequency, leading to reduced sync times at increased S/No ratios. GPS time can also simplify the most complicated functions used in bandwidth efficient Time Division Multiple Access (TDMA) communications: precise and real-time synchronization of receive reference timing, transmit-timing control, acquisition control, unique synchronization word (UW) detection, and elastic buffering (Ref. 5).

5.2 DATA SOURCE SYNCHRONIZATION

Independent redundancy requires some form of synchronization among the independent data sources. In the past, continuous timing references were maintained using fault-tolerant redundant clock distribution, based on majority logic voting algorithms. As used in the NASA Fault Tolerant Multiprocessor (FTMP) and other high reliability concepts for manned aerospace vehicles, the timing reference must continuously remain within tolerances for time-correlated data transfer.

A set of GPS disciplined voltage-controlled phase-locked crystal oscillators can ensure that failure of one of the oscillators does not destroy the phase lock of the survivors. Normally all clock receiver outputs are in phase lock with each other and with all the oscillators.

6. IN-FLIGHT CLOCK CALIBRATION FOR REDUCED REACQUISITION "TIME TO FIRST FIX"

Many aerospace vehicles carrying atomic clocks to provide precise time to their communications systems need to be calibrated before take-off. With a GPS receiver on-board they can be reset or calibrated in flight. The clock can then be used to reduce time to first fix (TTFF) if the receiver has to reacquire satellites, and enables it to perform direct P-Code acquisition.

7. ENVIRONMENTAL STUDY REQUIREMENTS

Although widespread need for highest frequency and time accuracy may not be apparent, numerous applications for this accuracy will surface once the availability has been established. For instance, it may be of interest to review

requirements of a PLS-type vehicle for supporting various environ- mental study programs.

7.1 GRAVITY MODEL REFINEMENT

Precise GPS techniques are making possible new kinds of large scale surveys of the earth and its resources using aerospace platforms. The use of GPS receivers in space vehicles for decimeter-level space navigation and geophysics - including gravity field determination - will map broad global features of the gravity field, while future aircraft equipped with GPS will chart fine, regional details of the same field.

Gravity Probe B (GP-B) used GPS as a sensor for a closed loop space guidance system, from its location after initial insertion to a very precise low earth orbit, to within 0.001 eccentricity, and 0.001 degree inclination, and to alignment with the star Rigel to within 0.001 degree (Ref. 2).

7.2 RELATIVISTIC CLOCK EXPERIMENTS

Gravity Probe-B was also part of a NASA project that tested two aspects of Einstein's relativity theory, to verify two gyroscope spin axis motions in earth orbit not predicted by Newtonian analysis. In a precisely polar orbit the two effects would be orthogonal, with magnitudes calculated to be 6600 and 42 milliarc-sec/ year respectively (Ref. 2).

The required long term clock stability is about 10 E-13, or an accuracy of 10 ns, to successfully accomplish the relativistic clock experiment (Ref. 4).

7.3 NASA CRUSTAL DYNAMICS PROGRAM FOR EARTHQUAKE MECHANISMS STUDY

Movements of the earth's crust include extremely slow centimeters per year movements of the continents relative to one another, and also their slow internal deformations. Measuring distances between points up to thousands kilometers apart with accuracies exceeding a few parts in 100 million enables such movements to be detected.

Precision and accuracy of groundbased GPS determined interstation vectors for 50-450 km baselines were recently reported as better than 30 mm horizontally and 80 mm vertically, by comparing with very long baseline interferomerty measurements. Such measurements provide details about strain across faults of plate boundaries (Ref. 6).

7.4 SPACE GEODETIC TRACKING FOR GLOBAL CLIMATIC CHANGE STUDIES

The Topex/Poseidon oceanographic satellite will measure the topography of the open oceans using radar altimeters to a precision of a few cm. Measuring open ocean circulation and sea level as an indicator of global warming depends on detecting global sea level changes of about 1 mm/year. Tide gauges presently used measure relative change between the land surface and sea levels. This relative change may be due to sea level rise or land subsidence or both. Satellite altimetry measurements will have to be related to a well defined terrestrial and space reference frame that must remain valid for decades, with respect to a network of earth fiducial points (Ref. 7)

8. ENVIRONMENTAL STUDY SUPPORT

Reviewing the capabilities of a GPS-equipped PLS- type vehicle, potentially to support "Mission to Planet Earth" environmental studies, the following presently appear unlikely: 1) Precision orbit injection to support GP-B-type gravity model refinement, or precision to support 2) GP-B-type relativistic clock experiments, 3) Monitoring earth crust movements, or 4) Monitoring sea level changes for tracking climatic changes.

However years after dedicated environmental satellites have completed their missions, a PLS-type vehicle will be able to measure long-term drifts, and to monitor relative movements of earth fiducial points.

9. SUMMARY AND CONCLUSIONS

GPS onboard time, synchronized to the GPS space segment, provides the extremely precise reference needed for navigation, positioning purposes, and can serve as primary time transfer system, as well as for communications, tracking, electronic intelligence, metrology, & data acquisition.

Discretionary use of the PLS vehicle, with GPS on board, as a flexible platform to perform environmental studies will depend on the PTTI availability, and - at the very least - could identify long-term relative movements of earth fiducial points established for local or global reference frames.

Continous onboard filtering and orbital element updating based on GPS can eliminate groundbased state vector updates, provide earlier payload deployment opportunities, save fuel due to reduced dispersion rendezvous, make up for loss of TACAN by the year 2000, and also reduce communications bandwidth needs.

10. ACKNOWLEDGEMENTS

The author thanks Carl F. Ehrlich, AMLS/PLS Program Manager at the Rockwell Space Sytems Division, and Howard Stone, of the AMLS/PLS Program Office at NASA Langley Research Center, for permission to include Fig. 3, based on NASA Study Contract NAS1-18975 (Ref. 8). However the views expressed in this paper are for information only, and have not been endorsed by RIC or NASA.

11. REFERENCES

[1] Anderman A.,"Impact of AI, Fuzzy Logic, & GPS on Future Avionics", 5-12-92, RIC, Downey, CA

[2] Axelrad P.& Parkinson B.W.,"Closed Loop Navigation and Guidance for Gravity Probe B Orbit Insertion", *Navigation*, Vol. 36, No. 1, Spring 1989, P. 45-61

[3] Colomo O.L. & Peters M.F.,"Precision Long-Range DGPS for Airborne Surveys", *GPS World*, April 1992, P.44-50

[4] Green G.B.& Axelrad P.,"Space Applica- tions of GPS", *Proc.ION National Technical Meeting*, Jan. 1989, San Mateo, CA, P.89- 104

[5] Kato S. et al,"Application of Advanced Microelectronics to Large-Scale Communication Equipment -

requirements of a PLS-type vehicle for supporting various environ- mental study programs.

7.1 GRAVITY MODEL REFINEMENT

Precise GPS techniques are making possible new kinds of large scale surveys of the earth and its resources using aerospace platforms. The use of GPS receivers in space vehicles for decimeter-level space navigation and geophysics - including gravity field determination - will map broad global features of the gravity field, while future aircraft equipped with GPS will chart fine, regional details of the same field.

Gravity Probe B (GP-B) used GPS as a sensor for a closed loop space guidance system, from its location after initial insertion to a very precise low earth orbit, to within 0.001 eccentricity, and 0.001 degree inclination, and to alignment with the star Rigel to within 0.001 degree (Ref. 2).

7.2 RELATIVISTIC CLOCK EXPERIMENTS

Gravity Probe-B was also part of a NASA project that tested two aspects of Einstein's relativity theory, to verify two gyroscope spin axis motions in earth orbit not predicted by Newtonian analysis. In a precisely polar orbit the two effects would be orthogonal, with magnitudes calculated to be 6600 and 42 milliarc-sec/ year respectively (Ref. 2).

The required long term clock stability is about 10 E-13, or an accuracy of 10 ns, to successfully accomplish the relativistic clock experiment (Ref. 4).

7.3 NASA CRUSTAL DYNAMICS PROGRAM FOR EARTHQUAKE MECHANISMS STUDY

Movements of the earth's crust include extremely slow centimeters per year movements of the continents relative to one another, and also their slow internal deformations. Measuring distances between points up to thousands kilometers apart with accuracies exceeding a few parts in 100 million enables such movements to be detected.

Precision and accuracy of groundbased GPS determined interstation vectors for 50-450 km baselines were recently reported as better than 30 mm horizontally and 80 mm vertically, by comparing with very long baseline interferomerty measurements. Such measurements provide details about strain across faults of plate boundaries (Ref. 6).

7.4 SPACE GEODETIC TRACKING FOR GLOBAL CLIMATIC CHANGE STUDIES

The Topex/Poseidon oceanographic satellite will measure the topography of the open oceans using radar altimeters to a precision of a few cm. Measuring open ocean circulation and sea level as an indicator of global warming depends on detecting global sea level changes of about 1 mm/year. Tide gauges presently used measure relative change between the land surface and sea levels. This relative change may be due to sea level rise or land subsidence or both. Satellite altimetry measurements will have to be related to a well defined terrestrial and space reference frame that must remain valid for decades, with respect to a network of earth fiducial points (Ref. 7)

8. ENVIRONMENTAL STUDY SUPPORT

Reviewing the capabilities of a GPS-equipped PLS- type vehicle, potentially to support "Mission to Planet Earth" environmental studies, the following presently appear unlikely: 1) Precision orbit injection to support GP-B-type gravity model refinement, or precision to support 2) GP-B-type relativistic clock experiments, 3) Monitoring earth crust movements, or 4) Monitoring sea level changes for tracking climatic changes.

However years after dedicated environmental satellites have completed their missions, a PLS-type vehicle will be able to measure long-term drifts, and to monitor relative movements of earth fiducial points.

9. SUMMARY AND CONCLUSIONS

GPS onboard time, synchronized to the GPS space segment, provides the extremely precise reference needed for navigation, positioning purposes, and can serve as primary time transfer system, as well as for communications, tracking, electronic intelligence, metrology, & data acquisition.

Discretionary use of the PLS vehicle, with GPS on board, as a flexible platform to perform environmental studies will depend on the PTTI availability, and - at the very least - could identify long-term relative movements of earth fiducial points established for local or global reference frames.

Continous onboard filtering and orbital element updating based on GPS can eliminate groundbased state vector updates, provide earlier payload deployment opportunities, save fuel due to reduced dispersion rendezvous, make up for loss of TACAN by the year 2000, and also reduce communications bandwidth needs.

10. ACKNOWLEDGEMENTS

The author thanks Carl F. Ehrlich, AMLS/PLS Program Manager at the Rockwell Space Sytems Division, and Howard Stone, of the AMLS/PLS Program Office at NASA Langley Research Center, for permission to include Fig. 3, based on NASA Study Contract NAS1-18975 (Ref. 8). However the views expressed in this paper are for information only, and have not been endorsed by RIC or NASA.

11. REFERENCES

[1] Anderman A.,"Impact of AI, Fuzzy Logic, & GPS on Future Avionics", 5-12-92, RIC, Downey, CA

[2] Axelrad P.& Parkinson B.W.,"Closed Loop Navigation and Guidance for Gravity Probe B Orbit Insertion", *Navigation*, Vol. 36, No. 1, Spring 1989, P. 45-61

[3] Colombo O.L. & Peters M.F.,"Precision Long-Range DGPS for Airborne Surveys", *GPS World*, April 1992, P.44-50

[4] Green G.B.& Axelrad P.,"Space Applica- tions of GPS", *Proc.ION National Technical Meeting*, Jan. 1989, San Mateo, CA, P.89- 104

[5] Kato S. et al,"Application of Advanced Microelectronics to Large-Scale Communication Equipment -

Compact and Maintenance Free TDMA Equipment", *IEEE Jour on Selected Areas in Comm.*, Vol. 8, No. 8, Oct 1990, P. 1551-156[illegible]

[6] Larson K.M. et al,"GPS Application to Crustal Deformation Measurement: 1.Precision and Accuracy, and 2.Influence of Errors in Orbit Determination Networks", *Jour. Geophys. Res.*, Vol. 96, No. B10, 10 Sept 1991, P. 16547-16584

[7] McDoran P.F.& Born G.H.,"Time, Frequency, and Space Geodesy: Impact on the Study of Climate and Global Change", *Proc.IEEE*, Vol. 79, No. 7, July 1991, P. 1063-1069

[8] Rockwell SSD,"Advanced Manned Launch System/Personnel Launch System (AMLS/ PLS) Study": a)"AMLS Final Technical Review", Aug. 1991, & b)"PLS Hardware/ Software Design Description", Aug.1990

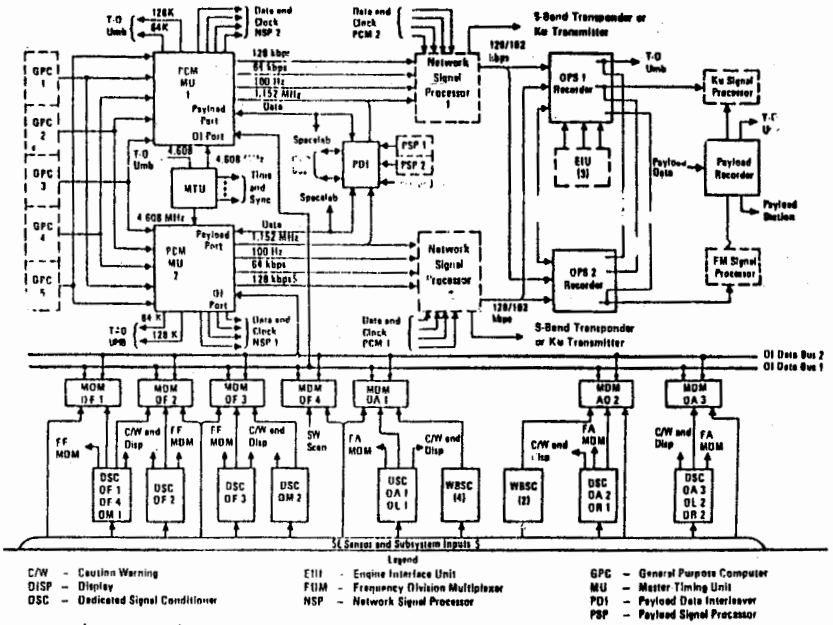


Figure 1. Time/Clock/Sync Inputs to Operational Instrumentation (OI)

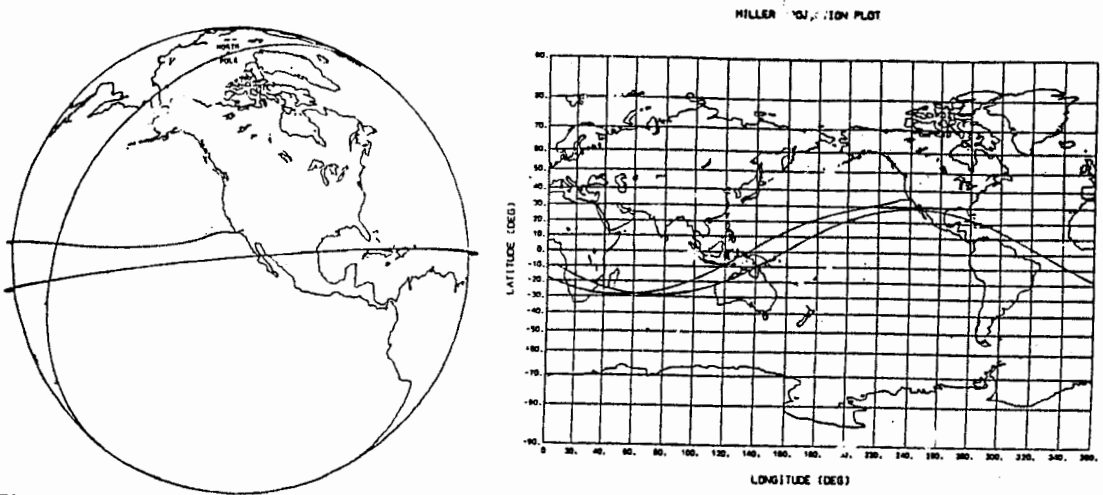


Figure 2. Space Shuttle GPS Developmental Flight Test (DFT) Profile: Orbit, Deorbit, and Entry

Compact and Maintenance Free TDMA Equipment", *IEEE Jour on Selected Areas in Comm.*, Vol. 8, No. 8, Oct 1990, P. 1551-1564

[6] Larson K.M. et al,"GPS Application to Crustal Deformation Measurement: 1.Precision and Accuracy, and 2.Influence of Errors in Orbit Determination Networks", *Jour. Geophys. Res.*, Vol. 96, No. B10, 10 Sept 1991, P. 16547-16584

[7] McDoran P.F.& Born G.H.,"Time, Frequency, and Space Geodesy: Impact on the Study of Climate and Global Change", *Proc.IEEE*, Vol. 79, No. 7, July 1991, P. 1063-1069

[8] Rockwell SSD,"Advanced Manned Launch System/Personnel Launch System (AMLS/ PLS) Study": a)"AMLS Final Technical Review", Aug. 1991, & b)"PLS Hardware/ Software Design Description", Aug.1990

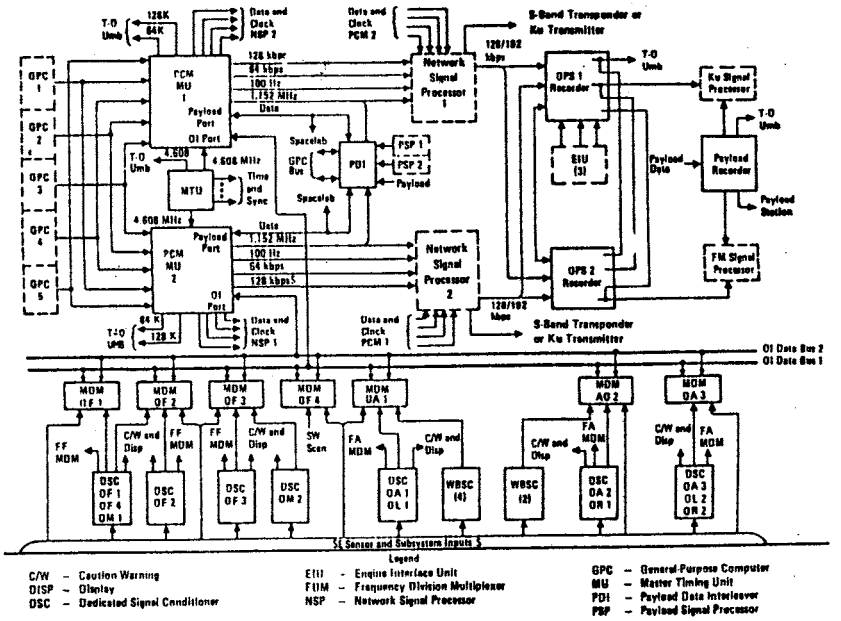


Figure 1. Time/Clock/Sync Inputs to Operational Instrumentation (OI)

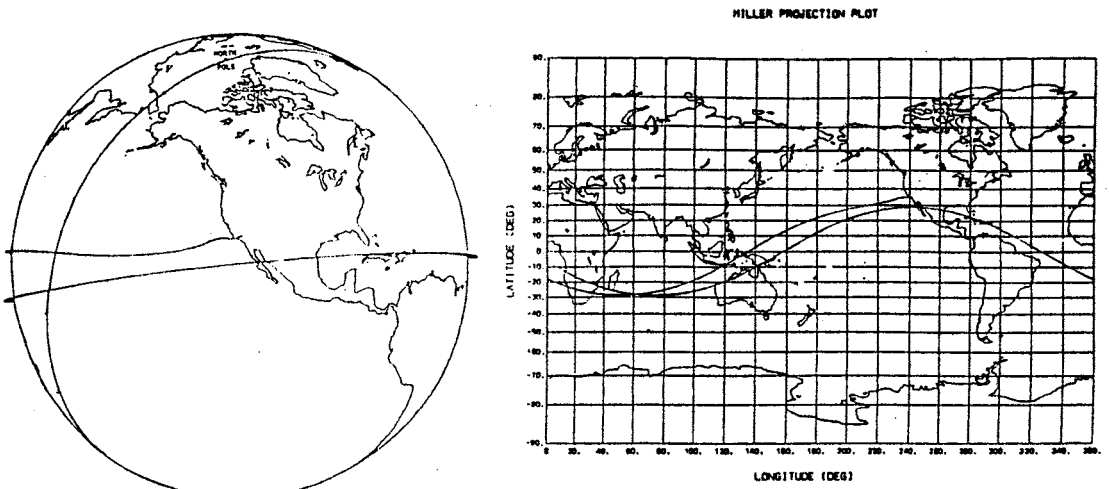


Figure 2. Space Shuttle GPS Developmental Flight Test (DFT) Profile: Orbit, Deorbit, and Entry

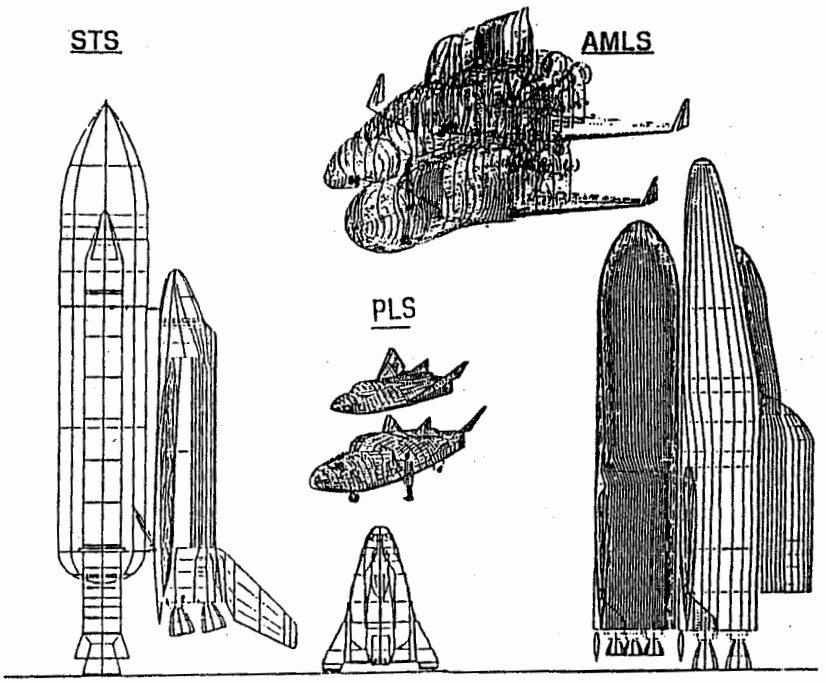


Figure 3. Relative Sizes of STS, AMLS, and PLS

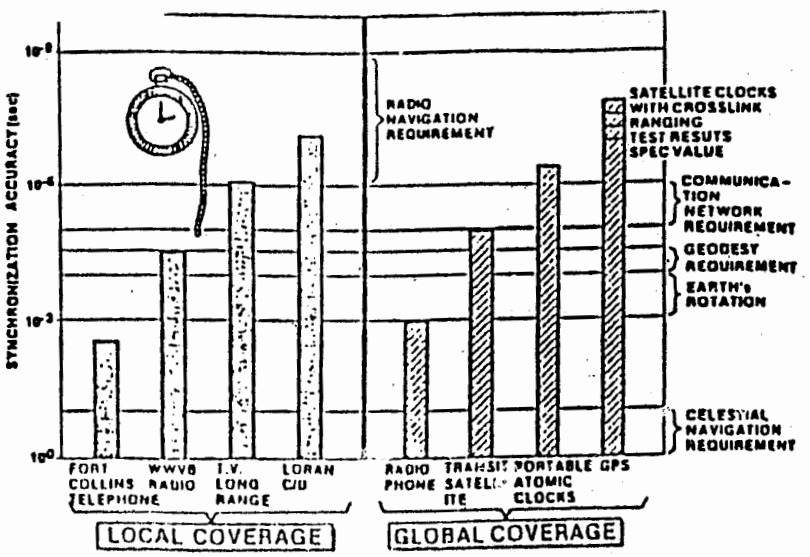


Figure 4. Time Synchronization Capabilities of Various Systems

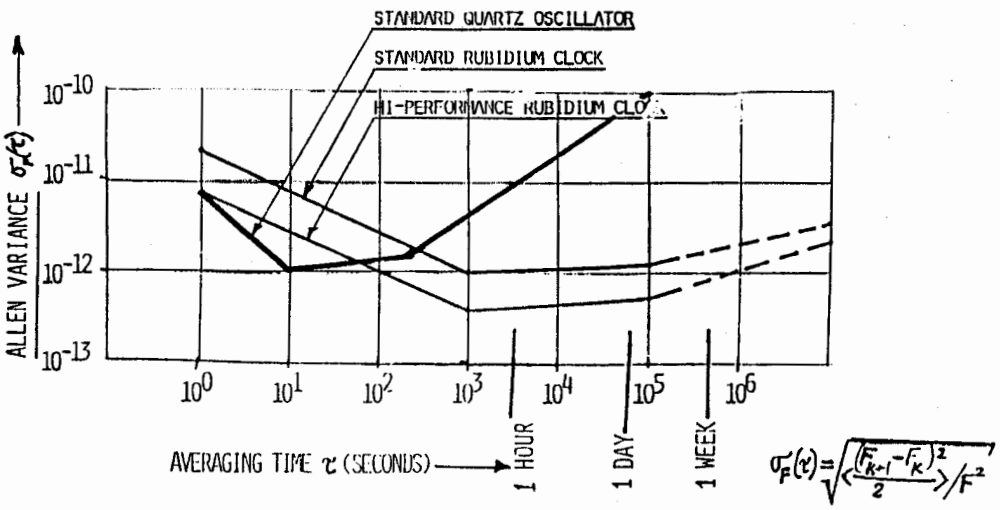


Figure 5. Comparison of Quartz Crystal Oscillator & Rubidium Clock Frequency Stability

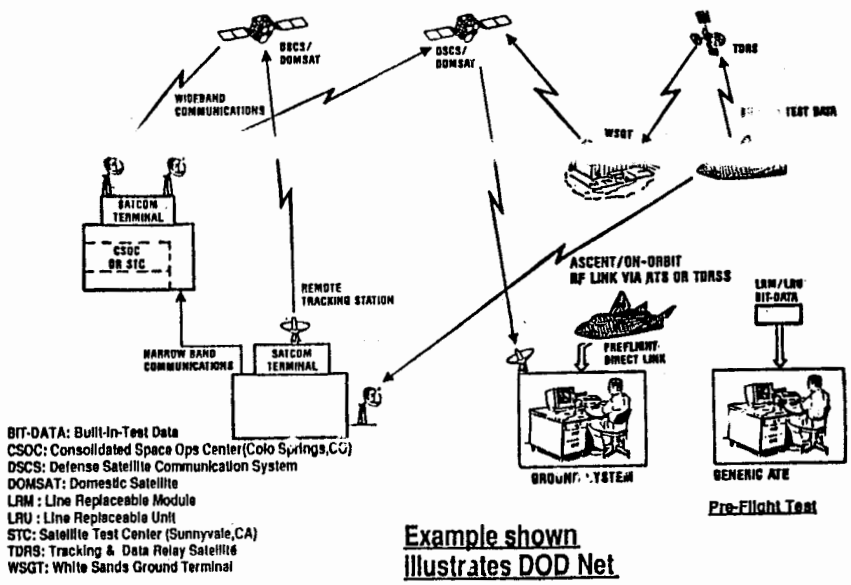


Figure 6. Universal GPS Time Reference Usage Will Aid Communications Synchronization

DESIGN OF A HYDROGEN MASER FOR SPACE and PROGRESS ON CRYOGENIC HYDROGEN MASER RESEARCH

R.F.C. Vessot, E.M. Mattison, G.U. Nystrom, L.M. Coyle,
R.L. Walsworth, R. Nicoll, and D. Graveline
Harvard-Smithsonian Center for Astrophysics
60 Garden Street, Cambridge MA 02138 U.S.A.
and
R. Decher
George C. Marshall Space Flight Center
Huntsville, AL 35812 U.S.A.

1. INTRODUCTION

We report progress at the SAO Hydrogen Maser Laboratory during the past year in two areas of work: development of an atomic hydrogen (H) maser for use in space, and initial results from successful oscillation of a cryogenically cooled H maser operating at 0.5 K.

2. SPACE MASER DEVELOPMENT

2.1 Project Overview

At last year's EFTF we described the conceptual design of the SAO/NASA Hydrogen Maser Clock (HMC) program[1]. The HMC program is a technology demonstration experiment in which a hydrogen maser frequency standard will be designed, built, and tested for long-term use in space. The maser will be flown on a European Space Agency (ESA) EURECA spacecraft for approximately six months in a 500 km altitude, 28° inclination orbit. The maser's frequency will be compared with clocks on the ground by means of laser pulse time transfer using an earth-based satellite laser ranging station located on Mt. Haleakala, Hawaii. When laser pulses arrive at the spacecraft they will be reflected back to the laser station by an array of corner cube retroreflectors mounted on the spacecraft, and will also be detected by a photodetector colocated with the retroreflector array. The arrival time of a pulse at the spacecraft, expressed in terms of the *space clock's* time scale, will be determined by a high-resolution event timer that uses the spacecraft maser's output signal as a reference. The pulse's arrival time at the spacecraft, expressed in terms of the *ground-based clock's* time scale, will be determined from the time the pulse left the ground, plus one-half the round trip travel time, both measured by an event timer similar to that on the spacecraft, but in terms of time kept by the ground-based clock. These two measures of the same event – arrival of a pulse at the spacecraft – provide a comparison of the rates of the space and ground clocks, and ultimately a measure of the space clock's frequency in terms of international time scales. We expect a precision of better than 100 picoseconds in these comparisons. This level of precision will reveal day to day systematic variations of the space clock's frequency with a resolution of about 1 part in 10^{15}. During the mission, a number of maser operating parameters will be varied by command to assess the performance of the maser system in space. A schematic diagram of the space and ground system is shown in Figure 1.

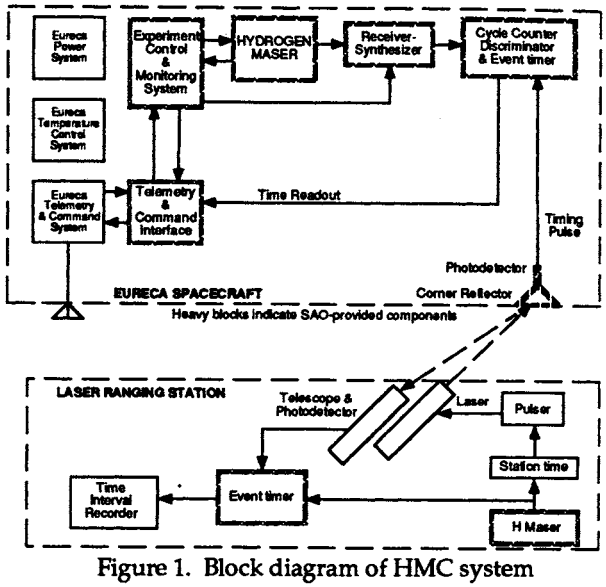


Figure 1. Block diagram of HMC system

The space maser is operated by a microprocessor that controls and monitors the experiment and serves as a data interface to the EURECA telemetry and command system. The microprocessor is programmed to perform automatically a number of maser diagnostic operations and adjustments, such as the varying the dissociator oscillator frequency and power level, and determining the magnetic field in the maser's storage bulb using a swept frequency technique to measure the frequency of the (F=1, m_F = +1) to (F =1, m_F = 0) Zeeman transition of atomic hydrogen.

2.3 Retroreflector, photodetector, and event timer

The EURECA spacecraft is a microgravity platform that is continuously oriented in a fixed direction relative to the sun during its orbit around the earth; as a result it presents a changing aspect to the ground throughout its orbit. In order to ensure that laser reflection is possible regardless of the spacecraft's attitude relative to the earth, EURECA will be equipped with two retroreflector arrays, each having an approximately hemispherical field of view. Each 12-cm diameter array will contain approximately 20 corner cubes of 1.3 cm diameter. At the apex of each reflector array will be mounted an omnidirectional laser light collector consisting of a bundle of about 60 optical fibers. At the input end of the bundle, the fibers fan out into a hemispherical shape, giving a field of view of approximately 2π steradians; at the exit end of the bundle, the fibers come together and connect to a silicon photodetector. Within the photodetector housing are focusing lenses; a nar-

row-band optical filter centered on the laser wavelength of 532 nm to exclude noise produced by ambient light; and a neutral density filter to bring the laser intensity in the linear range of the photodetector. Figure 2 shows a conceptual view of the reflector-detector system.

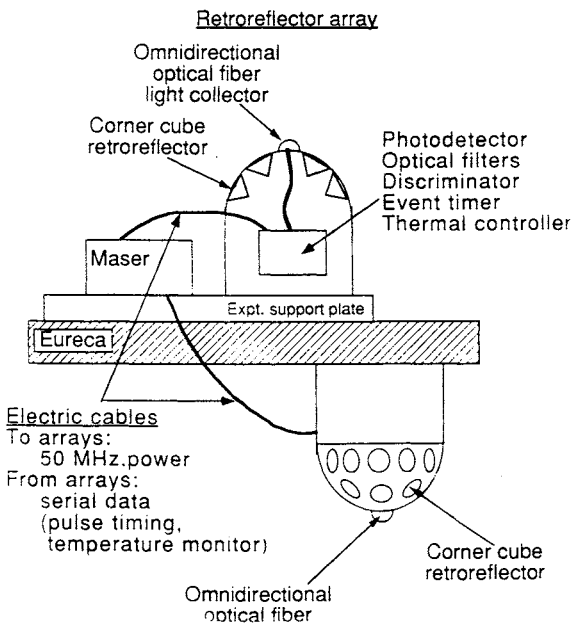


Fig. 2. Conceptual diagram of reflector-detector system. Component dimensions not to scale.

The output pulse from the photodetector is sensed by a constant fraction discriminator [2] that compensates for the variation of pulse shape with amplitude, and its arrival time is determined by a precision event timer. The event timer's clock is derived from the 50 MHz output from the maser's phase-lock receiver, whose frequency is adjustable by microprocessor command of a direct digital synthesizer in increments of 8 parts in 10^{18}. The heart of the event timer is an analog time interpolator [3] whose capacitive charge–discharge circuit interpolates the clock signal's 20 nanosecond period by a factor of 1000, resulting in a timing resolution of 20 picoseconds.

2.4 The atomic hydrogen maser physics package

The basic design of the maser follows in many aspects the design of the maser we developed for the SAO/NASA Gravitational Redshift Test (GP-A) flown in 1976. [4] Because the objective of this program is to qualify an H-maser suitable for a wide range of space applications, we have designed the maser to cope with the vibrational, thermal and magnetic environments of a variety of launch vehicles and orbits. The maser is shown in cross-section in Figure 3. It is mounted on a standard EURECA 70x70 cm experiment support plate by means of a triangular support structure. Five layers of magnetic shielding and seven thermal control zones surround the resonant cavity and hydrogen storage bulb. The outermost magnetic shield encloses the entire maser.

The hydrogen gas needed to operate the maser is obtained by heating a small amount (about 25 g) of lithium aluminum hydride ($LiAlH_4$). The resulting H_2 passes through a temperature-controlled palladium-silver diaphragm that purifies the gas and controls its flow to an r.f. discharge tube that dissociates it into atomic hydrogen (H). The dissociator is a cylindrical glass vessel surrounded by a coil driven by about four watts of power at a frequency of about 70 MHz. A beam of atomic hydrogen emerges from the dissociator through a multitube collimator and enters a hexapole state selection magnet. Atoms is the upper (masing) state are focused into the collimated aperture of the Teflon-lined quartz storage bulb. The bulb is located in a cylindrical TE_{011}-mode microwave resonant cavity made of low-expansion Cervit coated on its inside with copper and tuned to the hydrogen maser transition frequency of 1.42 GHz. The cavity's resonance frequency is tunable over a range of about 10 kHz by a varactor diode coupled to the cavity's field by a wire loop; a second loop couples the maser signal at a level of about 10^{-12} watts out of the resonator. The output signal is amplified by a phase-lock receiver that produces high-level signals at standard frequencies such as 50 MHz for use by the experiment's electronics.

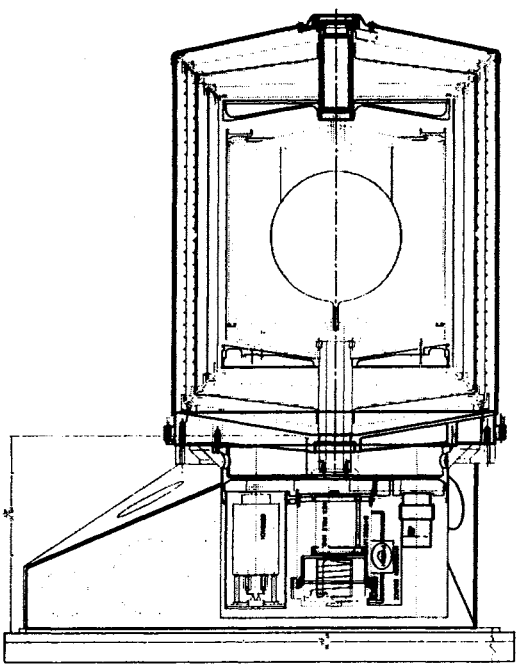


Fig. 3. Cross section view of space maser

Two sorption cartridges located below the main mounting plate scavenge expended hydrogen. A small ion pump removes non-hydrogen species and monitors the residual gas pressure. Metal vacuum seals are used throughout the maser.

Mechanically, the maser has been designed to cope with dynamic loads in excess of 9.5 g rms. Figure 4 is a representation of the finite-element mathematical model used to characterize the behavior of the maser under dynamic loads.

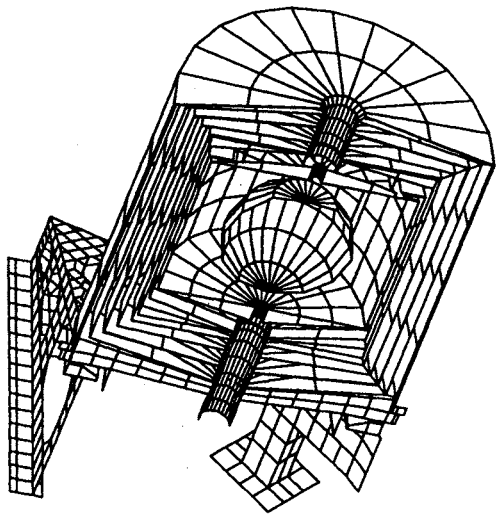


Fig. 4. Finite-element structural model of space maser and support base.

2.5 Thermal design

The maser and electronics will dissipate about 64 watts under steady-state operation. Because the entire package will be surrounded with multilayer reflective insulation, nearly all of this heat will be conducted by the ESP to a plate cooled with Freon that is circulated by the EURECA payload system. The housings that contain the retroreflectors and enclose the photodetectors and event timers will dissipate about 9 watts of power and are designed to operate with full exposure to space.

To make realistic estimates of the behavior of the thermal control system we built an engineering demonstration model of the space maser and equipped it with all systems necessary to operate it in space. We constructed a finite element thermal model of the maser to predict its performance, and we measured the maser's thermal behavior both in air and in a vacuum tank under a variety of conditions. Figure 5 shows the system prior to its installation in the vacuum tank, which is shown in the background of the picture. Temperature was measured at 14 locations. Thermal conduction to the outside of the tank to simulate thermal contact to the ESP is provided by two aluminum channels, one of which is shown in Figure 5. Table 1 gives the measured and predicted heat dissipations of the maser's major temperature control zones under three conditions. The results verify the validity of the computer model.

Table 1 – Measured and Predicted Thermal Control Power for Demonstration Maser

Zone	Air (W)		Vacuum (W)	
	Meas	Pred	Meas	Pred
Inner	9.0	8.1	2.5	2.8
Outer	8.4	8.5	2.5	2.6

We expect to dissipate a total of about 25 watts of thermal control power in all zones, including the outer zone (which is not included in the demonstration model shown in Figure 5) under conditions of the coldest expected cooling fluid temperature. Under nominal fluid conditions the expected power would be somewhat less than 20 watts, and around 12 watts at the maximum expected coolant temperature.

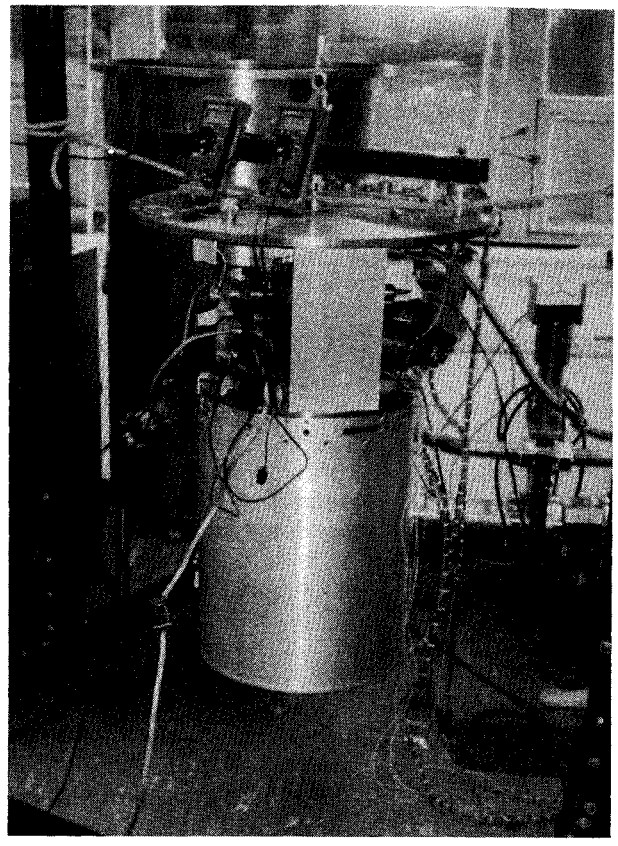


Fig. 5. Photograph of demonstration model of space maser

2.6 Magnetic fields and shielding

The maser's frequency can be affected by variations in the ambient magnetic field. In order to keep magnetically-induced frequency changes below 1 part in 10^{15}, the magnetic field within the maser's storage bulb must be maintained constant to less than 0.5 microgauss. This is accomplished in part by a five-layer cylindrical magnetic shield array that surrounds the maser's resonant cavity and storage bulb. Frequency shifts can also result from variations in the populations of the F=1 magnetic sublevels that enter the bulb, coupled with inhomogeneity of the magnetic field within the bulb. To prevent these shifts, we have designed the outermost magnetic shield to enclose the entire maser, thus preventing transitions among the F=1 sublevels that could be induced by changes in the magnetic field along the beam axis.

We have identified two main sources of ambient magnetic field variation: the magnetic torquers used to orient EURECA toward the sun, and the earth's varying magnetic field direction and intensity as the probe orbits the earth.

The worst-case magnetic variations expected in EURECA's 525 km altitude, 28.5° inclination orbit

due to the combination of the two effects is ±0.5 gauss parallel to the maser's axis. The required axial shielding factor S to maintain $\Delta f/f \sim 1 \times 10^{-15}$ is $S \sim 2 \times 10^6$.

A five-layer shield provides ample transverse shielding, but its axial shielding factor is approximately 3×10^5, low by a factor of about 10. To offset the effects of axial magnetic field variations, an active field compensation servo will be used. This system consists of a magnetic sensor located at the apex of the 4th shield, and a field compensation solenoid wound on the 4th shield. Such a system is capable of increasing the degree of axial field cancellation by a factor of up to 30 for an external magnetic field perturbation of ±1 gauss [5]; such an increase would reduce the effect of magnetic fields on the maser's frequency to a level well below a part in 10^{15}.

3. CRYOGENIC HYDROGEN MASER RESEARCH

3.1. Introduction

A cryogenic hydrogen maser (CHM) operating at temperatures below 1 K could have a frequency stability that is two to three orders of magnitude better than a room temperature H maser due to: (i) lower thermal noise; and (ii) a reduced H-H spin-exchange relaxation rate and the resulting higher operational signal power. In 1986 three groups [6,7,8], including ours, demonstrated oscillation of a CHM using superfluid helium-lined containers. It has subsequently been predicted [9,10] that new spin-exchange effects may make it more difficult to achieve a radical improvement in frequency stability using a CHM. In practice, the stability that can be realized by the CHM remains to be determined.

3.2. Harvard-Smithsonian CHM

The Harvard-Smithsonian CHM is currently operated in a new, high-cooling power recirculating ^{3}He refrigerator. This refrigerator provides 2 mW of cooling power at 0.5 K in the recirculating mode and with the CHM operating. A cross-sectional view of the CHM is shown in Figure 6; a detailed discussion of the CHM's design and operation has been published previously [3]. In the CHM, H atoms from a room temperature microwave discharge enter the ^{3}He refrigerator and are cooled to ~10 K. The atomic beam passes through a hexapole magnet that focusses the two upper hyperfine states into a microwave resonant cavity (at 1420 MHz) that is maintained ~ 0.5 K in a region of stable and homogeneous magnetic field. The cold H atoms are stored within the cavity for times $\gtrsim$ 2 seconds in a chamber coated with a superfluid helium film. With a modest input flux ($\gtrsim 10^8$ sec^{-1}) the Harvard-Smithsonian CHM operates with continuous maser oscillation. The CHM signal is detected inductively, isolated, amplified, and sent to a phase-locked heterodyne receiver. The frequency and power of the output signal are precisely measured. Room temperature H masers in our laboratory serve as frequency comparison standards for the CHM.

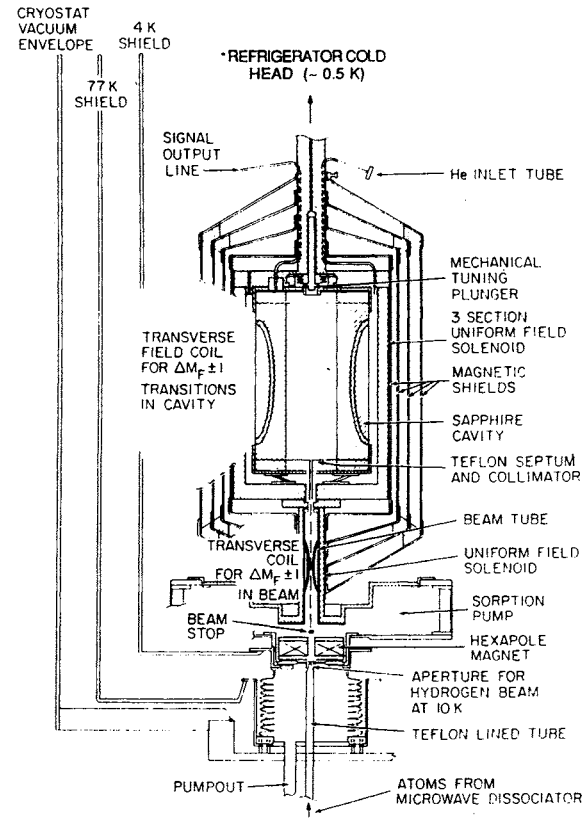


Fig. 6. Cross section view of CHM.

3.3. Recent results

The Harvard-Smithsonian CHM/^{3}He refrigerator system has been fully operational since early 1993. In this section we descibe our experimental results as of mid-April 1993.

We have operated the CHM with a maximum H input flux $\approx 10^{11}$ sec^{-1}, resulting in a maximum signal power $\approx 10^{-13}$ W and a maximum H density $\approx 10^9$ cm^{-3}. We have measured the CHM output frequency as a function of temperature (Fig. 7). The CHM signal frequency is decreased by H collisions both with the He-film-coated walls of the storage chamber and with the accompanying He vapor. These two negative frequency shifts have opposite trends with temperature, since the H-wall residency time decreases with temperature and the He vapor pressure increases with temperature. Thus there is a stable minimum in the magnitude of the total H-He frequency shift. For the current CHM geometry this minimum is near 0.53 K.

Our measurements show a stronger dependence of CHM frequency on temperature than that reported by the University of British Columbia (UBC) group [11]. We believe that this larger "wall shift" in our system is due to H interactions with paramagnetic impurities in the sapphire substrate that lies beneath the He film. (Note that there is also a 50 μm layer of Teflon between the sapphire and the He.) The hypothesis of strong H-wall interactions in our CHM is supported by two other measurements:

1) We have determined the absolute value of the H-wall frequency shift to be – 0.297 (5) Hz at T = 0.490 (2) mK. The results of the UBC group, however, imply a much smaller frequency shift ≈ – 0.1 Hz.

2) Our CHM's measured line Q is approximately 2.3×10^9 with insignificant magnetic and collisional relaxation, implying $T_2 \approx 0.52$ seconds. The contribution to T_2 from H escape from the storage bulb, as calculated from geometry, is approximately 2.5 seconds

Thus we conclude that the observed T_2 is due primarily to wall relaxation.

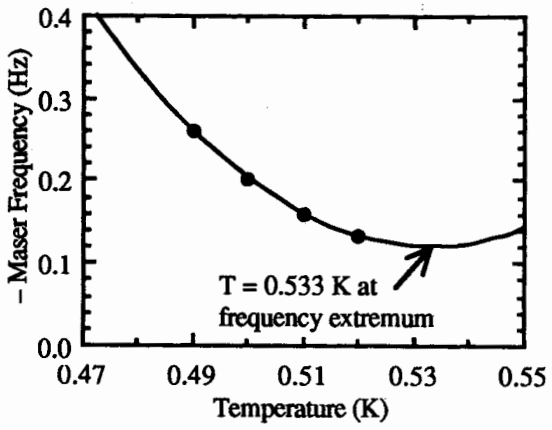


Fig. 7. Measured dependence of CHM signal frequency on temperature, and quadratic fit to data.

We have also measured the CHM's frequency stability relative to room temperature H masers, albeit for less than ideal operational conditions. At T ≈ 0.51 K, with 0.2 mK temperature control and a signal power ≈ 10^{-14} W, we found $\sigma_y(10\ \text{sec}) \approx 5 \times 10^{-13}$, where $\sigma_y(t)$ is the Allan, or two-sample, deviation [12], a measure of fractional frequency stability over an averaging interval τ. This measurement is consistent with the CHM frequency variations arising from the system's temperature fluctuations, given the results shown in Fig. 1.

3.4. Work in progress

We have recently made modifications to our CHM/^{3}He refrigerator system to improve the maximum H input flux into the maser by 2-3 orders of magnitude, and to maintain temperature control better than 0.05 mK. These system improvements are currently allowing us to do the following:

1) Investigate H-H spin-exchange collisions at temperatures ~ 0.5 K, including hyperfine-induced frequency shifts predicted [4,5] to be large at low temperatures and to affect the performance of new frequency standards like the CHM and the cesium fountain.

2) Determine the CHM's frequency stability while operating with good temperature control at the stable minimum of the H-He frequency shift.

4. ACKNOWLEDGEMENTS

The NASA/SAO Hydrogen maser clock program is supported by the George C. Marshall Space Flight Center of the National Aeronautics and Space Administration under contract number NAS8-39194.

The work on the cryogenic hydrogen maser is supported by the United States Air Force Office of Scientific Research under grant number AFOSR-91-0282.

5.0 REFERENCES

1. R.F.C. Vessot, E.M. Mattison, G.U. Nystrom, L.M. Coyle, R. Decher, S.J. Feltham, G. Busca, S. Starker, and S. Leschutta, Hydrogen masers for space. Proc. 6th Annual European Frequency and Time Forum, Noordwijk, NL, 17-19 March 1992. Pp 19-25.

2. B.T. Turko and R.C. Smith, A precision timing discriminator for high density detector systems. Lawrence Berkeley Laboratory report LBL-30602. Presented at IEEE Nuclear Science Conference, 1992; to be published in IEEE Trans. Nucl. Sci.

3. B. Turko, Multichannel interval timer. IEEE Trans. Nucl Sci., **NS-31, 167-171 (1984).**

4. R.F.C. Vessot, M.W. Levine, E.M. Mattison, et.al., Tests of relativistic gravitation with a space-borne hydrogen maser. Phys. Rev. Lett. **45**, 2081-2084 (1980)

5. H.E. Peters, et al., Atomic hydrogen masers with self auto-tune systems and magnetic field cancellation servo. Proc. 20th Annual Precise Time and Time Interval Applications and Planning Meeting, 337 (1988).

6. H.F. Hess, G.P. Kochanski, J.M. Doyle, T.J. Greytak, D. Kleppner, Spin-polarized hydrogen maser. Phys. Rev. A **34**, 1602 (1986).

7. M.D. Hürlimann, W.N. Hardy, A.J. Berlinsky, R.W. Cline, Recirculating cryogenic hydrogen maser. Phys. Rev. A **34**, 1605 (1986).

8. R.L. Walsworth, I.F. Silvera, H.P. Godfried, C.C. Agosta, R.F.C. Vessot, E.M. Mattison, Hydrogen maser at temperatures below 1 K. Phys. Rev. A **34**, 2550 (1986).

9. B.J. Verhaar, J.M.V.A. Koelman, H.T.C. Stoof, O.J. Luiten, S.B. Crampton, Hyperfine contribution to spin-exchange frequency shift in the hydrogen maser. Phys. Rev. A **35**, 3825 (1987).

10. J.M.V.A. Koelman, S.B. Crampton, H.T.C. Stoof, O.J. Luiten, B.J. Verhaar, Spin-exchange frequency shifts in cryogenic and room-temperature hydrogen masers. Phys. Rev. A **38**, 3535 (1988).

11. M. Morrow, R. Jochemsen, A.J. Berlinsky, and W.N. Hardy, Zero-field hyperfine resonance of atomic hydrogen for 0.18<T<1 K: the binding energy of H on liquid ^{4}He. Phys. Rev. Lett. **46**, 195 (1981).

12. D. Allan, Statistics of atomic frequency standards. Proc. IEEE **54**, 105 (1966).

POSTDEADLINE POSTER

TEMPORAL CHARACTERIZATION OF 100 MHz PICOSECOND LASER PULSES. APPLICATION TO PHASE NOISE MEASUREMENT OF A CRYSTAL CONTROLLED OSCILLATOR.

B. Cunin, C. Eckert, A. Haessler, F. Heisel, J.A. Miehé,
Groupe d'Optique Appliquée, Centre de Recherches Nucléaires, 67200 Strasbourg, France.
M. Planat,
Laboratoire de Physique et Métrologie des Oscillateurs du C.N.R.S.
associé à l'Université de Franche-Comté, Besançon, 25000 Besançon, France.

1. MEASUREMENT UNDER INVESTIGATION.

Statistics of the time intervals between the arrivals of ps light pulses I(t) and the instants of zero crossing of a sine wave V(t) [Fig.1] :

$$T_1(t) = T(t) - \Theta(t)$$

$$T(t) \rightarrow I(t) = \sum_{k=-\infty}^{+\infty} A(t)\delta\left[t + kt_0 + T(t)\right]$$

$$\Theta(t) \rightarrow V(t) = \sin\omega_0\left[t + \Theta(t) + \Theta_0\right]$$

$$\text{with:} \quad t_0 = \frac{2\pi}{\omega_0}$$

In the deterministic case T(t) = 0 and θ(t) = 0.

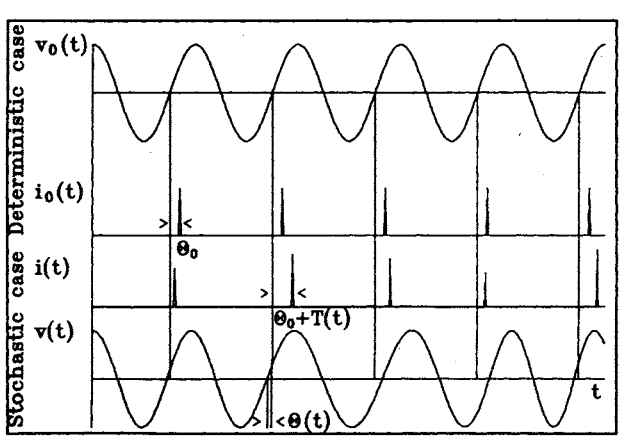


Fig. 1 : LASER BEAM I(t) AND SINE WAVE V(t)

2. INTEREST OF THE MEASUREMENT.

Commonly, the phase noise variance $\sigma^2(\tau,\tau')$ is expressed by :

$$\sigma^2(\tau,\tau') = \frac{1}{2}E\left[\bar{\Theta}_2 - \bar{\Theta}_1\right]^2$$

with :

$$\bar{\Theta}_k = \frac{1}{\tau}\int_{k\tau'}^{k\tau'+\tau} \Theta(t)dt$$

where τ is the measurement duration and τ'-τ is the dead time between two measurements.

The Allan variance is obtained with $\tau = \tau'$.

In the present investigation the variance is determined by measuring the phase at the instants t and t + τ' wich corresponds to the case τ = 0 so that :

$$\sigma^2(\tau') = E\left[\Theta(t+\tau') - \Theta(t)\right]^2$$

3. EXPERIMENTAL ARRANGEMENT [Fig. 2 and 3].

The oscillator under test drives the acousto-optic modelocker (G) of a cw Nd:YAG laser (L) associated with a synchronously pumped dye laser (DL). The ps pulses are recorded by a streak camera (SC) connected to an optical multichannel analyser (OA) and to a computer (CC). The time gate (T_G) is fixed at 10 ns. Two synchronization set-ups are examined :

a) in the PASSIVE SYNCHRONIZATION [Fig. 2], the deflection voltage is directly obtained from the oscillator signal after passing through a frequency doubler (FD) ;

b) in the ACTIVE SYNCHRONIZATION [Fig. 3], a mirror (M) reflects a part of the laser beam on a photodiode (D) whose output triggers a tunnel diode (TD) generating the sweeping voltage.

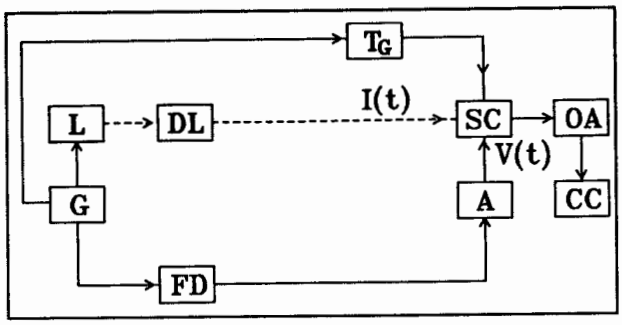


Fig. 2 : PASSIVE SYNCHRONIZATION.

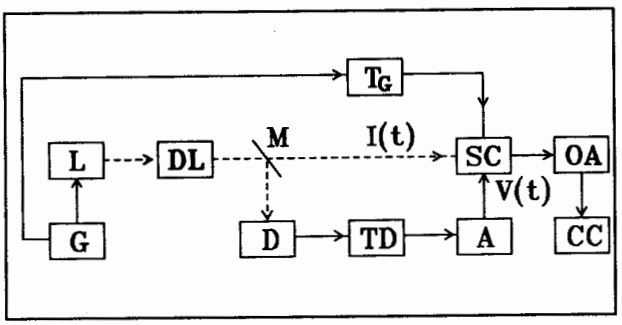


Fig. 3 : ACTIVE SYNCHRONIZATION.

In both configurations the sine wave is directed to an amplifier (A) and applied to the deflectors of the streak camera.

4. CHARACTERIZATION OF THE SERIES $T_1(k\Delta t)$ (PHASE NOISE OF THE OSCILLATOR).

1) Statistics of $T_1(t)$ for both synchronizations

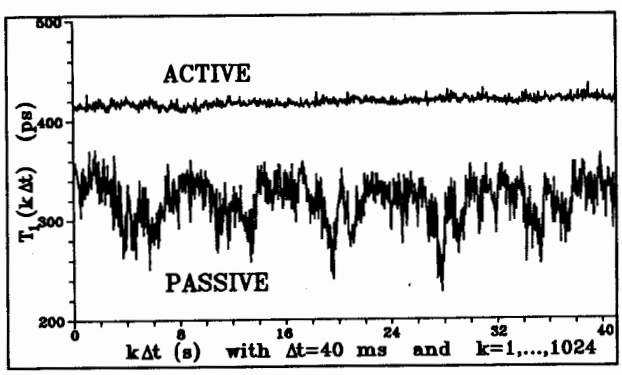


Fig. 4.

2) Probability distribution of $T_1(k\Delta t)$:

Prob[t < $T_1(k\Delta t)$ <t + δt] with δt = 1 ps.

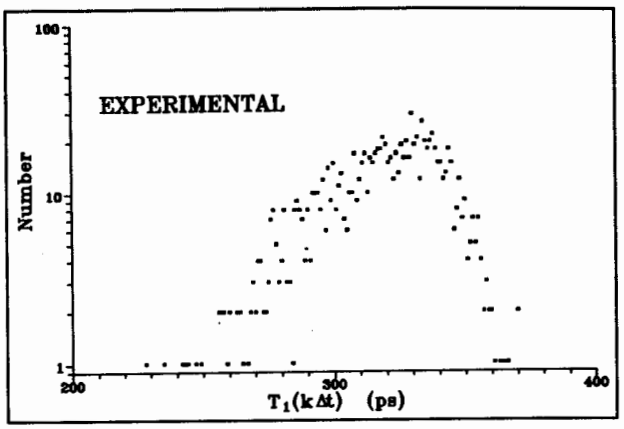


Fig. 5.

3) Experimental autocorrelation function.

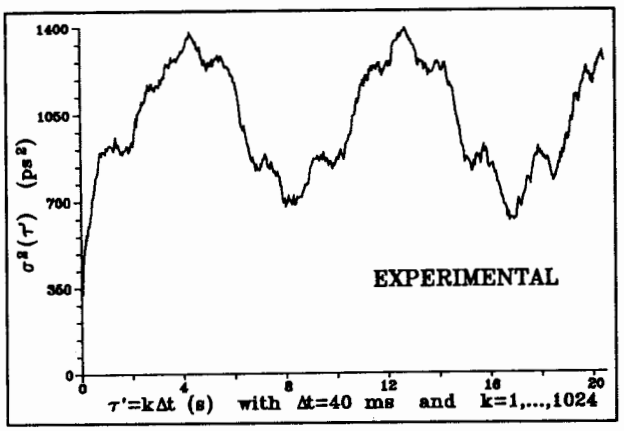


Fig. 6.

4) Autocorrelation function calculated for $S_\Theta(f) = 1/f$ (flicker phase noise) :

$$\sigma^2(\tau') = 4\int_{f_i}^{f_s} S_\Theta(f)\sin^2(\pi\tau' f)\,df$$

with : f_i = 0.12 Hz and f_s = 0.8 Hz.

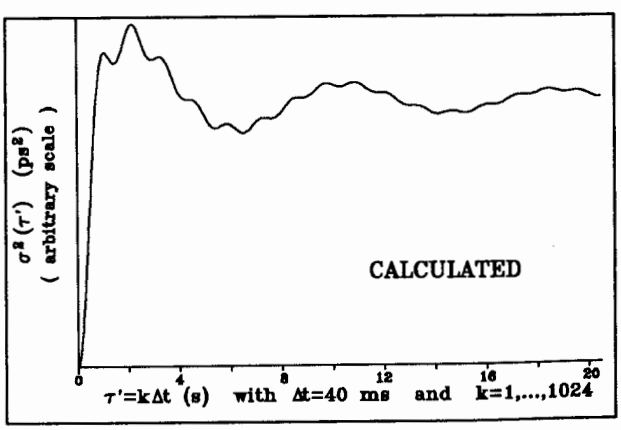


Fig. 7.

5. CHAOTIC OR STOCHASTIC BEHAVIOR ?

1) Poincaré map.

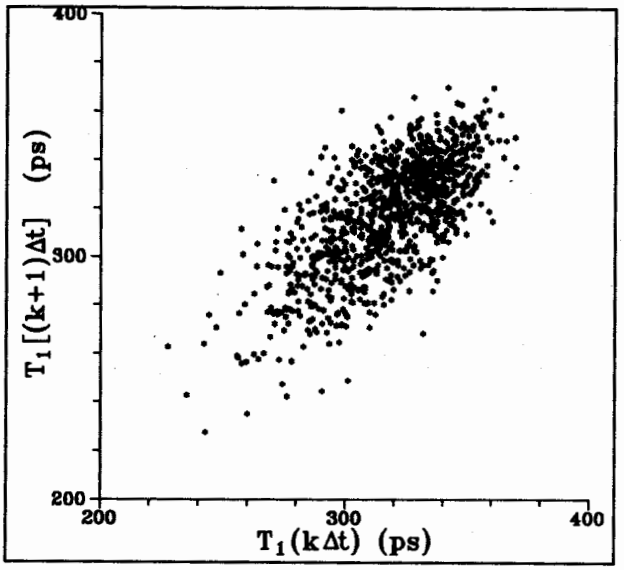


Fig. 8.

2) Correlation integral (Ref. 1) :

$$C(r)=\frac{2}{N(N-1)}\left[\text{number of pairs (i,j) for which } |Xi - Xj| < r\right]$$

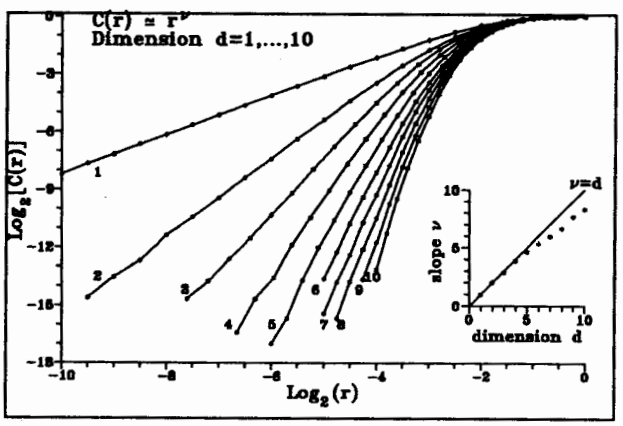


Fig.9.

3) Non-linear forecasting (Ref.2)

$$\text{Error}(p\Delta t)=\sqrt{\frac{\sum_i\left[P_i(p\Delta t) - T_i\right]^2}{\sum_i\left[T_i - \langle T\rangle\right]^2}}$$

where $T_k = T_1(k\Delta t)$, $\langle T\rangle$ is their mean value and $P_k(p\Delta t)$ is the predicted value of $T_1(k\Delta t)$.

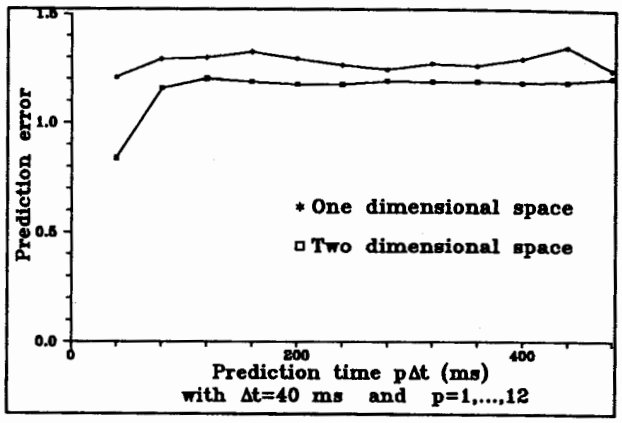


Fig.10.

6. CONCLUSION.

1) The experimental method allows the measurement of the instantaneous phase fluctuation.

2) There is no unambigous distinction between low-dimensional chaos and stochastic behavior.

7. REFERENCES.

[1] P. Grassberger, I. Procaccia :
Physica D9 (1983) 189.

[2] G. Sugihara, R.M. May :
Nature 344 (1990) 734

BANQUET

Summary of the banquet speech given by

Catherine Cardinal, Directeur scientifique
Institut "L'Homme et le Temps", Rue des Musées 29,
CH-2301 La Chaux-de-Fonds



The art in making Swiss timepieces - the artistic craftsmen

Ever since the Renaissance, the art of making timepieces has required the dedication of artisans skilled in embellishing the metal of clock and watch cases: specialists in increasingly complex techniques.

From the 17th century to the present, the artistic workmanship of timepieces has favored the watch. Before the second half of the 19th century, the reputation of Swiss timepieces was built not by its technicians but its decorative artists.

The goldsmiths and silversmiths

The first craftsmen to devote themselves to timepiece artistry. In 1703, there were 195 gold-and-silversmiths working in the town of Geneva - and about as many makers of timepieces.

The engravers

Fashion in the 17th century called for fancy watch and clock dials, cases - ornately engraved.

Under the influence of France's Romantic style in the 18th century, engravers often joined with enamellers who entrusted them with adorning the frame of their painting.

The last third of the 18th century was marked by a new technique called guilloche: this new breed of engravers were then able to repeat geometrical patterns mechanically on timepiece cases.

1830 - 1900: Engraving becomes a dominant art, equal to enamelling. In Geneva and Neuchâtel Canton, hundreds of engravers emerge, more and more skilled.

The future of engraving seems promising at the beginning of the 20th century - until the economic aftermath of the War and the advent of the wristwatch.

The enamellers

Enamelling in watch settings is, at first, part of the gold-and-silversmith's trade. Master craftsmen in the 16th and 17th centuries can do both.
c 1650: a new technique, enamelled painting, creates a specially artistic

class of craftsmen. The new medium gives birth to innumerable shades of color and drawing.

From 1730 to 1820, the craft of clocks and watches continues to develop and the constant demand for enamelled watch settings favors Geneva enamelling more than any other decorative art.

The enamellers (2)

At the start of the 19th century, enamelled painting reaches a peak: flattened cases, as large as 6 cm in diameter, are ornamented with remarkable dexterity.

Towards the end of the 19th century, translucent enamels are often used on guilloche settings with precious stones, or on chiseled patterns of Art Nouveau.

The widespread popularity of the wristwatch accounts for the decline of enamel work in timepieces except for pocket watches and small collectors' clocks. The 1925 Exhibition of Decorative Arts in Paris attests to the resounding success of enamelling as used in geometric compositions.

The Jewelers

Conscious of the market value of time, Geneva jewelers create watches of all sorts at the start of the 19th century.

The advent of wristwatches had no ill-fated effect on the jewellery trade, as it had on timepiece engraving and enamelling.

From 1920, the jewellery and timepiece crafts seem so linked in their development that the Geneva Association of Watches and Jewellery is founded in 1942.

The cabinetmakers, the bronzesmiths

In the second half of the 18th century, Geneva excels in watch artistry; Neuchâtel Canton gains renown for its clock settings. The Neuchâtel artisans, anxious to imitate Parisian clock settings but often unable to afford the techniques of marquetry and gilt bronze, create a style of their own: wood is sculpted, brass is spun; tortoiseshell and marble are replaced by painting.

At the end of the 18th century, the huge, ornate Parisian-style clocks give way to simpler ones derived from the Louis XV clocks. (Hundreds of these Neuchâtel timepieces are produced in the 19th century).

The cabinet painters

Mid-Eighteenth century: painted decors are in vogue using a Parisian technique of exceptional colour transparency and velvety luster.

This increasingly popular style is soon applied to the Neuchâtel clocks, triggering a new field of specialists.

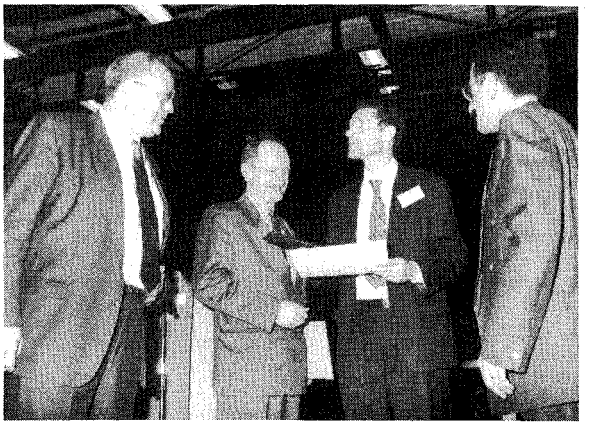
Conclusion

From the beginning of the 20th century, the Swiss timepiece industry has proven its mastership in the artistic and technical craftsmanship of the watch. Through their versatile ability to explore diverse materials and the audacity of their work, Swiss artisans have ensured the renown of their watches worldwide.

Award of the "European time and frequency prize"
made by the Société Française des Microtechniques et de Chronométrie (SFMC) to

Dr. Christophe SALOMON
Laboratoire de Spectroscopie Hertzienne, Ecole Normale Supérieure,
24, rue l'Homond, F-75005 Paris





Ladies and Gentlemen, Dear Colleagues,

Time is a very special field : when you get trapped in it, it's for ever. Time and Frequency Community is a very special Family; when you get accepted there, it's for ever. This Family has got special events. European Time and Frequency Award is one of these.

• The European Frequency and Time Forum has triggered the creation of the : European "Time and Frequency" Award

This award is attributed to a Scientist in our common field for outstanding contributions in the domain whatever is the nationality of the Award Winner. The award is attributed in principle each two years period.

Anyone in the Community may propose anyone by writting to the Executive Committee or to myself. The Award recognizes exceptionnal contributions providing worldwide recognition in the Community either for fundamental advances or for future or present applications. The goal is to encourage scientific efforts and competition. I already address the Community for proposals to be made for the 95 Award. The Award (together with a check of FF 10 000) is given by the "Société Française des Microtechniques et de Chronométrie" (SFMC) whose past President is A. REMOND organizer of 3 EFTF (1987,88,91).

This year, the Award has been attributed to a young, smart and, may I say, brillant physicist for "research work on physics of cesium laser cooled atoms ans its applications to atomic resonator using a cesium fountain".

The Award Winner is : ***Dr Christophe SALOMON*** of the "Laboratoire de Spectroscopie Hertzienne in Ecole Normale Supérieure" of Paris.

Before calling Dr SALOMON for the Award, let me say some words about his career.

Christophe SALOMON is not yet fourty years old, he got his Engineer Degree from "Ecole Centrale des Arts et Manufactures" de Paris in 76. After a Degree preparing to research in ORSAY, he obtained his first Thesis in 79 in the Laser Laboratory of North Paris University. He then started a career in CNRS in 1980 and obtained his Ph D Degree in 1984 on the following subject :

"Ramsey Fringes and ultra high resolution spectrocopy in the infrared".

In 1985, he was a post Doc. at JILA in Boulder, Colorado where he got involved in laser cooling of atoms and ultrastable lasers. Since 1986, he serves at "Ecole Normale Supérieure", rue Lhomond, where I spent some time myself long time ago. He is now "Directeur de Recherches" in CNRS. He has published 68 scientific papers including 21 invited papers.

I am probably not the right person to comment recent results on Ramsey resonance in Zacharies Fountain or quantized Atomic Motion in 1D césium molasses. Many of you probably know the importance of the work of Christophe SALOMON. Let me simply call attention on a recent paper entitled : "Laser cooling of atoms in microgravity" which many people believe extremely important for the future. The Executive Committee that ultimately attributed the Award was very sensitive to the extraordinary potential of Christophe SALOMON's research and to the fact that our Community always needs physicists to launch ideas and stimulate the Community before we finally obtain, one day, new frequency standards ready to be sold.

Now, I would also like to call Michel FROELICHER of the "Société Française des Microtechniques et de Chronométrie" and Bernard SCHLUETER of the EFTF Committee to participate in Award ceremony and of course to call Christophe SALOMON for presentation of the Award.

Prof. Raymond J. BESSON
Président of SFMC

Author index

I.V. Abramson 287
M. Aida 409
D.W. Allan 221
A. Al'shina 211
J. Amagai 341
A. Anderman 699
V.V. Andreyanov 33
J.P. Aubry 161
C. Audoin 479/537
R. Avery 89
G. Avitabile 465
I.D. Avramov 459

J.P. Bachheimer 671
P. Baeriswyl 119
H. Bahadur 247
A. Ballato 93-231
W. Balzer 311
J. Barbier 623
F. Bastien 155
A. Bauch 472/483/531/583
C. Baumer 349
F. Baumont 641
E. Bava 489
A.A. Belyaev 353
L.G. Bernier 373
M. Berte 567
A. Berthaut 57
G. Beutler 119
E. Bigler 235
Y. Billon 567
P. Blind 601
P.P. Bogdanov 633
R. Bourquin 607
J. Bourrieau 269
J.J. Boy 57
M. Brunet 269/681

A. Cachard 601
M. Calhoun 399
C. Caliendo 465
O. Cambon 161
V. Candelier 319
B. Capelle 561/647/659
Ph. Cathelin 319
P. Cérez 479/537
D.P. Chen 455
G.T. Cherenkov 147
M. Chiampi 367
A. Choujaa 601
M. Chowdhury 525
M. Christen 613
F. Collin 429
D.T. Comer 77
C.J. Cooper 525
G. Costanzo 367
L.M. Coyle 707
M.P. Cracknell 327
B. Cunin 715

G. Daigne 433
S. Dalla Piazza 613
P. Daly 211
A. Davis 135
D. Davis 391
J.A. Davis 113
A. De Marchi 367/489/541
R. Decher 707
C. Déjous 615
A. Del Casale 367
M. Deleuze 255
N.A. Demidov 353
J. Détaint 561/647/659
F. Deyzac 291
M. Dienert 345
N. Dimarcq 479/537
M.P. Dinca 359
J. Dowsett 181
R.A. Dragonette 379
B.G. Drakhlis 315
R.E. Drullinger 549/593
B. Dulmet 607
P.A. Durrant 511
J.F. Dutrey 269

C. Eckert 715
A. El Habti 155
L.A. Epictetov 413

B. Fischer 531
C.J. Foot 525
E. Fossler 387
A.D. Freik 261
P. Fridelance 637/641
M. Froelicher 601

C. Gelle 99
A.G. Gevorkyan 105/623
O.C. Gheorghiu 359
V. Giordano 479/537
L.C. Giurgiu 359
D.J. Glaze 549
A. Goasguen 681
A. Godone 589
A. Goiffon 255
T. Goto 409
Y.G. Gouzhva 633
M. Granveaud 129/409
D. Graveline 707
P. Grüdler 641
J.G. Gualtieri 231
G. Guitel 319
W. Gurtner 119

A. Haessler 715
S. Hama 341
C.S. Hartmann 455/497
K. Hashimoto 517
D. Hauden 601
J. Heighway 455/497
T. Heindorff 473/531
F. Heisel 715
H. Hellwig 9
P. Hetzel 125
G. Hillenbrand 525
K. Hilty 3
J.M. Hodé 499/647
W. Hofman 267
M. Hoummady 235/615
D. Howe 391

A. Ibanez 255/659
M. Imae 341/409
A. Imaoka 83

G. Jeandel 667
G. Jenni 297
H. Jinlun 335
D. Jinquan 363
L. Jinting 363
C. Joly 561
J.L. Josset 171

B. Kalinowska 323
K. Kalliomäki 383-393
I. Kawasaki 239
H. Kawashima 51
M. Kihara 83
D. Kirchner 135
H. Kiuchi 341
W. Klepczynski 135
S.N. Klyushnikov 633
N.B. Koshelyaevsky 147/211
J.A. Kosinski 93/231
S.P. Kovita 105
P. Krempl 653
B.N. Krutov 261
N. Kuramochi 239

B. Laporte 129/269
F. Lardet-Vieudrin 155
I.M. Larionov 555
J.L. Le Corre 567
S. Lechopier 567
A. Lepek 221
L. Levine 99
W. Lewandowski 147/409//417/641
A. Lisowiec 283
C. Lombardi 367
M. Lombardi 391
C. Longet 57
V. Lorent 525
J.P. Lowe 549/593

F. Maddaleno 367
L. Maleki 679
G. Manes 465
J. Mannermaa 393
G.D. Mansfeld 261
T. Mansten 383/393
G. Martin 235/601
A. Masiukiewicz 323
I. Mateescu 63
I. Mateescu 63
H. Matthey 177
E.M. Mattison 707
J. Mc Nabb 387
A.V. Medvedev 555
H.J. Megens 577
R.R. Merzakreev 413
F. Meyer 129/305
A. Michaud 525
J.P. Michel 161/567
J.A. Miehé 715
C. Miki 409
Y. Minguy 423
G.K. Montress 439
P.E. Morley 181
G. Morlot 667
P. Moussay 147
D. Mouton 567
A. Myasnikov 105

T. Nakase 239
B. Nasreddine 319707
R. Nicoll 589
C. Novero 593
A. Novick 707
G.U. Nystrom

N. Oura 239

P. Paquet 429
T.E. Parker 439
J. Pavlovec 41
P.R. Pearce 113
G. Petit 409/417/641
P. Petit 537
I.G. Petrosian 315
E. Philippot 255/659
W.D. Phillips 577
O. Piller 7
J. Pistré 615
M. Planat 349/715
V.P. Plessky 505
C. Poignon 667
J.C. Pommier 601
G. Portas 681
W. Powell 135
S.B. Pushkin 147/211
I.G. Pushkina 633

K. Rahimullah 589
D. Rebière 615
Ch. Reichinger 243
A. Reid 23
G. Renard 57
M. Repetto 367
H. Ressler 135
G. Robichon 57
R. Robnik 135
S.L. Rolston 577
L. Roselli 465
F. Ross 181
M.E. Roulet 171
R. Roy 57
E. Rubiola 367

S.A. Sakharov 555
B.A. Sakharov 353
Y.A. Samsonov 555
E. Sappl 693
W. Schäfer 693
T. Schildknecht 119
D. Schnier 583
R. Schröder 473/483/531
M.A. Schwab 511
L. Schwartz 567
J. Schwartzel 561/647/659
J. Sedlacek 41
D. Sennedot 423
A. Shaofeng 363
N. Shibanova 607
J. Shirley 549
J. Shuguang 335
V.V. Sidorov 413
A. Smolarski 283
M. Solal 499
A. Söring 125
T. Soubirous 567
J. Stadler 653
S. Starker 693
A. Steane 525
J. Suchanek 41

J.J. Suter 379
R.L. Sydnor 399
C. Szekely 593

M. Takahashi 239
Ch. Tamm 583
P. Tavella 489/589
G. Théobald 479/537
D. Thiébaud 57
B. Thomas 423
C. Thomas 141/409/417
T. Thorvaldsson 505
N. Tirole 601
T. Trémas 623
K. Trümpy 45
K.K. Tuladhar 297

A.A. Uljanov 353

L. Veenstra 135
C. Veillet 637/641
F. Vernotte 129
E. Verona 465
R.F. Vessot 707
M. Vincent 129
A.A. Volkov 315
Y.L. Vorokhovsky 315

M. Walhout 577
W. Wallnöfer 653
F.L. Walls 189/353/593
R.L. Walsworth 707
M.A. Weiss 199/391
T.P. Weissert 199
A.L. Welti 69
S. Weyers 473
P. Wheeler 135
J. Wiant 641
A. Witte 577
M. Wojcicki 283

Z. Xiwen 363

M. Yamaguchi 517
J.S. Yerna 279
T. Yongfang 363
T. Yoshino 341

M. Zafrany 161/305/567
G. Zampetti 89
A. Zarka 155/561/647/659
J. Zelenka 41
K.P. Zetie 525
Y. Zheng 659
W. Zingg 45

Meeting Report

Event:	7th European Frequency and Time Forum
Date:	16 - 18 March 1993
Location:	University of Neuchâtel, Faculté des Lettres
Participants:	334 incl. exhibitors
Countries:	24 (details see Table 1)
Papers:	(see Table 2)
- *Invited*:	11 (incl. 4 for the opening and 1 banquet speaker)
- *Contributed:*	67
- *Poster*:	45
Sessions:	19 (Opening + 18 sessions led in parallel)
Exhibitors:	16

For the first time a **Tutorial session** was held the day before the beginning of the Forum i.e. the **15th March.**

Participants:	54
Countries:	9
Speakers:	4

Support:

Financal support:

Republic and Canton of Neuchâtel; City of Neuchâtel; Swiss Academy of Engineering Sciences (SATW); ESA-ESTEC; United Tobacco Factories; Federal office of education and science; Swiss federal office of metrology.

Other support:

University of Neuchâtel, Swiss Bank Corporation, Imprimerie Krattiger, Jean-Claude Facchinetti, Café au Moka.

1) Introduction

The European Frequency and Time Forum (EFTF) is an international conference and exhibition, providing information on recent advances and trends of scientific research and industrial development in the fields of Frequency and Time.

The EFTF was inaugurated in 1987 as a meeting and discussion point for the European Time and Frequency community. In the six years of its existence, the reputation of the EFTF's high standards has spread around the globe and the Forum now attracts participants from all over the world.

2) Purpose

The fact that time can be determined and measured so precisely, far better than any other physical parameter, represents a technological asset of great importance. Modern telecommunication systems, space programmes and navigation require precise and stable timing for successful operation. For various scientific research, like very long baseline interferometry gravity and relativity experiments, geodesy, timing is an essential parameter.

3) Programme

The programme is given in the introduction and the distribution of the contributed papers and poster papers to the different topics is shown in table 2. A special session was dedicated to telecommunication applications as well as one of the invited papers at the opening.

For the first time EFTF was combined with a Tutorial session held the day before the Forum. Three working sessions were dedicated to the problems of quartz oscillators, their applications in control loops and appropriate measurement techniques. With more than 50 participants this first tutorial was quite successful.

4) World-wide participation

334 participants from 24 countries having a wide variety of backgrounds including national observatories, research institutes, universities and industries attended the Forum (table 1). This diversity was reflected in the wide spectrum of topics addressed, ranging from fundamental physics and advanced research to materials, applications and measurement techniques.

5) Facilities

The Aula of the University of Neuchâtel is a perfect location for the symposium. The services, microphones, slide projection etc were excellent.

The auditorium had reached the limits of its capacity with 334 participants devided into two rooms (2 parallel sessions).

Unfortunately the poster exhibition was somewhat disconnected from the rest of the activities of the Forum.

6) To remember

- The tutorial should be organized for at least another 2 to 3 years in order to get experience.

- The tutorial should treat every year a different, specialized subject.

- Review papers should give an overview of the state of the art; these are foreseen to update the knowledge of the participants.

- Short papers were an excellent idea.

- The posters should be shown during the whole symposium. At a specific time the authors should be present in order to discuss topics in details (inform authors in advance of the procedure).

- The papers should be collected during registration at the reception desk, deadlines should be fixed and the authors informed in advance.

- A sightseeing tour in the evening, would have been appreciated by a lot of participants.

- As the whole symposium was held in English, the banquet speaker ought to have his/her talk in English.

- A cultural happening along with the banquet would have been appreciated.

- Speakers should be informed about the following procedure that could be applied:
  a) a registration deadline
  b) a payment deadline
  c) a cancellation deadline.

7) Acknowledgements

I acknowledge:
- the great support and professionalism of Mrs Rutschmann and her team for their excellent preleminary work and the organisation of the Forum
- the given support of Dr. Ecabert and Mr. Schlueter during preparation of the Forum
- the generosity of the University of Neuchâtel for letting the location
- the kindness of Mr. Pfyffer, Director, and Dr. Nadler, head of division of the R&D Laboratory of Swiss PTT Telecom allowing me the time to prepare the meeting.

8) Conclusions

- The EFTF 1993 including the Tutorial was a great success. The large number of participants reflects the worldwide interest in this Forum.

- The next EFTF will be held on 9-11 March 1994 in Weihenstephan near Munich, Germany.

K. Hilty, Scientific Chairman

Participants by country

Table 1

Country	No of participants
Australia	1
Austria	5
Belgium	2
Bulgaria	1
Canada	1
China	2
Finland	3
France	115
Germany	39
Italy	12
Japan	7
Netherlands	6
New Zealand	1
Norway	2
Poland	2
Portugal	1
Rumenia	2
Russia	11
Spain	5
Sweden	2
Switzerland	67
United Kingdom	19
United States	27
Yugoslavia	1
Total: 24 countries	**334 participants**

Table 2

EFTF 93 - Contributed papers and posters

Sessions		Number of papers
Opening session + evening conference		5
Quartz Crystals (2 sessions)		10
Telecom Applications		4
Oscillators		4
Time Transfer		7
Time scales and instrumentation		4
SAW (2 sessions)		9
Cesium (2 sessions)		9
Filters		3
New atomic frequency standards		4
Sensors		4
Space (2 sessions)		10
Materials		5
Posters (2 sessions) + 1 postdeadline poster		45
- Frequency control components and materials	21	
- Frequency control and timing systems	24	
Total		**123**