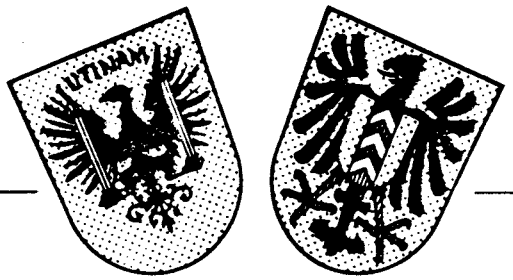


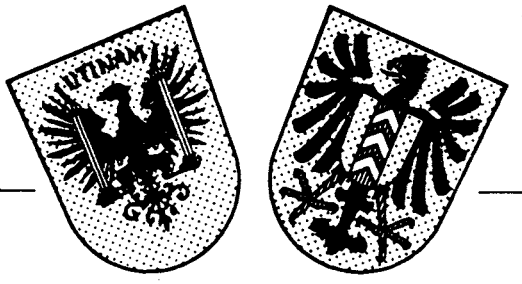
EFTF 88

Second European Frequency and Time Forum
Deuxième Forum Européen Fréquence et Temps

ACTES

PROCEEDINGS

16, 17, 18 March 1988
NEUCHÂTEL
Switzerland



E F T F 1 9 8 8

Second European Frequency
and Time Forum

Deuxième Forum Européen
Fréquence et Temps

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

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

March 16, 17, 18, 1988

16, 17, 18 mars 1988

Neuchâtel University, Aula
Quai Robert-Comtesse 2

Neuchâtel

2 ND EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 16 - 17 - 18 MARCH 1988

Table of contents

Introduction

Founding Committee, Organization, Support XV

Science Committee XVII

Session chairmen XIX

Exhibitors XXI

Opening message XXIII
G.F. Bauer

* Abstracts only

Invited Papers

0. Significance of Metrological Measurements of Frequency and Time 3
O. Piller
Office fédéral de métrologie, Wabern - Switzerland

1. Advances in Science, Emerging Technologies and Global Competition : Opportunities and Challenge 13
H.W. Hellwig
National Bureau of Standards, Gaithersburg - USA

2. Laser Frequency Standards and Frequency Measurement for the Optical Fibre Bands 23
D.J.E. Knight
National Physical Laboratory, Teddington - UK

3. A review on ion traps frequency standards 35
G. Werth
Universität Mainz, Mainz - FRG

4. Time and Frequency Accuracy from GPS 45
A.R. Pratt
Polytechnik Electron PLC - Daventry, North - UK

5. Timing Requirements of Future ISDNs 71
R.D. Calvert
Cable and Wireless PLC, London - UK

6. Timing of a European Worldwide Satellite-Based Navigation System 99
H.A. Laue
ESA, Darmstadt - FRG

7. Very Long Baseline Interferometry (VLBI) 117
R. Spencer
University of Manchester, Macclesfield - UK

Time and Frequency

TF-01 Characterisation of Oscillators

1. Linear Prediction of the Non-Stationary Clock Error Function 125
L.G. Bernier
OSA, Neuchâtel - Switzerland

2. Oscillator Characterization by Correlation Method 139
A. Mahmoud
CNRS, Besançon - France

3. Measurement of Phase and Frequency Fluctuations of Microwave Oscillators 151
J. Groslambert, M. Olivier
CNRS, Besançon - France

4. Specifications and performances of ultra-stable oscillators in the Doris orbitography and positioning system 163
J.B. Laporte
Centre National d'Etude Spatiales, Toulouse - France

TF-02 G P S

1. Nanosecond time comparisons in Europe using the GPS 187
B. Guinot, W. Lewandowski
Bureau International des Poids et Mesures, Sèvres - France

2. GPS High Precision Time Transfer Methods and Related Equipment 195
R. Gounon
Sercel, Carquefou - France
J.N. Lelong
Osafrance SA, Boulogne - France

3. Operational Highly Precise Time Transfers in France by GPS and by the TV Method 213
M. Granveaud, P. Parcelier, G. Freon, R. Tourde
Laboratoire Primaire du Temps et des Fréquences, Paris - France
M. Vincent
Observatoire de Besançon - France
P. Grudler
Centre d'Etudes et de Recherches Géodynamiques

4. Improvement of international time comparison of OFM using GPS common view satellite observations 231
L. Prost, H. Brandenberger
Swiss Federal Office of Metrology, Wabern - Switzerland

TF-03 Timing and Dissemination

1. Present Status of UTC (NPLI) 243
B.S. Mathur, P.N. Taneja, G.K. Goel, A. Sen Gupta, P. Banerjee, A.K. Hanjura
National Physical Laboratory, New-Delhi - India

2. Some sources of timing for communication systems* 251
J.L. Jespersen, D.W. Hanson
National Bureau of Standards, Boulder - USA

3. Measurement of Delay Variation in Digital Communication Networks 253
J.P. Mellana, K. Hilty
Swiss Post Office, Bern - Switzerland

4. Short Distances Frequency Distribution Systems 263
E. Rubiola
Politecnico di Torino, Torino - Italy

5. Slave Time Code Generators* 271
D. Ceppi
CIR, Gals - Switzerland

TF-04 Various Satellite Timing Experiments

1. Timing Data Reduction and Processing for the NASA - Goddard Laser Tracking Network 275
E. Detoma, A. Clinton, W. Hanrahan, P. Klushmeider, R. Price, P. Dachel
Bendix Field Engineering Corp., Columbia - USA

2. Synchronization of Ground Station Atomic Clocks Using Weighted Transportable Cesium Time Standards in the Space Shuttle NAVEX Experiment 293
H. Schild
DFVLR, Wessling, FRG

3. Time Scales and their Estimated Uncertainties in the Space Shuttle Experiment NAVEX 307
S. Starker, E. Sappl
DFVLR, Weissling - FRG

4. Loran-C GPS Hybridization in Time Keeping 317
M. D'Amico
Techniphone, Le Puy Sainte-Reparade - France

TF-05 VLF and LF

1. Envelope Stability of Loran-C Pulses 333
E. Rubiola
Politecnico di Torino, Torino - Italy

2. A simple Loran-C timing receiver for supporting VLBI 341
J.E.B. Ponsonby
University of Manchester, Macclesfield - UK

3. Time dissemination via the LF Transmitter DCF77 using a pseudo-random phase-shift keying of the carrier 351
P. Hetzel
Physikalisch-Technische Bundesanstalt, Braunschweig - FRG

4. Precise determination of UTC in remote regions using OMEGA signals with reference to monitoring performance of a hydrogen maser 365
H.J. Gaede
Hartebeesthoek Radio Astronomy Observatory/NITR, Johannesburg - Republic of South Africa

Atomic Frequency Standards

AFS-01 Various Atomic Frequency Standards

1. The Recirculating Cryogenic Hydrogen Maser* 383
M. Hürlimann, W.N. Hardy
University of British Columbia, Vancouver - Canada

2. Studies of Dielectrically Loaded Cavities for Small Size Hydrogen Masers 385
N. Yahyabey, P. Lesage, C. Audoin
Laboratoire de l'Horloge Atomique, Orsay - France

3. Doppler Effect in Mercury Ion Clock 399
C. Meis, M. Jardino, M. Desaintfuscien
Laboratoire de l'Horloge Atomique, Orsay - France

4. Miniaturization Versus Performances in Rubidium Frequency Standards 409
G. Busca, L. Pan
Oscilloquartz SA, Neuchâtel - Suisse

5. Performance of Disciplined Rubidium Oscillators* 419
G. Huebner
Ball Efratom Elektronik GmbH, München - FRG

6. Frequency shift of an atomic or molecular resonance, bound to the interaction time between an atom and the electromagnetic field* 421
J. Balbi
Laboratoire d'Hélioénergétique, Université de Corte - France
D. Gillet
Laboratoire de Piézo-électricité et d'Electronique, La Bouloie, Besançon - France
G. Lomaglio, J.G. Théobald
Laboratoire de Spectoscopie Hertzienne et d'Electronique, La Bouloie, Besançon - France

AFS-02 Cesium Beam Frequency Standards

1. Long-term performance of the PTB's primary clock CS2 425
A. Bauch, H. de Boer, B. Fischer, T. Heindorff, R. Schröder
PTB, Braunschweig, FRG

2. The Effect of Feeding Junction Asymetries on End to End Phase Difference in Ramsey Cavities 437
A. De Marchi
Universita di Ancona, Ancona - Italy

3. Characterization of a Cesium Beam Detector 435
H.S. Schweda, P. Thomann, G. Busca
Oscilloquartz S.A., Neuchâtel - Switzerland

4. Full Digital-Processing in a New Commercial Cesium Standard 461
J. Rabian
Sté Sercel, Carquefou - France
P. Rochat
Oscilloquartz SA, Neuchâtel - Switzerland

5. Improving the Long-Term Stability of a Time Scale Generated by Commercial Cesium Clocks 469
F. Cordara, V. Pettiti
Istituto Elettrotecnico Nazionale Galileo Ferraris, Torino - Italy
P. Tavella
Associazione Nazionale Galileo Ferraris Grant, Torino - Italy

AFS-03 Optically Pumped CS Beam Frequency Standards

1. Frequency Stability of an Optically Pumped Cesium Beam Frequency Standard : Preliminary Results 483
V. Candelier, V. Giordano, A. Hamel, G. Theobald, P. Cerez, C. Audoin
Laboratoire de l'Horloge Atomique, Orsay - France

2. Light Pressure effects on an Optically Pumped Cesium Beam Frequency Standard 499
A. Clairon, E. de Clercq, B. Dahmani, A.H. Gérard
Laboratoire Primaire du Temps et des Frequences, Paris - France

3. Commercial Size Optically Pumped Cesium Beam Resonator 513
P. Thomann, H. Schweda, G. Busca
Oscilloquartz S.A., Neuchâtel

4. Experiments on optically pumped cesium beam frequency standard* 521
J. Umezu, B. Komiyama, H. Saitho, Y. Ohta
Radio Research Laboratory Koganei-Shi, Tokyo - Japan

5. A FM-stabilized semiconductor laser for optical pumping of Cesium atoms and length measurements 523
H.R. Telle
Physikalisch-Technische Bundesanstalt, Braunschweig - FRG

6. The Effect of Laser Line-Norrowing on the Performance of Optically Pumped Cesium Atomic Beam Frequency Standards 531
S. Ohshima, Y. Koga, Y. Nakadan
National Research Laboratory of Metrology, Ibaraki-ken - Japan
L. Hollberg, R. Drullinger
National Bureau of Standards, Boulder - USA

AFS-04 Atomic Frequency Standards

1. Ultra-Narrow Semiconductor Laser Linewidth for Optically Pumped Cesium Clock 537
A. Clairon, B. Dahmani
Laboratoire Primaire du Temps et des Fréquences, Paris - France
Ph. Laurent, Ch. Breant
Laboratoire Lasers Ultra-Stables, E.T.C.A., Arcueil - France

2. Extended-Cavity Semiconductor Lasers for Optical Pumping 547
M. de Labachelerie, K. Diomande, N. Dimarcq
Laboratoire de l'Horloge Atomique, Orsay - France

3. Frequency-Stabilized Laser for Atomic Frequency Standards Using Saturated 555
J. Deng, Y. Tan, Y. Xiang, X. Zhu
Wuhan Institute of Physics, Academia Sinica, China

4. Phase-lock design of a 600 GHz BWO for the Mg frequency standard 557
E. Bava, A. Godone, C. Novero
Istituto Elettrotecnico Nazionale Galileo Ferraris, Torino - Italy

Resonators and Oscillators

RO-01 Resonators 1

1. Vibration Modes of Contoured Piezoelectric Resonators with Additional Film Layers 573
H. Nowotny,
Institut für Theoretische Physik, Wien - Austria
E. Benes, M. Schmid
Institut für Allgemeine Physik, Wien - Austria

2. Analysis of Partially or totally Contoured Resonators with Finite Lateral Dimensions 601
J. Detaint, H. Carru, J. Schwartzel, C. Joly,
CNET PAB/MCT/CMM, Bagneux - France

3. Lateral DC Electric Field in Quartz Plate Resonators 621
C.K. Hruska
Piezoeletricity Resarch Laboratory, Toronto - Canada

4. Non Piezoelectric Crystal Resonators 629
D. Mazerolle, E. Dieulesaint, D. Royer
Université P. & M. Curie, Paris - France

RO-02 Resonators 2

1. An LCC Monolithic Crystal Filter 637
R.C. Smythe
Piezo Technology, Orlando - USA

2. Force-Frequency Characteristics of a Plano-Convex Square AT-Cut Plate for Force-Sensor Use 645
N. Oura, N. Kuramochi, A. Okada
Tokyo Institute of Technology,Yokohama - Japan

3. Stroboscopic X-Ray Topography of Quartz Resonators 651
A. Zarka, B. Capelle
Université P. & M. Curie, Paris - France
J. Detaint, J. Schwartzel
Centre National d'Etudes des Télécommunications, Bagneux - France

4. Studies on Oscillating Quartz Resonators for Application in Frequency Control and Standardization 657
H. Bahadur, R. Parshad, B.S. Mathur
National Physical Laboratory, New-Delhi - India

5. The Development of a Transportable Hydrogen Maser to be used for the VLBI Network in China* 661
Z.C. Zhai, H.X. Huang, C.F. Lin, G.X. Jiang
W.H. Luo, L. Jia-Fu
Shanghai Observatory, Shanghai - China

RO-03 BVA

1. Systematic Characterization of a High Stability Resonator in Pre-Serie Production 665
R. Delaite, Th. Migeot
E.N.S.M.M., Besançon - France

2. The Radiation Susceptibility of BVA-AT Resonators to Simulated Medium Earth Orbit Radiation 679
J.J. Suter, A.G. Bates, J.M. Cloeren
The Johns Hopkins University, Laurel - USA

3. In Plane Orientation Optimization of QVA Resonators 687
J.P. Aubry, B. Bonnotte
CEPE, Argenteuil - France
J.C. Craveur, F. Deyzac, D. Janiaud
Onera, Châtillon - France

4. Space qualified 100 MHz BVA-VHF-Resonator 701
U.R. Peier, G. Jenni
Oscilloquartz SA, Neuchâtel - Switzerland

5. Results of the Qualification Programme of a BVA-Type Crystal Oscillator for Satellite use 715
W. Schäfer
University of Stuttgart, Stuttgart - FRG

6. Environmental Effects on the Medium and Long Term Frequency Stability of quartz Oscillators 719
F.L. Walls
National Bureau of Standards, Boulder - USA

RO-04 SAW

1. Novel Quartz Cuts for SAW Devices 731
D.E. Bower, T.N. Oliver
STC Components, Harlow - England

2. Aging in SAW Oscillators 739
A. D'Almeida, P. Leblois, D. Hauden
CNRS, Besançon - France
S. Calisti, J.Y. Duquesnoy, L. Penavaire
Thomson-Sintra ASMDTAS, Valbonne - France

3. An Analysis of the Normal Acceleration Sensitivity of ST-Cut Quartz Surface Wave Resonators with the Substrate Extending Beyond the Supports 753
H.F. Tiersten, D.V. Shick
Rensselaer Polytechnic Institute Troy, New-York - USA

4. Bulk Waves Excited into an Anisotropic Half Space by a Metallic Interdigital Transducer 767
Y. Zhang, M. Planat
CNRS, Besançon - France

5. Accoustic Surface Transverse Wave Filters at 2 to 2.8 GHz for Oscillators and Timing Recovery 779
B. Fleischmann, G. Riha, H.W. Woerz
Simens AG, Munich - FRG

RO-05 Oscillators

1. Microwave Oscillator Using Partly Metallized Dielectric Resonators 795
M. Valentin, R. El Cheikh
CNRS, Besançon - France

2. Thermal Hysteresis in Quartz Resonators and Oscillators 803
J. Beaussier
ONERA, Châtillon - France

3. Miniature Oscillators 817
K. Schwaller, K. Schulthess, W. Zingg
Micro Crystal, Grenchen - Switzerland

Materials and Processing

MP-01 Crystallography

1. Quartz-Homeotypic Crystals 827
G.F. Engel, P.W. Krempl
AVL List Ges. m.b.H, Graz, Austria

2. Berlinite Blank Absolute Orientation Determinations by X-Ray Diffraction 833
J.F. Darces, M. El Bellal, H. Merigoux
U.F.R. des Sciences et des Techniques, Besançon - France

3. A New Approach to Berlinite Crystal Growth 843
E. Philippot, A. Goiffon, J.C. Jumas
USTL, Montpellier - France
X. Buisson, R. Arnaud
SICN, Annecy - France
Y. Toudic
CNET, Lannion - France

4. Lattice Defects in Lithium Tantalate 851
P. Cordier, B. Boulogne, N. Doukhan, J.C. Doukhan
Université de Lille Flandres Artois, Villeneuve d'Ascq - France
M. Fragneau, J.P. Aubry
Compagnie d'Electronique et de Piezo Electricité, Argenteuil - France

5. Evaluation and Application of High Purity Berlinite Crystals..... 873
J. Detaint, J. Schwartzel, R. Lefevre
CNET PAB/BAG/MCT/CMM Bagneux - France
J.P. Aubry, L. Michel
CEPE, Argenteuil - France
B. Capelle, A. Zarka
Université P. et M. Curie, Paris - France
E. Philippot
Université de Montpellier, Montpellier - France

MP-02 Physical Properties

1. A Model for the Dependence of the Q-Factor of Quartz Resonators on the Surface Texture: Application to Etched Quartz Plates 893
T. Leblois, R. Bourquin, C.R. Tellier
Ecole Nationale Supérieure de Mécanique et des Microtechniques, Besançon - France

2. Nonlinear Properties of Piezoelectric Semiconductors 909
N. Daher, J.J. Gagnepain
CNRS, Besançon - France

3. Growth Defects in Cultured Quartz, a Microscopic Investigation 921
P. Cordier, J.C. Doukhan
Université de Lille Flandres Artois, Villeneuve d'Ascq - France
J. Bignon, X. Buisson
SICN, Annecy - France

4. The Orientation Dependence of the Dissolution Rate of Singly Rotated Quartz Plates Derived from Combined Numerical and Digital Techniques 937
C.R. Tellier, P. Blind, D. Jozwick
Ecole Nationale Supérieure de Mécanique et des Microtechniques, Besançon - France

Author index 959

* Abstracts only

2 ND EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 16 - 17 - 18 MARCH 1988

Founding Committee

R. Besson	*Directeur du Laboratoire de Chronométrie, Electronique et Piézoélectricité (LCEP)* *Professeur à l'Ecole Nationale Supérieure de Mécanique et des Microtechniques (ENSMM), Besançon - France*
M. Ecabert	*Directeur de la Fondation suisse pour la recherche en microtechnique (FSRM), Neuchâtel - Suisse*
J.J. Gagnepain	*Directeur du Laboratoire de Physique et Métrologie des Oscillateurs (LPMO) du CNRS, Besançon - France*
P. Kartaschoff	*Ingénieur, Chef de Section, Recherche et Développement de la Direction Générale des PTT, Berne - Suisse*
A. Rémond	*Ingénieur de Recherche et Chef du Service Chronométrique de l'Observatoire de Besançon - France*
B. Schlueter	*Directeur d'Oscilloquartz SA, Neuchâtel - Suisse*

Organization

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

2000 N e u c h â t e l - Switzerland (Tel. ++38 24 52 00)

With support of

République et Canton de Neuchâtel / Ville de Neuchâtel / Communauté de travail du Jura (CTJ), Suisse - France / Hasler AG, Berne / Oscilloquartz SA, Neuchâtel / Rolex Montres SA, Genève / Brasserie Muller, Neuchâtel / Facchinetti SA, Neuchâtel / Helvetia assurances, Neuchâtel / PTT, Neuchâtel / Neuchâteloise assurance, Neuchâtel / Société de Banque Suisse, Neuchâtel / Université de Neuchâtel / Winterthur assurances, Neuchâtel

2 ND EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 16 - 17 - 18 MARCH 1988

Science Committee

P. Kartaschoff	*PTT, Berne*	*Switzerland*
C. Audoin	*C.N.R.S.-L.H.A., Orsay*	*France*
A. Ballato	*ERDACOM, Ft. Monmouth*	*USA*
J. Beaussier	*ONERA, Châtillon*	*France*
R. Besson	*E.N.S.M.M.-L.C.E.P., Besançon*	*France*
M. Brunet	*C.N.E.S., Toulouse*	*France*
G. Busca	*Oscilloquartz SA, Neuchâtel*	*Switzerland*
H. de Boer	*PTB, Braunschweig*	*Germany*
H. Duchaussoy	*D.R.E.T., Paris*	*France*
J.J. Gagnepain	*C.N.R.S.-L.P.M.O., Besançon*	*France*
E. Graf	*Oscilloquartz SA, Neuchâtel*	*Switzerland*
M. Granveaud	*Observatoire de Paris, Paris*	*France*
B. Guinot	*Bureau International des Poids et Mesures, Paris*	*France*
P. Hartl	*Universität Stuttgart, Stuttgart*	*Germany*
J. Henaff	*C.N.E.T.-D.I.T.-A.S.C., Issy-les-Moulineaux*	*France*
C. Hruska	*York University, Downsview*	*Canada*
D. Kirchner	*Technische Universität Graz, Graz*	*Austria*
S. Leschiutta	*Ecole Polytechnique, Torino*	*Italy*
Ch. Maerfeld	*Thomson-Sintra-DTAS, Valbonne*	*France*
K. Nakagiri	*Radio Research Laboratory, Tokyo*	*Japan*
G. Nard	*Sercel SA, Carquefou*	*France*
B. Oehman	*Swedisch Telecom Radio, Farsta*	*Sweden*
L. Prost	*Office fédéral de Métrologie, Wabern*	*Switzerland*
J. Rutman	*C.E.P.E., Argenteuil*	*France*
B. Serene	*European Space Agency (ESA), Toulouse*	*France*
S. Starker	*D.F.V.L.R., Wessling*	*Germany*
J. Vanier	*National Research Lab., Ottawa*	*Canada*
F. von Willisen	*ASULAB SA, Neuchâtel*	*Switzerland*
F.L. Walls	*National Bureau of Standards, Boulder*	*USA*
G. Winkler	*US Naval Observatory, Washington*	*USA*
N. Yannoni	*Rome Air Development Center, Bedford*	*USA*
W. Zingg	*Micro Crystal, Grenchen*	*Switzerland*

2 ND EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 16 - 17 - 18 MARCH 1988

Session chairmen

Claude AUDOIN	*C.N.R.S. - L.H.A.* *Bât. 221, F-91405 ORSAY Cedex*	AFS-02
Raymond BESSON	*E.N.S.M.M. - L.C.E.P.* *Route de Gray-La Bouloie, F-25030 BESANCON Cedex*	RO-03
Michel BRUNET	*C.N.E.S.* *18, avenue E. Belin, F-31055 TOULOUSE Cedex*	TF-05
Giovanni BUSCA	*OSCILLOQUARTZ SA* *Rue Brévards 16, CH-2000 NEUCHATEL*	AFS-03
Hermann DE BOER	*PHYSIKALISCH-TECHNISCHE BUNDESANSTALT* *Bundesallee 100, D-3300 BRAUNSCHWEIG*	AFS-01
Jacques GROSLAMBERT	*C.N.R.S. - L.P.M.O* *32, av. de l'Observatoire, F-25000 BESANCON*	TF-01
Bernard GUINOT	*BUREAU INTERN. DES POIDS ET MESURES* *Pavillon de Breteuil, F-92312 SEVRES Cedex*	TF-04
Carl HRUSKA	*YORK UNIVERSITY* *4700 Keele Street, CDN-M3J 1P3 DOWNSVIEW, Ontario*	MP-02
Peter KARTASCHOFF	*P T T* *Ostermundigenstrasse 93, CH-3006 BERNE*	TF-03
Siafrido LESCHIUTTA	*ECOLE POLYTECHNIQUE* *I-10125 TORINO*	AFS-04
Charles MAERFELD	*THOMSON-SINTRA-DTAS* *399, route des Crêtes, F-06560 VALBONNE Cedex*	RO-04
Jacques RUTMAN	*C.E.P.E.* *B.P. 165, F-95100 ARGENTEUIL*	MP-01
Harry F. TIERSTEN	*RENSSELAER POLYTECHNIC INSTITUTE* *110 Eighth Street, USA-12180-3590 TROY, N.Y.*	RO-01
Gernot WINKLER	*US NAVAL OBSERVATORY* *34th Massachusetts Av NW, USA 20392-5100 WASHINGTON*	TF-02
Nicholas YANNONI	*ROME AIR DEVELOPMENT CENTER* *Hanscom AFB, USA 1731 BEDFORD, MA*	RO-05
Walter ZINGG	*MICRO CRYSTAL* *CH-2540 GRENCHEN*	RO-02

2 ND EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 16 - 17 - 18 MARCH 1988

Exhibitors

CABLES CORTAILLOD SA	*Av. François-Borel*	*CH-2016 Cortaillod*
C I R		*CH-2076 Gals*
CSEM SA	*Maladière 71*	*CH-2000 Neuchâtel*
ETA SA, DIV. MICRO-CRYSTAL	*Schild-Ruststr. 17*	*CH-2540 Grenchen*
FAVAG SA	*Monruz 34*	*CH-2000 Neuchâtel*
FSRM	*Orangerie 8*	*CH-2000 Neuchâtel*
GOTELE AG	*Bahnhofstrasse 76*	*CH-8500 Frauenfeld*
LANGE-ELECTRONIC GMBH	*Hauptstrasse 73*	*D- 8037 Olching*
OFFICE DU TOURISME	*Place-d'Armes 7*	*CH-2000 Neuchâtel*
OSCILLOQUARTZ SA	*Brévards 16*	*CH-2002 Neuchâtel*
SERCEL	*B.P. 64*	*F-44471 Carquefou*
SICN	*4, rue du Radar*	*F-74008 Annecy*
SOCIETE DE BANQUE SUISSE	*Fbg de l'Hôpital 8*	*CH-2001 Neuchâtel*
THOMSON-CEPE	*B.P. 165*	*F-95105 Argenteuil*
UNVERDROSS TECHNIK	*Tutzinger Hof Platz*	*D- 8130 Starnberg*
VIBRO-METER SA	*Moncor 4*	*CH-1701 Fribourg*

2 ND EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 16 - 17 - 18 MARCH 1988

Message d'accueil

de Monsieur Gérard F. Bauer

Mesdames et Messieurs,

C'est un devoir enviable et un honneur agréable pour le président du Comité de patronage du IIe Forum Européen Fréquence et Temps, de vous adresser une chaleureuse bienvenue en cette nouvelle Aula de l'Université de Neuchâtel qui nous accueille obligeamment aujourd'hui.

* * *

En premier lieu, j'aurai le privilège, Mesdames et Messieurs, de saluer nombre de personnalités, représentant les milieux politiques, les administrations nationales et régionales, universitaires de France, de Suisse, des Pays européens et non-européens, en particulier de nos voisins de la Franche-Comté, des Cantons jurassiens, des organisations transfrontalières de nos régions.

*

Tout à l'heure, nous accueillerons Monsieur Jean Cavadini, Conseiller d'Etat et Conseiller aux Etats de la République et Canton de Neuchâtel, qui, à son tour, vous adressera les paroles de bienvenue du Canton, dont nous sommes les hôtes.

*

Je m'en voudrais de ne pas exprimer dès maintenant nos sentiments de gratitude aux innovateurs et aux organisateurs des régions de Besançon et de Neuchâtel d'avoir convié pour la deuxième fois, après Besançon, en 1987, à Neuchâtel, leurs collègues scientifiques et industriels au Forum Fréquence et Temps.

*

C'est avec combien de satisfaction que nous constatons, en effet, que le Forum, âgé d'à peine deux ans, a acquis, d'ores et déjà, et à juste titre une réputation internationale.

Vous êtes venus, Mesdames et Messieurs, en grand nombre de la plupart des Etats européens comme aussi de l'Afrique du Sud, du Canada, du Japon, de la République populaire de Chine et des Etats-Unis d'Amérique.

* * *

Le Forum Européen Fréquence et Temps représente une initiative exemplaire aux plusieurs titres suivants:

- En effet, elle émane de deux Régions, qui ont, siècle après siècle, su cultiver et enrichir leur savoir-faire horloger inspirées qu'elles étaient et qu'elles sont aujourd'hui d'une tradition et d'une ambition de précision, développant des enseignements universitaires et techniques généraux et spécialisés de haute valeur.

- Ce capital de tradition et cette capacité d'innovation permanente, ces régions ont eu l'intelligence d'en faire usage pour multiplier les applications les plus modernes de la microtechnique contemporaine.

- Ce Forum trouve sa source dans deux régions transfrontalières, la Franche-Comté et les régions jurassiennes; c'est là un modèle démonstratif d'une communauté d'intérêt ignorant les frontières nationales.

- Le Forum constitue une rencontre bénéfique d'industriels et de scientifiques, animés des mêmes ambitions de mieux et de plus savoir.

- Enfin, très opportunément, le Forum a convié, pour la deuxième fois, à ses travaux des chercheurs et des développeurs scientifiques et des entreprises industrielles d'un grand nombre d'Etats européens et non-européens,

passant de la dimension interrégionale à la dimension européenne et mondiale, persuadés qu'ont été les organisateurs du Forum, dès son origine, que stimulation et émulation sont sources inégalées de richesses scientifiques et économiques.

Que les organisateurs français et suisses du Forum soient remerciés de l'ouverture d'esprit dont ils ont fait montre !

* * *

Mesdames et Messieurs, je forme les voeux les plus sentis pour le plein succès des journées de travail du IIe Forum Européen Fréquence et Temps.

Qu'il contribue à vous faire mieux apprécier les résultats les plus récents de la recherche et à mieux saisir les orientations prises par l'industrie et la science dans les domaines qui sont les vôtres.

Je souhaite enfin que vous puissiez aussi prendre plaisir aux heures de détente que vous vivrez à Neuchâtel.

Gérard F. Bauer

Président du Comité de patronage

Neuchâtel, le 16 mars 1988

Invited Papers

0. Significance of Metrological Measurements of Frequency and Time
O. Piller
Office fédéral de métrologie, Wabern - Switzerland

1. Advances in Science, Emerging Technologies and Global Competition : Opportunities and Challenge
H.W. Hellwig
National Bureau of Standards, Gaithersburg - USA

2. Laser Frequency Standards and Frequency Measurement for the Optical Fibre Bands
D.J.E. Knight
National Physical Laboratory, Teddington - UK

3. A review on ion traps frequency standards
G. Werth
Universität Mainz, Mainz - FRG

4. Time and Frequency Accuracy from GPS
A.R. Pratt
Polytechnik Electron PLC - Daventry, North - UK

5. Timing Requirements of Future ISDNs
R.D. Calvert
Cable and Wireless PLC, London - UK

6. Timing of a European Worldwide Satellite-Based Navigation System
H.A. Laue
ESA, Darmstadt - FRG

7. Very Long Baseline Interferometry (VLBI)
R. Spencer
University of Manchester, Macclesfield - UK

Session Invited Number 0

IMPORTANCE DES MESURES DE TEMPS ET DE FRÉQUENCE EN MÉTROLOGIE

O. Piller, Office fédéral de métrologie CH-3084 Wabern
P. Koch, Office fédéral de Métrologie Wabern

Introduction

Il semble naturel à l'homme du 20ème siècle que tout ce qui doit être mesuré puisse l'être et, de plus, avec la précision voulue. Il semble tout aussi naturel qu'un kilogramme, un kilowattheure, un mètre ou toute autre unité ait la même signification dans le monde entier. Pour apporter cette garantie, il faut qu'une institution en ait la charge dans chaque pays. Les institutions responsables de l'ensemble de la métrologie doivent maintenir les contacts nécessaires entre elles. Cette coopération existe dans le cadre de la Convention du mètre, accord fondé en 1875, qui réunit les états du monde entier. Ceux-ci s'engagent à appliquer et à employer les unités choisies au niveau international (Système International d'Unités SI).

Si, à l'origine, le mètre et le kilogramme, prototypes en platine iridié, couvraient les besoins métrologiques, il ne faut pas moins de 7 unités de base aujourd'hui. Ce sont celles des grandeurs physiques suivantes: longueur, masse, temps, intensité de courant électrique, température, quantité de matière et intensité lumineuse. Toutes les autres unités peuvent en principe être dérivées de ces sept unités de base. Les experts de la Convention du mètre se sont toujours efforcés de donner aux unités fondamentales des définitions basées sur les valeurs de constantes naturelles.

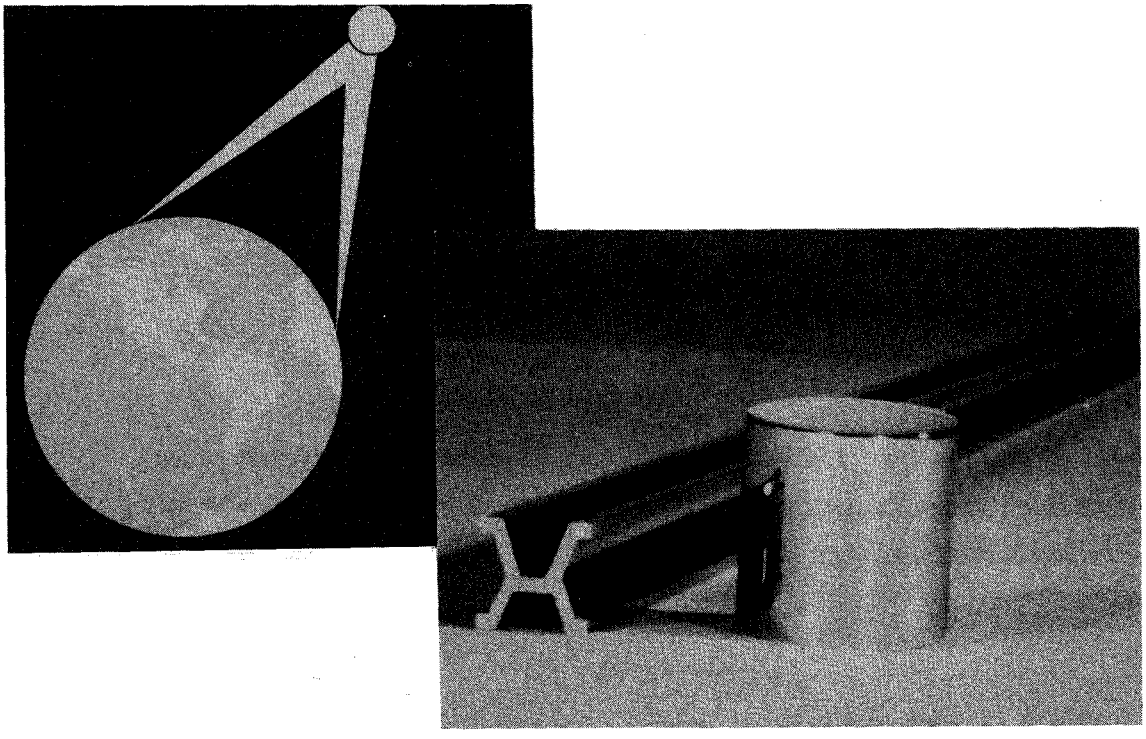


Fig. 1: Mètre et kilogramme prototypes des 18e et 19e siècles.

Dans cette optique, le 18ème et le 19ème siècle cherchèrent à définir le mètre par rapport à la circonférence terrestre, et le kilogramme par rapport à la masse d'un décimètre cube d'eau prise à sa masse volumique maximale. La seconde fut définie à partir de la vitesse de rotation de la terre sur elle–même. A cette époque déjà, des savants célèbres firent remarquer que la création d'unités vraiment universelles passait nécessairement par l'adoption de meilleures définitions. En 1870, James Clark Maxwell écrivait ceci:

> *The earth has been measured as a basis for permanent standard of length, and every property of metals has been investigated to guard against any alteration of the material standards when made. To weigh or measure anything with modern accuracy requires a course of experiment and calculation in which almost every branch of physics and mathematics is brought into requisition.*
>
> *Yet, after all, the dimensions of the earth and its time of rotation, though relatively to our present means of comparison, very permanent, are not so by any physical necessity. The earth might contract by cooling, or it might be enlarged by a layer of meteorites falling on it, or its rate of revolution might slowly slacken, and yet it would continue as a planet as before. But a molecule, say of hydrogen, if either its mass or its time of vibration were to be altered in the least, would no longer be a molecule of hydrogen.*
>
> *If, then, we wish to obtain standards of length, time and mass which shall be absolutely permanent, we must seek them not in the dimensions, or the motion or the mass of our planet, but in the wave length, the period of vibration, and the absolute mass of these imperishable and unalterable and perfectly similar molecules.*

En 1889, Max Planck affirmait:

> *Ein physikalisches Einheitensystem sollte nicht auf Atomen und Molekülen basieren, sondern auf tiefer liegenden Realitäten, wie die Fundamentalkonstanten der Natur selbst. Solche Fundamental–*

konstanten wie h, c, k und e sind universell wegen ihrer Unabhängigkeit von bestimmten Atomen oder Substanzen und wegen ihrer Konstanz bezüglich Raum und Zeit.

Dans la sphère de la Convention du mètre, on s'est attaché jusqu'ici à oeuvrer dans le sens de Maxwell et Planck et à améliorer les définitions en fonction des dernières découvertes des sciences exactes, de manière à rendre les unités indépendantes du temps et de l'espace.

Outre le problème de leurs définitions, se pose le problème du nombre des unités nécessaires pour mesurer l'ensemble des grandeurs de notre monde physique. Nous sentons intuitivement le besoin de trois unités pour décrire les phénomènes dans un espace fini: une pour la longueur, une pour la masse, et une pour le temps. On peut démontrer mathématiquement que ces trois unités suffisent pour mesurer toutes les grandeurs mécaniques. En effet, il y a au total 36 équations indépendantes, mettant en relation 39 grandeurs différentes. Si l'on veut résoudre ce système d'équation, il faut fixer trois grandeurs.

Qu'en est-il de l'électricité?
Depuis Maxwell, nous savons qu'il y a quatre équations qui contiennent huit grandeurs physiques:

$$\text{rot } \vec{H} - \epsilon_0 \frac{\partial \vec{E}}{\partial t} = \vec{j}$$

$$\text{rot } \vec{E} - \mu_0 \frac{\partial \vec{H}}{\partial t} = 0$$

$$\text{div } \vec{E} = \frac{1}{\epsilon_0} \delta$$

$$\text{div } \vec{H} = 0$$

avec $\vec{H}$: champ magnétique
$\vec{E}$: champ électrique
μ_0 : perméabilité magnétique du vide
ϵ_0 : constante diélectrique du vide
$\vec{j}$: intensité de courant électrique

t : temps

$\vec{X}$: distance

δ : charge électrique volumique.

Ces quatre équations fondamentales permettent de décrire tous les phénomènes électromagnétiques dans lesquels il n'y a pas de réaction atomique. C'est une loi naturelle bien connue qu'un champ magnétique exerce une force sur une charge électrique en mouvement.

La force elle-même est définie par les unités mètre, kilogramme et seconde. Par conséquent, ces trois grandeurs fondamentales sont implicitement contenues dans les équations de Maxwell. On en déduit qu'il faut quatre unités de base pour décrire tous les phénomènes mécaniques et électriques, la quatrième étant l'ampère, comme on le sait. La poursuite de ce raisonnement aboutit au Système International d'Unités, avec 7 unités de base dont les définitions valables aujourd'hui, avec en plus l'indication de la précision de leur réalisation actuellement atteignable, sont les suivantes:

Le mètre m — 10^{-9}

Le mètre est la longueur du trajet parcouru dans le vide par la lumière pendant une durée de 1/299 792 458 de seconde.

Le kilogramme kg — 10^{-9}

Le kilogramme est la masse du prototype international du kilogramme.

La seconde s — 10^{-13}

La seconde est la durée de 9 192 631 770 périodes de la radiation correspondant à la transition entre les deux niveaux hyperfins de l'état fondamental de l'atome de césium 133.

L'ampère A — 10^{-6}

L'ampère est l'intensité d'un courant constant qui, maintenu dans deux conducteurs parallèles, rectilignes, de longueur infinie, de section circulaire négligeable, et placés à une distance de

1 m l'un de l'autre dans le vide, produirait entre ces conducteurs une force égale à 2×10^{-7} N par mètre de longueur du conducteur.

Le kelvin K 10^{-5}

Le kelvin est la fraction de 1/273,16 de la température thermodynamique du point triple de l'eau.

La mole mol 10^{-6}

La mole est la quantité de matière d'un système contenant autant d'entités élémentaires qu'il y a d'atomes dans 0.012 kg de carbone 12.

La candela cd 10^{-4}

La candela est l'intensité lumineuse dans une direction donnée, d'une source qui émet un rayonnement monochromatique de fréquence 540×10^{12} Hz et dont l'intensité énergétique dans cette direction est de 1/683 W/sr.

La précision atteinte dans la réalisation de la seconde est frappante; elle est vraiment extraordinairement élevée. Citons un exemple: en automne, alors que les feuilles tombent des arbres et se rapprochent du centre de la terre, celle–ci accélère sa rotation sur elle–même comme une patineuse faisant une pirouette lorsqu'elle ramène les bras le long du corps.

Physiquement, on peut bien se représenter un tel phénomène; mais est–ce possible de mesurer un effet aussi insignifiant? Pour y parvenir, il faudrait utiliser une horloge capable de mesurer la durée d'un jour à un millionième de seconde près: ceci est en fait possible avec une horloge à césium.

De ce fait, la mesure du temps et des fréquences revêt une importance toute particulière dans la métrologie moderne. La définition du mètre, basée sur la vitesse de la lumière dans le vide, s'appuie sur la seconde. Différentes méthodes physiques, visant à réaliser des unités dérivées ou à mesurer des constantes fondamentales, sont fondées sur la mesure du temps ou des fréquences. Citons deux exemples.

La nouvelle définition du mètre

La 17e Conférence générale des poids et mesures a décidé, le 21 octobre 1983, de donner au mètre sa nouvelle définition. Simultanément, une deuxième résolution invitait le Comité international des poids et mesures (CIPM) à formuler des instructions sur la manière de traduire cette définition dans les faits, sur les longueurs d'ondes lumineuses à choisir pour les mesures de longueurs par interférométrie, et de plus, à promouvoir des recherches visant à améliorer les étalons des longueurs d'ondes choisies. Sur la base de cette deuxième résolution, le CIPM a publié une recommandation N^0 1 (CI–1983) dans laquelle figurent des méthodes de réalisation pratique de la nouvelle définition du mètre. Elle recommande en particulier d'employer les trois raies d'absorption suivantes du iode 127:

F_1 : 520 206 806,8 MHz
λ_1 : 576 294 760,27 fm

F_2 : 473 612 214,8 MHz
λ_2 : 632 991 398,1 fm

F_3 : 489 880 355,1 MHz
λ_3 : 611 970 769,8 fm

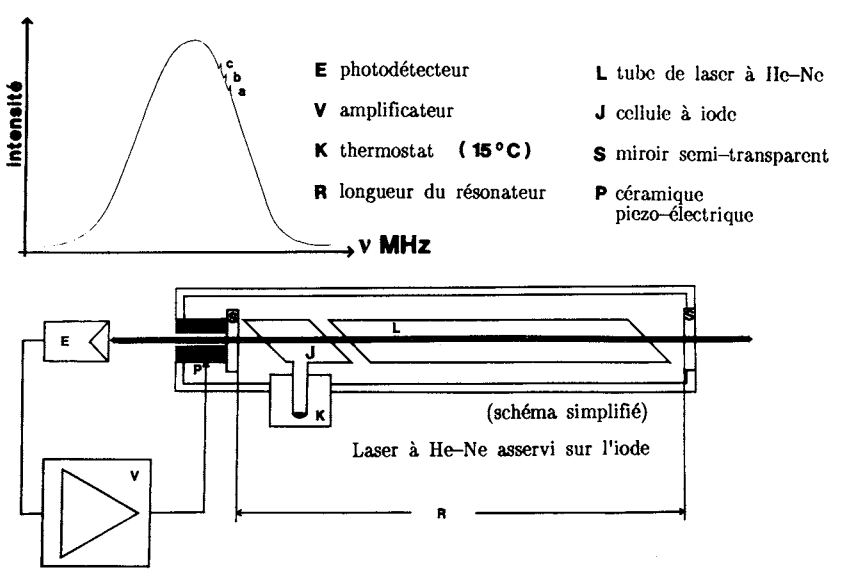


Fig. 2:
Laser stabilisé à l'iode.

Le laser stabilisé à l'iode est l'instrument de travail qui permet de réaliser cette recommandation. A l'aide d'une cellule d'absorption à iode, on stabilise un laser

He–Ne. Un amplificateur "lock–in" règle la longueur du résonateur de manière à ce que la fréquence moyenne du laser corresponde à la raie d'absorption. Ce processus est analogue à celui de l'horloge au césium où la fréquence de l'oscillateur à quartz est réglée sur celle d'une transition entre deux niveaux d'une structure atomique hyperfine. Ainsi, la stabilité de fréquence du laser passe de quelque 10^{-9} sans cellule à 10^{-12} avec cellule. Mais pour pouvoir utiliser cette stabilité, et pour la lier directement à la définition du mètre, il faut que la fréquence du laser stabilisé soit connue avec une meilleure précision. Pour cela il faut employer des chaînes d'étalonnage de fréquence à multiplicateurs. Celles–ci permettent de rattacher les fréquences laser dans le spectre visible à l'étalon primaire de la fréquence qu'est l'horloge à césium. La longueur d'onde du laser, utilisée pour la mesure interférométrique de longueurs, se déduit de la fréquence ν mesurée et de la vitesse de la lumière c. La nouvelle définition du mètre permet de tirer profit de la très haute exactitude de réalisation de la seconde pour transmettre cette précision à la réalisation du mètre. Aujourd'hui déjà, les principaux instituts nationaux de métrologie assurent la transmission de l'unité de longueur avec des lasers stabilisés à iode.

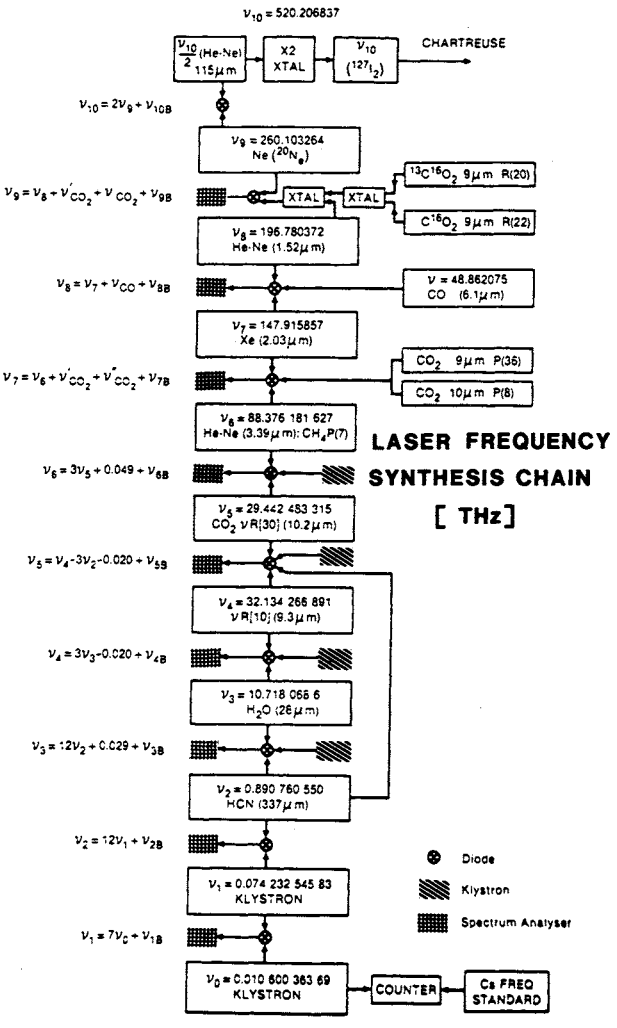


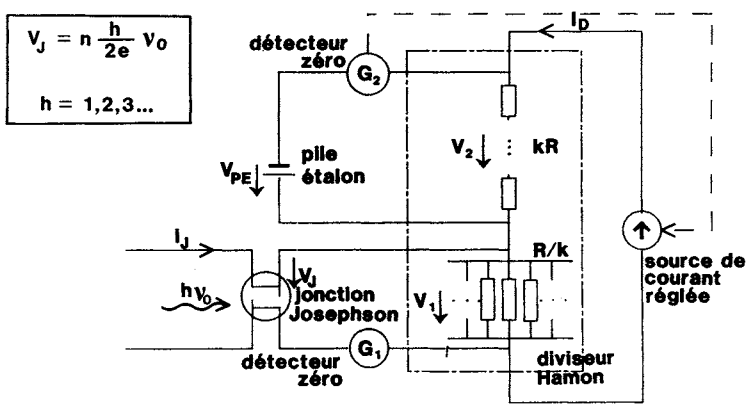
Fig. 3:
Chaîne d'étalonnage de fréquences à multiplicateurs.

Certains disposent de chaînes d'étalonnage de fréquences qui permettent de comparer régulièrement les lasers aux horloges à césium. D'autres se basent sur les fréquences recommandées par le CIPM.

Le volt Josephson

La découverte au début des années 60 d'effets quantifiés macroscopiques comme la quantification du flux et l'effet Josephson dans le domaine de la supraconductivité a contribué à approfondir et à élargir nos connaissances sur l'état de la matière. Ces nouveaux développements jouèrent aussi leur rôle dans la technique de mesure. A titre d'exemple, l'effet Josephson en courant alternatif permet de reproduire l'unité de tension électrique volt, indépendamment de l'espace et du temps, avec une précision qui ne dépend que de celle de la fréquence d'une micro-onde. Le facteur de proportionnalité en jeu est un multiple entier du quotient de deux constantes fondamentales, soit 2e/h. La valeur du quotient 2e/h a déjà été fixée en 1972 dans le cadre de la Convention du mètre pour assurer une réalisation uniforme du volt

$$2e/h = 4{,}8359767 \cdot 10^{14}\ \text{Hz} \cdot \text{V}^{-1}.$$



Ces dernières années, on a fait de grands progrès dans la réalisation directe de 1 volt sans passer par les diviseurs de tension, en multipliant le nombre de jonctions Josephson.

Fig. 4: Circuit classique de la comparaison tension Josephson et pile étalon.

Le PTB–Braunschweig a réussi l'exploit de connecter en série 1440 éléments, montés en un bloc monolithique, ce qui fournit une source de 1 volt. Cette réalisation rend superflu l'étalonnage de diviseurs de tension. Une relation directe entre tension et fréquence peut être établie de manière optimale. On a pu créer ainsi, pour la première fois, une tension de 1 volt avec une incertitude relative de seulement $3 \cdot 10^{-10}$. La stabilité en fréquence est contrôlée soit directement avec un étalon à césium, soit, pour les sources de tension mobiles, par référence aux signaux de l'émetteur DCF 77. Ici aussi, il devient possible de beaucoup mieux cerner la technique de mesure en électricité grâce à la très haute précision de la seconde et à l'application judicieuse de quelques effets physiques.

Ces deux exemples illustrent bien l'importance que la mesure du temps et des fréquences a prise en métrologie. D'autres pourraient être cités, en particulier l'effet Hall quantifié. Celui–ci, qui permet la réalisation de l'ohm, fournira à la technique des mesures en électricité une base solide reposant sur des constantes naturelles.

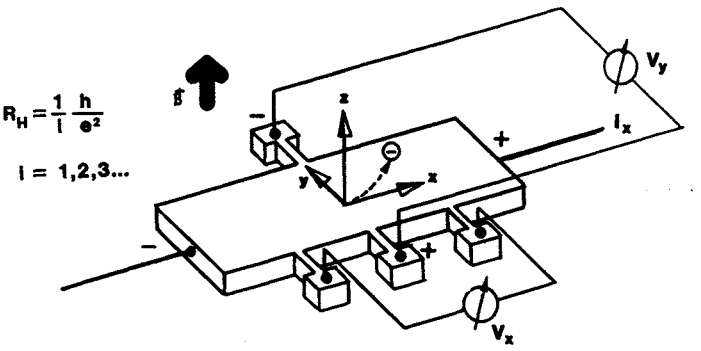


Fig. 5:
Schéma d'un échantillon Hall.

En 1980, le physicien Klaus von Klitzing découvrit comment déterminer avec précision la valeur de h/e^2. Il fit des mesures de tension Hall sur des éléments MOSFET à très basse température (< 4.2 K) et sous champ magnétique très élevé (15 tesla). Le résultat surprenant de cette expérience fut que la résistance Hall des couches d'électrons, que l'on peut admettre bidimensionnelles, varie de manière quantique, en sauts égaux à un multiple de la valeur h/e^2. De la même manière que l'effet Josephson en courant alternatif pour le volt, l'effet Hall quantifié pourrait être utilisé pour réaliser l'ohm avec une très haute précision.

Malheureusement, la précision souhaitée pour la réalisation de l'ampère n'a pas pu être atteinte jusqu'à maintenant. La question se pose de savoir s'il ne faudrait pas envisager, ici aussi, une nouvelle définition, dans l'intérêt de la technique métrologique. En combinant les effets Josphson et Hall quantifiés sous le chapeau de la loi d'ohm, on obtiendrait une source de courant dont l'intensité serait directement proportionnelle à une fréquence. Le facteur de proportionnalité est un multiple de la charge élémentaire de l'électron e.

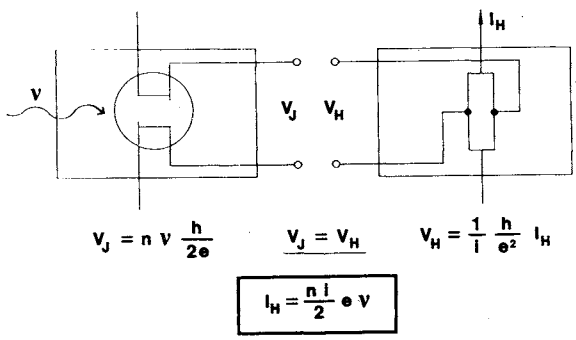


Fig. 6 :
Suggestion pour la combinaison de l'effet Josephson et de l'effet Hall quantique pour une redéfinition possible de l'ampère.

Références

U. Stille — Messen und Rechnen in der Physik, Vieweg und Sohn, Braunschweig (1961).

A. Sommerfeld — Vorlesungen über theoretische Physik, Akademische Verlagsgesellschaft (1964).

B. W. Petley — The fundamental physical constants and the frontier of measurements, Adam Hilger Ltd (1985).

O. Piller — Einführung in die elektrische Messtechnik, AT-Verlag (1984).

P. Kartaschoff — Frequency and time, Academic Press (1978).

J. Niemeyer, L. Grimm — Present Status of Josephson Voltage Metrologie, Interne Mitteilung, PTB Braunschweig (1988).

Codata Bulletin — The 1986 adjustment of the fundamental physical constants, Pergamon Press, number 3, Nov. 86.

Session Invited Number 1

ADVANCES IN SCIENCE, EMERGING TECHNOLOGIES AND GLOBAL COMPETITION: OPPORTUNITIES AND CHALLENGE

Helmut Hellwig
National Bureau of Standards
Gaithersburg, Maryland 20899 USA

For the remainder of the century, the time and frequency community is faced with significant new opportunities and challenges. The roots of current time and frequency technology date back half a century; scientific advances have been largely evolutionary and have gradually expanded the boundaries of these known foundations. In stark contrast to this, the technological world is undergoing nearly revolutionary changes which are driven by three interrelated forces: Global Competition, Emerging Technologies, and Advances in Science.

In the following, specific examples of these three areas are described as they relate to time and frequency. The opportunities which can be derived from emerging technologies and scientific advances are quite large, but seemingly require more efficient methods of moving an idea from the laboratory to market readiness and a more sophisticated pooling of resources. Thus, winning ideas and organizations probably will be based on significant levels of cooperation both vertically (research, engineers, marketeers) as well as horizontally (between organizations, nationally and internationally).

Global Competition: We find ourselves in a world characterized by intense competition between a multitude of countries wishing to attain, maintain, or regain satisfactory socioeconomic positions [1]. In this context, a revolution has taken place in the approach towards the generation of technology and its products. Knowledge is converted to process and product with increasing speed and, for the benefit of everyone, with increased benefit-cost ratios. A key factor is high quality performance at time of introduction of the product or service. Advantages in global competition are derived from early integration of process, product and market objectives into the research and development phase and, in general, from a compression of the time-line from laboratory to market.

Figure 1 depicts what we may call "product-genesis," i.e., the conceptual steps typically found in the creation of a marketed product as a result of ideas generated in the pursuit of basic research.

Historically, product-genesis happened as a result of contributions from individuals or teams who were (nearly) independently active in each of the listed conceptual steps. Barriers existed between these steps related to organizational, funding, educational, and social differences as well as to the tools and equipment available to the "players." The resulting difficulties in handing over the results from one step to the next (in Fig. 1) have been a major cause for time-delays, costly redesigns, and the like:

GLOBAL COMPETITION

Product Genesis

- Basic Research
  - Applied Research
    - Product-directed Research
      - Product Development
        - Product Design
          - Prototyping
            - Production
              - Market Penetration

FIGURE 1

The ability to improve, correct deficiencies, and react to opportunities is substantially inhibited if the effort requires significant and coordinated contributions from several steps of the product-genesis. Furthermore, if organizational barriers are high, this sequential, compartmentalized approach to generate product often shows paralysis in the face of new opportunity. We are all familiar with the "not invented here" inhibition which is especially painful at junctions between product-directed research, product-development, product-design and prototyping. In the different approaches taken by the various firms, organizations, and societies, a general trend has developed giving those who most thoroughly follow it a competitive edge. This trend is the integration of the product-genesis sequence by adopting techniques ranging from improved communication through personal involvement (e.g. marketeers and manufacturing engineers becoming part of an R&D team) to structured techniques aimed at the removal of barriers.

Japan is commonly credited to be the leader in this trend. The result is a shortening of the time required to move an idea from the laboratory to the market; the associated expenses are reduced and an efficient internalization of external ideas occurs. A study by the University of Pennsylvania [2] has found that both time and cost associated with converting external research to product are, in Japan, half of that for the U.S. Furthermore, the associated costs (and time) in the U.S. are higher for external research results as compared to what internal research requires to reach product status; the reverse is true for Japan.

Not surprisingly, quality improves in parallel with the degree of implementation of these new approaches. It allows the producer to be first to market with a "debugged" product and a harmonious relationship between customer expectations and design and performance features of the product.

The trends towards improved quality through integration of the product-genesis, aided by advanced manufacturing technologies and automation, has led to a shift of focus away from the product to the process creating the product. In other words, quality and reliability at low costs are competitively stronger parameters than offering unique performance features.

The foregoing is of particular importance to the time and frequency community because of the extraordinarily strong incentives to the customer to obtain reliability in time and frequency references. This has led to an unusually high barrier to the market introduction of new time and frequency technologies and devices. Users have repeatedly experienced lack of quality in new products (and services). Thus, the competitive gradient is here even steeper than in other areas of modern technology. If one considers that nearly all organizations which provide time and frequency products and services are relatively small and without access to major resources, the advantages which can be gained from cooperative and interactive approaches may have nearly revolutionary impact on our field. In the author's opinion, the escalating user demands driven by the Global Positioning System and by the requirements of other technologies will force such developments.

Emerging Technologies: Emerging technologies are those technologies with a clear and substantial potential for new products or advanced production processes, but which do not yet have significant markets. In other words, they are advanced options and tools which have grown out of scientific knowledge by being associated with substantial economic potential. Figure 2 is a summary of emerging technologies as compiled by a task force in the

MAJOR EMERGING TECHNOLOGIES

- **Advanced Ceramics**
- Polymer Composites
- Applications of Computers
- **Optoelectronics**
- Artificial Intelligence
- **Advanced Microelectronics**
- Biotechnology
- Medical Technology
- **Thin-Layer Technology**
- Automation
- **New Superconducting Materials**

FIGURE 2

U.S. Department of Commerce [3]. Highlighted are those technologies with substantial potential for the time and frequency community.

Advanced ceramics are of importance for essential components of time and frequency products such as cavities, resonators, and structural components. In addition, advanced ceramics are of great importance to dielectrics and dielectric substrates and thereby to all applications of electronics, especially at very high frequencies. Also, advanced ceramics can be designed to feature certain properties; thus, they offer the potential of specifying particular materials characteristics as required by the designer of a product.

Optoelectronics is important for signal detection inside of frequency standards as well as signal transmission, especially with very low noise characteristics. Artificial intelligence and expert systems are of great importance in integrating time and frequency systems.

Advanced microelectronics, especially very large scale integration, is of enormous potential for miniaturization of time and frequency devices. Developmental circuit integration is now being pursued at feature dimensions of a few tenths of micrometers offering complete integration of analog, digital and computer functions on one small chip. Thus, the prospect to integrate the complete electronics of any frequency standard or oscillator onto one chip will create pressure to substantially reduce the size of the resonator component of oscillators and frequency standards.

Thin-layer technology is a persuasive generic technology for a great number of applications. In time and frequency, these include surfaces of detectors, optical emitters, and quartz crystals. Metalization of quartz crystals, coatings of storage vessels, and gettering surfaces are but a few examples where a better understanding may lead to control of the interaction between solid and liquid, solid and gas, and solid and solid phases.

Finally, we have to mention the new high temperature superconductors. The potential of these materials for time and frequency is as revolutionary as for any other of the numerous fields discussed in the public and technical press. The prospect of operation at room temperature offers a great number of options ranging from low loss cavities and resonators to advanced high-speed electronics and, very important for our field, to the possibility of efficient magnetic shielding and shielding against electromagnetic interference and electromagnetic pulses. The time and frequency community could contribute to the understanding of these new superconducting materials using time and frequency measurement technology, such as the exploration of nonlinear electromagnetic phenomena (mixing and multiplication) as well as studies, at high frequencies, of the characteristics of superconducting resonators.

An exciting example where the application of several of these emerging technologies has led to a breakthrough in measurement science is the Josephson-junction one-volt array [4]. Microelectronics, optoelectronics, advanced ceramics and superconducting technology were combined to link a macroscopic electric quantity to a quantum phenomenon using a micro array of more than a thousand Josephson-junctions fed by a microwave signal. This technology is now the de-facto standard for the volt, and, through the microwave signal, it is linked to the standard for frequency and time.

Advances in Science: The scientific ancestry of today's time and frequency technology reaches back about 50 years: Atomic and molecular

beams, spectroscopy and the understanding of crystal behavior were the subjects of intense investigation. The related scientific advances not only generated today's spectrum of time and frequency technologies but also were the underpinning of related technologies which are now important for time and frequency. Lasers, radiation and particle beams and, of course, modern electronics are now available as powerful investigative tools. We are, apparently, at the threshold of a new wave of scientific opportunities.

Scientific advances of importance to time and frequency are taking place in atomic, molecular, optical and solid state science. In the following, examples have been selected from prior publications:

Our ability to study and manipulate crystal surfaces has become substantial. We can grow gallium arsenide on a silicon wafer using laser technology as the probing and controlling means [5].

We can determine the structure and bonding of atoms in a host lattice using optical techniques such as measuring spacial directivity of emitted radiation [6, 7].

A powerful tool for the study of magnetization on an atomic scale is provided by the scanning electron microscope using polarization analysis of the emitted electrons. High resolution imaging of the magnetic domains and their structure and polarization is now possible [8].

The use of small angle neutron scattering allows depth profiling; i.e., the precise measurement of the penetration of hosted material into a host lattice as a function of distance from the top surface [9].

Applying analytical, spectroscopic techniques combined with digital data processing allows surface imaging of chemical composition. This technique, called compositional mapping [10], uses electron or ion probe beams in order to achieve a spacial resolution at the micrometer scale.

Resolution of atomic dimensions can be obtained by field ion microscopy which has recently been applied to complex materials such as the high-T_c superconductors [11].

Precision spectroscopy has been the foundation of accurate physical and chemical analytical methods, and, in particular, has been the roots of atomic frequency standards. It is important that not only the stabilization of oscillators (such as lasers) by atomic and molecular resonances is possible as has been practiced for some time, but that external stabilization of any optical signal is now possible. Experiments have demonstrated [12, 13] that the signal from a dye laser can be conditioned so that its spectral width is less than 1 Hz. Stabilization is done by a combination of acousto-optic frequency shifting and rapid electro-optic phase shifting.

The concept of squeezed states [14] is important especially to precision spectroscopy. Electromagnetic fields can be described by two canonically conjugate variables whose product cannot exceed certain values. However, we can manipulate this product by allowing one of the variables to have relatively large values while squeezing the other variable to feature reduced variability. Using this technique, a reduction of, for example, phase variations is possible with obvious benefits to optical spectroscopy and derivative technologies, such as optical communication and storage.

Finally, we must mention the fundamental advances possible in the storing and cooling of atoms and ions. A single ion can be stored in a trap and cooled by using properly aligned laser resonances [15]. Effective temperatures below one millikelvin have been reported. Not only does this achievement allow the study of free ions, nearly at rest, it also results in very long storage times, i.e. very high Q values of the resonance. Furthermore, the study of weak interactions between the ion and other ions in the trap becomes possible and the formation of ion clusters has been observed [16].

The cooling of neutral atoms to below one millikelvin has also been demonstrated. A near motionless cloud of sodium ions results from the slowing of atoms in an atomic beam through photon recoil from a counter-propagating laser beam (termed "optical molasses") and further, three-dimensional laser cooling of the atoms. Additional confinement in a laser trap is possible [17]. Again, we have not only the opportunity to observe atoms at rest, but also to observe their interaction over large distances with other atoms. Atomic cross-sections become very large and interactions can be optically manipulated [18].

The foregoing is merely a selection of some of the exciting developments in science during recent years. These and other developments impact not only the technologies we are currently using but they offer opportunities for new and different approaches to time and frequency devices. Introduction of these new ideas and their conversion to practical devices (or to new methods in processing and manufacture) can revolutionize our field. The development of the electrodeless quartz crystal (the "BVA" type) [19] is a good example of the advances possible with the merger of contributions from different disciplines and several organizations. Further improvements of the BVA (and of quartz crystal technology, in general) could result from concerted use of scientific advances as outlined above offering the prospect of eliminating aging and/or performance variability altogether.

The step from scientific opportunity to applied technology requires increasingly a multidisciplinary capability. Such capability not only demands cooperation between greater numbers of people from the various disciplines, but also requires access to resources and investigative tools which, typically, are not available to any particular organization. These new opportunities and advances seem to indicate that in the future a much higher degree of cooperation between people and organizations will be required to survive in global competition and to retain leadership.

References

1. Young, John A. (Chairman), Global Competition: The New Reality, Report of the President's Commission on Industrial Competitiveness, U.S. Government Printing Office.

2. Testimony before the Joint Economic Committee of Congress by Edwin Mansfield, Director, Center for Economics and Technology, and Professor of Economics, University of Pennsylvania, December 2, 1987.

3. The Status of Emerging Technologies: An Economic/Technological Assessment to the Year 2000, U.S. Department of Commerce, NBSIR 87-3671, June 1987 (issued November 1987).

4. Hamilton, C.A., Kautz, R.L., Lloyd, F.L., Steiner, R.L., and Field, B.R., The NBS Josephson Array Voltage Standard, IEEE Transactions of Instrumentation and Measurement, IM-36 (2), 258-261 (June 1987)

5. Bourguignon, B., Smilgys, R.V., and Leone, S.R., AES and LEED Studies Correlating Desorption Energies with Surface Structures and Coverages for Ga on Si(100), submitted to Surface Science (1988)

6. Egelhoff, W.G., Jr., A New Tool for Studying Epitaxy and Interfaces: The XPS Searchlight Effect, J. Vac. Sci. Technol. A, 3(3), 1511-1513 (1985)

7. Madey, T.E., Electron- and Photon-Stimulated Desorption: Probes of Structure and Bonding at Surfaces, Science, 234, 316-322 (1986)

8. Celotta, R.J. and Pierce, D.T., Polarized Electron Probes of Magnetic Surfaces, Science, 234, 333-340 (1986)

9. Maki, J.T., Fleming, R.F., and Vincent, D.H., Deconvolution of Neutron Depth Profiling Spectra, Nuclear Instruments and Methods in Physics Research, B17, 147-155 (1986)

10. Marinenko, R.B., Myklebust, R.L., Bright, D.S., and Newberry, D.E., Digital X-Ray Compositional Mapping with "Standard Mapping" Corrections for Wavelength Dispersive Spectrometer Defocusing, J. Microscopy, 145, 207-223 (1987).

11. Melmed, A.J., Shull, R.D., Chiang, C.K., and Fowler, H.A., Possible Evidence for Superconducting Layers in Single Crystal $YBa_2Cu_3O_{7-x}$ by Field Ion Microscopy, Science, 239, 176-178 (1988)

12. Hall, J.L. and Hansch, T.W., External Dye-laser Frequency Stabilizer, Optics Letters, 9, 502-50 (1984)

13. Hall, J.L., Zhu, M., Shimizu, F.J., and Shimizu, K., External Frequency Stabilization of a Commercial Dye Laser at the Hertz Level, Abstract for the International Quantum Electronics Conference IQEC-88, Tokyo, Japan (1988).

14. Wu, L-A., Kimble, H.J., Hall, J.L., and Wu, H., Generation of Squeezed States by Parametric Down Conversion, Physical Review Letters, 57 (20), 2520-2523 (1986)

15. Bergquist, J.C., Itano, W.M., and Wineland, D.J., Recoilless Optical Absorption and Doppler Sidebands of a Single Trapped Ion, Phys. Rev. A, 36 (1), 428-430 (1987)

16. Wineland, D.J., Bergquist, J.C., Itano, W.M., Bollinger, J.J., and Manney, C.H., Atomic-Ion Coulomb Clusters in and Ion Trap, Phys. Rev. Lett, 59 (26), 2935-2938 (1987)

17. Phillips, W.D., Gould, P.L., and Lett, P.D., Cooling, Stopping, and Trapping Atoms, Science, 239, 877-883 (1988)

18. Gould, P.L., Lett, P.D., Julienne, P.S., and Phillips, W.D., Observation of Associative Ionization of Ultra-cold Laser-Trapped Sodium Atoms, Phys. Rev. Lett., 60, 788 (1988)

19. Reference is made to papers published in these Proceedings (Proc. 2nd European Frequency and Time Forum); especially the two papers authored by Delaite, R., and Migeot, T.; and Peier, U.R., and Jenni, G., respectively.

2 ND EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 16 - 17 - 18 MARCH 1988

Session Invited Number 2

LASER FREQUENCY STANDARDS AND FREQUENCY MEASUREMENT FOR THE OPTICAL FIBRE BANDS

D J E Knight

National Physical Laboratory, Teddington, Middx. TW11 OLW, UK.

ABSTRACT
Existing laser frequency measurements in and near the 1.3 - 1.6 μm fibre-optic communication bands are discussed and tabulated. Possibilities for further laser frequency standards and for their measurement by frequency techniques are considered. Work towards this end at NPL is described, including experiments with the 1.52 μm HeNe laser.

Introduction

Techniques for laser frequency measurement are now established (1), and an introductory network of accurately measured frequency standards extends from the microwave to the visible region of the spectrum (2).

The use of coherent communication techniques in the silica optical fibre bands (180 - 240 THz, 1.6 -1.3 μm) is a likely development in the next few years (3). Such techniques will be needed to exploit the very large (60 THz) bandwidth available (4), and become possible because suitable solid-state lasers such as the distributed-feedback type are becoming commercially available.

Suitable means of frequency measurement or frequency-referral are needed in these bands. Fabry Perot etalon fringes can provide local frequency references and means of stabilising tunable lasers (5), and there is a further need for absolute frequency references such as by locking to gaseous absorption spectra (6).

High-resolution fundamental physical experiments (7) and spectroscopy activities (8) are already utilising optical fibre band frequencies in order to take advantage of techniques and devices developed for communications, such as tunable diode lasers (8). Such experiments both help provide, and themselves need, accurate frequency standards.

Also, a stabilised laser's signal, if in the low absorption region of a suitable fibre, might be distributed via an optical fibre system for use as a frequency standard both outside and within the system.

Further, frequency standards in the optical fibre bands, if accurate enough (1 part in 10^{11}), could be placed so as to be useful stepping stones to the measurement of visible stabilised-laser frequencies needed as frequency standards. Such standards are essential for realisation of the metre and for fundamental physical experiments. At present their accuracy limits that of determinations of the Rydberg constant (9), for example.

In the UK, collaborations are therefore needed between standards work for optical fibre measurements, for wavelength standards and for time and frequency measurements together with the research activities of the optical communication industry and research in the universities on

physics, chemistry and optoelectronics.

Until now, the developments in optical communication show a conflict of techniques with developments in optical frequency standards, because the existing communication method is by fast pulse modulation, involving large bandwidths, to about 20 GHz, with uncertain centre frequency (3), where by contrast frequency standards are essentially cw based, and are aimed at absolute accuracy better than, say 100 kHz. However, both activities have a common need for wide bandwidth devices such as modulators, mixers and detectors (4). An interesting evolution of frequency standards and techniques to meet the needs of the optical fibre bands can thus be expected as the development of coherent optical communication proceeds.

In this paper I point out the main spectral features of the silica optical fibre bands, show the existing set of known frequencies in and near the bands, indicate where useful developments for frequency standards might be found, and consider new routes for measurement using frequency mixing techniques. Some NPL work, mainly on the 1,52 μm HeNe laser and laser frequency measurement techniques, is discussed and other work on optical fibre measurements, fast detectors and mixers (4) and on stabilised diode lasers (BAR 87, 10) (*1) is drawn upon. Work towards absolute frequency measurement of a visible frequency at 576 nm is now suspended in favour of the work reported here (11) and towards a good frequency standard in the 10 μm band (12), and work towards improved visible standards using cooled ions in traps continues (13).

1. The optical communication bands

An approximate representation of the silica optical fibre transmission bands (Figure 1a) has two dips, separated by an O-H absorption band at 1.4 μm. The minimum absorption occurs in one dip at 1.55 μm, and a remarkable minimum of dispersion, which passes through zero, occurs in the other dip at 1.312 μm. To further reduce losses in transmission, longer wavelengths, e.g. 1.8 - 3 μm, offer reduced scatter, but different fibre material will be needed, e.g. fluoride glass, so the use of such fibres is further off (3). In silica glass the 1.4 μm absorption peak can be low, as depicted, or up to 10 dB/km, depending on the water content.

2. Existing spectral frequency references

Spectra of potential interest have been scanned using diode lasers, e.g. of NH_3 in the 1.5 μm band (14), and of CH_4 in the 1.3 μm band (15). The region between 1.5 and 1.3 μm, corresponding to the absorption hump in the fibre material from OH radical spectra, is populated with spectral lines from H_2O (16).

The only frequency between 1.5 and 1.3 μm so far measured by frequency mixing techniques is that of the 1.52 μm HeNe gas laser, for which Evenson's group (JEN 79) measured the gain-profile centre frequency to be

$$f = 196\ 780\ 372(25)\ \text{MHz}.$$

Other measurements have been made with lambda meters to an accuracy of the same order, however, and wavelength and frequency measurements are formally equivalent following the 1983 Definition of the metre (BI 84),

---

(*1) The Table references, e.g. BAR 87, are listed before the text references.

in which a fixed value (299 792 458 m/s) was adopted for the speed of light in vacuum.

A list of frequencies measured using either frequency-beating or interferometric methods is shown in Table 1. This ranges from 3.39 μm to visible standards to include neighbouring frequencies of interest for measurements in the fibre bands.

The measurements listed offer little coverage of the 1.5 and 1.3 μm bands themselves, see Figure 1b, and thus need to be augmented by new work. NH_3, CH_4 and H_2O appear to offer good scope for this, and perhaps HCN (17) and more complex molecules also (18). A coincidence between the Er laser and methane at 1.645 μm has been pointed out (19). A good summary of high resolution absorption spectra in this region is provided by NASA (16), and a new edition is in preparation. A limited set of molecules is covered however. In order of absorption strength these are, at 1.5 μm, HI, CO (λ> 1.557 μm), H_2O and CO_2, and at 1.3 μm, HF, H_2O, HBr, CO_2 and O_2. Notably the diatomic molecule HI has a set of well spaced and strong transitions at 1.3 μm (G H C Freeman, private communication). Atomic transitions have been very thoroughly listed (20), and there are a number in this region, including those used for the parity-violation experiments near 1.3 μm (7).

3. Lasers suitable for frequency standards work

Spectral widths comparable with those of cw gas lasers (10^{-8} or better) are needed for coherent communication work, or for sub-Doppler spectroscopy in gases. Also power levels above 10 mW are needed for frequency mixing in nonlinear devices, such as metal insulator metal (MIM) diodes or bulk nonlinear crystals. Diode lasers having external cavities (21-24,10) can nearly achieve these requirements and are also tunable, but may suffer from multimoding, mode-hopping, marginal power or electrical fragility.

Tunable, narrow cw radiation is also available from optically-pumped colour centre lasers such as from NaCl::OH at 1.47 to 1.87 μm (25) or from the Co:MgF_2 optically pumped laser at 1.52 - 1.98 μm (26). The optically-pumped 1.319 μm Nd laser tunes from 1.312-1.334 μm when a $NaNdP_4O_{12}$ host crystal is used (27), so that it might access several methane lines ((15), and R H Partridge, private communication).

The relatively narrow "fixed" lasers in each band comprise the diode-pumped Nd-YAG laser at 1.319 μm and the HeNe laser at 1.52 μm. It should be noted that the 1.15 μm HeNe laser emission also falls in the low-loss optical fibre region, and that its wavelength, although short, may be long enough for single-mode transmission (3).

4. Stabilisation of the 1.52 μm HeNe laser

The Zeeman mode-splitting technique offers a simple means of stabilising small HeNe lasers to the centre of the Ne line in the lasing plasma. A stability of 1.25 MHz, 6 parts in 10^9, has been reported from China (28). Experience at NPL (W R C Rowley, private communication) is that the laser line Zeeman splits in an axial magnetic field quite well, but that with a cavity length of 340 mm weak modes in the wings impede stabilisation of the central mode, implying a full lasing width exceeding 800 MHz.

In the NPL laser frequency laboratory we are operating a 4-m long 1.52 μm HeNe laser, first to realise the running conditions for the measured centre frequency in Table 1, and second to investigate a suitable gas

cell reference line. It should be possible to use an intra-cavity gas cell to find weak absorptions. The gain of this laser is low. A 2% output coupling is used and we obtain 3 mW single-mode output with a metal film mode selector placed 285 mm from the output mirror. The gas mix is best at about 30:1 He:Ne at 270 - 530 Pa (2 - 4 Torr), compared with the 10:1 used by Jennings et al. (JEN 79), but further investigation is needed.

Stabilisation of the 1.52 μm HeNe laser by locking to an absorption in NH_3 has been described (29), but the nearest NH_3 absorption was outside the laser's Doppler tuning range, about 600 MHz, so that saturated absorption stabilisation was not possible without frequency shifting the laser. $^{15}NH_3$ was offered as possibly having a coincidence.

The advantage of a relatively "fixed" frequency standard, such as the stabilised HeNe laser is in its simplicity of operation. The classic small methane stabilised HeNe laser of Barger and Hall (30) at 3.39 μm uses an accidental coincidence between a high-gain Ne laser emission and a strong absorption. At 1.52 μm the gain is lower and the absorptions are weaker, but it remains worth looking for a similar coincidence. Many telecommunications experiments have already used the 1.52 μm HeNe laser as a heterodyne reference. A tunable laser might be locked to many spectral references, but will need a dedicated lambda meter to identify the transition to which it is locked, apart from being intrinsically a more complex device.

5. Synthesis of frequencies to and from the optical fibre bands

It is useful to examine a linear scale of laser frequencies extending to the visible region at 500 THz, Figure 2. On this scale the optical-fibre bands are seen to be both large in extent and strategically placed for mixing with CO_2 laser frequencies en route to the visible region.

To reach the 1.52 μm laser directly from 9 μm, however, generation of the 6th harmonic is required. The present published limit of harmonic mixing in metal-insulator-metal (MIM) diodes is 2.3 μm, 130 THz (POL 83), at which frequency the 5th order ($2f_1 + 3f_2$) of CO_2 frequencies were mixed with that of a colour-centre laser. Recently, MIM point contacts having equally high frequency response but with increased CO_2 laser power handling, 0.25 W per laser, have been developed (31). Thus by using a powerful HeNe laser at 1.52 μm (JEN 79) and "blunt" MIM diodes there is a prospect of direct 6th-harmonic mixing with a CO_2 frequency (K M Evenson, private communication). In the previous experiments (JEN 79) summation of a 2.03 μm (Xe) and a 6.1 μm (CO) frequency was achieved in a MIM diode to reach 1.52 μm.

Alternatively, a frequency may be synthesised from known stabilised laser frequencies in the visible, e.g. by subtracting the frequency of HeNe/I2 at 1.15 μm from that of HeNe/I2 at 612 nm using non-critically phase-matched difference frequency generation in LiNbO3 (32). This process is complicated by needing two reference frequency standards.

The existence of a wide range of accurate frequency standards in the CO_2 band at 9-11 μm makes upward multiplication from this region to 1.5 μm attractive if achievable. It is convenient that the difference frequencies between 1.5 μm, 1.3 μm and 1.15 μm also fall in the CO_2 band.

Frequencies at both 1.26 μm and 1.15 μm can be doubled to beat with visible-region I2 standards, so that optical-fibre band standards could if accurate enough (1 part in 10^{11}) provide stepping stones for the

better measurement of the visible standards. Alternatively, lasers at these frequencies can be locked to the visible standards via frequency doubling, to provide standard emissions, or to calibrate any new local standards.

6. Discussion and conclusions

Frequency standards that would be useful in the 1.5 and 1.3 μm optical-fibre bands also fit quite conveniently into the existing patterns (2) of absolutely-measured laser frequency standards, primarily because of their relation to frequencies in the CO_2 laser band. In the wider frequency-synthesis picture, the connection with CO_2 frequencies is important too because of the hope that direct frequency mixing from the microwave to CO_2 frequencies may become possible. This might be through extension of high-frequency mixing in Josephson devices from the present limit at 4.25 THz (1,33) by the development of point-contact junctions made from high-Tc materials (34). (The electron pair binding energy in superconductors is proportional to Tc and for niobium corresponds to a photon frequency near 0.9 THz.)

The list of measured frequency references in Table 1 shows that a number of useful references exist at the 1 part in 10^7 level or better. An apparent gap in the 1.3 μm band is likely to be temporary, because high-resolution tunable diode laser spectra of methane have been observed (15) and because 1.26 μm is at half the frequency of the well-known 633 nm stabilised laser frequency. A diode laser at 1.26 μm may be used in a frequency chain (A Clairon, private communication). Lambda meters will continue to be useful for measuring frequency references in the near infrared, because their accuracy exeeds 1 part in 10^7.

At NPL, experiments will include stabilising 1.52 μm HeNe lasers and an investigation of harmonic generation in MIM diodes from CO_2 towards this region. Work relevant to accurate laser frequency standards in the 10 μm region (12) is being performed at the Cavendish Laboratory, Cambridge, towards our future needs. The general aim is to support optical communications in the 1.5 and 1.3 μm bands whilst developing techniques and devices also useful for supporting the needs of physics and for realisation of the metre from its definition. The continuation of work on visible ion-trap standards (13) implies a future need to make frequency comparisons with the Cs standard to very high accuracy.

The optical-fibre bands are extremely wide (60 THz) and present a large task in providing suitably spaced working standards for the region. This is complicated by the relative paucity of spectra, mainly of overtone bands, which may be drawn upon. There is nonetheless a certain urgency in the work because long-distance communication by optical fibre is becoming practical and international agreement on suitable frequency standards for coherent transmission may be required.

Aknowledgements

I thank G Duxbury, G H C Freeman, P Gill, D A Humphreys, M Lawrence, M J Padgett, R H Partridge, K I Pharaoh and W R C Rowley for discussions and material towards this paper.

Table References

AV 86. G Avila, P Gain, E de Clercq and P Cerez Metrologia 22 1986 pp 111-114.

BAR 87. G P Barwood, P Gill and W R C Rowley "A rubidium stabilised laser diode for interferometry" NPL Report MOM 85, November 1987 (National Physical Laboratory, Teddington, Middx. TW11 OLW, UK). Note: revised measurements for linear absorption in a gas cell at room temperature are quoted (unpublished).

BI 84. BIPM Metrologia 19 1984 pp 163-177.

ENG 75. R S Eng and D L Spears Appl. Phys. Lett. 27 (12) 1975 pp 650-652.

FU 84. K Fukuoka, M Ohtsu and T Tako Jpn J. App 1. Phys. 23 (2) February 1984 pp L117-L120.

GOL 85. V G Gol'dort, G S Gutov, V F Zakhar'yash, Yu F Kirillov, Yu A Matyugin, B A Timchenko and V P Chebotaev Kvant. Elektron. 12 October 1985 pp 2176-9; English transl.: Sov. J. Quant. Electron. 15 (10) October 1985 pp 1435-7.

JEN 75. D A Jennings, F R Petersen and K M Evenson App 1. Phys. Lett. 26 (9) 1975 pp 510-1.

JEN 79. D A Jennings, F R Petersen and K M Evenson, Optics Lett. 4 (5) May 1979 pp 129-130, see note 10.

OHT 83. M Ohtsu, H Kotani and H Tagawa Jpn. J. App 1. Phys. 22 (10) October 1983 pp 1553-7.

POL 83. C R Pollock, D A Jennings, F R Petersen, J S Wells, R E Drullinger, E C Beaty and K M Evenson, Optics Lett. 8 (3) March 1983 pp 133-135.

Text References

1. D A Jennings, K M Evenson and D J E Knight "Optical frequency measurements" Proc. IEEE 74 (1) January 1986 pp 168-179.

2. D J E Knight "A tabulation of absolute laser-frequency measurements" Metrologia 22 1986 pp 251-257.

3. P S Henry "Lightwave primer" IEEE J. Quant. Electron. QE-21 (12) December 1985 pp 1862-1878.

4. D A Humphreys and R A Lobbett "Investigation of an optoelectronic nonlinear effect in a GaInAs photodiode, and its application in a coherent optical communication system" IEE Proc. 135 Pt. J (1) February 1988 pp 45-51.

5. B Glance, P J Fitzgerald, K J Pollack, J Stone, C A Burrus, G Eisenstein and L W Stulz "Frequency stabilisation of FDM optical signals" Electronics Lett. 23 (14) 2 July 1987 pp 750-752.

6. T Yanagawa, S Saito, S Machida, Y Yamamoto and Y Noguchi "Frequency stabilisation of an InGaAsP distributed feedback laser to an NH_3 absorption line at 15137 Å with an external frequency modulator" Appl.

Phys. Lett. 47 (10) 15 November 1985 pp 1036-8.

7. T P Emmons, J M Reeves and E N Fortson "Parity nonconserving optical rotation in atomic lead" Phys. Rev. Lett. 51 (23) 5 December 1983 pp 2089-2092.

8. J C Camparo "The diode laser in atomic physics" Contemp. Phys. 26 (5) 1985 pp 443-477.

9. Ping Zhao, W Lichten, H P Layer and J C Bergquist "Remeasurement of the Rydberg constant" Phys. Rev. A 34 (6) December 1986 pp 5138-5141.

10. P B Coates "Use of gain-guided laser diodes for metrological purposes" J. Phys. E, in the press.

11. D J E Knight, K I Pharaoh and M J Padgett "Laser frequency measurement at NPL", IX Vavilov Conference, Novosibirsk, June 1987; Kvantovaya Elektronika, in the press.

12. M J Padgett, J A Golby, D D Arnone and R J Butcher "An ultra-high resolution offset-locked carbon dioxide laser spectrometer" preprint, February 1988.

13. R C Thompson, G P Barwood and P Gill "Progress towards an optical frequency standard based on ion traps" Appl. Phys. B 45 1988, in the press.

14. T Yanagawa, S Saito and Y Yamamoto "Frequency stabilisation of 1.5-μm InGaAsP distributed feedback laser to NH_3 absorbtion lines" Appl.Phys. Lett. 45 (8) 15 October 1984 pp 826-828.

15. K Uehara "Signal recording and averaging in diode laser spectroscopy" Optics Lett. 12 (2) February 1987 pp 81-3.

16. J H Park, L S Rothman, C P Rinsland, M A H Smith, D J Richardson and J C Larson "Atlas of absorption lines from 0 to 17900 cm^{-1}" NASA Reference Publication 1084 December 1981 pp 1-206.

17. K N Rao, C J Humphreys and D H Rank "Wavelength standards in the infrared" (Academic Press New York,London 1966).

18. T Hirschfeld and A Z Hed "The atlas of near infrared spectra" (Sadtler Res. Labs. Philadelphia, 1981).

19. K Fox "Coincidence of Er:YAG laser emission with methane absorption at 6077 cm^{-1}" Appl. Phys. Lett. 24 (1) 1 January 1974 pp 24-6.

20. M Outred "Tables of atomic spectral lines for the 1000 Å to 4000 Å region" J. Phys. Chem. Ref. Data 7 (1) 1978 pp 1-262.

21. V L Velichansky, A S Zibrov, V I Molochev, V V Nikitin, V A Sautenkov, D A Tyurikov and G G Kharisov "Asymmetry of some characteristics of tunable injection lasers with an external resonator" Kvant. Elektron. 8 September 1981 pp 1925-1934: English transl. : Sov. J. Quant. Electron. 11 (9) September 1981 pp 1165-1171.

22. R Wyatt and W J Devlin "10 kHz linewidth 1.5 μm InGaAsP external cavity laser with 55 nm tuning range" Electron. Lett. 19 1983 pp 110-112.

23. M DeLabachelerie and P Cerez "An 850 nm semiconductor laser tunable

over a 300 Å range" Optics Commun. 55 (3) 1 September 1985 pp 17408.

24. B Dahmani, L Hollberg and R Drullinger "Frequency stabilisation of semiconductor lasers by resonant optical feedback" Optics Lett. 12 (11) November 1987 pp 876-878.

25. K R German and C R Pollack "$(F_2^+)_H$ Color-center laser operation in NaCl:OH with simultaneous 1.06- and 0.53-μm pumping" Optics Lett. 12 (7) July 1987 pp 474-6.

26. A M Fox, A C Maciel and J F Ryan "Efficient cw performance of a Co:MgF2 laser operating at 1.5 - 2.0 μm" Optics Commun. 59 (2) 15 August 1986 pp 142-4.

27. K Otsuka, H Li and H R Telle "CW Nd-lasers with broad tuning range" Optics Commun. 63 (1) 15 July 1987 pp 57-60.

28. Wu Y, Guan K, Yu T and Tan F "Experiment on 1500 nm stabilised Zeeman laser" Electron. Lett. 23 (7) 26 March 1987 pp 318-9.

29. J W Eerkens, W Lee, R Smith and R S Vodhanel "Ammonia stabilisation of the 1.52 μm HeNe laser for IR fibre-optic applications" Proc. SPIE Int. Soc. Opt. Eng. (USA) Optical Fiber Sources and Detectors, Cannes, France, Nov. 1985, 587 1986 pp 8-10.

30. R L Barger and J L Hall "Pressure shift and braodening of methane line at 3.39 μm studied by laser-saturated molecular absorption" Phys. Rev. Lett. 22 1969 pp 2-4.

31. K M Evenson, M Inguscio and D A Jennings "Point contact diode at laser frequencies" J. Appl. Phys. 57 (3) 1 February 1985 pp 956-960.

32. G J Edwards and M Lawrence Opt. & Quant. Electron. 16 1984 pp 373-5.

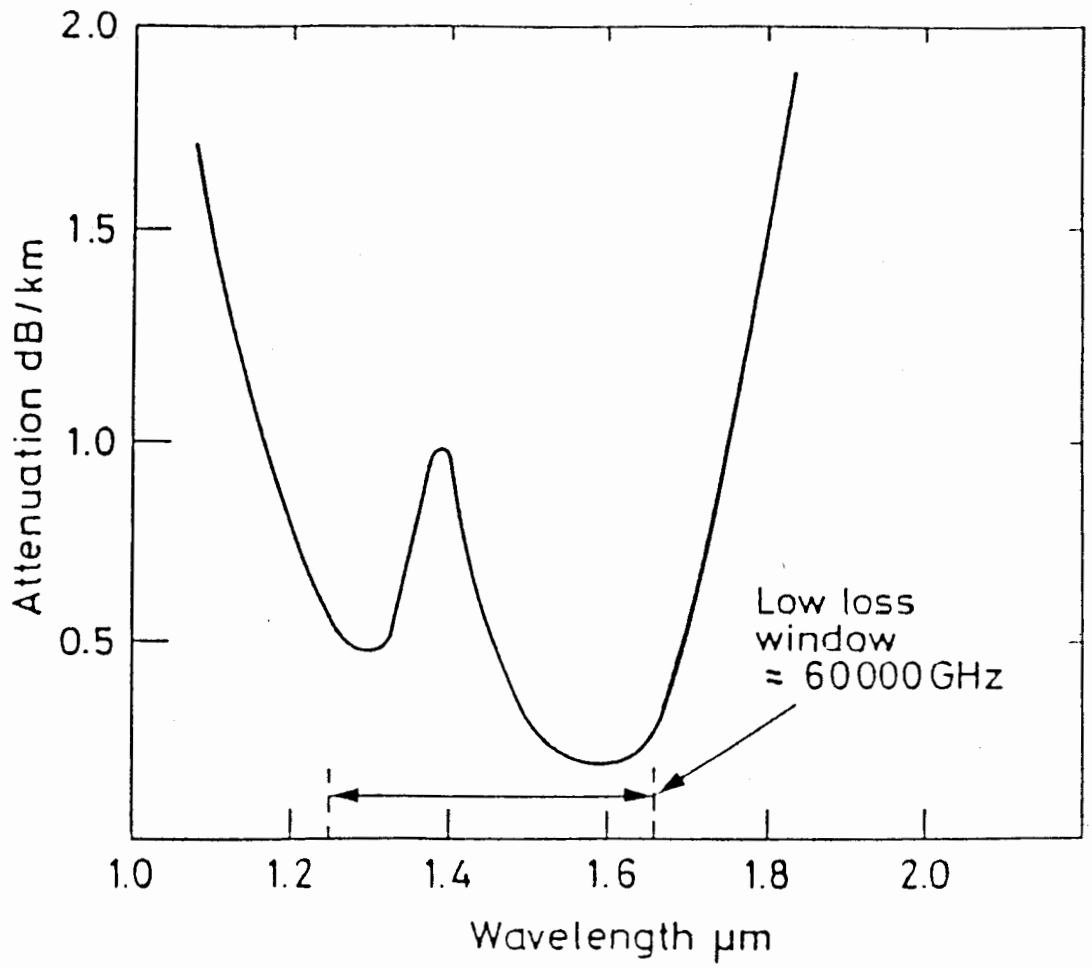
33. Y Miki and E Sakuma "Frequency mixing of 50 GHz microwave and 4.25 THz FIR light by a Josephson point contact" Jpn J. Appl. Phys. Pt.2 26 (9) September 1987 pp L1482-3.

34. J Moreland, L F Goodrich, J W Ekin, T E Capobianco, A F Clark, A I Braginski and A J Panson "Josephson effect above 77 K in a YBaCuO break junction" Appl. Phys. Lett. 15 (7) 17 Aug 1987 pp 540-1.

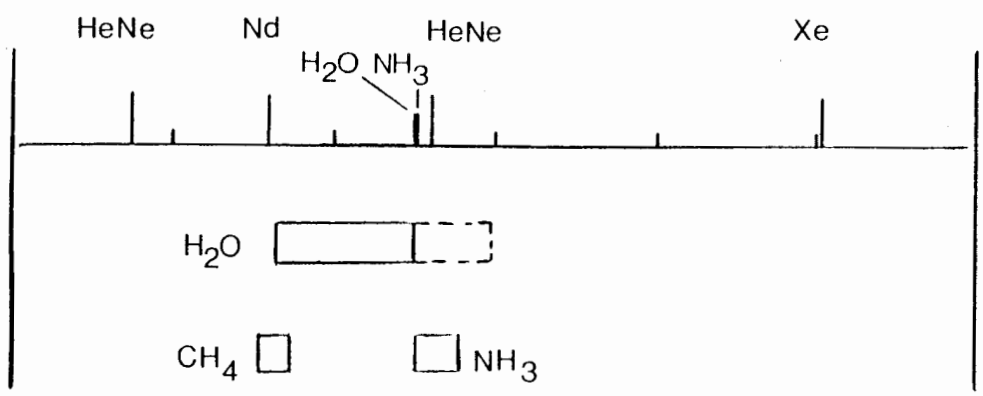
Table 1. Measured frequency references in and near the optical fibre bands.

SYSTEM molecule	laser	L/μm	MEASUREMENT f/MHz	df/MHz	ref	yr	METHOD mixing/notes	frequ.ref.used
CH4 (F)	HeNe	3.39	88 376 181.608	.004	BI	84	f mean of 6	Cs;1-sigmaREC
HF	HF	2.71	110 725 697.4	>7	ENG	75	f	CO
Ar	HeAr	2.39	125 053 923	10	GOL	85	f mim	CO2
Xe	HeXe	2.03	147 915 850	15	JEN	75	f mim	CH4,CO2
20Ne	HeNe	1.523	196 780 372	25	JEN	79	f mim	Xe(2μm),CO2
NH3	diode	1.504	199 396 142	147	OHT	83	I L-meter	633nm
NH3	diode	1.503	199 406 076	93	OHT	83	I L-meter	633nm
NH3	diode	1.503	199 416 382	133	OHT	83	I L-meter	633nm
NH3	diode	1.503	199 417 894	120	OHT	83	I L-meter	633nm
NH3	diode	1.503	199 432 487	67	OHT	83	I L-meter	633nm
NH3	diode	1.503	199 433 269	133	OHT	83	I L-meter	633nm
NH3	diode	1.503	199 435 870	120	OHT	83	I L-meter	633nm
NH3	diode	1.503	199 436 745	213	OHT	83	I L-meter	633nm
NH3	diode	1.503	199 437 395	173	OHT	83	I L-meter	633nm
NH3	diode	1.503	199 445 131	67	OHT	83	I L-meter	633nm
NH3	diode	1.503	199 446 750	107	OHT	83	I L-meter	633nm
NH3	diode	1.503	199 450 465	107	OHT	83	I L-meter	633nm
NH3	diode	1.503	199 452 084	120	OHT	83	I L-meter	633nm
NH3	diode	1.503	199 456 463	160	OHT	83	I L-meter	633nm
NH3	diode	1.503	199 459 024	133	OHT	83	I L-meter	633nm
NH3	diode	1.503	199 461 532	80	OHT	83	I L-meter	633nm
NH3	diode	1.499	200 021 923	107	OHT	83	I L-meter	633nm
NH3	diode	1.499	200 029 864	53	OHT	83	I L-meter	633nm
NH3	diode	1.499	200 033 855	53	OHT	83	I L-meter	633nm
NH3	diode	1.499	200 047 563	93	OHT	83	I L-meter	633nm
H20	diode	1.497	200 328 511	120	OHT	83	I L-meter	633nm
NH3	diode	1.496	200 354 364	107	OHT	83	I L-meter	633nm
20Ne LD	HeNe	1.153	260 103 249.260	.081	POL	83	f mim,xtal	CH4(F)[..609k]
Cs D2	diode	0.852	351 725 768.2	12	AV	86	I L-meter	633nm 2-mode
H20	diode	0.828	361 979 571	43	FU	84	I L-meter	633nm LD
H20	diode	0.828	362 124 685	20	FU	84	I L-meter	633nm LD
H20	diode	0.826	362 792 716	25	FU	84	I L-meter	633nm LD
H20	diode	0.823	364 093 801	44	FU	84	I L-meter	633nm LD
H20	diode	0.823	364 279 030	62	FU	84	I L-meter	633nm LD
H20	diode	0.823	364 322 977	39	FU	84	I L-meter	633nm LD
Rb D1 b	diode	0.795	377 106 004	13	BAR	87	I L-meter	633nm
Rb D1 c	diode	0.795	377 109 264	13	BAR	87	I L-meter	633nm
Rb D2 b	diode	0.780	384 229 173	13	BAR	87	I L-meter	633nm
Rb D2 c	diode	0.780	384 232 061	13	BAR	87	I L-meter	633nm
127I2(i)	HeNe	0.633	473 612 214.80	.16	BI	84	f,I(4)meanof5	CH4,CO2(1);REC
127I2(o)	HeNe	0.612	489 880 355.10	.18	BI	84	I mean of 5	0.633μm(i);REC
127I2(o)	dye	0.576	520 206 808.51	.11	BI	84	f,I mean of 2	CH4,1-sigmaREC
127I2(a3)	Ar ion	0.515	582 490 603.60	.25	BI	84	I mean of 4	0.633μm(i);REC

Symbols and abbreviations:- f: frequency, df: 1-sigma uncertainty in f,
I: interferometric, L: lambda,
LD: Lamb dip, REC: recommended value.



a



b

Figure 1. (a) Silica fibre transmission bands, after reference (4).
(b) Existing narrow-spectrum laser frequency references and laser-investigated spectra.

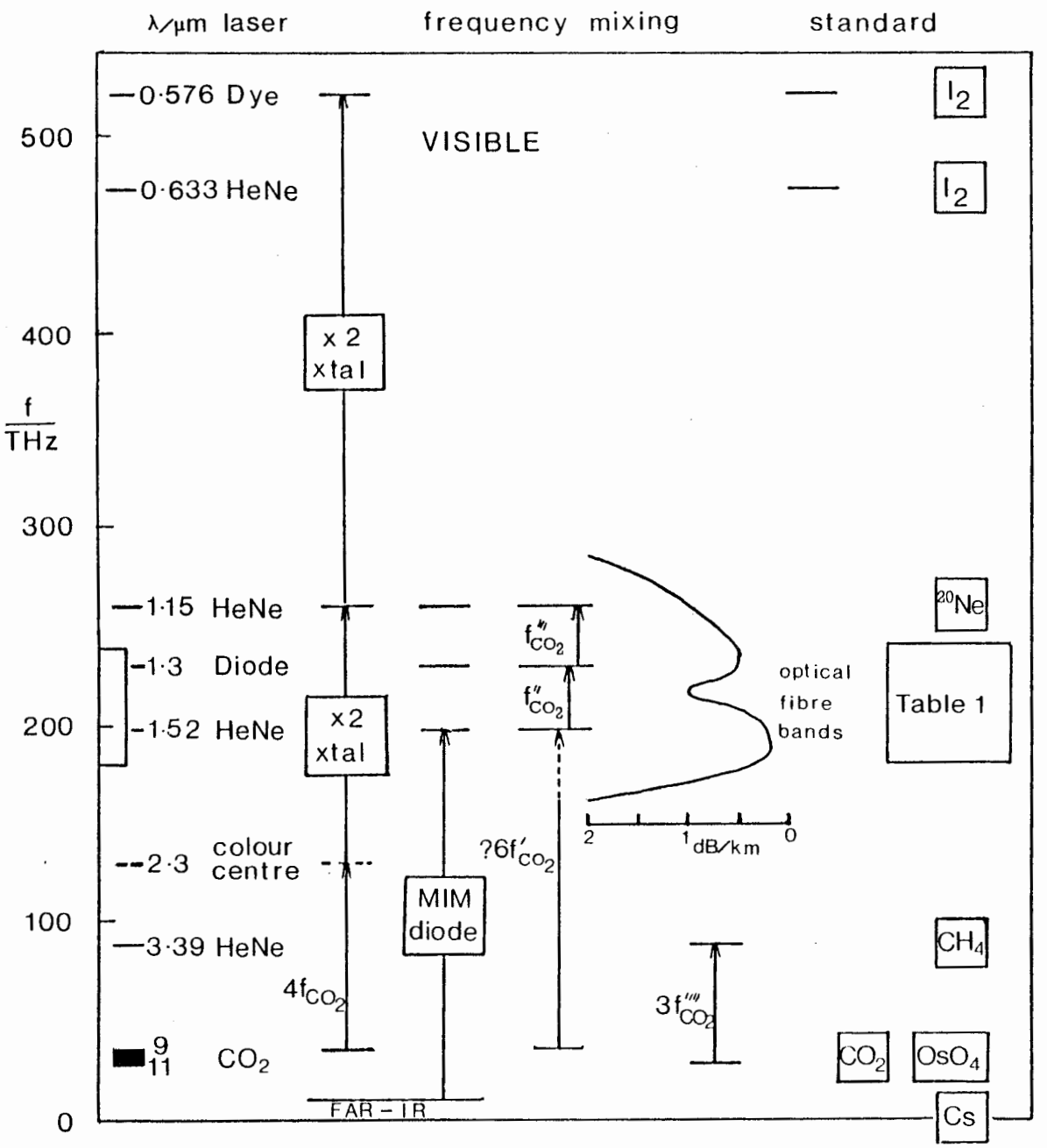


Figure 2. Relationships between frequency standards in the 0 -500 THz region.

2 ND EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 16 - 17 - 18 MARCH 1988

Session Invited Number 3

A review of ion trap frequency standards

G. Werth
Institut für Physik, Universität Mainz
D-6500 Mainz, Germany

Summary

Remarkable progress has been made in performance of ion trap frequency standards in the microwave domain. Experimental results about long-term frequency stability and precision are available from different laboratories. A $^{199}Hg^{+}$ ion trap operating with buffer gas cooling and a spectral lamp for optical pumping has reached a fractional frequency stability below $1 \cdot 10^{-14}$. Errors of systematic frequency shifts are below $1 \cdot 10^{-13}$ in an experiment on laser cooled $^{9}Be^{+}$ ions. They are mainly determined by residual second order Doppler shifts, which are expected to be substantially smaller in other ions like Ba^{+} or Yb^{+}. First operational experience of prototype ion trap standards is remarkably good and a substantial impact on long-term timekeeping can be expected.

Introduction

Substantial progress has been made in recent years since early proposals for frequency standards based in ion traps. These proposals cover the microwave frequency domain as in presently available atomic frequency standards like the Cs- or Rb-clock /1,2/ as well as the optical domain /3/. In both cases long-term frequency instabilities and irreproducibilities as low as 10^{-15} have been envisaged. Since the main obstacle to achieve this goal, the second order Doppler shift due to the fairly large ion velocities in traps, has been overcome by laser cooling a few years ago /4,5/, some experiments have been performed to determine the final uncertainty and the fractional frequency stability, which can actually be obtained in microwave standards. Moreover, first steps have been made towards the stabilization of a laser to a narrow transition of a trapped ion, which finally may lead to a frequency standard in the optical domain. In both cases there may be a trade off between a standard based on a single trapped ion, which offers the highest expectations with respect to stability and reproducibility, and a large ion cloud with better signal-to-noise ratio. For practical purposes it is of interest the degree of experimental complexity, and again one might have to compromise between very good performance, which could be difficult to maintain for a long measuring time, and simpler and more reliable set-ups at the expense of figures of merit.

Candidate ions

Optical-microwave double resonance techniques similar to those applied in Rb-cell clocks have been used in all experiments on trapped ion microwave frequency standards. Consequently an essential requirement for an ion to be considered as a possible candidate is a resonance transition which can easily be excited by lamps or preferentially by c.w. lasers, if laser cooling shall be applied. Moreover, the ground state hyperfine separation (hfs), which serves as clock transition,

So far the experiments have concentrated on isotopes of the elements Ba, Yb and Hg, whose main characteristics are listed in Table 1. The relevant parts of the energy level schemes are shown in Fig. 1,a,b,c. Ba^+ offers a wavelength which is most easily produced by c.w. lasers, but has the problem of very long-living metastable states. This requires a second laser to repump those ions which decay into these states after laser excitation of the resonance transition. It has been shown, however, that the associated difficulties can easily be overcome /5/. Production by surface ionisation is easy because of the low ionisation potential and separate isotopes are available in sufficient quantities. Disadvantages of the odd stable Ba^+ isotopes are a somewhat smaller hfs compared to other candidates and a nuclear spin of 3/2.

Most interesting from a principle point of view are the Hg^+ isotopes, in particular $^{199}Hg^+$ because of its large hfs of 40.5 GHz and the nuclear spin of 1/2. No problem arises from metastable states. The main obstacle is the pumping wavelength of 194 nm which at present cannot be produced on a routine basis by c.w. lasers. It has been obtained by frequency mixing of dye and Ar^+ lasers by a group at NBS, Boulder /6/ with sufficient output power to excite and cool the ions. It seems, however, difficult to imagine a continuous operation of this system as a part of a frequency standard. As shown below a good fractional frequency stability might be obtained without laser cooling of the ions. In this case Hg^+ has the special feature that the resonance line of $^{202}Hg^+$ at 194 nm coincides with one of the hf-components of the corresponding line in $^{199}Hg^+$ and thus might be used for optical pumping /7/. Although the practical difficulties in driving a lamp with sufficient output power at this wavelength are not easy to overcome, a breakthrough has recently been reported by a group at HP company which operated a r.f. excited $^{202}Hg^+$ lamp in an ion trap experiment continuously for several months /8/. In addition the problem of a large stray light background from the lamp was sufficiently reduced by a carefully designed detection optic.

should be as high as possible, and the nuclear spin small to simplify the hf-Zeeman spectrum. Less important but not completely negligible are questions of isotopic abundance, vapor pressure or ionisation potential of the parent atom.

Ion	Hyperfine splitting (GHz)	Nuclear spin	Pumping and cooling wavelength (nm)	Branching ratio $\frac{P\text{-}D}{P\text{-}S}$
$^{199}Hg^+$	40.5	1/2	193	10^{-7}
$^{201}Hg^+$	30.0	3/2	193	10^{-7}
$^{171}Yb^+$	12.6	1/2	370	1/290
$^{173}Yb^+$	10.5	5/2	370	1/290
$^{137}Ba^+$	8.0	3/2	493	1/3
$^{135}Ba^+$	7.2	3/2	493	1/3

Table 1 Candidates for microwave ion-trap frequency standards

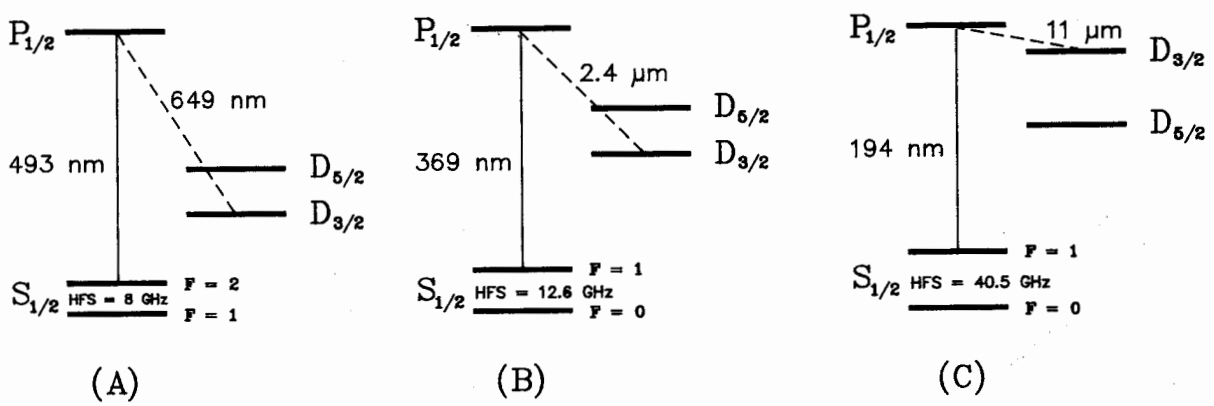


Fig. 1 Partial energy level diagram for (A) $^{137}Ba^+$, (B) $^{171}Yb^+$, (C) $^{199}Hg^+$. The radiative lifetimes for the metastable $D_{3/2}$ states are 17 s (Ba^+), 52 ms (Yb^+), and 90 ms (Hg^+)

Yb^+ represents in many ways a compromise between Ba^+ and Hg^+: The size of the hfs is in between the values of the other candidates and so is the wavelength of the resonance transition which can be produced by frequency doubled c.w. lasers. The problem of pumping the ions into metastable states is less severe as in Ba^+ because of the small branching ratio, moreover the ions can be pumped back by a diode laser. In $^{171}Yb^+$ the nuclear spin in 1/2, and an experiment on this has produced the highest line-Q ($\nu/\delta\nu$) of $3 \cdot 10^{11}$ /9/.

Apart from the ions mentioned above a few other candidates have been condidered by different authors. Ions like Mg^+, Be^+, Zn^+ and Cd^+ offer similar pumping and detection schemes as in the other cases /10/. In fact experiments on $^{25}Mg^+$ and $^{9}Be^+$ have been performed by the NBS group /11,12/. Although the precision obtained is remarkable, they may not be considered as a final choice of a clock ion because of their low hfs of a few hundred MHz, but serve as test ions since the wavelength for optical pumping and cooling are rather easily obtainable. Strumia proposes $^{205}Pb^+$ /13/, which has a ground state hfs of 14.2 GHz and a fine structure resonance transition at 710 nm. The lifetime of the excited state, however, is 38.8 ms /14/ and thus too long for efficient laser cooling.

Apart from stable isotopes one might consider also radioactive ones. The isotope $^{133}Ba^+$ has of course the same optical properties as the stable isotopes, but a nuclear spin of only 1/2 and a somewhat larger hfs of 9.1 GHz. The lifetime is 11 years and radioactive decay does not represent a practical problem. It can be produced by high energy reactions at accelerators in sufficient quantities for long term operation. In a recent experiment a precise hfs value has been obtained and estimates of the required number of atoms for frequency standard applications have been given /15/.

Experiments

Because of its most favourable properties, as mentioned above, most attention as a possible frequency standard has received $^{199}Hg^{+}$ /16-19, 8/. For all these experiments a $^{202}Hg^{+}$ spectral lamp for optical pumping has been used. A slight reduction of the ion energy was achieved by collisions with He buffer gas at about 10^{-5} mbar, where the hf-pressure shift is of no importance. The group at Laboratoire de l'Horloge Atomique, Orsay, reports a fractional frequency stability of $3.6 \cdot 10^{-11} \cdot \tau^{-1/2}$ for measuring intervals between 10 s and 3500 s /18/. Similar results have been obtained at the University of Mainz /20/. Most recently a group at Hewlett Packard has operated a $^{199}Hg^{+}$ trap almost continuously over a period of about one year. The $^{199}Hg^{+}$ hfs was measured with respect to two hydrogen masers for 23 hours per day, while one hour was used to monitor the signal and background photon counts and other parameters. During the measurements a pulsed mode was employed with a cycle time of 2.5 s. It starts with pulsing an electron gun and optical pumping to attain the initial population distribution. The 40.5 GHz microwave interrogation signal is then brought to a position on one side of the resonance line and the pumping light is turned off to avoid light shift. After 1.0 s the microwave is moved off the line and the pumping light is turned on again. The photon counter is then activated for 0.5 s to measure the ion population. The cycle is then repeated on the other side of the resonance line, and the frequency error is determined from successive photon counts on alternating sides of the resonance line, taking into account possible drifts in the background light level from lamp degradation or ion loss. Operating in a residual magnetic field of 100 nT and temperature stabilized to 0.01 °C the group reports standard deviations of 21 minutes averages between 35 and $70 \cdot 10^{-15}$, strongly correlated with the signal-to-noise ratio. A comparison to two of the USNO H-masers (M18, M19) yields an average daily frequency scatter of $7.6 \cdot 10^{-15}$ for interrogation times of one day from a total observation time of a 115 days.

One should point out that this remarkable result has been obtained without laser cooling of the ions. The ions temperature is estimated from a model suggested by Knight /21/, which assumes that the ions macromotion in the pseudopotential well created by the r.f. trapping field is damped by collisions with light background atoms to a value corresponding to the background gas temperature, while the r.f. induced micromotion is always present and its amplitude depends on the ion's position in the trap and the trapping parameters. For a spherical cloud of 50 khz macromotion frequency the relation between the ion cloud temperature T and the ion number N is given by

$$T = 0.436 \cdot N^{2/3}$$

For $N=2 \cdot 10^6$ as estimated in the reported experiment we have T=6000K. This corresponds to a second order Doppler shift of $2 \cdot 10^{-12}$. In order to obtain the reported frequency variance the temperature has to be stable to better than 1 part in 100.

Of substantial practical interest is the fact that the r.f. excited $^{202}Hg^+$ spectral lamp, which has to be driven at high temperatures to produce sufficient output of the ionic spectral line at 194 nm, can be operated for a period of several months. A slow decay is observed, but it is gradual and replacement can conveniently be planned in advance.

Only a very rough estimate is given for the precision of the mercury trap clock. A second device has been put to operation, which differs significantly in construction from the first one and preliminary measurements show a frequency difference which is between 1 and $2 \cdot 10^{-13}$.

A more detailed discussion of absolut frequency errors, which might be obtainable is given in an experiment on $^9Be^+$, performed at NBS, Boulder /22,12/. A cloud of a few hundred to 2000 ions was stored in a static Penning trap. Although the presence of a strong magnetic field as necessary in Penning traps seem prohibitive for frequency standard applications, there is the possibility to choose values of the field, where

the transition frequency between two hf-Zeeman levels is independent of magnetic field strength fluctuations in first order. The advantage might be that because of the absence of r.f. heating, which occurs in radio-frequency traps, the final temperature of a laser cooled ion cloud is smaller. In case of $^{9}Be^{+}$ a "field independent" transition occurs at B=0.82 T at a frequency of 303 MHz. The ion cloud has been cooled by laser irradiation. Although the final ion cloud temperature can be brought to values below 1 mk, in a practical device, operated as frequency standard, the cooling laser has to be turned off during the microwave interrogation time to avoid light shift and some heating occurs. The temperature has been measured and rises up to 30 K for interrogation times of 20 s. The fractional frequency stability has been measured to vary between $1.3 \cdot 10^{-11} \cdot \tau^{-1/2}$ and $4 \cdot 10^{-11} \cdot \tau^{-1/2}$ for total cycling times of 20 s, while the measuring time was extended to 2 hours. Of interest is the estimate of different contributions to the systematic error as given in Table 2.

Systematic Effect	size of effect	Uncertainty
Second-order Doppler shift	$-3.8 \cdot 10^{-14}$	$9 \cdot 10^{-14}$
Magnetic field fluctuations	$-1.3 \cdot 10^{-14}$	$2.5 \cdot 10^{-14}$
Rotation of earth	$-1.8 \cdot 10^{-14}$	$0.2 \cdot 10^{-14}$
Servo offsets	$<5 \cdot 10^{-15}$	
Background slopes	$<5 \cdot 10^{-15}$	
Blackbody radiation drift	$-3 \cdot 10^{-16}$	$3 \cdot 10^{-16}$
Background gas collision	$<1 \cdot 10^{-15}$	
Stark shift	$<1 \cdot 10^{-15}$	
First-order Doppler shift	$<1 \cdot 10^{-15}$	
ac Zeeman shift	$<1 \cdot 10^{-15}$	
Total systematic offset	$-41.1 \cdot 10^{-14}$	$9.4 \cdot 10^{-14}$

Table 2 Estimated systematic errors in laser cooled $^{9}Be^{+}$ clock. The numbers are given as fraction of the 303 MHz transition frequency (from ref.12)

The most important one is the second order shift which contributes to 90% of the total uncertainty of $9.4 \cdot 10^{-14}$. It may possibly be reduced at smaller ion clouds, if ion-ion collicion play a dominant role in the heating process which is at present not completely understood.

Conclusion

Remarkable progress has been made in recent years towards the realization of ion trap frequency standards. Even without laser cooling long-term fractional frequency stabilities have been obtained which are superior to Cs-beam standards and are approaching the level of hydrogen masers. Errors of systematic frequency shifts have been determined in the case of a $^{9}Be^{+}$ clock to be less than $1 \cdot 10^{-13}$. Considering the fact that this is determined almost completely by second order Doppler shifts; a further decrease of this uncertainty can be expected for ions which have clock transition of 10 GHz or above in contrast to the low 300 MHz transition in $^{9}Be^{+}$. At present experiments in different frequency standard laboratories are under way which concentrate on $^{171}Yb^{+}$.

References

/1/ F.G. Major: NASA Report X-52-69-167 (1969)
/2/ H.A. Schuessler: NSF Report PQ 19 166-000 (1969)
/3/ H.G. Dehmelt: Bull.Am.Phys.Soc. 18, 1521 (1973)
/4/ D.J. Wineland, R.E. Drullinger and F.L. Walls, Phys. Rev.Lett. 40, 1639 (1988)
/5/ W. Neuhauser, M. Hohenstatt, P. Toschek and H. Dehmelt, Phys.Rev.Lett. 41, 233 (1978)
/6/ H. Hemmati, J.C. Bergquist and W.M.Itano, Opt.Lett. 8, 73 (1983)
/7/ F.G. Major and G. Werth, Phys.Rev.Lett. 30, 1155 (1973)
/8/ L.S. Cutler, R.P. Giffard, P.J. Wheeler and G.M. Winkler, 4th Ann.Frequ.Contr.Symp., Philadelphia, 1987
/9/ R. Blatt, H. Schnatz and G. Werth, Z.Phys. A312, 143 (1983)
/10/ J.L. Hall, Science 202, 147 (1978)
/11/ W.M. Itano and D.J. Wineland, Phys.Rev. A24, 1364 (1984)
/12/ J.J. Bollinger, W.R. Itano and D.J. Wineland, Proc.37th Ann.Frequ.Contr.Symp. Philadelphia (1983)

/13/ F. Strumia, Proc. 32nd Ann.Frequ.Contr.Symp., Philadelphia (1978)
/14/ Ch. Gerz, A. Roths, D. Wilsdorf and G. Werth, to be published
/15/ H. Knab, M. Schupp and G. Werth, Europhys.Lett. 4, 1361 (1987)
/16/ F.G. Major and G. Werth, Appl.Phys. 15, 201 (1978)
/17/ M.D. McGuire, R. Petsch and G. Werth, Phys.Rev. A17, 1989 (1978)
/18/ M. Jardino, M. Desaintfuscien, R. Barillet, J. Vrennet, P. Petit and C. Audoin, Appl.Phys. 24, 107 (1981)
/19/ L.S. Cutler, R.P. Giffard and M.D. McGuire, Proc. 37th Ann.Frequ.Contr.Symp., Philadelphia (1983)
/20/ G. Werth, Proc. 10th Int.Congr. of Chromometry, Geneva (1979)
/21/ R.D. Knight, Ph.D. Thesis, Univ.of California, Berkeley (1979);
Lawrence Berkeley Lab Publication LBL 9082
/22/ D.D. Bollinger, J.D. Prestage, W.M. Itano and D.J. Wineland Phys.Rev.Lett. 54, 1000 (1985)

Session Invited Number 4

Time and Frequency : Accuracy for GPS

Dr. A. R. Pratt,
Science Director,
Polytechnic Electronics PLC,
Royal Oak Way,
DAVENTRY, Northants.,
NN11 5PJ, England.

The advent of GPS has posed new challenges for the frequency and time industry. Solutions to problems both old and new have been induced by the development of GPS receivers and associated systems. The needs of navigation receivers especially GPS (but also Transit NNSS) in respect of frequency and time accuracy are established. Practical solutions to the receiver requirements are described. One section of the paper describes the consequences of imperfect time and frequency input - which lead to increased search time when seeking satellite signals, increased time re-acquiring a faded satellite signal or poor velocity estimation.

The final section of the paper addresses the problem of vibration on frequency sources and the effects of this type of instability on GPS signals. The results of system simulations are presented for this case.

Session Number

1. GPS Introduction

The Global Positioning System (GPS) relies, for its operation, on a receiver, located in a host vehicle, aquiring signals from up to four satellites on frequencies in L-band. There are two main frequencies in use L1 (1575.42 MHz) and L2 (1227.60 MHz) although it seems that only L1 will be finally available to the civilian community. The carriers are BPSK modulated by sequences which have good cross-correlation properties. One is a short 1mS sequence with a chipping rate of 1.023 MHz called the C/A code. This is available on L1 and can be transmitted on L2 but this is not expected to be a normal feature. The so-called P-code is a much longer sequence lasting 38 weeks at a chipping rate of 10.23 MHz. Separate one week sections of this sequence are allocated to each satellite. The sequence in each satellite is restarted at Saturday midnight (GPS time). The P-code is transmitted on both L1 and L2 and in phase-quadrature to the C/A code. The codes contain embedded timing marks (e.g. the epoch instant of the C/A code) which enable the user with accurate knowledge of GPS time to establish the time delay imposed on the signals in transmission from the satellite to the user.

The measurements are made by generating a replica of the code in use (either C/A or P for the chosen satellite) and then aligning the generated with the code signal received from the satellite. There are many references available which describe the mechanisation of this process in detail (for example, GPS Signal Structure and Performance Characteristics: J. J. Spilker, Vol. I, GPS Papers reprinted by The Institute of Navigation, Washington, U.S.A.). In a few words, the process is similar to that of a phase lock loop but instead of deriving an error signal from the phase offset between local oscillator and received carrier, the delay lock loop (from tracking psedo-noise signals like the C/A code) derives a time-error signal from a comparison of two cross-correlation estimates of locally generated code and received code from the satellite. The two locally generated codes are distinguished by a small time delay of the order of one chip (977 nS for C/A code). This will enable the delay lock loop to track the peak of the cross-correlation estimate. The epochs of the locally generated sequence represents the time at which the GPS signals are received. Transit time, and hence, distance from the user to the satellite may then be established by the difference between GPS time and the signal received

time. Three such measurements thus enable the user to establish his location by triangulation from known satellite locations. The satellite locations are transmitted by an ephemeris message at 50 bps also BPSK modulated on each satellite signal.

A consideration of measurement error budgets for the GPS system (see figure 1 for example) suggests that in order for a user to make best use of the accuracy potential of GPS, GPS time must be known to within a few nS (perhaps 10 nS maximum). Even if a user had knowledge of GPS time to this accuracy at a certain instant in time, it would quickly be lost (within a few hours or days) even with the best atomic standards. It is normal, therefore, for a user to make measurements to a fourth satellite by which GPS time can be established with the required accuracy.

Position solutions with less than four satellites are possible but only with input of additional information - e.g. input of height above the reference spheroid on which the users position is to be established - will enable a three satellite solution. In addition, input of precise time (to height data) will allow a 2-satellite solution. This is practical because the user has the possibility of aligning a local atomic standard to GPS time in periods of three or four satellite coverage. The atomic clock will then represent GPS time with reasonable accuracy for some hours during periods of reduced satellite coverage.

2. The effects of Frequency and Time Uncertainty

Frequency uncertainty in GPS receivers arise primarily from three sources:

(a) Satellite Doppler
(b) User Clock Frequency Offset
(c) User Velocity

A minor source of uncertainty is the frequency of the satellite atomic or quartz frequency standard.

Lack of precise knowledge of the frequency setting to which a user's receiver must be adjusted to obtain GPS satellite signals causes difficulties in obtaining the code and data messages. In fact, a feature of BPSK modulation is that all the transmitted energy goes into information bearing sidebands and there is no energy at the carrier frequency at all. It is possible,

therefore, to detect the presence of a satellite only through a correlation detector using the correct C/A or P-code reference which is approximately aligned with the received C/A or P-codes (within 977 nS or 97.7 nS respectively). The output of such a correlation detector must be measured during a 20 mS time period or less since the 50 bps data signal caused phase reversal of any correlation estimate with 50% probability every 20 mS.

The worst case search situation occurs when the user has no knowledge of his location, velocity or time. The first consequence of this is that the user does not know which satellites, if any, are above the local horizon. It is therefore necessary to search a range of frequencies covering the maximum uncertainty due to (a), (b) and (c) above, as well as every C/A code phase (1023) for every satellite which might be emitting usable signals. The P-code signal could not be used due to its 1 week repetition time.

In order to establish the orders of magnitude of the search time issue, we will make some assumptions on user velocity and on various structures for the code correlator.

(a) The satellite orbital velocity induces by the Doppler effect a frequency uncertainty of +/- 3.75 KHz depending on the location of the satellite relative to the user. For example, a satellite just rising above the horizon and going to pass through the user's zenith will have a Doppler offset near +3.75 KHz.

(b) The user will have some frequency offset in the receiver reference oscillator. In most cases, this will be a quartz device and perhaps some years old and and possibly without any frequency adjustment since new. Such an oscillator might easily be 1 ppm - off-frequency. This error corresponds to nearly 1.6 KHz at L1. The effect of this error might be reduced in a sophisticated receiver by a suitable combination of software and hardware resources, and recent previous exposure to a GPS constellation. If the receiver had been switched off since last exposure to GPS, there will be an additional oscillator retrace error.

(c) The host vehicle containing the GPS receiver may be moving and in most cases will be subsonic (less

than 1000 km/hr). Such vehicles will have maximum Doppler offsets of +/- 1.5 KHz approximately. A small range of host vehicles (e.g. Supersonic aircraft, missiles, space vehicles) will have larger relative velocities and thus Doppler offset - but the same general principles will apply to these.

It is, therefore, reasonable to bound the frequency uncertainty by +/- 7.5 KHz with the exception of those specialised host vehicle classes above. The search strategy in frequency is to establish the presences of energy in a set of frequency bands or bins which is significantly larger than the noise level. The methods of statistical hypothesis testing enable probabilities of detection and false alarm rates to be established given data on the wanted signal to unwanted noise ratio.

The search for alignment between the locally generated replica of the C/A code and the version received from the satellite depends upon the sophistication of the GPS receiver. The simplest receiver will search 1 chip at a time spending between 1 and 10 mS accumulating energy in the correlator before making the signal present/absent decision. The range of frequency errors which the process is able to accept without significant attenuation is the reciprocal of the accumulation time (i.e. 1 KHz for 1 mS correlation and 100 Hz for 10 mS correlation). However, the shorter correlation intervals give poor detection characteristics (probability of detection at or below 50% with a false alarm probability at or above 10 power 3 for normal GPS signal levels).

The search time calculations for a 1 mS correlator are:-

Accumulation time	:	1 mS = T
Frequency Bins size	:	+/- 5000Hz = +/- f/2
No of frequency Bins	:	15 for a 15 kHz (F) frequency uncertainty
No of correlation Bins	:	1023 = N
Search Time	:	$= T \times \frac{F}{f} . N$
		= 15.345 seconds

In order to improve the detection characteristics, it would be good practice to repeat the correlation measurement in each code and frequency bin. A 6 dB

equivalent signal to noise ratio improvement results from a 4 times application of the search. This is sufficient to convert poor detection characteristics to accpetable ones. The search time in this case is therefore about 60 seconds for each satellite. When the full constellation of 18 satellites and 3 active spares is deployed, it may be expected that the maximum search time for a chosen satellite is 21 x 60 seconds - about 20 minutes by this method. However, the average time to find the first satellite is about 5 minutes.

A second calculation for a 40 mS correlation interval split into 4 groups of 10 mS each gives a much longer search period because the number of frequency bins is much increased.

$$\text{Search time} = T \cdot \frac{F}{f} \cdot N$$

$$= 40 \times 10^{-3} \times \frac{15 \times 10^{3}}{100} \times 1023$$

$$= 102.3 \text{ minutes}$$

This is unacceptably long on a basis of both use and the fact that the Doppler offset of the satellite would have changed significantly during the search period.

One solution to such long search periods to be embodied in the use of multi-tap correlators. These can be mechanised directly (a possibility with the use of large gate errays) or indirectly (for example, using the FFT process to transform into an orthogonal domain such as code phase). An example is a 16 mS correlator split into 4 groups of 4 mS each with a +/- 125 Hz frequency acceptance band:

$$\text{Search time} = T \cdot \frac{F}{f} \cdot N$$

$$= 16 \times 10 -3 \times \frac{15 \times 10^{3}}{250} \times 1$$

$$= 0.96 \text{ seconds}$$

The probability of detection and false alarm rate remains the same in this case as for previous ones.

The search senario remains unchanged even with

knowledge of position if the time is unknown or if the time is known but position is not. An important change in strategy occurs when both time and position are known approximately. For example a position uncertainty of 300 km and a time uncertainty of 10 mins will allow a user equipped with satellite almanac data to determine the visible satellite constellation, their location and Doppler frequency offset. The position uncertainty of 300 km is equivalent to a time uncertainty of about 1 minute (at satellite orbital velocity) so it is the fine uncertainty which is more important. Over a 10 minute period, a satellite's Doppler offset could change as much as 360 Hz for a satellite near the user's zenith. A user's speed uncertainty of up to 100 km/hr will generate up to +/- 140 Hz additional Doppler offset.

These constraints would require the use of a single frequency bin search using a 4 mS correlation interval split into 4 separate sections. A single tap correlation would have to search for a maximum of

$$T \cdot \frac{F'}{f} \cdot N$$

$$= 4 \times 10^{-3} \quad \frac{1000}{1000} \quad 1023$$

$$= 4.092 \text{ seconds}$$

in order to find any satellite in the visible constellation. Similarly a 16 mS, 1023 tap correlation (sectioned into 4 x 4 mS accumulation intervals) would require:

$$T \cdot \frac{F'}{f} \cdot N = 16 \times 10^{-3} \times \frac{1000}{250} \times 1$$

$$= 64 \text{ mS}$$

In this case 4 frequency bins will have been searched.

As soon as the position and time of the user are known to better than 300 km and 1 mS respectively - that is smaller than the C/A code distance in space or repetition time - then two effects occur. Firstly, the frequency search requirement is no longer required unless the user's speed uncertainty is large - 30 km/hr for a 10 mS accumulation period in the correlator.

Secondly, only a portion of the C/A sequence appropriate to the time and space uncertainty may require to be searched.

When the position and time uncertainty together are smaller than 1 mS (or 300 m), then direct lock to the C/A code is possible because the user's receiver is able to estimate its distance to the chosen GPS with less than 1 chip error. Similarly, if the position and time uncertainty are known to less than 100 ns (or 30 m), then direct lock to the P-sequence is possible because there is less than 1 chip error in the user-satellite distance. The means by which such knowledge of the user's position may be obtained are too detailed to discuss in this paper and are briefly mentioned in the figures.

3. Determination of User Velocity from GPS Signals

GPS signals may be used to determine the velocity of the host vehicle by one of two means. The first measures the rate of change of position. The method is shown in the figures. The pseudo-range vector at time i is Pi. The geometrical transformation between the satellite to user unit vectors (skew co-ordinate system) and the user's location co-ordinate system (e.g. ECEF or lat/long which are orthagonal co-ordinate frames) is H.

Hence the user's velocity vector Vn is :

$$Vn = H. \frac{\Delta Pi}{T}$$

In practice this method produces relatively noisey results. The solutions are noisiest when the measurements are made over an interval DT which is equal to or less than the de-correlation time of the Psuedo-range measurements. This time may be as long as 4 seconds in receivers with code-loop aiding and maybe as short as $\frac{1}{2}$ - 1 second in those which are not and have 2nd-order code trackers. Clearly such a system is not very responsive to rapid velocity changes.

A second and more accurate method of velocity determination is available by use of the rate of change of carrier phase (or Doppler offset frequency). As an observation this is considerably less noisey than the psuedo-range measurements with measurement noise on the order of a 4-5 mm in comparison with 1.5 - 2 metres for

psuedo-ranges. The tracking of carrier phase is more or less common in modern GPS receiver architectures. The measurement noise in the Costas (carrier) tracking loop will de-correlate in 40 to 150 mS depending upon detailed design considerations. Such a system is thus capable of being more responsive to rapid velocity changes. The method of calculation is similar to that for pseudo-ranges. A vector of host vehicle induced Doppler offset frequencies, Di, at time i is derived from the carrier tracking loops. The velocity vector is therefore:

$$Vn = k\,H.\,\Delta i$$

where H is the same transformation matrix as above and k relates velocity to frequency Viz:

$$k = c/f0 = \quad \text{(for L1 or L2 as need)}$$

c is the velocity of light.

The determination of Δi is a little more complex and is set out in detail in detail in the figure. The model used does not include the effects of atmospheric noise (which is easy to account for in the error analysis) but does include the effects of local oscillator instability. Instinctively, one would expect local oscillator instability to be important because of its role as the reference against which these small Doppler offsets are measured. The equations show that the fractional oscillator instability (Allan Variance) propogates directly into the velocity estimate with a multiplex having the value of the speed and light. Hence, the use of a local reference oscillator with an Allan Variance of 1 part in 10 power 11 implies a velocity measurement noise floor of 3 mm/sec. The level of Allan Variance may not be too difficult to achieve under static conditions with an ovenised quartz oscillator.

The drive to low cost GPS receivers - for example, 300 US Dollar equipments are expected by 1995 - implies that non-ovenised low cost oscillators will be needed. One casuality of the low cost GPS designs may be good velocity performance. One market motivation for low cost GPS receivers is their potential for use in land vehicles (for example, cars, public transport, goods vehicles, security vehicles, etc.). This category implies high vibration levels with $1\frac{1}{2}$ g at low frequencies and up to 5 g at the higher frequencies

enveloping the vehicles characteristics.

4. Performance of Quartz Oscillator in Vibrating GPS Receivers

The land vehicle use above motivated a study of the effects of vibration on a quartz oscillator and the consequence for the GPS carrier tracking loops. The results of simulations are presented below. Additional data has been published - the interested reader is refered, for example, to "Phase Noise Under Vibration in Crystal Oscillators" by Glen R. Kurzenknabe of the Pierzo Crystal Company in Carlisle, Pa, U.S.A.

A third-order carrier tracking loop as shown in the figures was set up with suitable multipliers K1, K3 and K4 to give a damped response to transients. A typical transient response obtained to a frequency step is shown.

Noise and vibration at a critical response frequency, added to the tracking loop as shown and the performance of the system established by measuring the probability with which carrier cycle slips occured. The results of three simulations are shown - one without any regular vibration and two with $\frac{1}{2}$ radian and $\frac{1}{4}$ radian peak carrier phase error due to the vibration. It can be easily seen that the vibration leads to important increases in the probability of cycle slips. The numbers chosen would be typical for a quartz oscillator with a g-sensitivity of around 7 pts in 10 power 10 per g and vibration level of about $1\frac{1}{2}$g at 5 Hz - typical for some land vehicles. The loss in performance is equivalent to 1 to 2 dB of signal to noise ratio in the received signal.

5. Summary

The paper has covered three main areas by which GPS performance is affected by imperfect local frequency and time references. These are GPS signal search time, velocity estimation using GPS, rate of cycle slips in the carrier tracking loops due to vibration on the (usually) quartz oscillator.

These are real problems which demand rather good and certainly expensive oscillators for good performance. At a time when the market pressure is for reducing the cost of ownership of GPS receivers for all classes of user, innovative solutions are needed to meet the oscillator stability requirements but without the excessive cost penalty which present technology demands.

Session Number

DETERMINATION OF Di

- DOPPLER OFFSET FREQUENCY VECTOR

  RECEIVED FREQUENCY = $fo\left(1 \pm \frac{Vj}{C}\right)$ FROM j th SATELLITE

  DOPPLER OFFSET, $fj = fo\left(1 \pm \frac{Vj}{C}\right) - Nfr\ (1+a)$

  DOPPLER OFFSET COUNT

  $fj \triangle t = \triangle tfo\left(1 \pm \frac{Vj}{C}\right) - Nfr \triangle t\ (1+a)$

fo : TRANSMITTED FREQUENCY FROM SATELLITE

L1 : 1575.42×10^6 Hz

L2 : 1227.6×10^6 Hz

fr : GPS RECEIVER FREQUENCY REFERENCE

N : NON-INTEGER SYNTHESISER MULTIPLIER CONSTANT

a : FRACTIONAL OSCILLATOR INSTABILITY OVER INTERVAL $\triangle t$

- SHORT TERM OSCILLATOR INSTABILITY PROPAGATES INTO VELOCITY ESTIMATES

  $Vj = fj\lambda = Vj - \lambda foa = Vj - c.a$

- MEASUREMENT ERRORS - Vj DEPEND UPON INTERVAL $\triangle t$
  - a INDEPENDENT OF $\triangle t$

Session Number

GPS CORRELATOR

TYPICAL PERFORMANCE DRIVERS

GOOD TRACKING REQUIRES LOOP SNR > 20dB

CORRELATOR ACCUMULATION INTERVAL:

MUST BE LESS THAN 20mS

(DATA RATE LIMIT)

MUST BE LARGER THAN 100uS

(SIGNAL TO NOISE LIMIT)

TYPICAL TIMES : 1,4 OR 10mS

CORRELATOR TAPS : SINGLE TAP

: MULTITAP - 6

- 32

- 1023 (FFT)

Session

Number

GPS SEARCH TIME

SEARCH FREQUENCY AND CORRELATOR PHASE BINS
FOR CHOSEN SATELLITE SIGNAL

1. POSITION AND TIME UNKNOWN

STRATEGY: SEARCH ALL C/A CODE PHASE IN EACH FREQUENCY BIN

STARTING AT 0 DOPPLER OFFSET

(1mS, 1 TAP CORRELATOR 15 x 1023 x 4mS = 60 SECS)

(10mS 1 TAP CORRELATOR 150 x 1023 x 40mS = 100MINS)

(4mS 1023 TAP CORRELATOR 60 x 1 x 16mS = 0.96 SEC)

2. (a) POSITION KNOWN, TIME UNKNOWN
   (b) POSITION UNKNOWN, TIME UNKNOWN

STRATEGY : AS ABOVE

3. POSITION KNOWN WITH IN 300m - 300km
   TIME KNOWN TO WITHIN 1mS - 10 MINUTES
   VELOCITY KNOWN TO WITHIN 0 - 5 m/SEC (18km/hr ± 25Hz)

STRATEGY : DETERMINE VISIBLE SATELLITES AND LOCATION
: DETERMINE SATELLITE AND VEHICLE RELATIVE DOPPLER OFFSET
: SEARCH ALL CODE PHASES IN CHOSEN FREQUENCY BIN

10mS 1 TAP CORRELATOR 1 x 1023 x 40mS = 40 secs

4mS 1023 TAP CORRELATOR 1 x 1 x 16mS = 16mS

4. POSITION KNOWN TO WITHIN 300m - 300km
   TIME KNOWN TO BETTER THAN 1mS
   VELOCITY KNOWN TO WITHIN 5m/s

STRATEGY : DETERMINE VISIBLE SATELLITES AND LOCATION TO BETTER THAN VEHICLE ACCURACY (EPHEMERIS)
: DETERMINE RELATIVE SATELLITE - VEHICLE DOPPLER
: SEARCH LIMITED RANGE OF C/A CODE PHASES

5. POSITION KNOWN TO BETTER THAN 300m
   TIME KNOWN TO BETTER THAN 1μS
   VELOCITY KNOWN TO WITHIN 5m/S

   DIRECT C/A CODE LOCK (LESS THAN 1 CHIP ERROR)

6. POSITION KNOWN TO BETTER THAN 30m
   TIME KNOWN TO BETTER THAN 100nS
   VELOCITY KNOWN TO WITHIN 5mS

   DIRECT P-CODE LOCK (LESS THAN 1 CHIP ERROR

Session Number

VELOCITY DETERMINATION

1. DERIVE VELOCITY FROM MEASURMENTS OF PSEUDORANGE CHANGE OVER SAY 0.1 OR 1 SECOND INTERVAL

HOST VEHICLE VELOCITY VECTOR V_h

PSEUDORANGE VECTOR AT TIME I P_I

GEOMETRICAL TRANSFORMATION H

$$V_h = H\,(P_{I+1} - P_I).\left(\frac{1}{T_{I+1} - T_I}\right)$$

2. DERIVE VELOCITY FROM DOPPLER MEASUREMENTS FROM SATELLITE TO HOST VEHICLE OVER 0.1 OR 1.0 SECOND INTERVAL

DOPPLER OFFSET FREQUENCY VECTOR D_I

$$V_h = (H.D_I)\,k$$

$$\text{CONSTANT } k = \frac{C}{f_o} = \lambda(\text{L1 OR L2})$$

Session **Number**

GPS ERROR BUDGET
(metres)

	PHASE I	PHASE II	Differential	GT
SATELLITE EPHEMERIS	3.6	1.5	0	0
SATELLITE GROUP DELAY & CLOCK	2.7	0.9	0	0
IONOSPHERIC DELAY				
(L1 only with mode)	6.4	6.4	1.0 (0.5)	0
(L1 + L2)	(3.0)	(3.0)		
TROPOSPHERIC ERROR (5° ELEVATION)	1.0	1.0		0
RECEIVER PSEUDORANGE NOISE (38 dBHz)				
(2 MHz)	2.4	2.4	2.4	2.4
(20 MHz)	(0.24)	(0.24)	(0.24)	(0.24)
MULTIPATH (C/A code)	3.0	3.0	3.0	3.0
(P code)	(1.0)	(1.0)	(1.0)	(1.0)
UERE (rms)	8.77 (5.60)	7.73 (3.76)	3.97 (1.14)	(3.84) (1.03)

Session Number

CODE LOOP

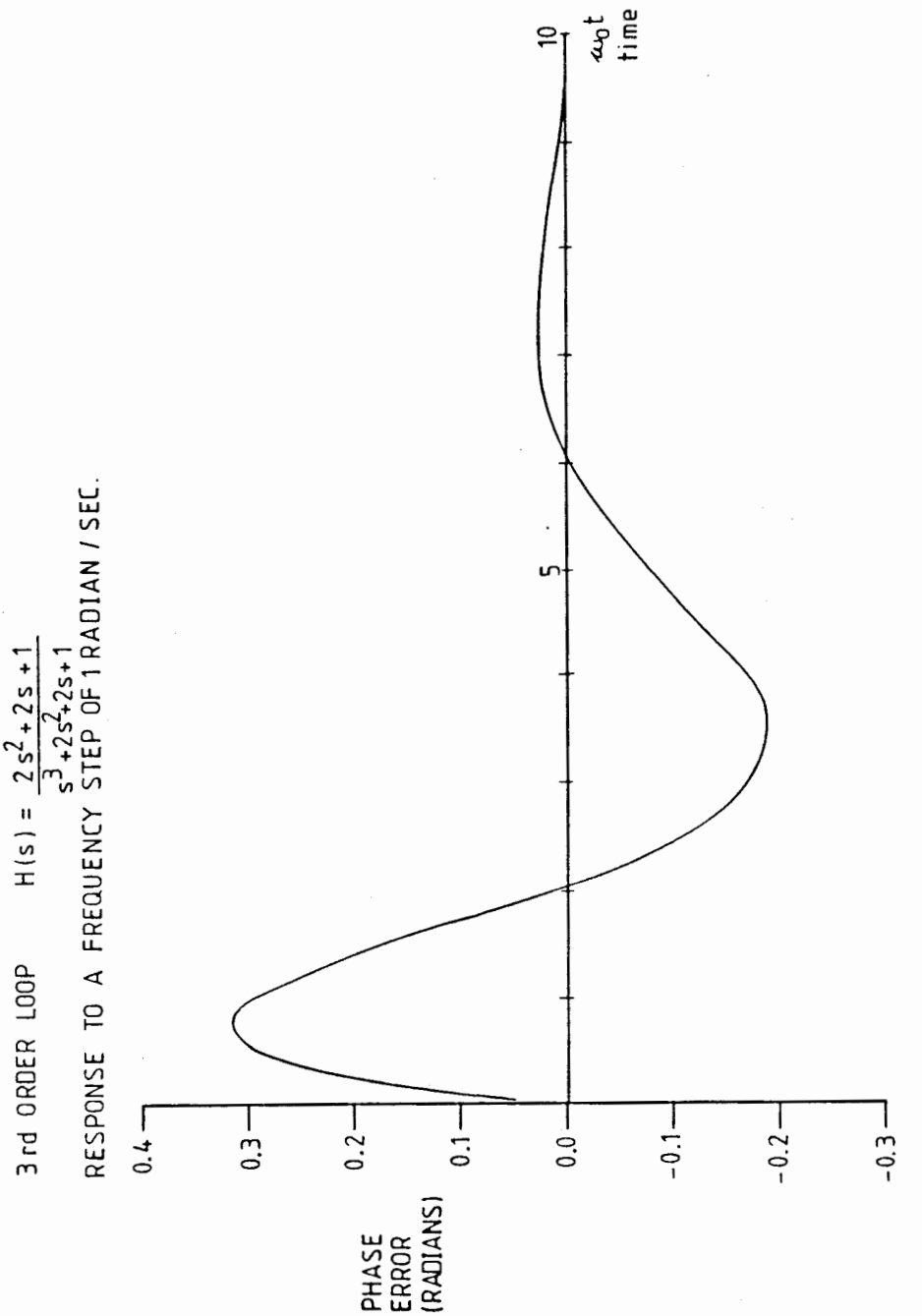
TYPICAL BANDWIDTH : 1/4 - 2Hz

DEPENDS UPON : UNAIDED - 1st ORDER
2nd ORDER
3rd ORDER
- VEHICLE DYNAMICS

: AIDED - USUALLY 1st ORDER
- BY INS
- BY CARRIER PHASE MEASUREMENTS
- DOPPLER SPEED & AHRS
- PROPAGATION OF RESIDUAL VEHICLE DYNAMICS

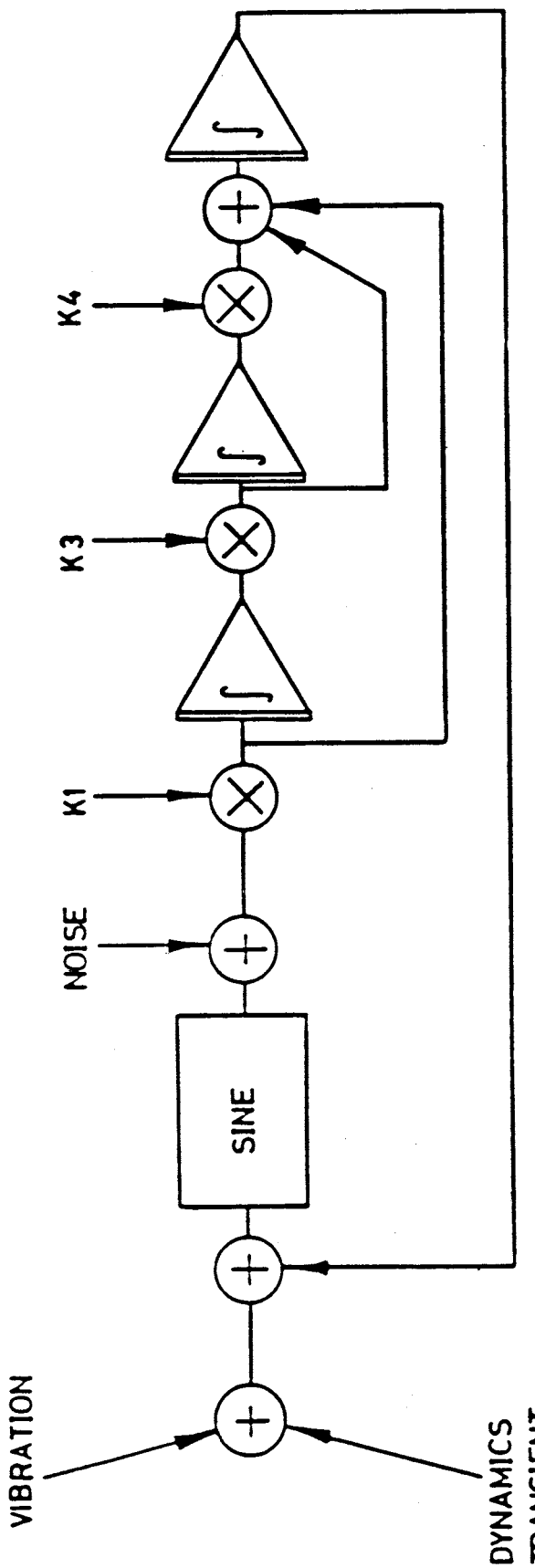
Session **Number**

LOOP TRANSIENT RESPONSE



Session Number

CARRIER LOOP PERFORMANCE ANALYSIS



$K1 = Z1\,\omega n / Kpd\,K\,nco$

$K3 = Z2\,\omega n / Z1$

$K4 = \omega n / Z2$

$$H(s) = \frac{\frac{Z1}{\omega n^2}\,S^2 + \frac{Z2}{\omega n}\,S + 1}{\frac{1}{\omega n^3}\,S^3 + \frac{Z1}{\omega n^2}\,S^2 + \frac{Z2}{\omega n}\,S + 1}$$

Session Number

CARRIER LOOP

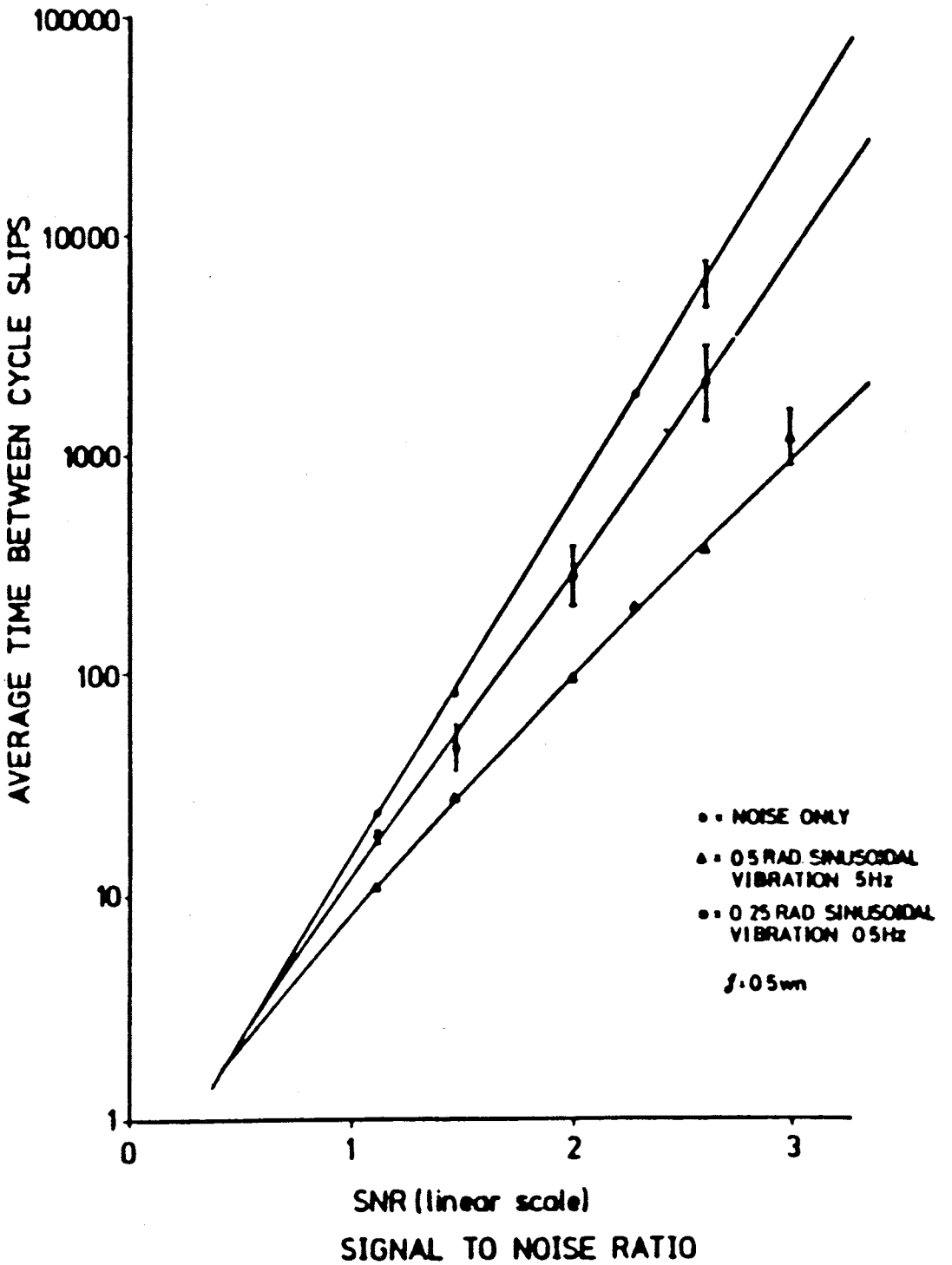
TYPICAL BANDWITH : 2 - 25Hz

- DEPENDS UPON : LOOP ORDER 2nd, 3rd, 4th ORDER
  : VEHICLE DYNAMICS
  : OSCILLATOR NOISE
  : SIGNAL TO NOISE RATIO

- AIDING : MAY BE AIDED BY HIGH QUALITY INS

Session Number

CARRIER LOOP CYCLE SLIPS (SINUSOIDAL VIBRATION)



Session Number

OSCILLATOR CHARACTERISTICS FREQUENCY ACCURACY

MAIN OSCILLATOR EFFECTS ARE - CALIBRATION ACCURACY AGEING

DOPPLER EFFECT ON RECIEVED SIGNALS AT L1 : ± 7.5kHz
(VEHICLE AND SATELLITE VELOCITY)

MAXIMUM FREQUENCY UNCERTAINTY IN 1mS CORRELATION INTERVAL
: ± 500Hz

No. OF FREQUENCY SEARCH BINS : 15

MAXIMUM FREQUENCY UNCERTAINTY IN 10mS CORRELATION INTERVAL
: ± 50Hz

No. OF FREQUENCY SEARCH BINS : 150

AGEING OF QUARTZ OSCILLATOR : 3 PARTS IN 10^7 PER ANNUM TYPICAL
AT L1 : 47QHz / ANNUM

POSSIBLE SOLUTIONS : LOW AGEING RATE QUARTZ OSCILLATOR

: ATOMIC OSCILLATORS

: FREQUENT RECALIBRATION OF OSCILLATOR

: SOFTWARE MEMORY OF OSCILLATOR FREQUENCY OFFSET

Session Number

OSCILLATOR CHARACTERISTICS TIME ACCURCY

USE OF TIME IN GPS RECEIVER : CORRELATOR INTERVAL GENERATION

: INTERRUPT GENERATION

: CALIBRATION AGAINST GPS TIME OR UTC

: DETERMINATION OF SATELLITE LOCATIONS

SATELLITE TRANSMITTS 3 CLOCK CORRECTION PARAMETERS

a0 TIME OFFSET 22 BITS ± 976.6uS

a1 FREQUENCY OFFSET 16 BITS ± 3.725nS/SEC

a2 FREQUENCY RATE 8 BITS ± 3.553fS/SEC 2

TYPICAL GPS SATELLITE CLOCK ACCURACIES ARE 5nS RMS USE

USER ACCURACY DETERMINED BY : ORIGINAL TIME CALIBRATION

: FREQUENCY UNCERTAINTY

: AGEING RATE OF OSCILLATOR

POSSIBLE SOLUTION : MODEL USER OSCILLATOR

Session Number

OSCILLATOR CHARACTERISTICS SHORT TERM STABILITY AND NOISE

TYPICAL FREQUENCY REFERENCE FOR GPS : 5 - 100 MHz

CARRIER TO NOISE RATIO IN 2Hz CARRIER LOOP:

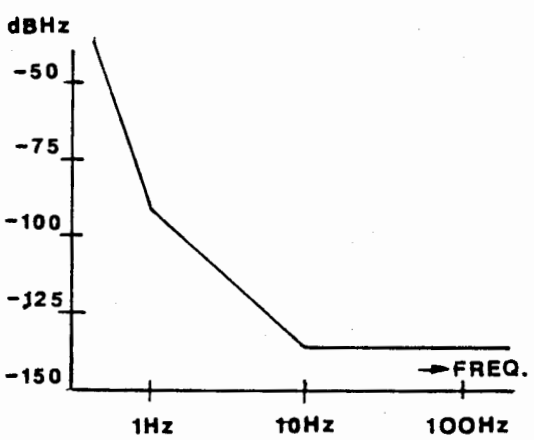
THERMAL NOISE LEVEL -204 dBWHz

GPS SIGNAL STRENGTH (L1 C/A CODE)-160 dBW

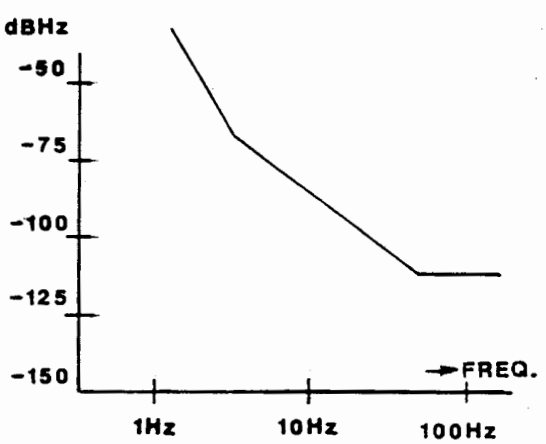
CARRIER LOOP BANDWIDTH (2Hz) 3dBHz

NOISE FIGURE OF RECEIVER 2dB

C/No 39 dB



OSCILLATOR SPECTRUM



SYNTHESISER SPECTRUM

REQUIREMENT IS : INTEGRATED NOISE IN CARIER LOOP BW

SUBSTANTIALLY SMALLER THAN REFERRED LEVEL

OF INPUT THERMAL NOISE

Session Number

OSCILLATOR
G-SENSITIVITY

- SIMULATE CARRIER LOOP RESPONSE
  - (A) DYNAMICS INPUT
  - (B) NOISE INPUT
  - (C) SINUSOIDAL VIBRATION INPUT WITH NOISE
  - (D) RANDOM VIBRATION INPUT WITH NOISE
  - (E) COMBINATIONS OF (A) and (C) ; (A) and (D)
- MEASURE CARRIER LOOP RESPONSE USING PROBABILITY OF CYCLE SLIP OR MEAN TIME BETWEEN CYCLE SLIPS
- ESTIMATE EQUIVALENT SNR DEGRADATION

OSCILLATOR G-SENSITIVITY

- STUDY INDICATES THAT ANTENNA MOTION-INDUCED CARRIER PHASE MODULATION AND OSCILLATOR G-SENSITIVITY INDUCED CARRIER PHASE MODULATION IN PHASE QUADRATURE

- TABLE OF RESULTS (sinusoidal vibration model without dynamics)

Double Amp Vibration (inches)	0·5	0·25	Peak Carrier Phase Modulation Radians
1·0	7pts in $10/g^{10}$		
0·5	1pt in $10/g^{9}$	35pts in $10/g^{10}$	
Performance Loss	1·6 dB	0·5 dB	B_l Tcs ≈ 2500

2 ND EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 16 - 17 - 18 MARCH 1988

Session Invited Number 5

TIMING REQUIREMENTS OF FUTURE ISDNs

R.D. CALVERT
PRINCIPAL ENGINEER

ENGINEERING AND PROJECTS DIVISION
CABLE AND WIRELESS PLC
MERCURY HOUSE, THEOBALDS ROAD,
LONDON WC1X 8RX,
ENGLAND

CONTENTS

1. INTRODUCTION

2. NETWORK TIMING
2.1 Performance Requirements for NSRF
2.1.1 The Use of Caesium
2.1.2 The Use of GPS
2.1.3 The Use of Loran-C
2.1.4 Network Standard Reference Frequency (NSRF) Architecture
2.2 Performance Requirements for Slave Clocks at Lower Hierarchical Levels
2.3 Timing Distribution in a Synchronous Network

3. SLIPS

4. BUFFERING
4.1 Plesiochronous Buffering
4.2 Doppler Buffering

5. TIMING REQUIREMENTS FOR B-ISDNs EMPLOYING THE SONET TECHNOLOGY

6. REFERENCES

7. ACKNOWLEDGEMENTS

1. INTRODUCTION

Telecommunications Networks are evolving into National Synchronous Networks which interwork plesiochronously, and the foundations are being laid for the further evolution of these networks into the Integrated Services Digital Networks (ISDNs) of the future. For this reason, Network Synchronization is receiving increasing attention and a great deal of research is being undertaken by Telecommunications Administrations and Registered Private Operating Agencies (RPOA).

In addition to current work on the 64 kbit/s Narrow Band or Traditional ISDN performance, studies are progressing on future Broadband ISDN (B-ISDN) concepts. From the studies so far undertaken, there appears to be general agreement that the long term target network for the B-ISDN should be based on Asynchronous Transfer Mode (ATM) Principles and that a realistic date for the implementation of this target network would be 1995. It is accepted that the basic principles embodied in the 64 kbit/s based ISDN should also apply to the B-ISDN. The B-ISDN is considered applicable to services at bit rates above the primary rate.

It has been suggested that the transmission infrastructure of the B-ISDN should be based on the new synchronous digital hierarchy developed from the North American Synchronous Optical Network (SONET). However, the switching systems are expected to be ATM based, as this is better suited to traffic which has the bursty characteristics expected of Broadband Services.

The new synchronous digital hierarchy, internationally based at 155,520 kbit/s (although within the USA it may have a basic building block at 51840 kbit/s), brings with it a possible requirement for new synchronization performance specifications. This aspect is discussed in the last section of this paper.

Guidance in the requirements of Network Synchronization is given in CCITT Recommendations G811, G81X (new), G822, G823 and G824, and these recommendations are reviewed by CCITT Study Group XVIII Working Party 6.

The main focus of this paper is a consideration of the effect which CCITT Recommendations have on the requirements of Network Synchronization.

In the interests of simplicity, the term synchronization has been used throughout the paper even where, from a purist viewpoint, syntonization is the more correct term.

2. NETWORK TIMING

Timing can be considered from three aspects; firstly, the requirements of the Network Standard Reference Frequency (NSRF), secondly, the performance of clocks at lower levels in the hierarchy and, finally, the distribution of timing from the NSRF to the slave clocks at the lower hierarchical levels, which is essential in order to achieve full network synchronization.

For political as well as technical reasons, network synchronization does not usually transcend national boundaries. Therefore interworking between synchronous national networks is achieved plesiochronously. For this reason, very accurate reference standards are necessary to enable the CCITT objectives to be met. Network synchronization has a considerable influence on the design of a digital network and should be taken into consideration during the inception and planning stages of such a network.

The concept of international digital interworking between synchronous national networks is shown in Figure 1. It is clear from the figure that the timing of all equipment within a network is traceable to the Network Reference Clock. Since the two networks are timed independently, there will be a requirement for plesiochronous buffering at the international interface. This is discussed in more detail in subsequent sections of this paper.

The synchronization of national digital networks can be achieved by a number of methods. The main alternatives are; Hierarchical Master-Slave Synchronization, Mutual Synchronization and a Hybrid form combining the two.

Of these three alternative approaches, Master-Slave synchronization is the most universally preferred, as it is relatively less complex to implement and maintain. Timing is distributed in a hierarchical manner with the higher level clocks feeding the lower ones. The direction of timing flow is such as to avoid closed loops in the synchronization network which could cause instability.

For hierarchical Master-Slave synchronization, the network operator must introduce a hierarchy of clock performance which ensures that the network achieves the timing requirements stipulated by the CCITT. The synchronization network is embedded within the telecommunications network and the hierarchical timing plan is used to define the required clock performance at each location. A simple network hierarchical timing plan is shown in Figure 2. This may result in the network synchronization plan shown in Figure 3.

As synchronization reference is distributed from the primary reference source through intermediate clocks and distribution facilities, the phase stability of the reference continues to degrade. This is because wander from the clocks and from the network transmission systems passes through the clock recovery circuits and accumulates in synchronization distribution paths.

In general, phase-time variations experienced within digital networks can be considered to result from three contributary sources. These are jitter, wander and frequency departure. The effect of combining these three components is represented in Figure 4. Wander is used as a general term to include slow phase-time variations resulting from temperature changes in transmission systems, cyclic-variations in clock frequency and integrated frequency departure of clocks.

It is important that jitter and wander accumulation through a digital network is restricted. Therefore, the choice of phase lock loop (PLL) time constant at slave clock locations has an important impact as it determines the PLL cut-off frequency and, hence, the extent of rejection achievable.

2.1 Performance Requirements for NSRF

To meet the requirements stated in CCITT Rec. G811 concerning the performance of clocks at network nodes that terminate international digital links, it is necessary to use either caesium technology in the NSRF clock system or timing derived from a suitable off-air source. The essentially drift-free characteristics of caesium coupled with its high accuracy make it a prime candidate for Reference Clock applications. Alternatives to caesium are the reception of either Loran-C or GPS and the use of such off-air standards to discipline a quartz or even a rubidium oscillator.

A comparison of some performance parameters for these three options is shown in Table 1. A Level 1 Reference Clock System can be configured using any of these sources to achieve the availability specified in CCITT G.811.

Table 1
Comparison of Off-Air Standards with Caesium

Source	Departure	Time Accuracy (Per Year)	Typical Output Jitter at 2048kHz	MTTR	MTBF	Availability
Telecoms Caesium Standard	7×10^{-12}	220 us	less than 10 ps	30 days	6 years	98.6%
Loran-C Reception	5×10^{-12} to N Parts x 10^{-13}	150 us	less than 0.5 ns	-		99.9% Note 1
GPS Reception	3×10^{-14}	1 us	-	10 months per satellite	4.5 months per satellite	99% Note 2

Note 1 This figure is quoted by the US Naval Observatory as the typical availability for reception of three simultaneous sources of Loran-C for navigation purposes. For Telecommunications timing applications the requirement is to receive from one or two sources, but not necessarily simultaneously.

Note 2 This figure is based on the full GPS constellation of satellites which will provide 4 satellites in view to all locations.

2.1.1 The Use of Caesium

CCITT Rec. G811 specifies a long term frequency departure for all clocks at network nodes which terminate international links. This long term frequency departure should be better than 1 part in 10^{11}. Current caesium clocks exhibit a departure which is better than 7 parts in 10^{12}, and so can achieve this objective.

Cable and Wireless are successfully using caesium reference clocks to time their networks in both the UK and Hong Kong. In the latter case the caesiums are distributed at two locations. Comparisons of Time Interval Error, between the two reference clock sub-systems over the intervening transmission facility, provided results better than 1 part in 10^{12} performance. In the former case a classic triplicated caesium system is employed with the standards now entering their fifth year of operation.

2.1.2 The Use of GPS

In many small national networks throughout the world, the expense of introducing caesium based reference clock systems is prohibitive. For their network operators to offer multidestination international digital services to their customers, a cheaper alternative has to be found. Full digital interworking between national networks in remote geographical locations, such as island nations in the Pacific and Indian Oceans and Caribbean sea, requires the use of an off-air standard which can provide timing meeting the requirements of G311 and which is not geographically limited in coverage area. Research has shown that the GPS satellite navigation system may provide the solution to this problem and give cost effective reference timing to small national networks.

The two or more GPS receivers used to control either quartz or rubidium oscillators can be suitably packaged into a reference clock system which should meet the timing, phase discontinuity and availability requirements of G811. Telecommunications operating companies are showing increasing interest in the possibilities offered by this technology.

It appears that even with the present satellite configuration GPS can provide reference timing with an accuracy of 10 ns per day, which relates to 1 part in 10^{13}. However, when the full constellation of satellites are in orbit, it is expected that GPS will achieve an accuracy of 3 parts in 10^{14} per day which is between 10 and 100 times more accurate than the performance achievable from Loran C.

2.1.3 The Use of Loran C

Loran-C also offers a cost effective alternative to caesium for achieving network reference timing. Two or more Loran-C receivers used in conjunction with quartz oscillators can provide the basis of a system that will give the user the performance and long term stability of a dedicated caesium system as long as reception is good.

The timing accuracy obtainable from this source is better than 5 parts in 10^{12} per day after 24 hours tracking and can be typically N parts in 10^{13} per day after one weeks tracking, where N is between 1 and 10.

Cable and Wireless have sucessfully used a Loran-C based system to time their digital network in Bermuda. The performance exceeds 5 parts in 10^{12}.

2.1.4 Network Standard Reference Frequency (NSRF) Architecture

The NSRF occupies the highest level (Level 1) of the synchronization hierarchy and can be packaged to achieve the performance requirements of CCITT Rec. G.811 in many forms. The most expensive is a triplicated caesium system, whilst the cheapest is a system based on either Loran-C or GPS receivers.

Some typical system configurations are shown in Figures 5 and 6.

Distributing the oscillators between a number of locations provides greater network operational security. This concept is shown in Figure 7. The Distributed Clock System incorporates all the performance requirements of G.811, and has the additional advantage that the availability of synchronization to any node in the network can be improved.

2.2 Performance Requirements for Slave Clocks at Lower Hierarchical Levels

The performance of clocks at lower levels in the synchronization hierarchy is typically determined by network availability considerations. Such clocks are known as slave clocks. It is their performance in free-run or holdover mode which is all important as, during periods of loss of synchronization, this will determine the slip performance of any end to end 64 kbit/s service.

The CCITT in Rec. G81X has now defined the performance of slave clocks at both Transit and Local Nodes. Such clocks should comply with the following performance:

	Frequency Departure	Drift per day
Transit Node	2 parts in 10^{10}	2 parts in 10^{9}
Local Node	1 part in 10^{8}	2 parts in 10^{8}

However, at the latest meeting of CCITT SGXVIII in Seoul, there was a proposal to make the Transit Node specification more stringent:

Frequency Departure	5×10^{10}
Drift per day	5×10^{10}

A decision will be taken on this proposal in June 1988. Although it is not specifically defined in the text of Rec. G81X, these figures should identify the performance achievable taking into account ageing and temperature effects. Such systems should include oven control.

Rec. G81X defines three operating modes for slave clocks. These are ideal operation, stressed operation and holdover (or free-run) operation. Ideal operation represents a situation where the clock experiences no disruptions of input timing reference and so the slave clock is phase locked with the reference.

Stressed operation reflects performance in the presence of interruptions which can result in a frequency offset between the slave clock and its reference. The implication is that there is an accumulation of frequency offset in a chain of slave clocks such that they may actually be operating plesiochronously with respect to each other in the network.

Holdover Operation accounts for periods when a clock loses reference for a significant time interval. In holdover, or free-run operation, the important components determining TIE performance are the frequency drift and initial frequency offset of the slave clock.

Slave clock systems should be located at nodes allocated to Level 2 and Level 3 in the synchronization hierarchy. A typical block diagram is shown in Figure 8.

The system includes timing extraction equipment which will enable the disciplined quartz oscillator to synchronize to a traffic carrying 2048 kbit/s stream dedicated to synchronization distribution. The synchronization stream selected must originate from a source which is of higher level in the synchronization hierarchy.

The timing recovered from the 2048 kbit/s stream is used to phase lock a quartz oscillator. The phase lock loop time constant should be selected to achieve optimum jitter and wander rejection. The performance required of the quartz oscillator is defined by the synchronization network hierarchical timing plan and the position of the slave clock at a particular hierarchical level.

At level 2, the PLL time constant chosen by Cable and Wireless is typically 1000 seconds, which relates to a cut-off frequency of 1.6×10^{-4}Hz.

2.3 Timing Distribution in a Synchronous Network

Within Digital Networks, timing is distributed by using traffic carrying 1544 kbit/s or 2048 kbit/s Primary Rate digital streams. However, one of the current problems is that there is no agreement on how such streams should be marked to indicate timing quality. For the 2048 kbit/s streams, it is possible to use any of the bits in time slot zero of the Not Frame Alignment frame, which are allocated for National use. These are bits 4 to 8, and of these Cable and Wireless supports the use of bit 5 for time marking purposes. A method of marking primary rate streams for quality indication is shown in Figure 8.

For synchronization availability performance reasons, primary rate streams used for timing distribution are usually duplicated, with one stream designated as the primary distribution stream and the other as secondary distribution. Where possible, the two streams exhibit route diversity, the most reliable streams being chosen for this application.

Using these primary rate streams, timing is distributed throughout the network from equipment timed from clock systems at the highest hierarchical level to slave clocks at lower levels. By this means all equipment in a network is timed traceable to the Network Standard Reference Clock.

3. SLIPS

A lack of synchronization in a digital network will result in slips. A slip is the term used to identify the repetition or deletion of bits from a digital stream. It is the result of a buffer under-fill or over-fill state. To avoid slip impairments caused by plesiochronous interworking of equipment within the network, timing synchronization is required. Buffers are used within the network to absorb the difference between node clock frequency and the timing of received primary rate streams.

Slips occur for two basic reasons:

a) A lack of frequency synchronization among the clocks at various nodes within the network.

b) The existence of phase wander and jitter in the digital bit streams, caused by various mechanisms within the network.

Therefore, even with a completely synchronous network where there is no frequency departure between two timing signals at a buffer, slips can still occur because of the accumulation of jitter and wander at that location. When the peak-to-peak phase-time variation exceeds the buffer capacity (the level that the buffer can absorb) the buffer will introduce a slip. The use of a buffer has the effect of controlling slips to be of specific size and to occur at frame boundaries.

The size of the buffer together with the phase time error between the incoming bit stream or write-in rate and the local clock or read-out rate, will determine the size of the slip, the interval between slips and therefore the slip rate.

The impact of slip impairments is service dependent. For example, voice is relatively insensitive to slips, whereas data, facsimile, encrypted voice and video services are much more sensitive, since each bit of data is a discrete piece of information. A high slip rate can adversely affect the quality, the information transfer rate, and the throughput of data and other services.

This has been taken into account in Recommendation G.822 which provides for a 64 kbit/s end-to-end slip rate objective of less than 5 slips in 24 hours. This actually corresponds with the North American objective of 1 slip in 5 hours or 4.8 slips in 24 hours. In the North American system half of the slip rate objective is allocated to uncontrollable impairments such as Wander, which is caused by daily temperature variations, and Jitter, which is caused by equipment operation factors.

Acceptable slip rate performance for various services is summarized in Table 2. It can be seen that non-error protected facsimile with a slip interval requirement of one slip in six hours is the most sensitive to this impairment. CCITT have used this figure as the acceptable slip performance for an end-to-end connection when referenced to the standard digital hypothetical reference connection (HRX), which is 27500 km in length, and defines 13 nodes within the connection.

Table 2 Acceptable Slip Rates for Various Services

Service		Objectives	Slip Interval Objectives(Hrs)
PCM Voice	Telephony	1 click/min Note 1	0.01
	Sound Programme	1 click/hour Note 1	1
Data	Fixed block length	99.5% efficiency	1
	Variable block length		3
Facsimile	Non-error protected	1% probability of picture degradation	6
	Error protected	1 line replacement every 160 seconds	0.04

Note 1. A click is an audible sound caused by a frame slip.

In summary, therefore, we can say that network slip rate objectives are achieved by ensuring buffers at digital interfaces are of adequate size, and that all clocks within the network are synchronized.

4. BUFFERING REQUIREMENTS IN DIGITAL NETWORKS

4.1 Plesiochronous Buffering

For plesiochronous interworking on switched traffic, buffering will be provided integral to the International Switching Centre (ISC). These buffers are usually of two-frame capacity and introduce slips of a single frame size.

For plesiochronous interworking on the non-switched traffic via satellite, buffering must be provided at the synchronization interface, which will be the initial primary multiplexer accessed by the traffic. Synchronization timing will be fed to the buffer's external clock input and also to the primary multiplexer's external clock input.

Plesiochronous buffers are a standard 2-frame 250 us capacity for 2048 kbit/s primary rate data.

4.2 Doppler Buffering

For satellite paths, additional buffering is required. The maximum peak-to-peak diurnal satellite path delay variations for INTELSAT 4A and 5 satellites is 0.6 ms and the maximum rate of change is 18 ns/s. This Doppler shift delay cannot be absorbed by use of an 8 bit buffer and special buffer arrangements are needed. Doppler buffers are normally supplied as standard 2 ms size, but can be strapped for the required capacity to ensure the Doppler shift can be absorbed. An example of the effect of Doppler on Time Interval Error for a 56 kbit/s satellite service is shown in Fig. 9. It can be seen from this chart recording that Doppler manifests itself as a low frequency wander, which has a diurnal cycle approximating to 10^5 seconds and a rate of change of approximately 2 parts in 10^8.

These buffers can be thought of as containing both a Doppler and a plesiochronous section. The purpose of the Doppler section is to absorb the diurnal variations in path delay caused by the movement of the satellite relative to the receiving station. The purpose of the plesiochronous section is to re-time the bit stream to the national network standard. The buffer's plesiochronous section will be capable of introducing slips of single frame size should any under-fill or over-fill occur. This can best be visualised by referring to Figure 10.

The Doppler buffers will normally be installed in the satellite Earth stations although, as an alternative, the Doppler buffer could be located adjacent to the ISC as the network can handle the maximum rate of change of Doppler. Regardless of its location, this buffer must be installed prior to a synchronization interface, whether this is an exchange for the switched services or a primary multiplexer for the non-switched services.

5. TIMING REQUIREMENTS FOR B-ISDNs EMPLOYING THE SONET TECHNOLOGY

The fundamental research on SONET has come from North America where the work of the past five years has now resulted in a Draft American National Standard on the subject. This has been considered in the international forum provided by the CCITT and, as a consequence, has resulted in the emergence of three new Draft Recommendations numbered G70x, y and z, which form a coherent set of recommendations for a new synchronous digital hierarchy.

The original SONET concept is based on a Synchronous Transport Signal, known as STS-1, which has a bit rate of 51,840 kbit/s and a frame with a 125 microsecond period. The Synchronous Transport Signal is made up of two separate parts, a Transport Overhead (TOH) of 27 bytes and a Synchronous Payload Envelope (SPE) of 783 bytes. It is proposed to byte interleave the STS-1 level signals to form higher bit rate signals. A clearer understanding of the basic frame structure can be gained from Figure 11.

The SPE is frequency justified with respect to the TOH, using a payload pointer mechanism. Pointer adjustment is allowed every four frames, which results in a maximum adjustment rate of 16 kHz, and creates an 8 unit interval phase hit. The phase hit is leaked out over a suitable time interval, chosen so that the resulting frequency deviation falls within the half power bandwidth of the clock timing recovery circuit. An example of a positive and negative pointer adjustment operation is shown for clarity in Figures 12 and 13 respectively.

The concept envisaged in the CCITT Recommendations on the new synchronous digital hierarchy differs from that discussed above, as the basic bit rate is 155520 kbit/s which is equivalent to STS-3 rate. There is also an additional level of pointers added.

The integration of all synchronous optical equipment into the present North American telecommunications network synchronization hierarchy has been specified in ANSI standard T1.101.1987. This recommends that the clocks used to synchronize the synchronous optical equipment should be of stratum 3 or higher quality. In addition to this, it is proposed that the short term stability of the associated synchronous optical equipment clock, internal or external, must achieve the provisional requirements shown in Figure 15. Internal oscillators must have a minimum accuracy of $\pm 2 \times 10^{-5}$.

The North American Synchronization Architecture is based on the Building Integrated Timing System (BITS) concept. This is illustrated in Figure 15.

This method allows a single clock system in a building to control the timing of all other clocks within the building. The clock selected as the BITS is of the highest stratum (level) within the building, and receives an external reference traceable to a primary reference source.

The current ANSI standard identifies the need for synchronous optical equipment to supply a timing reference to the BITS clock. This requires the recovery of a 1544 kbit/s primary rate stream from within the SPE, and can be achieved from either a Terminal Equipment, an Add/Drop Multiplexer or a Virtual Tributary Cross-Connect device. The minimum requirement is for the BITS to recover timing from two 1544 kbit/s primary rate streams, the required signal being a framed, all 'ones', 1544 kbit/s timing signal. In principle, it should be possible to recover timing from the incoming optical signal with the equipment operating in a loop-timed mode.

There is much discussion on the need for timing recovery from STS-1 and STS-3 rate streams. This means there is a possibility of a future requirement for clock recovery and timing distribution at either or both 51840 kbit/s and 155520 kbit/s. The significance of this with regard to future timing recommendations, in terms of the performance required, needs further study. Synchronous optical terminal equipment must be able to utilise loop timing without violating the stratum 3 clock requirement.

During periods of loss of timing reference, or loss of synchronization, the equipment internal clock should continue to operate in a free-running or holdover mode to generate proper maintenance signals.

Throughout this section, the author has maintained the original ANSI terminology rather than use the latest CCITT terminology relating to this subject. This has been done for clarity as the overwhelming amount of literature on this subject is written using the ANSI terminology.

6. REFERENCES

Reference Clock Parameters for Digital Communications Systems Applications - Peter Kartaschoff.

Frequency Control and Timing Requirements for Communications Systems - Peter Kartaschoff.

From Sundials to Atomic Clocks - Jesperson and Fitz-Randolph. National Bureau of Standards, US Dept of Commerce.

CCITT Recs. G.811, G.822, G.823, G.824.

IEEE Transactions on Communications Part 1. Synchronization
Vol. COM-28, No. 8. August 80.

Draft Proposal American National Standard - Synchronization Interface Standard for Digital Networks, TI Standards Committee.

Impact of Pointer Process on Phase-Time Variation, SONET Wander and Synchronization Study Group, American National Standards Institute.
Draft ANSI Document.

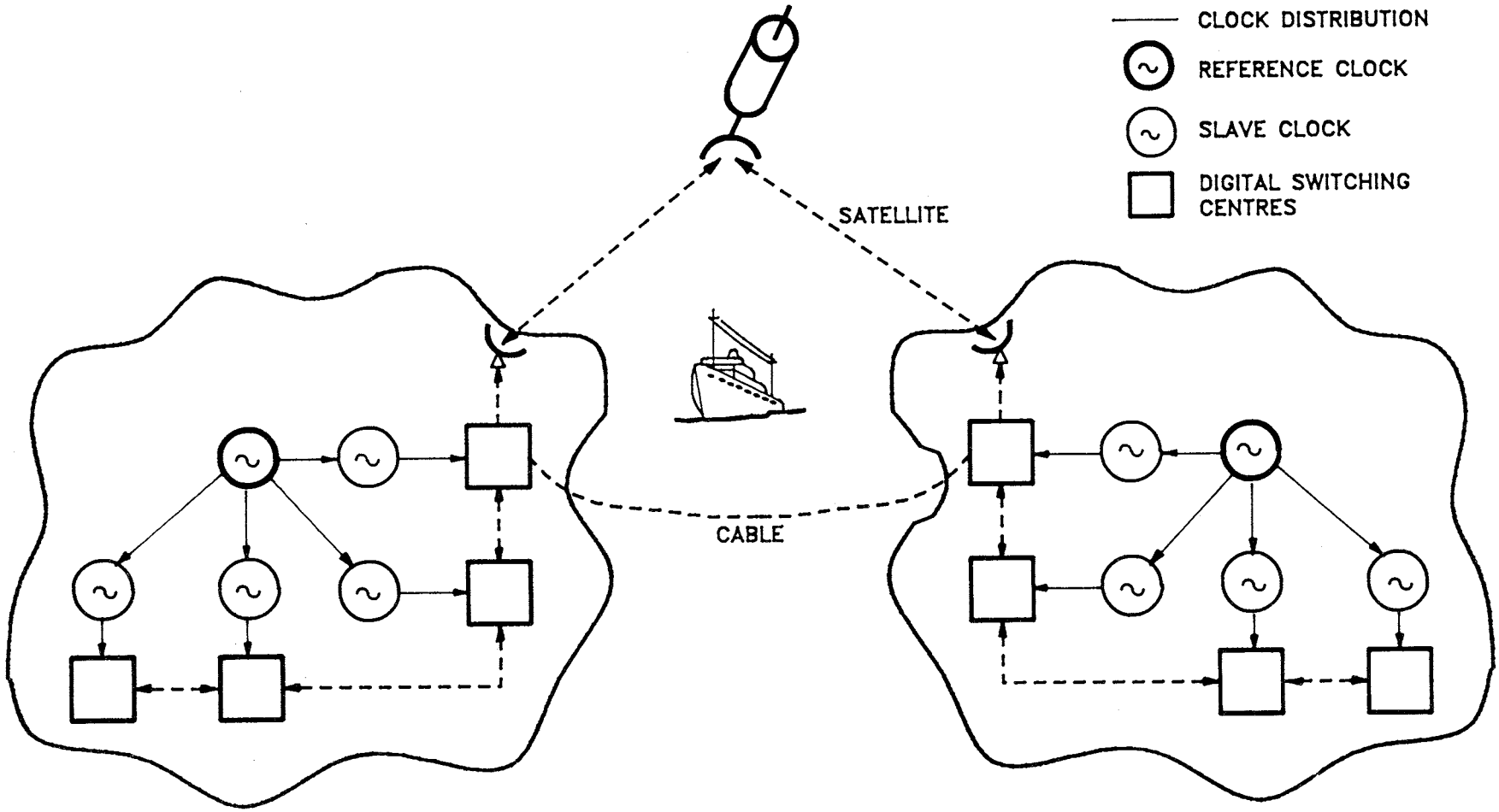
7. ACKNOWLEDGEMENTS

The author wishes to thank the Group General Manager (Engineering and Projects), Cable and Wireless plc., for permission to publish this paper.

The author wishes to acknowledge the use of American National Standards Institute (ANSI) material in Section 5 of this paper and in the Figures which relate to that section.

DEFINIION OF TERMS

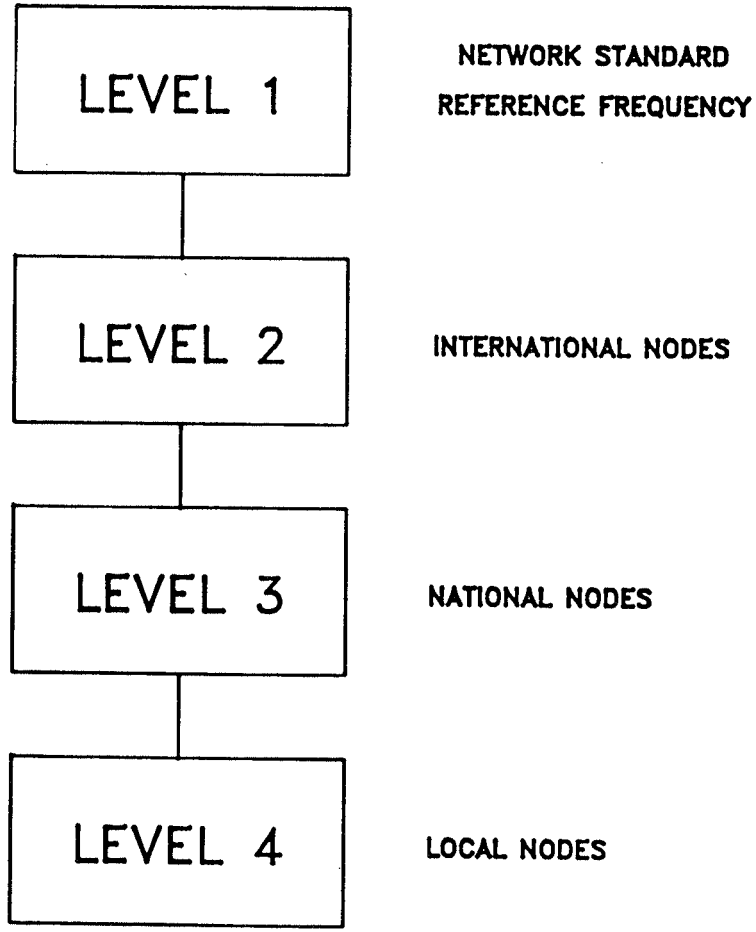
Asynchronous	The essential characteristics of time-scales or signals such that their corresponding significant instants do not necessarily occur at the same average rate.
Controlled Slip	The irretrievable loss or gain of a set of consecutive digit positions in a digital signal, in which both the magnitude and instant of that loss or gain are controlled, to enable the signal to accord with a rate different from its own. Note: Where appropriate the term may be qualified, for example: controlled octet slip, controlled frame slip.
Integrated Frequency Dep.	An underlying offset in the frequency of a timing signal from it's nominal value.
Plesiochronous	The essential characteristics of time-scales or signals such that their corresponding significant instants occur at nominally the same rate, any variation in rate being constrained within specified limits Note 1 Two signals having the same bit rate, but not stemming from the same clock are usually plesiochronous. Note 2 There is no limit to the time relationship between the corresponding significant instants.
Synchronization	The process of adjusting corresponding significant instants of signals to make them synchronous In this paper, synchronization has been used in the broader sense to include syntonization.
Synchronous	The essential characteristics of time-scales or signals such that their corresponding significant instants occur at precisely the same average rate. Note that the timing relationship between corresponding significant instants usually varies between specified limits.
Synchronous Network	A network in which the corresponding significant instants of nominated signals are adjusted to make them synchronous.
Uncontrolled Slip	The loss or gain of a digit position or a set of consecutive digit positions in a digital signal resulting from an aberration of the timing processes associated with transmission or switching of a digital signal, and in which either the magnitude or the instant of that loss or gain is not controlled.



INTERCONNECTION OF NATIONAL SYNCHRONOUS NETWORKS

Fig. 1
BLK 2584

CABLE AND WIRELESS PLC



SYNCHRONISATION NETWORK HIERARCHICAL TIMING PLAN

Fig. 2
BLK2590

CABLE AND WIRELESS PLC

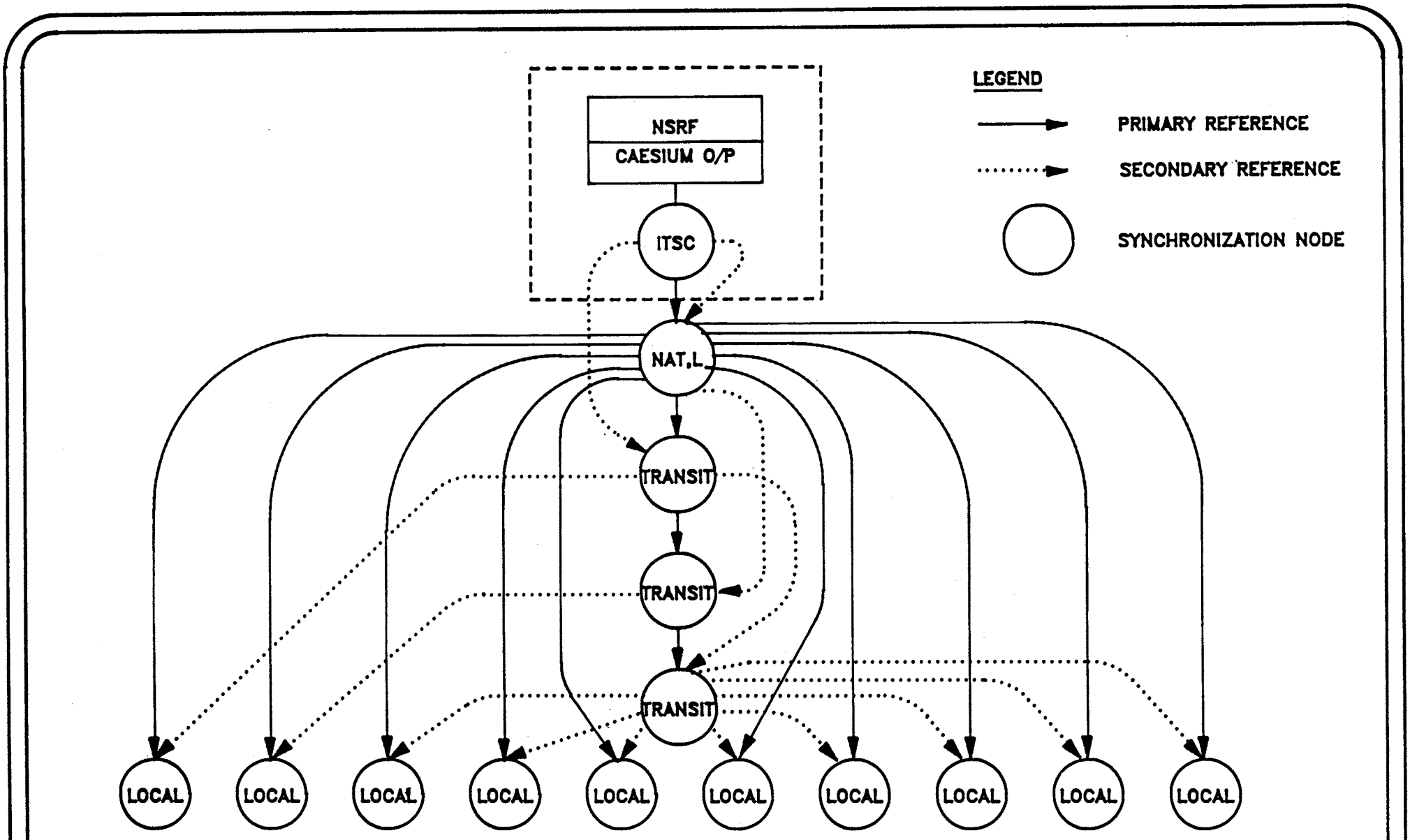
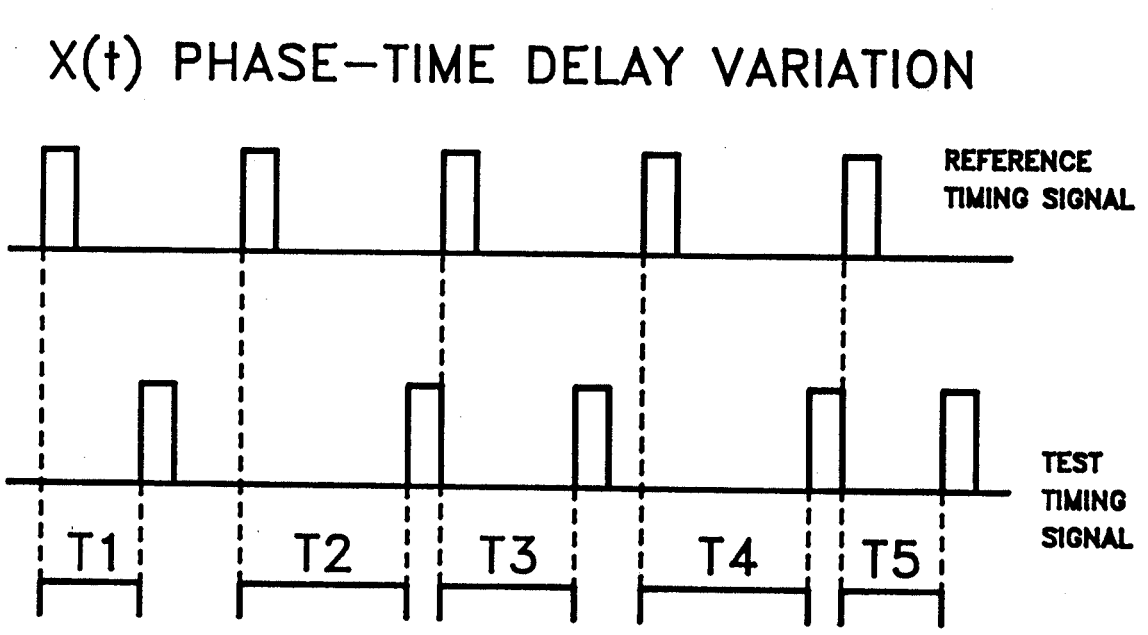
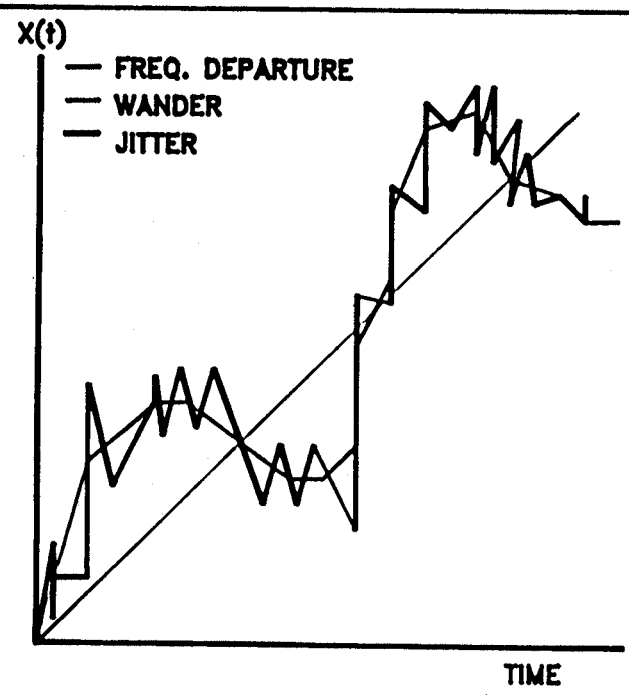


Fig. 3
BLK 2591

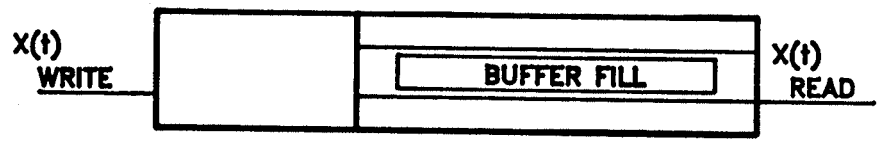
NETWORK SYNCHRONIZATION PLAN (SWITCHED SERVICES)

CABLE AND WIRELESS PLC





SLIP: REPETITION OR DELETION OF A BLOCK OF BITS AT A BUFFER



JITTER: GREATER THAN 10 Hz

WANDER: LESS THAN 10 Hz

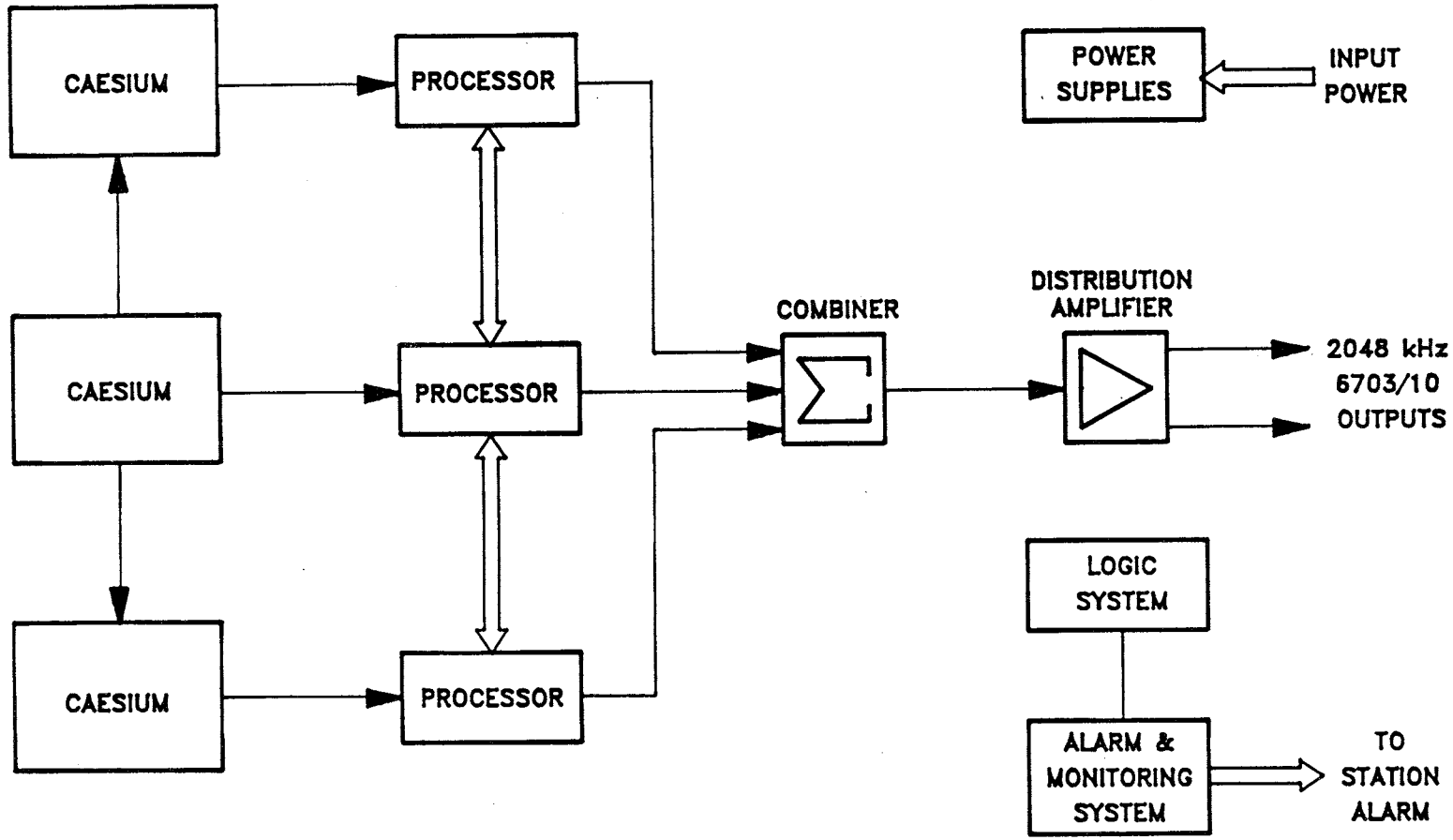
FREQUENCY DEPARTURE: TREND LINE

BASIC CONCEPTS

JITTER, WANDER, FREQUENCY DEPARTURE AND SLIPS

Fig. 4
BLK 2572

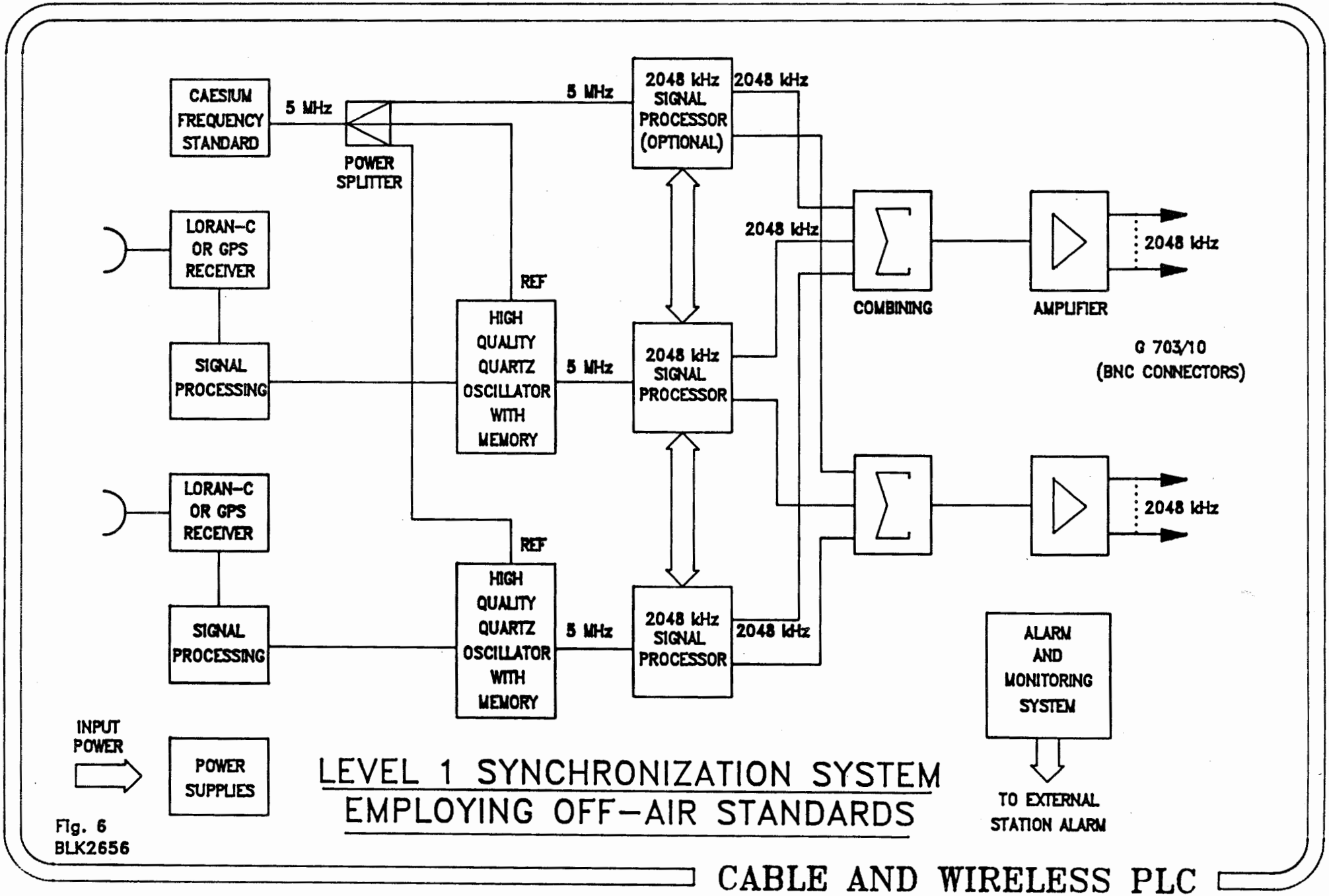
ACKNOWLEDGEMENT: REPRODUCED FROM ANSI DOCUMENT

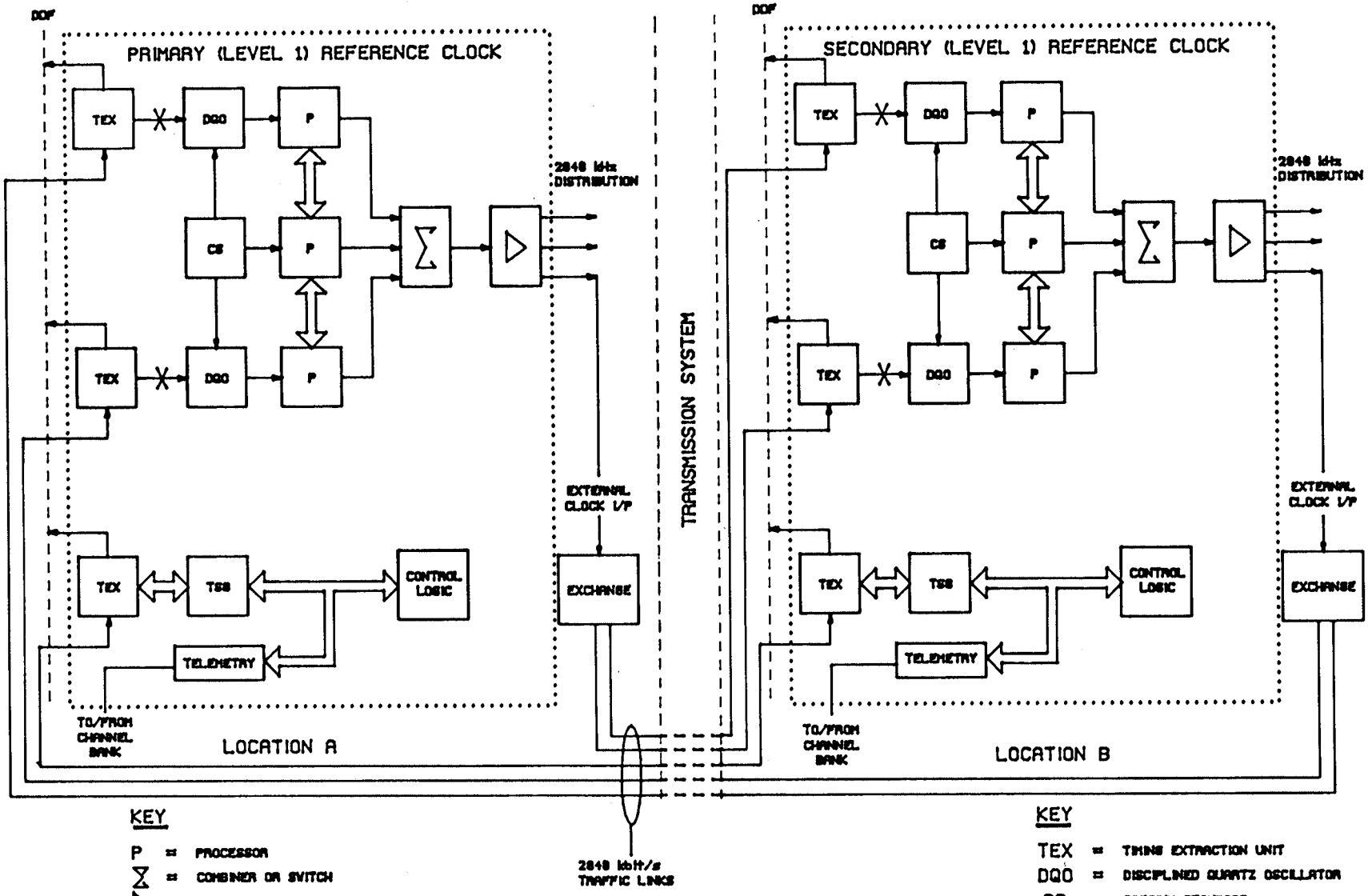


LEVEL 1 SYNCHRONIZATION (3 CHANNELS)

FIG. 5
BLK 2596

CABLE AND WIRELESS PLC


CAESIUM FREQUENCY STANDARD
5 MHz
POWER SPLITTER
5 MHz
2048 kHz SIGNAL PROCESSOR (OPTIONAL)
2048 kHz
2048 kHz
LORAN-C OR GPS RECEIVER
SIGNAL PROCESSING
REF
HIGH QUALITY QUARTZ OSCILLATOR WITH MEMORY
5 MHz
2048 kHz SIGNAL PROCESSOR
2048 kHz
COMBINING
AMPLIFIER
2048 kHz
G 703/10
(BNC CONNECTORS)
LORAN-C OR GPS RECEIVER
SIGNAL PROCESSING
REF
HIGH QUALITY QUARTZ OSCILLATOR WITH MEMORY
5 MHz
2048 kHz SIGNAL PROCESSOR
2048 kHz
2048 kHz
ALARM AND MONITORING SYSTEM
TO EXTERNAL STATION ALARM
INPUT POWER
POWER SUPPLIES
LEVEL 1 SYNCHRONIZATION SYSTEM
EMPLOYING OFF-AIR STANDARDS
Fig. 6
BLK2656
CABLE AND WIRELESS PLC

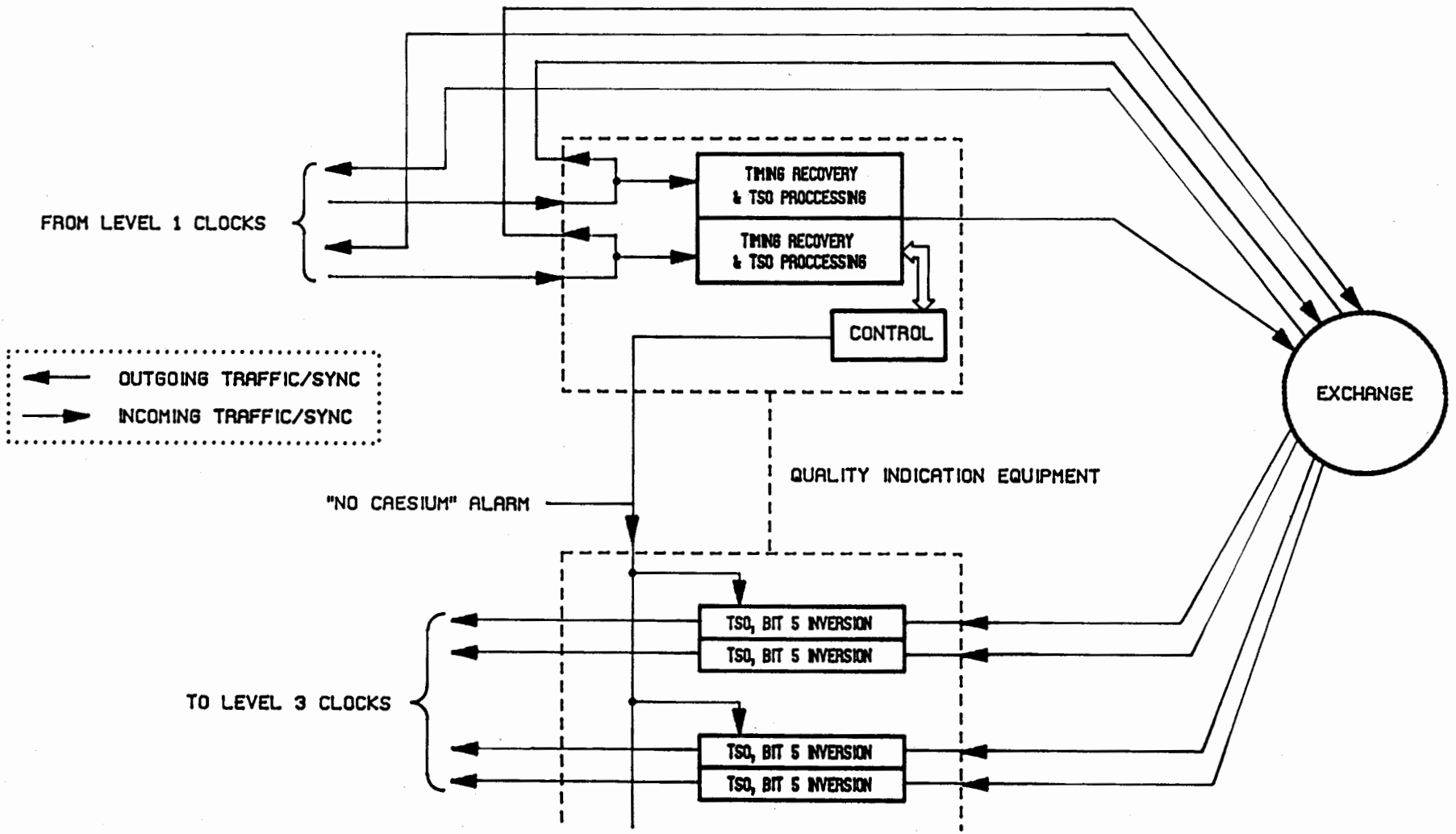


DISTRIBUTED CLOCK SYSTEM

Fig. 7

BLK 2593

CABLE AND WIRELESS PLC



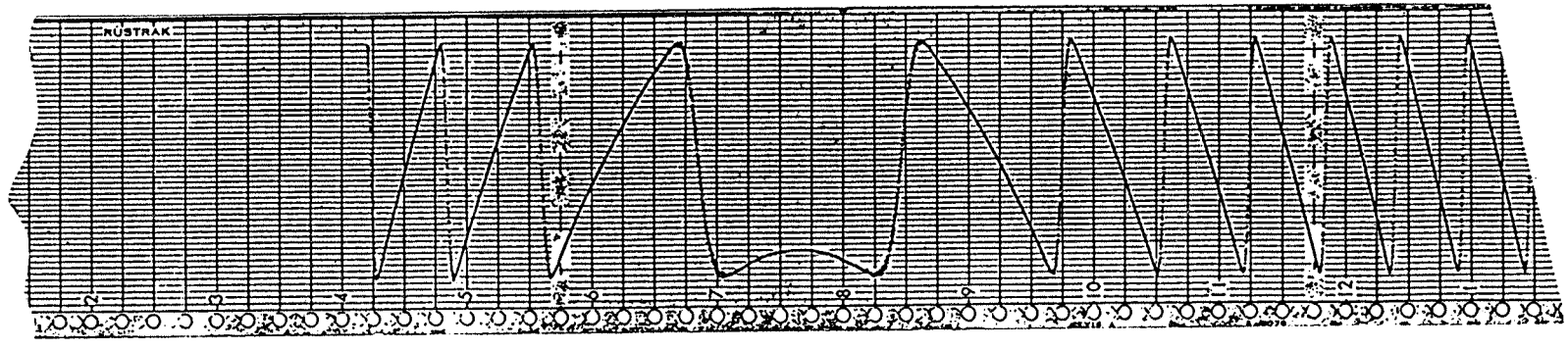
NOTE 1. THE TSO, BIT 5 INVERSION MODULE OF THE QUALITY INDICATION EQUIPMENT MAY BE PROVIDED AS AN OPTION WITHIN LEVEL 2 CLOCK SUBSYSTEMS FOR INDICATION TO LEVEL 3 CLOCK SUBSYSTEMS.

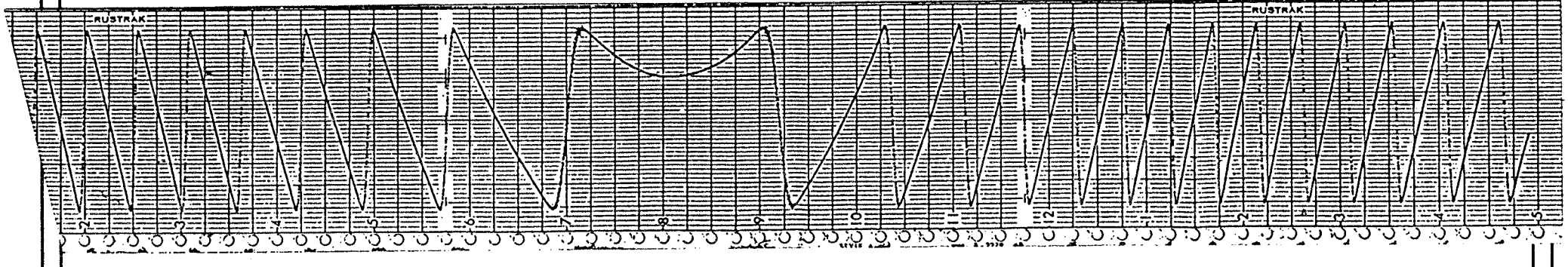
HIGH QUALITY OR MEDIUM QUALITY, LEVEL 2 OR LEVEL 3 CLOCK SUBSYSTEM

Fig. 8

BLK 2594

CABLE AND WIRELESS PLC





EFFECT OF DOPPLER ON TIME INTERVAL ERROR MEASUREMENT

Fig. 9
blk 2668

CABLE AND WIRELESS PLC

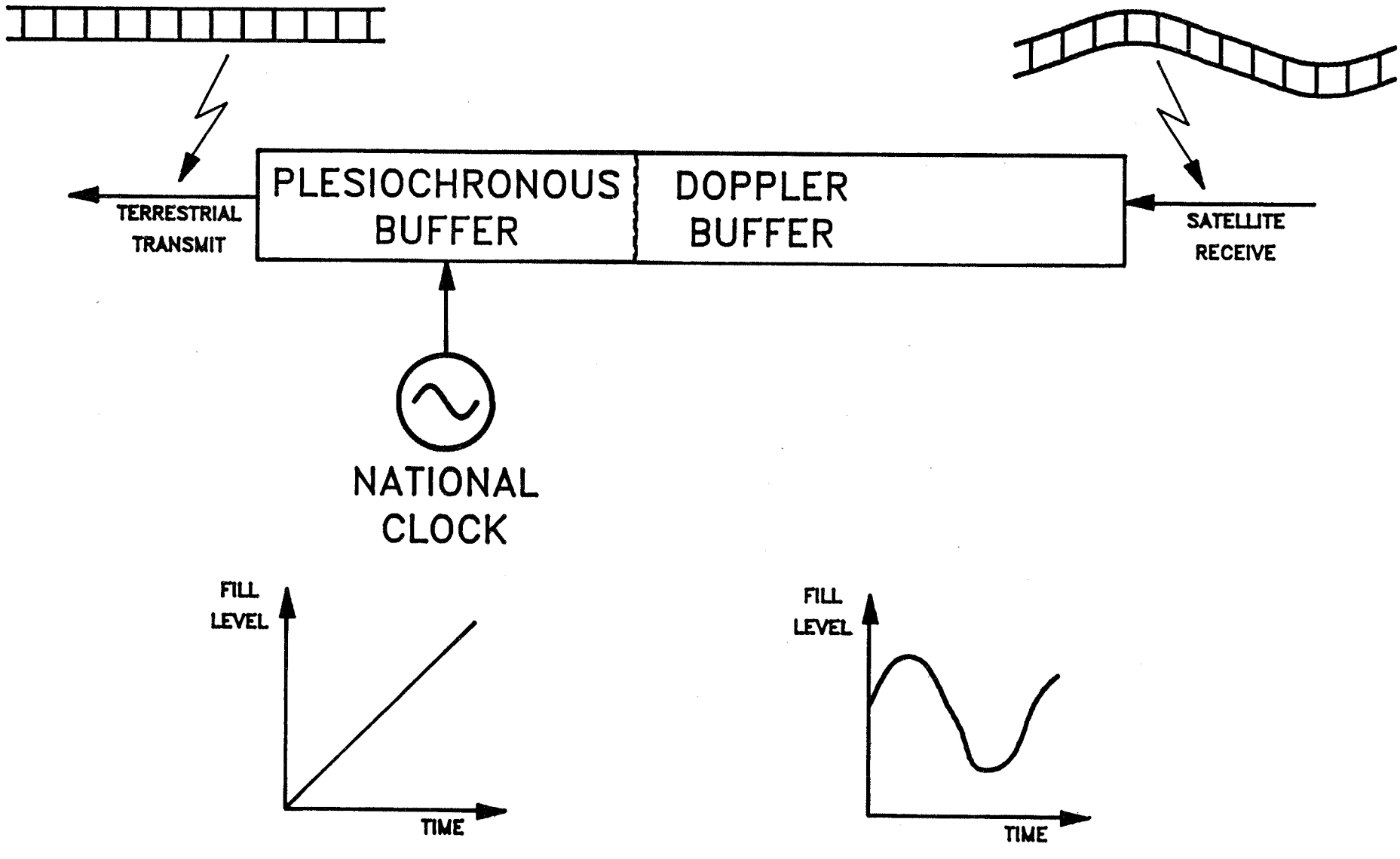
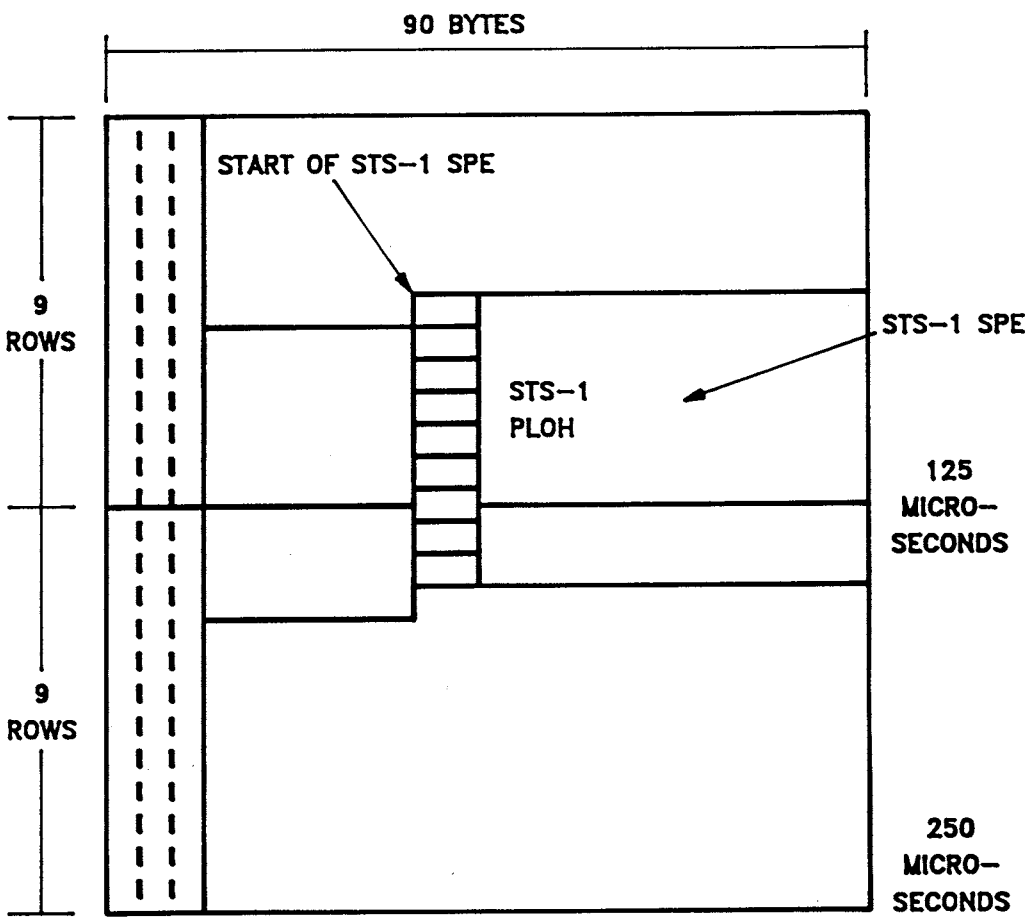


Fig. 10
blk 2667

BUFFERING

CABLE AND WIRELESS PLC



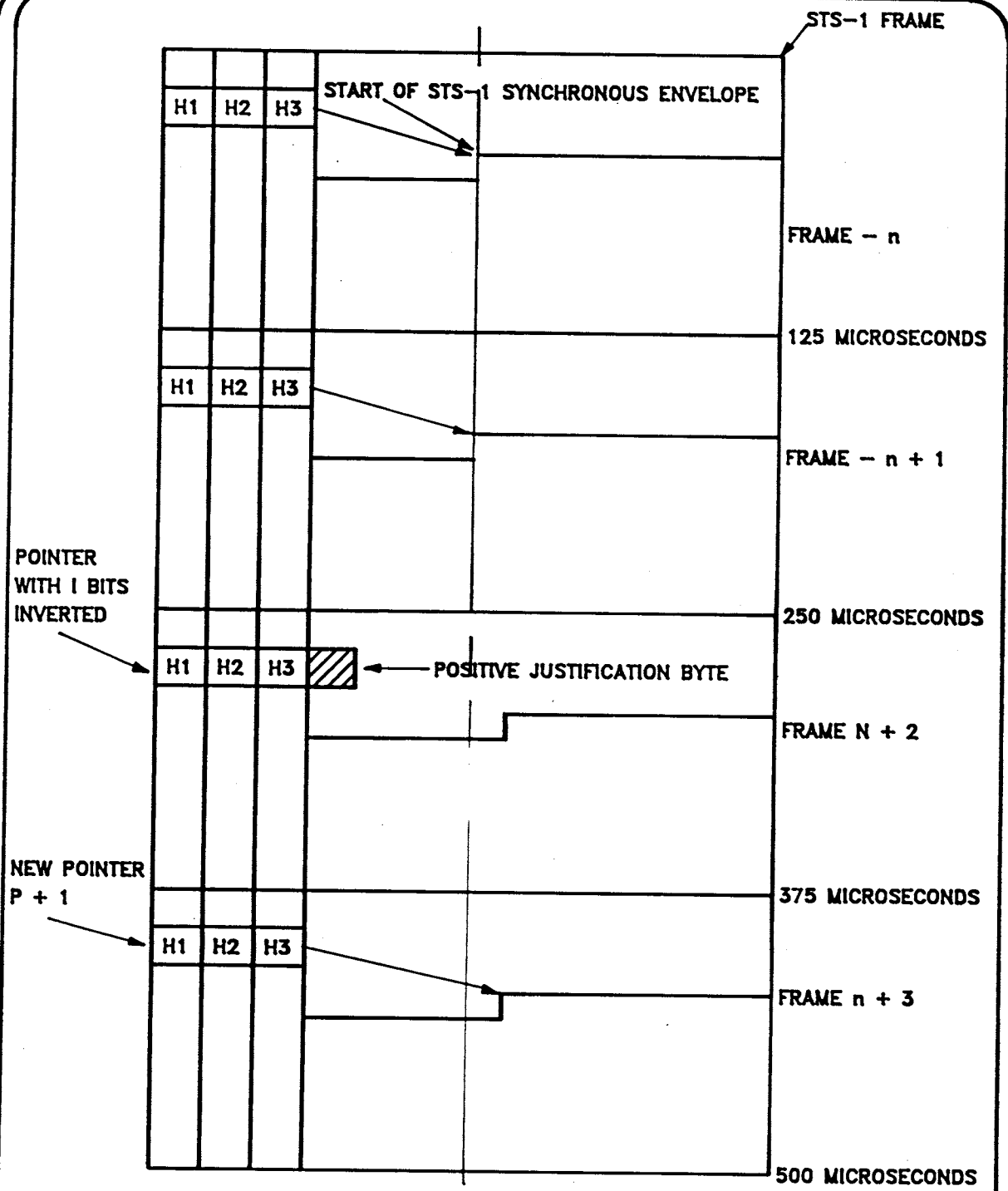
SPE–SYNCHRONOUS PAYLOAD ENVELOPE
PLOH–PATH LAYER OVERHEAD

START OF STS-1 SPE WITHIN STS-1 FRAME

FIG. 11
BLK 2568

CABLE AND WIRELESS PLC

ACKNOWLEDGEMENT: REPRODUCED FROM CCITT REC.G77x (NOW G70z)



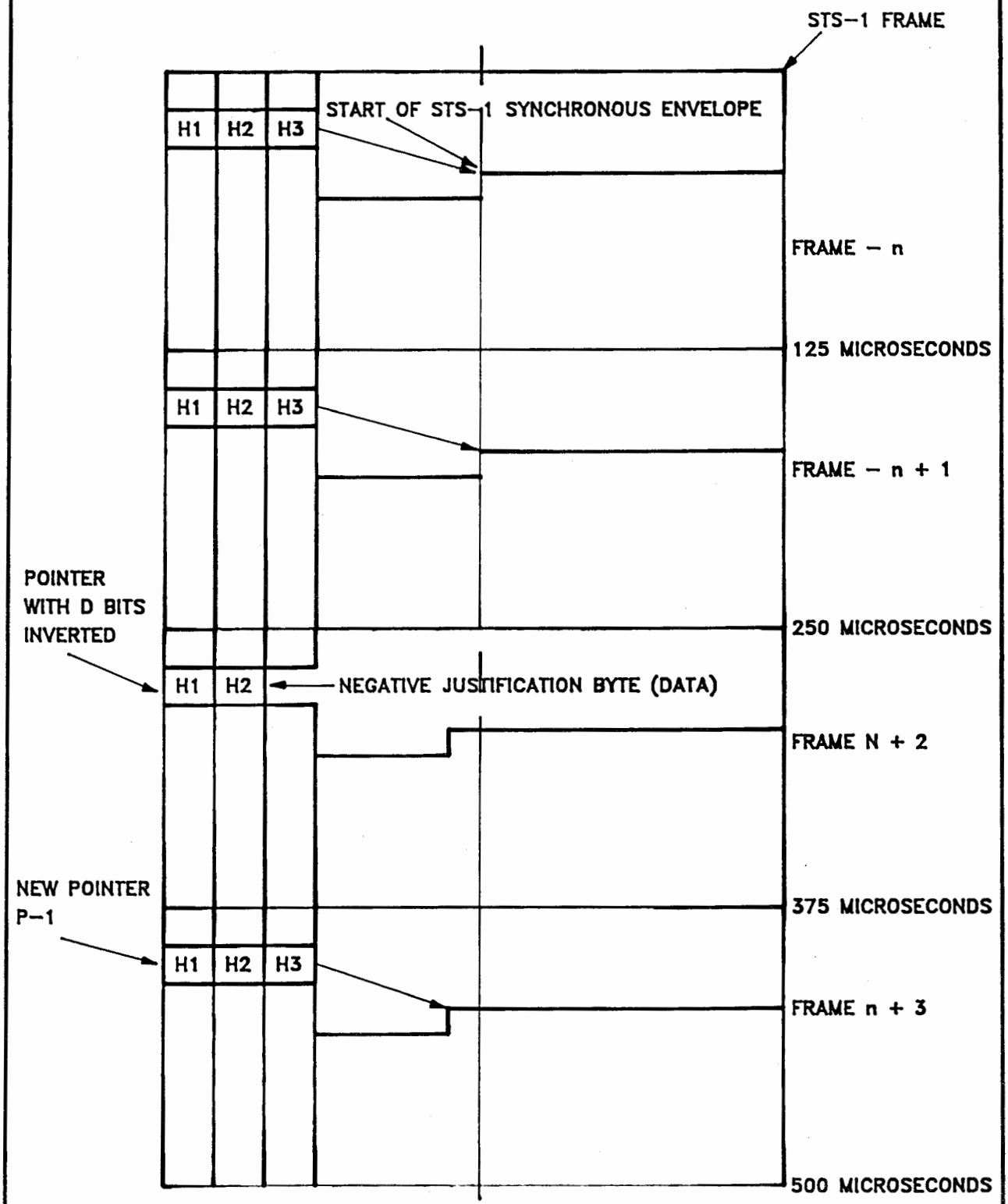
POSITIVE STS–1 POINTER ADJUSTMENT

OPERATION

FIG. 12
BLK 2570

CABLE AND WIRELESS PLC

ACKNOWLEDGEMENT: REPRODUCED FROM CCITT REC. G77x (NOW G70z)

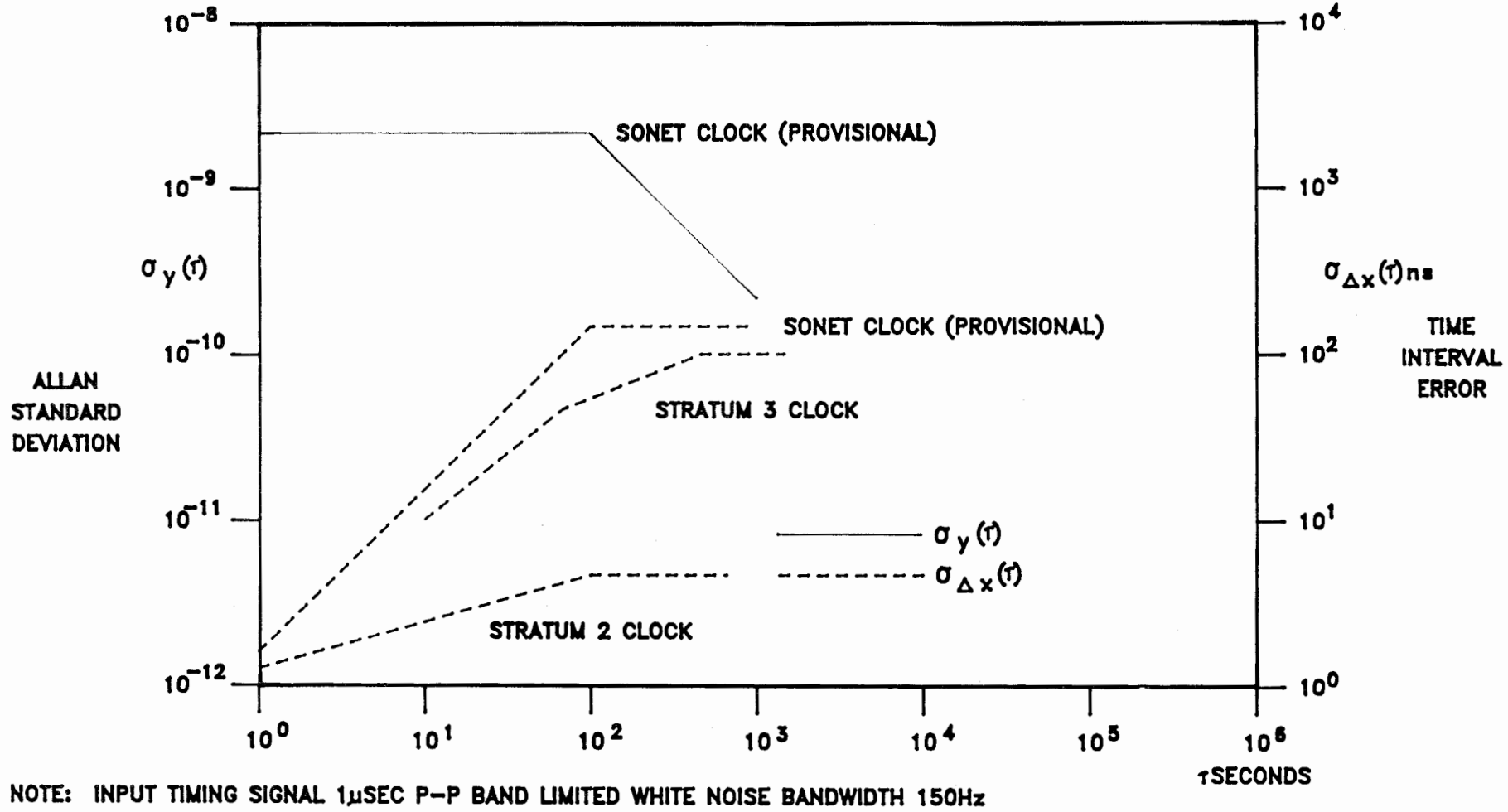


NEGATIVE STS–1 POINTER ADJUSTMENT OPERATION

FIG.13
BLK 2569

CABLE AND WIRELESS PLC

ACKNOWLEDGEMENT: REPRODUCED FROM CCITT REC. G77x (NOW G70z)



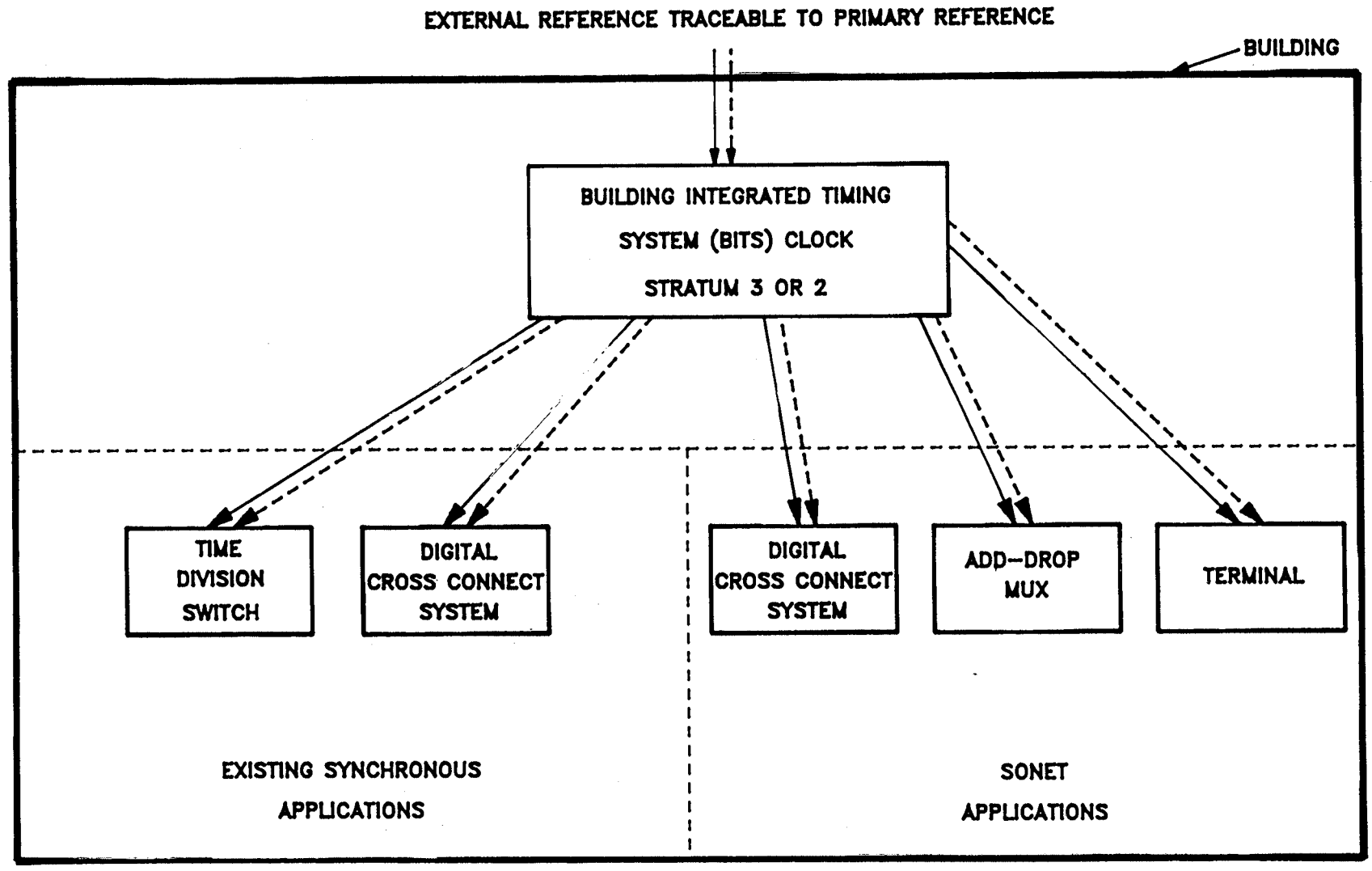
NOTE: INPUT TIMING SIGNAL 1μSEC P–P BAND LIMITED WHITE NOISE BANDWIDTH 150Hz

TYPICAL CLOCK SHORT TERM STABILITY

Fig. 14
BLK 2575

CABLE AND WIRELESS PLC

ACKNOWLEDGEMENT: REPRODUCED FROM ANSI DOCUMENT



EXAMPLE OF BITS ARCHITECTURE

Fig. 15
BLK 2573

CABLE AND WIRELESS PLC

ACKNOWLEDGEMENT: REPRODUCED FROM ANSI DOCUMENT

2 ND EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 16 - 17 - 18 MARCH 1988

Session Invited Number 6

TIMING OF A EUROPEAN WORLDWIDE SATELLITE-BASED NAVIGATION SYSTEM

H. A. Laue

ESA
European Space Operations Centre
Darmstadt

Table of Contents

1.0 INTRODUCTION
1.1 Background
1.2 Applications of Satellite Navigation

2.0 Navigation and Timing
2.1 Basic Principles
2.1.1 Active navigation systems
2.1.2 Passive navigation systems
2.2 Time Distribution Options
2.2.1 System time distributed over the space segment
2.2.2 System time distributed over the ground segment
2.2.3 System time distributed over space and ground segment
2.3 Position Determination Accuracy

3.0 NAVSAT System Concept
3.1 System Requirements
3.2 Orbital Configuration
3.3 Navigation Signal
3.4 Frequency Utilisation
3.5 Control Segment
3.6 Time-keeping Specifications

4.0 Conclusions

5.0 REFERENCES

List of Tables

Table 1. Navigation Error Budget
Table 2. NAVSAT HEO Orbit Characteristics
Table 3. System Build-up Options
Table 4. TDMA characteristics
Table 5. PN Signal Characteristics
Table 6. NAVSAT Frequency Plan

List of Illustrations

Figure 1. Passive Navigation Principle
Figure 2. Passive Navigation with Space-based System Time
Figure 3. Passive Navigation with Ground-based System Time
Figure 4. NAVSAT GEO + HEO Constellation
Figure 5. Control Segment Configuration

1.0 INTRODUCTION

1.1 Background

Since a number of years it has become apparent that demands for navigation aids are growing, in terms of increased numbers of potential users, more stringent requirements on achievable positioning accuracy ,and need for larger areas of availability.

Terrestrial radio-navigation aids have in the past been set up to cater for the requirements of specific user groups, applications, coverage areas and accuracies, resulting in the parallel existence of a number of disparate and largely incompatible systems.

The extension of those existing systems to meet the growing demands is generally not feasible due to the limitations of the terrestrial basing, which has as consequence that the requirements of wide area coverage and high positioning accuracy are mutually exclusive.

The logical solution to this dilemma lies in the space basing of the navigation signal transmission system to provide truly global availability without any constraints arising from geography, the achievable accuracies then being dependent only on the selected positioning scheme and system parameters selected.

Such a space-based radio navigation system is the object of the studies pursued by the European Space Agency since 1981, resulting in an early definition of the 'NAVSAT' system (ref 1).

The present paper, after highlighting the possible applications for an international navigation system, outlines the timing principles for the proposed system and gives an overview of the system architecture as presently envisaged.

1.2 Applications of Satellite Navigation

Civilian user communities with their requirements, as they emerged during the studies at ESA (ref 2), are outlined below:

- **Land Navigation:** low user dynamics but relatively high requirements on positioning accuracy. Two basic modes of operation are presently emerging:
  - Independent user navigation in unknown areas, usually to find the optimal route to a given target. Operation in conjunction with electronically stored maps; the satellite positioning service will be required to provide an inertial reference and drift compensation to acceleration sensors used in such systems.
  - Automatic position reporting for vehicles in large fleets to aid in vehicle dispatching and fleet usage optimisation. Applies to commercial fleets, e.g. taxis, cargo and delivery vans, as well to fleets belonging to law enforcement and other government agencies. For the latter category of users, satellite navigation will have particular applications for emergency services and disaster aid.
- **Marine Vessels:** low user dynamics, low to medium accuracy
  - optimal routing
  - secure navigation
  - collision avoidance in coastal waters and harbours,
  - vessel traffic management in congested areas

- pleasure craft: needs for accurate and safe navigation as important as for commercial vessels.

- **Aviation:** very high positional accuracy and user vehicle dynamics; high position fix update rates; uninterrupted availability.
  - Commercial aviation in high-density traffic areas, e.g. the North Atlantic: optimal route planning possible by a closer spacing of aircraft.
  - Possibility to replace with a single (redundant) system a number of presently used radio-navigation aids on board, such as Vertical Omnidirectional Range (VOR), Automatic Direction Finder (ADF), Distance Measuring Equipment (DME) and Marker Beacon (MKR). Integration with flight management systems; standardisation with respect to future services in air traffic control, surveillance and precision approach.
  - General aviation.

- **Space Applications:** very high dynamics, very high accuracy
  - tracking of low earth orbiting spacecraft with high position accuracy requirements (normally requires extensive networks of ground stations also in areas where no primary data contact exists).

- **Mining, Drilling and Survey Applications:** no dynamics; very high positional accuracy; low acquisition speed.
  - Exploitation of natural resources , both land-based and offshore, in regions where terrestrial surveying aids are not available.

2.0 Navigation and Timing

2.1 Basic Principles

Most radio-navigation systems are based on triangulation from range measurements between the unknown position of the user node and the known positions of several navigation nodes. These range measurements are effected by measuring the signal travel time between user and navigation node transponders.

There is a fundamental difference in the timing requirements between active and passive navigation systems. After briefly describing the timing requirements for active systems, we shall examine in more detail the passive navigation concept, which has been chosen for NAVSAT.

2.1.1 Active navigation systems

Active systems require receivers and transmitters at both ends of the system (i.e. user and navigation nodes), and the range measurement is based on the return trip time of the signal transmitted at one end, transponded at the other end and received again at the point of origin, with could be either the user or the navigation nodes.

This principle has the advantage that timing needs are limited to relatively low requirements on clock stability. Clocks at the user node and the navigation nodes have to have sufficient stability only to give the required timing accuracy over the return trip period, i.e. typically over a fraction of a second.

Moreover, the system can always be established such that requirements for synchronisation between user clocks and navigation node clocks are virtually eliminated.

The disadvantage of active systems is the limited number of range measurements within any given period. This will most commonly limit the number of users which can be serviced by the system, but will also influence the average period a user will have to wait to obtain access.

2.1.2 Passive navigation systems

While the basic problem of active navigation is just one of measuring time, passive navigation systems are based on distributing time.

Here the user is only equipped with a receiver, and he measures ranges by comparing the time of a 'system clock', broadcast via the navigation nodes, with the time of his own clock. Since no co-operation is required between navigation nodes and user nodes, the passive system cannot be saturated and will accommodate an unlimited number of users, who can access the system at any time.

For a space navigation system the navigation nodes are implemented by a number of satellites. Assuming a basic system time scale to be kept by a single 'clock' at C (conveniently located on ground), the navigation task is equivalent to transferring system time from that system clock via the multiple space navigation nodes S to the user nodes U (Fig. 1).

The user receiver measures the signal travel time C-S-U, or the sum of the rangesR_i and r_i. From the known positions to the satellites, the ranges R_i between system clock and satellites can be computed, and the equations for ranges r_i as a function of satellite positions and unknown user position can bewritten down. With at least 3 range measurements, these equations can be solved for three position coordinates.

This simple implementation above works under the assumption, that the user clock is indeed synchronised to the system time scale. If this constraint is dropped, the user clock will have a

time-variable (depending on its stability) offset from the system clock. Measurement of signal arrival times by the user will then not yield true range , but values different from the true range by the same bias error for all navigation nodes. These values are called 'pseudo-ranges'. In order to determine the user clock offset as a fourth unknown besides the three user position coordinates, pseudo-ranges to at least four different navigation nodes are required.

In practice, the position determination is not instantaneous, but subsequent pseudo-range measurements from all satellites in view will be input to a Kalman filter which models the relative motion between user and satellites and continuously updates the position coordinates.

2.2 *Time Distribution Options*

The generic navigation system configuration shown in Fig. 1 with both C and U in view of the satellites will not be possible for real system with global coverage; hence system time will have to be relayed and stored during the distribution process to the user.

The following options can be considered for passive navigation systems:

2.2.1 System time distributed over the space segment

Military navigation systems need to be as independent as possible from ground control and ground intervention. Therefore systems like GPS are designed such, that contact with the 'Master Control Station' is required only once every 24 hours (Fig. 2).

System time is distributed over all satellite navigation nodes, i. e. each satellite carries a highly stable clock which is compared with the master clock of the Master Control Station only during the daily passes. In order to limit the daily deviation of all on-board clocks from the master clock to 3 ns rms, the on-board clocks must have a stability of 10^{-14}.

2.2.2 System time distributed over the ground segment

A civilian system like NAVSAT is not constrained by number or complexity of ground segment nodes. Here the objective is rather to limit the performance requirements for in-orbit subsystems and the risks related to degradation or failure of on-board equipment.

In the NAVSAT system the system time is thus kept in a collection of ground control centres, one of which is always in view of any satellite of the NAVSAT space segment.

The master system clock is located in one specific ground control centre, and the clocks at the other centres are synchronised to the master clock by time transfer via satellite link (Fig. 3).

Provided the configuration of satellite orbits and ground station locations is properly designed, the synchronisation throughout the total ground station network can take place more or less continuously, so that requirements for long-term stability of Regional Control Station clocks are rather moderate.

The ground control station with the longest distance from the master clock location can be reached within a maximum of three satellite hops or successive synchronisation steps. The synchronisation procedure around the globe will however be adaptive by evaluating the time offsets after a full tour of the globe at the master clock again.

2.2.3 System time distributed over space and ground segment

A further development of this principle is proposed in the GRANAS concept (ref. 6), where all ground and space based clocks form a free-running, connected network coupled at one location on ground to the system master clock.

Synchronisation of the network is achieved by comparing the satellites measuring round-trip times of signals returned from simple ground-based transponders in their view. These two-way range

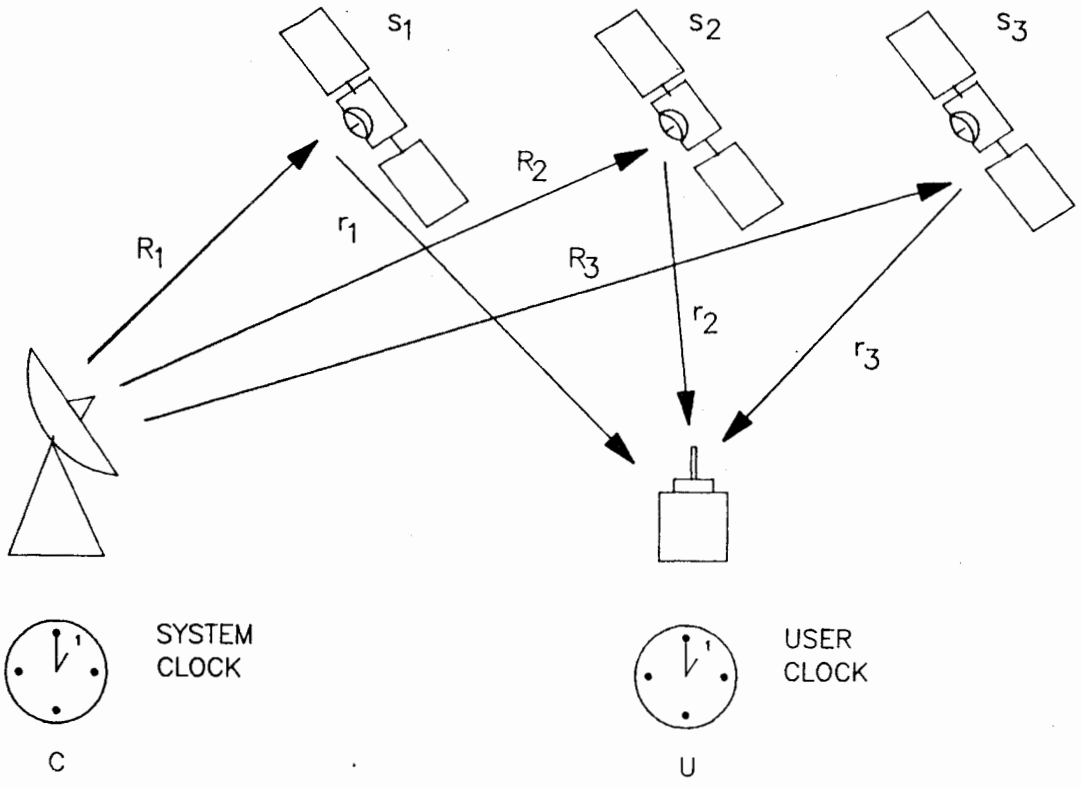


Figure 1. Passive Navigation Principle

measurements are also used for satellite autonomous position determination, so the ground stations must be configured in such a way that at least 3 stations are always in view of any satellite.

2.3 *Position Determination Accuracy*

Intuitively it is clear that the accuracy of the position determination must depend on the accuracy of the ranging, hence the timing process. It can indeed be shown that the rms error on the final user position σ_p is proportional to the radial range error σ_r, by a factor called 'Positional Dilution Of Precision', or PDOP.

The value of the PDOP factor depends on the actual geometry between the 4 satellites selected, and the users. The PDOP attains its minimum if the four satellites are spaced most evenly around the user location. For satellites located along a line with the user, in the plane of the user, or clustered close to one point, the system of equations obviously becomes ill-conditioned, and the PDOP becomes very large.

Typical operational values for the PDOP lie between 2 and 6 for the satellite configurations under consideration, depending on geographical location of the user and on the time.

The contributions to a typical error budget for a satellite-based navigation system is shown in the following table, together with the 1-σ values tentatively allocated to these sources in the NAVSAT system proposal.

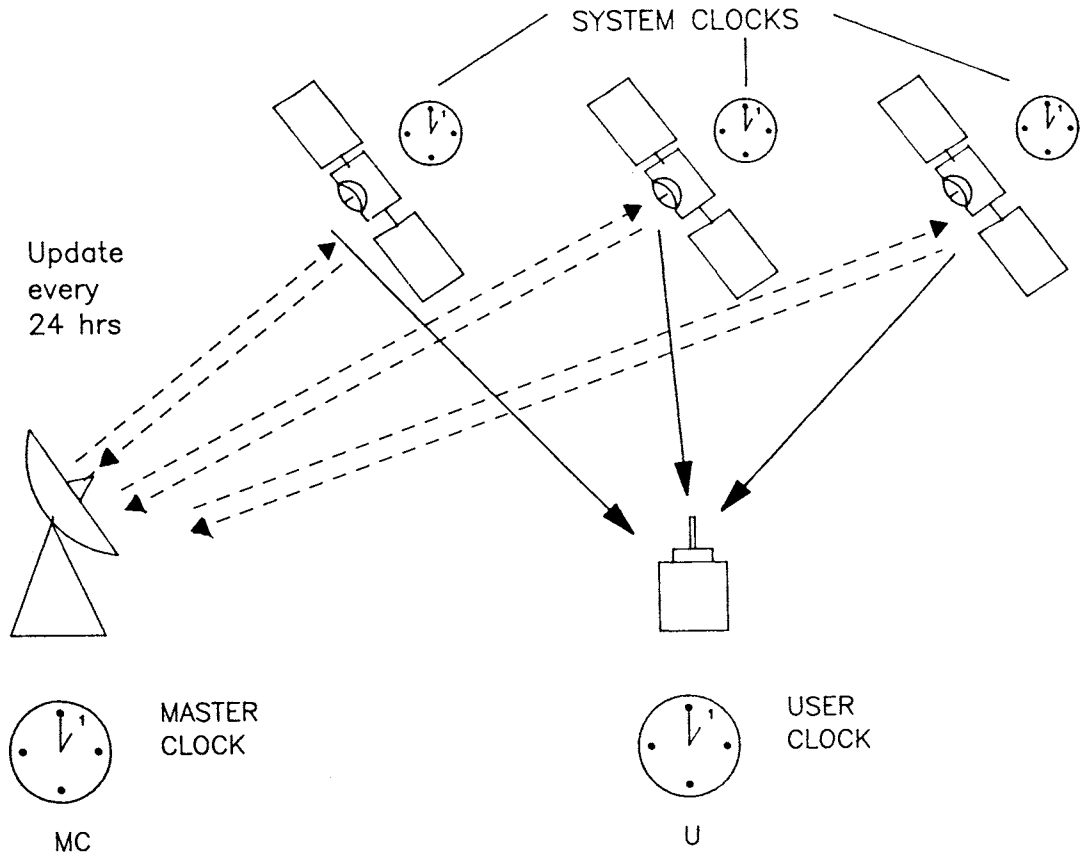


Figure 2. Passive Navigation with Space-based System Time

SOURCE / EFFECT	EQUIVALENT	USER ERROR
	time (ns)	range (m)
Control Segment		
Synchronisation to system clock	3	1
Satellite position knowledge	4.5	1.5
Space Segment		
Uplink delays C-band (residual ionospheric and tropospheric)	0.45	0.15
Satellite transponder delay	2.4	0.8
User Segment		
Residual ionospheric downlink delays	9	3
PN code acquisition resolution	3	1
User clock stability	3	1
User computation resolution	3	1
User dynamics contribution	4.5	1.5
RMS Total	12.6	4.2

Table 1. Navigation Error Budget

With a pessimistic PDOP value of 6, this translates into a 1-σ user position error of approximately 25 m.

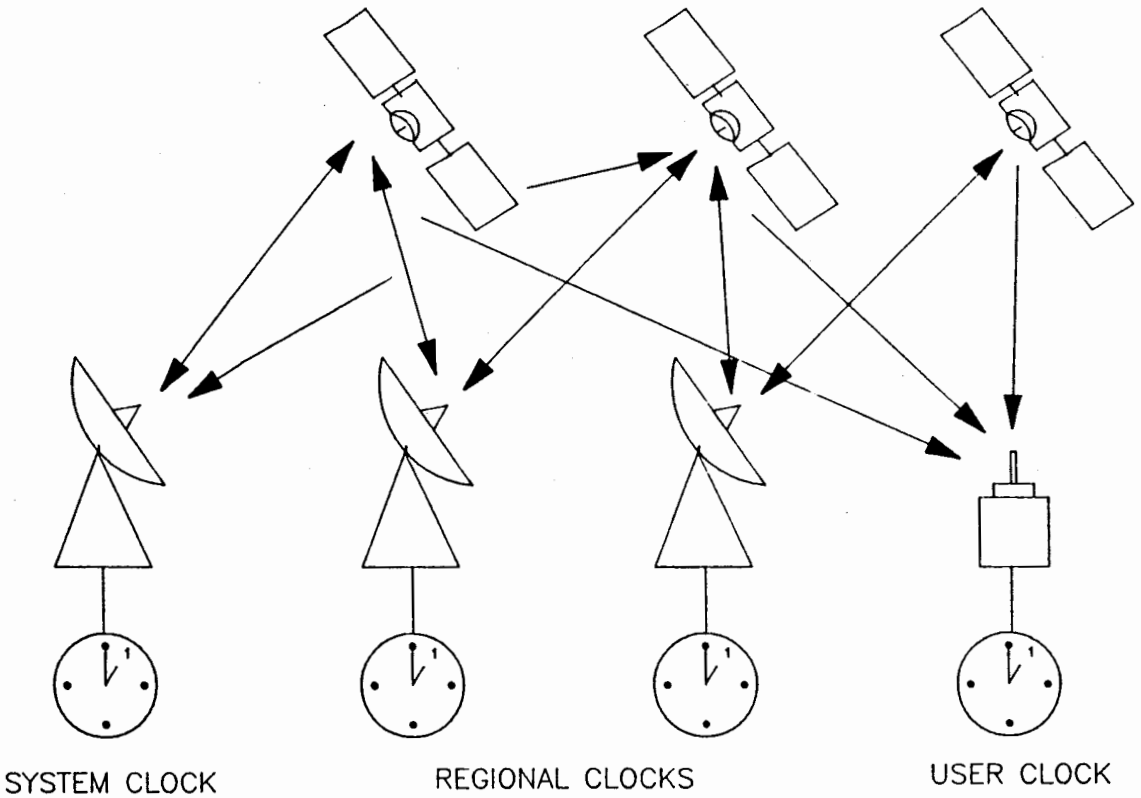


Figure 3. Passive Navigation with Ground-based System Time

3.0 NAVSAT System Concept

The concept chosen for the proposed NAVSAT system is based on the passive navigation principle with a distributed ground segment.

In order to allow the use of simple, low-weight and low-power navigation payload packages, the overall system complexity with respect to the generation and synchronisation of the navigation signals has been shifted from the space segment to the control segment as far as possible, leading to a dramatic reduction in space segment complexity and cost as compared with, e.g., systems like GPS.

Following this principle, the space-borne navigation payload will act simply as relay for the navigation signal generated in the ground stations, which requires all satellites to be continuously in view of at least one ground station, except for the period of perigee passage of the HEO satellites. This feature has a number of important operational advantages:

- the Control Segment can detect, diagnose and correct space segment problems in real time;
- users are in continuous contact with the Control Segment for communications, position reporting or distress signalling, even at polar latitudes;
- the navigation data can always carry the most up-to-date information for satellite ephemeris, clock correction and ionospheric modelling;
- users can be informed quasi-instantaneously of any degradation or malfunction in the system (this is a mandatory requirement for operational air navigation systems).

The system configuration summarised below (for more details see refs. 8 and 9) is the result of studies undertaken in 1986/1987 on the basis of earlier concepts developed by ESA (ref. 1) and by DFVLR (ref. 6).

3.1 System Requirements

In accordance with the needs of the various user communities outlined in Paragraph 2 above, the following functional requirements have been established for the proposed global international system:

- three-dimensional navigation plus time-synchronisation capability at any point on the globe for users with receive-only equipment
- continuous availability of the navigation service
- accommodation in the system design of global communications and search and rescue capabilities for users with the corresponding equipment
- accommodation in the system design of global high-precision positioning in deferred time for near-stationary users with the corresponding equipment.

Performance requirements have tentatively been set for three distinct groups of applications, in terms of three-dimensional positioning accuracy (rms):

1. Low user dynamics, low cost equipment (Doppler receiver): 100 m, acquisition time 1 min.
2. High user dynamics, full performance equipment: 10 m, acquisition time 2 sec.
3. (Near) stationary user with high-precision equipment: 10 cm, acquisition time 3 hrs.

An important additional requirement on the system is the capability of planned performance growth within the initial system architecture, i.e. the possibility to integrate future technological improvements in various subsystems to obtain improved system performance, without having to replace the complete system.

3.2 *Orbital Configuration*

The orbital configuration of NAVSAT, which originally was similar to that of GPS, was reassessed in order to identify configurations which could be built up in a modular fashion and which would take advantage of the existing and planned investment of the international space community in satellite systems in the geostationary orbit.

This has led to the definition of a configuration (Fig. 4) which consists of a combination of satellites in elliptical orbits at high inclinations, to cover the polar regions, together with geostationary positions, where the navigation package could be an additional payload on satellites otherwise used for communications (Ref. 5) . This concept allows to limit the total number of satellites to 18, 6 of which are geostationary and need not to be dedicated to the NAVSAT system, i.e. offer further potential for savings.

The elliptical orbits, which will have to be used for the dedicated NAVSAT spacecraft will have the following characteristics:

Perigee altitude	1250 km
Apogee altitude	39105 km
Period of revolution	12 h
Right ascension asc. node	100/220/340 deg
Inclination	63.45 deg
Argument. of perigee	90/270 deg
Operational altitude	> 10625 km
Operational period	20 h/day

Table 2. NAVSAT HEO Orbit Characteristics

Such a system would, with a limited number of spacecraft, provide a pre-operational, or demonstration capability over selected areas of coverage.

A minimum configuration in this concept consists of two geostationary payloads and three payloads in the elliptical orbits. With properly placed elliptical orbits such a configuration will cover Europe, the North Atlantic and the western US and Canada for 50% of the time. The addition of only two more geostationary payloads will extend this capability to the western Pacific area.

Further possibilities for system build-up in a modular way are shown in the table below, which lists coverage in terms of regions and of percentage of daily availability as a function of the number of satellites in the different orbit types.

Satellites in

GEO	HEO	Total	Region covered	Time %
2	3	5	North Atlantic /Europe	50
4	3	7	N Atlantic/Europe/NW Pacific	50
2	6	8	North Atlantic/Europe	100
6	3	9	All Northern Hemisphere	50
4	6	10	N Atlantic/Europe/NW Pacific	100
6	6	12	All Northern Hemisphere	100
6	12	18	Total Globe	100

Table 3. System Build-up Options

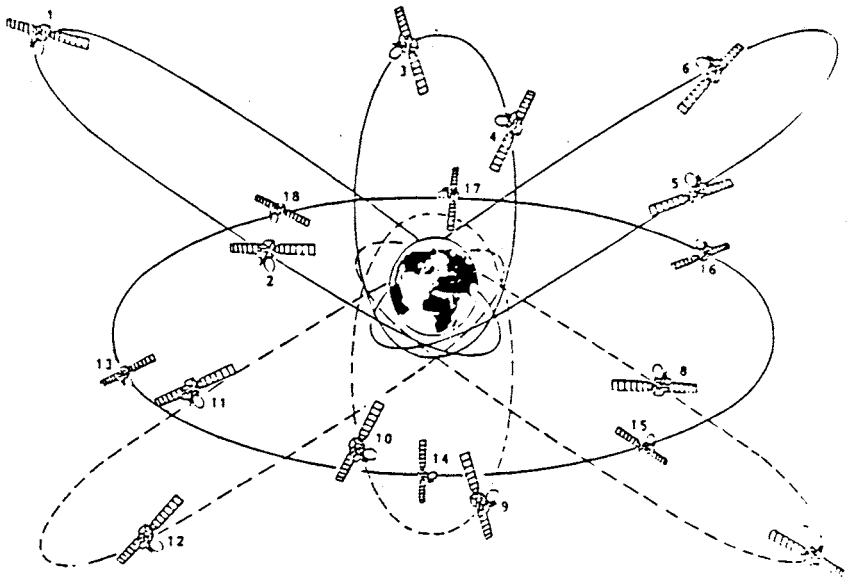


Figure 4. NAVSAT GEO + HEO Constellation

3.3 *Navigation Signal*

Candidates for transmitting navigation signals from different satellites to a user are:

- Frequency-Division Multiple Access (FDMA)
- Code-Division Multiple Access (CDMA)
- Time-Division Multiple Access (TDMA)

FDMA systems are not useful for the particular application because of their high bandwidth requirements (due to Doppler guarding requirements) and the necessary user receiver complexity. Military systems like GPS use a CDMA spread spectrum scheme to allow very fast position update, while providing high resistance against intentional interference.

For NAVSAT, the navigation signal is planned to use a TDMA channel sharing, in which each pair of satellites in opposite orbital positions is given a single time slot for transmission of the navigation signal within a total TDMA frame of 9 slots. A guard period with no transmission is foreseen between slots in order to avoid collisions due to slant range differences.

The characteristics of the TDMA scheme are given in Table 4 below:

Frame duration	3150 ms
Slots / frame	12
Slot duration	230.6 ms
Guard period	120 ms

Table 4. TDMA characteristics

Within each slot, the signal structure will consist of a CW carrier to aid signal acquisition and for Doppler ranging, and the high-bandwidth pseudo-noise (spread spectrum) signal for precise ranging, onto which the low bit-rate data message (70 bits per burst) will be modulated. It is expected that satellite orbits can be predicted accurately enough over periods of several weeks that the data can be distributed off-line to users with lower accuracy requirements who use the Doppler ranging mode only.

PN/data burst duration	217.6 ms
PN chip rate	5 Mcps
PN code period	2^{19}-1 chips
Data transmit rate	321.7 bps
Average data rate	22.22 bps

Table 5. PN Signal Characteristics

3.4 *Frequency Utilisation*

The frequency plan for the NAVSAT system takes account of the major functions of the system:

i) User navigation with dual-frequency ionospheric correction

ii) Communications relay between Control Segment (CS) and user

iii) Satellite Telemetry, Tracking and Control (TT&C).

The necessary RF links are summarised in Table 6 below.

SERVICE	LINK	BAND	DESIG-NA-TION	F (GHz)	BANDWIDTH
Navigation	CS --> Sat	C	C1	5	20 MHz
Navigation	Sat --> User	L	L1,L2	1.2-1.5	20 MHz
Communication	User --> Sat	L	L3	1.6	1.5 kHz
Communication	Sat --> CS	C	C2	4	1.5 kHz
Communication	CS --> Sat	C	C4	6	1.5 kHz
Communication	Sat --> User	L	L4	1.5	1.5 kHz
TT&C	CS --> Sat	C	C5	6	2 kHz
TT&C	Sat --> CS	C	C6	4	4 kHz

Table 6. NAVSAT Frequency Plan

It should be noted that the navigation signal Sat--> User is of course also received by the Control Segment for synchronisation and tracking purposes, and that in general the navigation and TT&C downlinks emanating from one satellite can be received from all ground stations in view.

3.5 *Control Segment*

The NAVSAT Control segment will consist of a number of geographically dispersed Regional Control Stations (RCS), which provide the global coverage, and a Mission Control Centre (MCC) controlling the overall conduct of the mission. Communication via the geostationary satellites of the NAVSAT space segment will be used for coordination of the Regional Control Stations and system-wide synchronisation.

For the purpose of keeping operational satellites in view of at least one RCS at all times, six such stations are required, and a tentative distribution is shown in Figure 5.

NAVSAT SYSTEM CONCEPT

The tasks of the control segment with respect to time measurement and time keeping in the widest sense are the following:

System Tasks (MCC)

i) establish a common system time scale and synchronise all Regional Control Stations to that time scale (this time scale should be synchronised to UTC to better than 100 ms);

ii) determine all satellite orbits, and distribute predicted satellite ephemerides to the Regional Control Stations;

iii)globally synchronise the TDMA transmission timing for all regional Control Stations and all satellites.

Local Tasks (RCS)

i) measure and predict slant range to controlled satellites, including tropospheric and ionospheric delays;

ii) synchronise the station clock with the clocks of the neighbouring stations;

iii)generate and uplink the navigation signal in the proper TDMA sequence and apply forward correction for time delays and Doppler;

iv) synchronise the handover of navigation signal generation with the neighbouring stations.

The time synchronisation of the different Regional Control Stations will be effected with the help of the geostationary satellites alone, each of which covers a longitude interval on the equator of ca. 160 deg.

3.6 Time-keeping Specifications

Neighbouring Regional Control Stations shall be synchronised with each other to better than 2 ns; the synchronisation of any RCS with the MCC shall be better than 3 ns. For a synchronisation using the C-band links, this appears to be no problem.

The detailed specifications for the clock performances in the different system elements are not complete. The preliminary assessment (ref 8) has led to the following assumptions for the types of clocks employed and their stabilities:

Satellite	VCXO (10^{-13} / 100s)
User set	VCXO (10^{-12}/ 100s)
Regional Control Station	Cs (10^{-13}/ 1 day)
Mission Control Centre	Cs/H (10^{-14}/ 1 day)

Whether Regional Control Station clocks would have to be equipped with Hydrogen masers as well, requires further assessment.

An alternative to the original NAVSAT concept with simple satellite repeaters was analysed in a study performed by Oscilloquartz SA of Switzerland (ref 3).

In this alternative time and frequency are in fact transferred from the ground station to the satellite VCXO for on-board navigation signal generation. OSA designed a synchronisation system based on hydrogen maser time references in the ground stations and up- and down-links in C-band. The study has shown that a time stability of better than 100 ps and a frequency stability of 10^{-14} can be transferred to the satellite clock in this way, while introducing no significant errors during handover from one RCS to the next.

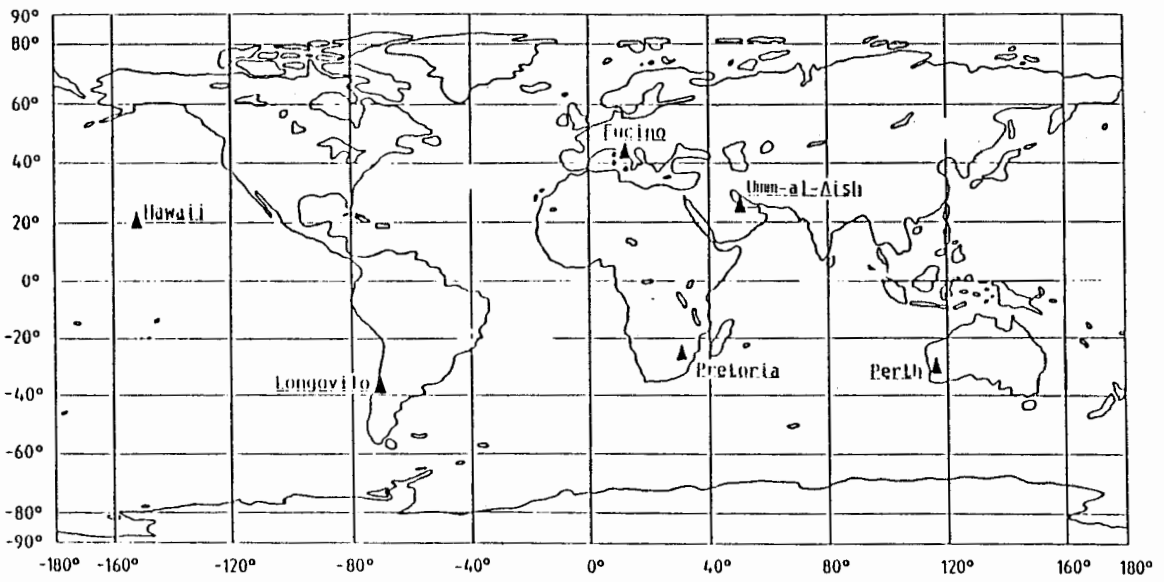


Figure 5. Control Segment Configuration

4.0 Conclusions

Global navigation by satellites is becoming a necessity for a growing number of civilian user communities. Time measurement and time keeping requirements are the foundations of any system design for such space navigation systems.

The European NAVSAT proposal presently under study could provide the basis for a worldwide civilian system, taking advantage in its system concept of the absence of design constraints imposed by a military usage.

The system concept could be demonstrated in a pre-operational regional configuration as a precursor to a complete global, international system. The system architecture will now have to evolve into detailed system specifications, in particular as regards the integrity of the overall timing concept.

5.0 REFERENCES

1. Diederich, P., Laue, H. A., Rosetti, C., 'NAVSAT - A Global Civil Navigation Satellite System', 1984 Conference of The Royal Institute of Navigation, London, May 1984

2. De Boer, G.A., Van der Matten, M. et al, 'Economic Study for NAVSAT', Final Report of ESA study contract 6022/84/F/RD, October 1984

3. Busca, G., 'NAVSAT Time Synchronisation Simulation Study', Final Report of ESA study contract 5804/84/F/RD, January 1986

4. Dekker, G.J., de Pagter, P.J. and Joosten, L.J.M., A Simulation Tool for the Study of Satellite Based Navigation Systems, Proceedings of 15th ISTS, Tokyo, in publication

5. Carnebianca, C., Mastini, G. , Rizzo, R., 'NAVSAT System Configuration Study , Final Report of ESA study contract 5754/84/F/RD, January 1985

6. SEL AG, GRANAS - Global Radio Navigation System, Final report of DFVLR study 01YY0335, Stuttgart, April 1984

7. Rosetti, C. , 'NAVSAT: A Worldwide Civil Satellite Navigation System , IAF-83-90

8. SEL AG (prime contractor), 'Comparative Study of NAVSAT Configurations , Final Report of ESA study 6802/86/F/RD, Stuttgart, November 1987

9. ITALSPAZIO, 'NAVSAT GEO + HEO Configuration', Final Report of ESA Study 6354/85/F/RD, Rome July 1985.

Session Invited Number 7

Very Long Baseline Interferometry (VLBI)

Ralph Spencer
University of Manchester
Nuffield Radio Astronomy Laboratories
Jodrell Bank
Macclesfield
Cheshire
SK11 9DL
U.K.

Secretary of the European VLBI Network Programme Committee

The advent of atomic clocks has directly led to the development of the powerful technique of VLBI. This method of radio observation is used for the study of the angular structures of celestial radio sources at high resolution, the measurement of accurate positions of such sources and the investigation of tectonic plate movement in geophysics.

The technique arose as a development of the radio interferometer, pioneered by Ryle & Vonberg 1952, where two radio telescopes are connected together. The resolution of such an instrument is λ/d where λ is the wavelength and d the separation of the two telescopes. The advent of high stability atomic clocks and sophisticated magnetic tape recording techniques has enabled the separation between the telescopes to be extended to trans-global baselines, and thereby achieving resolutions of better than 1 milliarcsec at wavelengths of a few cm. To put this figure in perspective, 1 milliarcsec corresponds to a 1 Swiss Franc piece at a distance of 3000 km, or in astronomical terms about 0.1 light year in galaxies at distances of about a hundred million light years.

The production of detailed images or radio maps of the sources needs simultaneous observations over several baselines of different lengths and orientations. This aperture synthesis method requires the collection of as much of the resolution plane data as possible, followed by Fourier transformation to give the final image. This obviously requires the use of many radio telescopes throughout the world and indeed the largest experiment to date involved the use of 18 telescopes in the USSR, Europe and the USA. The unusual radio galaxy M87 in the constellation of Virgo was observed in this experiment and a detailed map of the knots of radio emission in the nucleus and jet has been obtained. Unusually

the radio knots coincide with optical emission regions, suggesting that the radiating mechanism may be the same in both cases.

One of the most exciting discoveries in VLBI has been finding the so called superluminal sources. These are objects where the observed separation between clouds of emission in the radio appear to increase at a velocity which is greatly in excess of the speed of light. This is of course a trick, Einstein need not worry; it is an apparent effect and is due to a component of the motion being in the line of sight to the observer. A fast cloud can partially keep up with its radiation emitted at an earlier date and so the true separation is much less than that apparently observed. The velocity of the clouds is still high - some 90% or more of the speed of light, but not so high as to contravene a basic tenet of relativity! Several 10's of sources have been found to date to exhibit this 'superluminal' motion so it is just as well that they are reasonably well understood.

In addition to the investigation of fine scale structure in radio sources, VLBI can also be used for astrometry. Absolute positions to accuracies of a few milliarcsec, and relative positions to a few tens of microarcsec can be obtained. This is a much greater precision than can be obtained optically.

The measurement of the distance between the telescopes is related to astrometric analysis. Already the position of telescopes on the Earth is known to a few cm, and relative positions are known to a few mm. It is hoped that accuracies of around 1 mm will be obtained in the 1990's. The use of hydrogen masers with their excellent stability on all time scales are essential for this work. Other methods of geodetic measurement cannot yet approach the accuracy of VLBI and so the technique is being actively developed as a method for studying plate movement, mountain buuilding, earthquakes, Earth rotation and polar motion.

This all makes VLBI sound relatively easy to do given the necessary equipment. However as with most forms of human endeavour anything that can go wrong will, and we can devise 10 commandments of how not to do VLBI:

1) Observe at the wrong time.
2) Observe the wrong source.
3) Disagree on the observing frequency.
4) Disagree on the definition of circular polarization.
5) Get the polarization wrong anyway.
6) Miss-point the telescope.
7) Have an unstable clock.
8) Use a receiver which doesn't actually work.
9) Have a telescope full of snow.
10) Forget to record the data.

All of these faults have occurred at one time or another during the history of VLBI.

The setting up of Networks of observatories, in the US ('NUG') and in Europe (European VLBI Network (EVN)) has been an attempt to force disobedience of at least the first 4 commandments. These organisations receive observing proposals from scientists throughout the world, and the proposals are assessed on their scientific merit. The highest rated proposals are then scheduled within the next 6 months or year on the network telescopes. Members of the EVN include observatories from England, France, W. Germany, Italy, the Netherlands and Sweden, with Poland and the USSR as affiliate members. Some idea of the type of science being undertaken in VLBI astronomy can be obtained from the table, which shows the proposals received by the EVN programme committee in the last year grouped into headings.

EUROPEAN VLBI NETWORK (EVN)

PROPOSALS RECEIVED 1987

SUPERLUMINAL SOURCES	24
QUASAR NUCLEI	14
GALACTIC NUCLEI	22
GRAVITATIONAL LENSES	3
STEEP SPECTRUM CORES	7
POLARIZATION STUDIES	5
STARS - CONTINUUM STUDIES	8
STARS - LINE STUDIES	6
SUPERNOVA REMNANTS	2
PULSARS AND SCINTILLATION	6
	97

It can be seen that VLBI is very active and at the forefront of astronomical research.

VLBI is now at a crossroads. Future instruments are planned which will give us a major increase in capability. In the US there is the Very Long Baseline Array (VLBA) which has started construction and will involve 10 telescopes spread over the United States. This will be a powerful instrument allowing many observations to be made in a year, though not with as high a sensitivity as that obtainable in Europe with its large telescopes.

Various observatories in Europe are increasing the number of telescopes available, and there are also plans to set up a European Processor Centre in the Netherlands for the analysis of data recorded at each VLBI station. This will improve the quality and quantity of the data taken by VLBI.

Finally there are plans for what is perhaps the most exciting development, that is the launch of space vehicles carrying radio telescopes. The Soviet Institute for Space Research intend to launch a 10m radio telescope called 'RADIOASTRON' in a far Earth orbit to investigate the fine details close to the putative black holes which are believed to be powering Quasars and Active Galaxies. The European Space Agency (ESA) hopes to launch the 15m diameter QUASAT in a nearer Earth orbit, optimised for the best imaging capability. This should enable very high quality maps of radio sources to be obtained allowing us to see how the energy from the black holes is channelled into the larger scale radio emission seen by current ground based instruments.

Thus we are at a very busy time in VLBI, with exciting future prospects in the great adventure called astronomy.

Time and Frequency

		Chairman
TF-01	Characterisation of oscillators	J. Groslambert
TF-02	G P S	G. Winkler
TF-03	Timing and Dissemination	P. Kartaschoff
TF-04	Various Satellite Timing Experiments	B. Guinot
TF-05	VLF and LF	M. Brunet

TF-01

Characterisation of oscillators

1. Linear Prediction of the Non-Stationary Clock Error Function
   L.G. Bernier
   OSA, Neuchâtel - Switzerland

2. Oscillator Characterization by Correlation Method
   A. Mahmoud
   CNRS, Besançon - France

3. Measurement of Phase and Frequency Fluctuations of Microwave Oscillators
   J. Groslambert, M. Olivier
   CNRS, Besançon - France

4. Specifications and performances of ultra-stable oscillators in the Doris orbitography and positioning system
   J.B. Laporte
   Centre National d'Etude Spatiales, Toulouse - France

2 ND EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 16 - 17 - 18 MARCH 1988

Session TF-01 Number 1

LINEAR PREDICTION OF THE NON-STATIONARY CLOCK ERROR FUNCTION

Laurent-Guy Bernier
OSA, Neuchâtel, Switzerland*

*Part of the work reported in this paper was done at the Swiss PTT Research Center (Bern) while on leave from EPF (Lausanne).

SUMMARY

The clock error function x(t) of a frequency standard is defined as the integral of the normalized frequency fluctuations y(t). The time interval error TIE(t) is identical to x(t) except that two initial conditions are specified: x(0)=0 and y(0)=0.

In this paper a linear predictor for non-stationary processes is defined in such a way that the error on the prediction is stationary even if the original process is not stationary. The definition of the linear predictor is based on the theory of non-stationary random processes with stationary increments and does not need the usual separation between random and deterministic fluctuations.

As an application of the theory, it is shown that if a simulation of a non-stationary time error process is performed, starting the noise generator from a zero initial condition, then the output is not the expected non-stationary x(t) process but the error on a first order linear prediction of x(t). Despite of this, the simulation is still valid for a TIE estimation since the TIE is also shown to be identical to a first order linear prediction of x(t).

The proposed linear predictor can be defined up to an arbitrary order and enables the prediction of any process with stationary increments.

-I- INTRODUCTION

The clock error function, or time deviation, x(t) of a frequency standard is defined as the instantaneous difference between the time output of the frequency standard and the time output of a presumably perfect reference clock. The increment of x(t) over an interval τ after an initial synchronization (x(0)=0) and syntonization, or tuning, (y(0)=0) is defined as the time interval error (TIE) function TIE(τ) [1, 2].

The estimation and prediction of the TIE constitute an important issue in the fields of time scales generation, navigation and telecommunications. Possible applications of the TIE problem are the calculation of the time dispersion in an ensemble of independent clocks initially synchronized, the calculation of the time error accumulated between two successive synchronizations in a navigation network with distributed clocks, or the calculation of the rate of data stream collisions in a digital communication network when the clocks in the transmitter and receiver are free

running. Another example would be the calculation of the coherence time of an oscillator for VLBI or coherent radar applications.

Most of the previous approaches to the TIE estimation and prediction problem [3-7] assume that the observed divergent behavior of x(t) comes partly from deterministic causes such as the frequency offset and the frequency drift of the frequency standard. It is further assumed that these deterministic effects may be removed yielding a residual clock error function that is purely random and that may be modelled by a stationary process.

This paper, on the other hand, proposes another approach to the TIE problem based on the hypotheses that the clock error function is non-stationary and that the estimation-prediction problem is better defined if deterministic and random fluctuations are not treated separately. These assumptions may be justified by the following remarks.

Firstly, the distinction between deterministic and random fluctuations is always arbitrary because the exact deterministic part that is to be removed from the global random process is a function of the particular algorithm used in order to estimate the deterministic parameters against the non-stationary noise background. As a consequence, the statistical properties of the residual random process are also a function of the chosen estimation algorithm. The only way to define unambiguously the random process under study is, therefore, to treat the global time error function as a single non-stationary process.

Secondly, the conventional statistical parameters, for example the Allan variance, that are used in order to characterize the random fluctuations in oscillators are defined as average values taken over the ensemble of all the possible outcomes of the random process. The conventional estimation of these quantities by means of time averages is justified by the assumed ergodicity of stationary processes. Besides, it is often argued that a stationary model is adequate for the modelling of a non-stationary process such as a flicker, provided that the low cut-off frequency of the model is chosen to be smaller than the inverse of the useful observation time. In the context of a prediction problem, however, a single outcome of the random process is observed over a short time interval. Under such conditions, ergodicity is not achieved, even in the stationary case, because the random process does not go through all its possible internal states during the prediction interval. In other words, the initial conditions are not averaged out over the prediction interval and, therefore, they must be taken into account explicitly, just as if the process under study were not stationary. Therefore, it can be argued that, for the purposes of prediction, a non-stationary approach is relevant even under the hypothesis of a stationary process.

The problem of the TIE estimation was treated recently by Kartaschoff [6] and Greenhall [2]. Kartaschoff verified successfully an empirical bound on the TIE, proposed by Barnes et al. [7], using a simulation program based on the Barnes-Jarvis flicker noise generator [8]. Greenhall, on the other hand, proved that if the flicker generator is started from a zero initial state condition, as in Kartaschoff's simulation, the output is not the wanted flicker process but the error on the least squares linear prediction of the flicker process.

This paper begins with a description of the fractional integrator model which is necessary as a complement to the conventional power-law model for the analysis of non-stationary random processes (part II).

The theoretical analysis of non-stationary linear prediction is then stated using two different methods. The first method (part III) continues the work of Greenhall and is stated in terms of initial conditions and of the classical definition of the least squares linear prediction. The second method (part IV) is original and is stated in terms of the i first increments and structure functions of the random process. The respective merits of both methods are discussed. In the last part of the paper, finally, the general theoretical results are applied to the specific problem of the TIE estimation and compared with Kartaschoff's and Greenhall's results (part VI).

-II- THE FRACTIONAL INTEGRATOR MODEL

The conventional model of random frequency fluctuations describes the one sided power spectral density $S^+_{yy}(f)$ of the normalized frequency fluctuations y(t) as the sum of five independent power-law noise processes:

$$S^+_{yy}(f) \equiv \sum_{\alpha=-2}^{2} h_\alpha f^\alpha \qquad (1)$$

each process being characterized by an exponent α and a coefficient h_α. For the purpose of the present analysis, moreover, we are also interested in the time error function x(t) which may be defined as the integral of the y(t) [7]. Assuming that both x(t) and y(t) are stationary, their power spectral densities are linked by the following relationship:

$$S^+_{xx}(f) = \frac{S^+_{yy}(f)}{(2\pi f)^2} \qquad (2)$$

Therefore each noise process is characterized by two different exponents as summarized in table 1 below.

$S^+_{yy}(f)$	$S^+_{xx}(f)$		*Denomination*
$h_{-2}\, f^{-2}$	$(2\pi)^{-2} h_{-2}$	f^{-4}	Random walk of frequency
$h_{-1}\, f^{\alpha}$	$(2\pi)^{-2} h_{-1}$	f^{β}	Flicker of frequency
h_0	$(2\pi)^{-2} h_0$	f^{-2}	White frequency
$h_1\ f^{\alpha}$	$(2\pi)^{-2} h_1$	f^{β}	Flicker of phase
$h_2\ f^{2}$	$(2\pi)^{-2} h_2$		White phase

Table 1
Noise processes defined in the conventional power-law model

Following Keshner [9] we assume that the power spectral density of a flicker process is characterized by an exponent in the range]-2, 0[. Therefore we have the conditions

$-2 < \alpha < 0$ and $-4 < \beta < -2$

for a flicker of frequency and

$0 < \alpha < 2$ and $-2 < \beta < 0$

for a flicker of phase.

Unfortunately the power-law model does not make sense when the exponent is negative since a negative exponent yields a non-stationary process that, by definition, cannot be described by a power spectral density. A look at table 1 shows that the power law-model has a rigorous meaning only for a white phase noise process if the time error function is considered. This inconsistency of the power-law model can be circumvented by adopting an underlying dynamical model such as the fractional integration of a white noise.

Assuming that the power-law process n(t) characterized by the exponent u can be obtained by a stable linear filtering of a white noise process w(t) of power spectral density N_0, then the one-sided power spectral density of n(t) is given by

$$S^{+}_{nn}(f) = N_0 \left| H_u(j2\pi f) \right|^2 = N_0 (2\pi)^u f^u \qquad (3)$$

where

$$H_u(p) = p^{u/2} \qquad (4)$$

is the transfer function of the shaping filter [10]. If $u \geq 0$ the shaping filter is stable and the process n(t) is asymptotically stationary [10] in the sense that after a white noise is applied to the input of the shaping filter the output becomes stationary only when the filter reaches its steady state response. If $u < 0$, on the other hand, the filter is unstable, thus its output is forever non-stationary and equation (3) is not valid anymore.

However, the shaping filter defined by (4) exists for $u < 0$ and it is called a fractional integrator [11] because it is a generalization of the ordinary integrator ($u = -2$) for non-integer values of u. Its impulse response is obtained by taking the inverse Laplace transform of (4) [12] :

$$h_u(t) \equiv \frac{S(t)}{\Gamma(-u/2)} t^{-\frac{1}{2}u - 1} \qquad (5)$$

where S(t) is the step function and $\Gamma(x)$ the gamma function. If we consider that, for $u < 0$, equation (3) constitutes a formal definition of the power spectral density associated to the fractional integration of a white noise, then the power-law model becomes meaningful for every value of the exponent.

-III- LINEAR PREDICTION USING INITIAL CONDITIONS

Consider a non-stationary process n(t) produced by a linear combination of both stationary and non-stationary power-law processes. n(t) may represent either x(t) or y(t). It can be shown that there is an integer i such that the i^{th} derivative of n(t), noted $n^{(i)}(t)$, is stationary. The minimum number i of derivations needed in order to obtain a stationary

residual process is determined by the steepest negative power-law contained in n(t).

Assuming that $n^{(i)}(t)$ may be obtained by shaping a white noise w(t) through a linear filter of impulse response $h_i(t)$, the latter can be expanded into a past part and a future part

$$n^{(i)}(t) \equiv \int_{-\infty}^{0} d\tau\, w(\tau)h_i(t-\tau) + \int_{0}^{t} d\tau\, w(\tau)h_i(t-\tau)$$

$$n^{(i)}(t) \equiv p_{ls}\{n^{(i)}(t)\} + e_{ls}\{n^{(i)}(t)\} \tag{6}$$

The past part correspond to the classical definition of the least squares linear predictor $p_{ls}\{\ \}$ hence the future part yields the error $e_{ls}\{\ \}$ on the prediction [13]. The final process is then obtained by applying i successive integrations on $n^{(i)}(t)$

$$n(t) = \int_{-\infty}^{t} dt_i \int_{-\infty}^{t_i} dt_{i-1} \cdots \int_{-\infty}^{t_2} dt_1\, n^{(i)}(t_1) \tag{7}$$

as shown on figure 1. We also define h(t) as the impulse response of the global shaping filter that produces n(t) directly from w(t). The past part of each simple integration in equation (7) yields a single initial condition. Therefore equation (7) may be expressed as the sum

$$n(t) = p_{ic}^{(i)}\{n(t)\} + e_{ic}^{(i)}\{n(t)\} \tag{8}$$

where the first term is an IC predictor (for Initial Conditions) defined as the result of the integration of the initial condition on n(t) and of the i initial conditions on the derivatives of n(t):

$$p_{ic}^{(i)}\{n(t)\} = n(0) + \sum_{k=1}^{i} \frac{n^{(k)}(0)\, t^k}{k\ !} \tag{9}$$

The resulting IC error is given by the future part of (7):

$$e_{ic}^{(i)}\{n(t)\} = \int_{0}^{t} dt_i \int_{0}^{t_i} dt_{i-1} \cdots \int_{0}^{t_2} dt_1\, n^{(i)}(t_1) \tag{10}$$

The difference between the least squares linear predictor defined by (6) and the IC predictor defined by (9) is the following: the classical least squares definition assumes that the process to be predicted is stationary and that, consequently, $h_i(t)$ is a stable filter [13]. Under these conditions $p_{ls}\{\ \}$ yields a short-term prediction that goes rapidly to zero as the transient state of the shaping filter dampens out. The IC predictor, on the other hand, is based on the assumption that the process to be predicted is non-stationary. The multiple integration of the initial condi-

tions constitutes an unstable filter with a divergent transient state that enables a non-stationary prediction.

For a long-term prediction, i.e. for values of t larger than the duration of the transient state of hi(t), the least squares predictor $p_{ls}^{(i)}\{n^{(i)}(t)\}$ vanishes and the error $\epsilon_{ls}^{(i)}\{n^{(i)}(t)\}$ is equal to the process itself. Under these conditions (10) becomes

$$\epsilon_{ic}^{(i)}\{n(t)\} = \int_0^t dt_i \int_0^{t_i} dt_{i-1} \cdots \int_0^{t_2} dt_1 \int_0^{t_1} dt_0 \; w(t_0)h_i(t_1-t_0) \qquad (11)$$

which is equal to the output of the global shaping filter h(t) when it is started at t = 0 from a zero initial state.

A comparison between this last result and (6) shows that for a long-term prediction the IC prediction and the LS prediction are equivalent, since in both cases the error on the prediction is given by the output of the shaping filter started from a zero initial state condition. h(t) can be

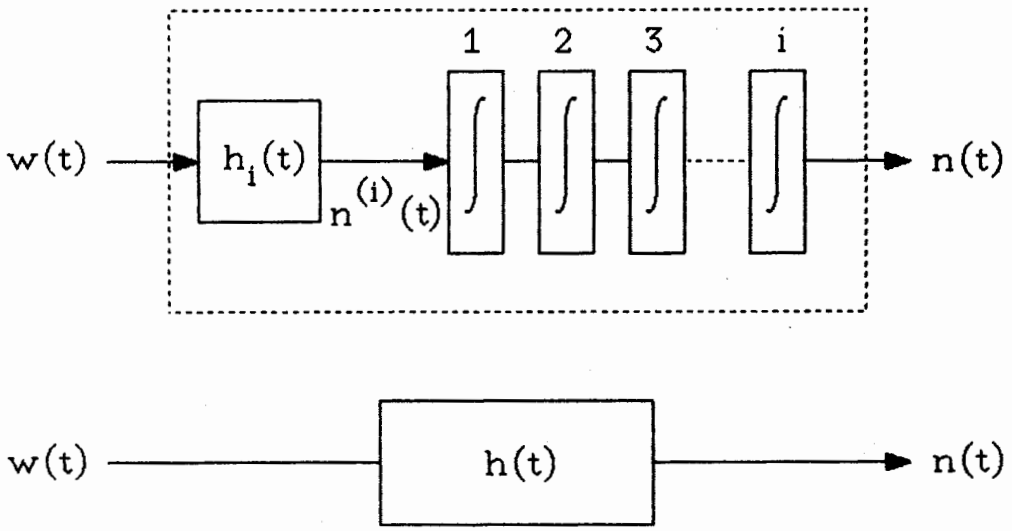


Figure 1
Model for the IC prediction

easily computed from the power-law model and the mean quadratic value of the output of the shaping filter yields the prediction error $\epsilon_{ic}^{(i)}\{n(t)\}$. The latter can be shown to be equal to

$$E\left[\left[\epsilon_{ic}^{(i)}\{n(t)\}\right]^2\right] = \tfrac{1}{2}N_0 \int_0^t d\tau\, h^2(\tau). \qquad (12)$$

under the zero initial state condition.

Estimation of the initial conditions

The main limitation of the IC approach to non-stationary linear prediction is the problem of the correct estimation of the initial conditions prior to the prediction itself. As a matter of fact the IC predictor is nothing but a deterministic extrapolation of a random process based on a Taylor power series expanded about t=0. If, for instance, there is a deterministic frequency offset or frequency drift in an oscillator, these parameters may be estimated with a good precision by observing the past behavior of the oscillator and by extrapolating to t=0. On the other hand, if the noise under study is a non-stationary random process, then its waveform can be best described as a fractal [14]. In such a waveform, the local mean value and derivatives are determined more by the high-frequency components of the process than by its long-term trends. On the other hand, it was shown that the error on the IC prediction is optimal in the least squares sense. We conclude that, in the random case, the initial conditions that must be put into the IC predictor are not the instantaneous derivatives of the process under study but its optimal trend parameters. Hence the IC prediction does not constitute a complete solution to the non-linear prediction problem in the sense that it gives the optimal prediction error without addressing the question of how the optimal trend parameters can be estimated prior to the prediction interval.

-IV- LINEAR PREDICTION BASED ON STRUCTURE FUNCTIONS

Definition of the general linear predictor

The correct approach to non-stationary linear prediction consists, then, of finding a global solution to the combined problems of the estimation of the initial trend parameters and to the prediction itself. A possible solution is the SF (for Structure Functions) predictor that is proposed in this section.

It can be shown that if the i^{th} derivative of n(t) is stationary then the i^{th} increment of n(t) is also stationary. The first order increment of a process n(t) may be defined as

$$\Delta(\tau)\{n(t)\} \equiv n(t) - n(t-\tau). \qquad (13)$$

The increment of order i is defined recursively by applying the rule:

$$\Delta^{(i)}(\tau)\{n(t)\} \equiv \Delta^{(i-1)}(\tau)\{n(t)\} - \Delta^{(i-1)}(\tau)\{n(t-\tau)\}. \qquad (14)$$

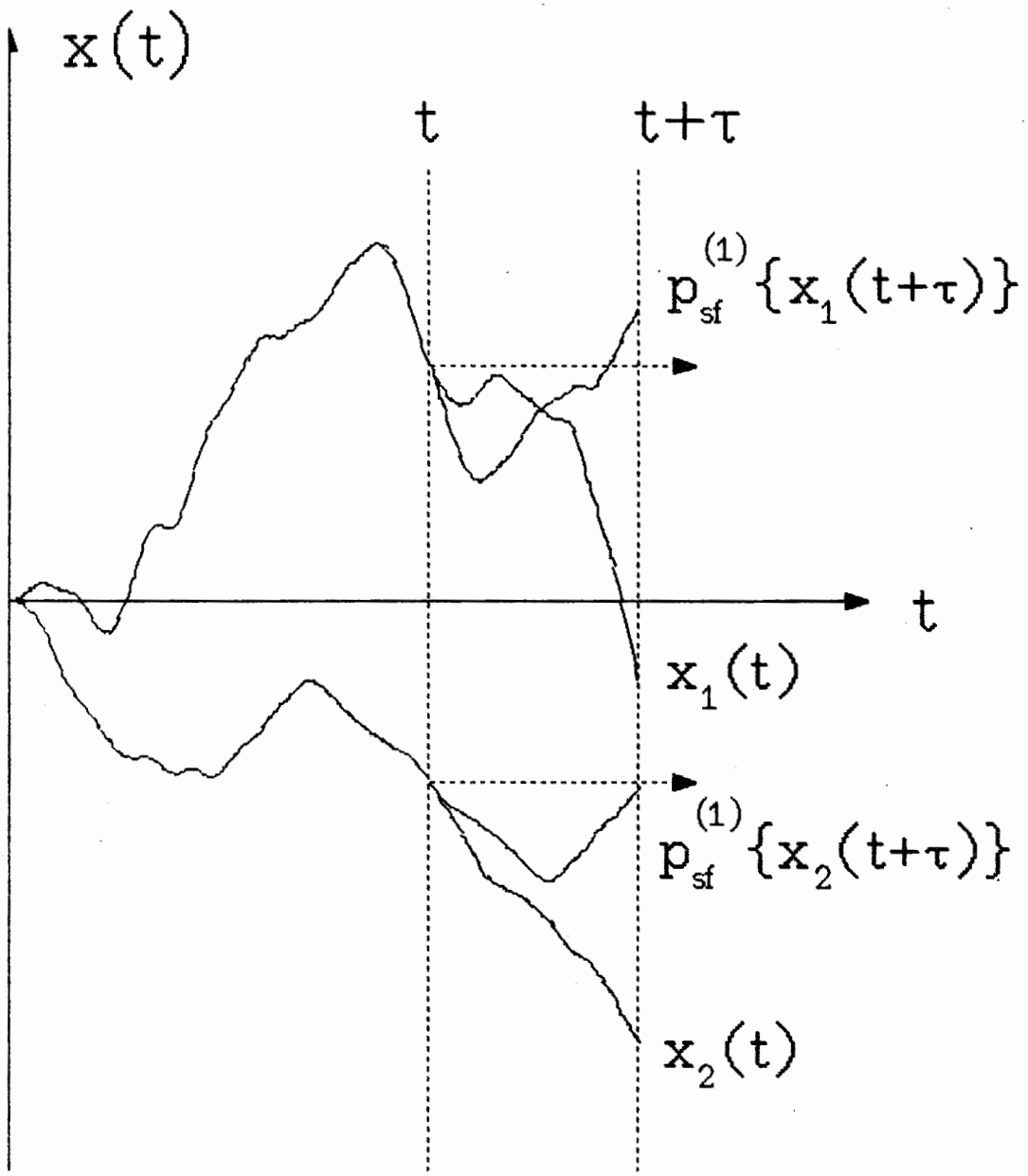


Figure 2
First order SF prediction of x(t) for a flicker of frequency

It is easy to show from the definitions (13) and (14) that

$$n(t+\tau) = n(t) + \sum_{k=1}^{i} \Delta^{(k)}(\tau)\{n(t)\} + \Delta^{(i+1)}(\tau)\{n(t+\tau)\}. \qquad (15)$$

It is clear that if t defines the present and t+τ some instant in the future then the first two terms on the right side of (15) depends only on the present and past values of the process n(t) while the third term depends on the future values. Thus the first two terms may be considered as the definition of a linear predictor

$$p_{sf}^{(i)}\{n(t+\tau)\} \equiv n(t) + \sum_{k=1}^{i} \Delta^{(k)}(\tau)\{n(t)\} \tag{16}$$

of order i and the third term yields the associated prediction error. The mean quadratic value of the error is equal to the ith structure function of n(t):

$$E\left[\left[\epsilon_{sf}^{(i)}\{n(t+\tau)\}\right]^2\right] = D_n^{(i+1)}(t+\tau,\ \tau) \tag{17}$$

where the structure function is defined as

$$D_n^{(k)}(t,\ \tau) \equiv E\left[\left[\Delta^{(k)}(\tau)\{n(t)\}\right]^2\right] \tag{18}$$

Assuming that n(t), which may be a superposition of deterministic and non-stationary random fluctuations, is a process with a stationary ith order increment, the ith structure function is stationary, i.e. independent of t. It is thus possible to define a linear predictor for a non-stationary process that yields a stationary error on the prediction.

On figure 2 are shown two possible outcomes of the f^{-3} x(t) process associated to a simulated f^{-1} flicker of frequency. Time t is the present and time t+τ represents an instant in the future. For each of the outcomes, the real future $x_1(t+\tau)$ and $x_2(t+\tau)$ are shown, as well as the associated first order SF linear predictions which, in this particular case, may be identified as mirror images of the past with respect to the present time.

Interpretation of the SF linear predictor in terms of initial conditions

The relationship between the IC and LC predictors is better understood if the latter is expressed in terms of the moving average operator. Let the causal moving average operator be defined as

$$\mu(\tau)\{n(t)\} \equiv \frac{1}{\tau}\int_{t-\tau}^{t} du\ n(u). \tag{19}$$

The i^{th} order moving average operator is then defined recursively using the rule:

$$\mu^{(i)}(\tau)\{n(t)\} \equiv \frac{1}{\tau}\int_{t-\tau}^{t} du\ \mu^{(i-1)}(\tau)\{n(u)\}. \tag{20}$$

Using the definitions (16), (18) and (19) it can be shown that the SF linear predictor may be expressed as

$$p_{sf}^{(i)}\{n(t+\tau)\} = n(t) + \sum_{k=1}^{i} \tau^k \, \mu^{(k)}(\tau)\{n^{(k)}(t)\} \qquad (21)$$

which means that the latter is similar to the IC predictor in the sense that it extrapolates n(t) by integrating the successive derivatives of the process. However the SF predictor differs essentially from the IC predictor because by averaging the kth derivative over an interval of duration $k\tau$ prior to the prediction, it correctly averages out the high-frequency noise in order to retain only the long-term trends of the process.

-IV- APPLICATION TO THE ESTIMATION OF THE TIE

IC estimation of the TIE

The time interval error (TIE) is defined as the increment of the time error function x(t) of an oscillator after an initial synchronization , i.e. x(t) = 0, and syntonization, i.e. y(0) = 0, with the reference. Clearly this definition corresponds to the error on the first order IC linear prediction of x(t):

$$TIE(t) \equiv \epsilon_{ic}^{(1)}\{x(t)\} = x(t) - x(0) - t\, y(0). \qquad (22)$$

Assuming that the noise source is a flicker of frequency then

$$S_{yy}^{+}(f) = h_{-1} f^{-1} \qquad (23)$$

from (1) and from (2). The density of the white noise and the impulse response of the fractional integrator that generate x(t) are computed using (2) and (3). The mean quadratic value of the TIE is then computed from (12) and yields

$$E\{TIE^2(t)\} = \frac{½\pi h_{-1}}{\Gamma^2(3/2)} t^2 = 2h_{-1} t^2 \qquad (24)$$

This last result is identical to Greenhall's calculation under the zero initial state condition in [2].

SF Estimation of the TIE

On the other hand, if the initial frequency offset y(t) is estimated with the SF method then the TIE is defined as

$$TIE(t, \tau) = x(t+\tau) - x(t) - \tau\, \mu(\tau)\{y(t)\} \qquad (25)$$

which corresponds to the error on the first order SF prediction. Using (17) the mean squared value of the TIE is found to be the second structure function of x(t):

$$E\{TIE^2(t, \tau)\} = D_x^{(2)}(t, \tau) = 2\, \tau^2\, \sigma_y^2(\tau) \qquad (26)$$

which, in the general case, can be expressed in terms of the Allan variance of y(t). Knowing that the Allan variance for a flicker of frequency is

$$\sigma_y^2(\tau) = 2\ \ln 2\ h_{-1}, \tag{27}$$

the SF estimation of the TIE is found to be

$$E\{TIE^2(t,\ \tau)\} = 2\ \ln 2\ h_{-1}\ \tau^2 = 2.7\ h_{-1}\ \tau^2 \tag{28}$$

The slightly larger factor of proportionality with respect to the IC estimation shows that the uncertainty on the TIE is partly determined by the uncertainty on the knowledge of the initial frequency offset as pointed out by Kartaschoff [6].

Estimation of the TIE without initial syntonization

If no initial syntonization is performed then the TIE is identical to the first increment of $x(t+\tau)$

$$TIE(t,\ \tau) \equiv x(t+\tau) - x(t) = \Delta(\tau)\{x(t+\tau)\}. \tag{29}$$

The mean quadratic value of the latter is thus simply given by the first structure function of $x(t+\tau)$

$$E\{TIE^2(t,\ \tau)\} \equiv D_x^{(1)}(t+\tau,\ \tau). \tag{30}$$

Besides it can be shown that if the fractional integrator that generates $x(t)$ is started at $t=0$ then the first structure function of $x(t)$ is given by

$$D_x^{(1)}(t,\ \tau) = \tfrac{1}{2}N_0 \int_0^{\min(t-\tau,\ t)} du\ h(t-\tau-u)h(t-u) \tag{31}$$

where $\min(a, b)$ is the minimum between a and b. Assuming as before that the noise source is a flicker of frequency, the mean quadratic value of the TIE is found to be

$$D_x^{(1)}(t+\tau,\ \tau) = h_{-1}\ \ln\left(\frac{4t}{\tau}\right)\tau^2 = h_{-1}\left[\ln\left(\frac{t}{\tau}\right) + 2\ln 2\right]\tau^2 \tag{32}$$

for $t >> \tau$. This result shows that the TIE without syntonization is non-stationary and matches the TIE with syntonization computed by Greenhall.

The expression (32) can be interpreted as follows. Using the definitions (13) (14) and (19) the TIE without synchronization may be expanded into

$$\Delta(\tau)\{x(t+\tau)\} = \tau\ \mu(\tau)\{y(t)\} + \Delta^{(2)}(\tau)\{x(t+\tau)\} \tag{33}$$

and if the two terms on the right side are assumed to be uncorrelated the mean squared value of the first structure function may be expanded as

$$D_x^{(1)}(t+\tau,\ \tau) = \tau^2\ E\left[\left[\mu(\tau)\{y(t)\}\right]^2\right] + 2\ \tau^2\ \sigma_y^2(\tau). \tag{34}$$

The second term corresponds to the TIE with syntonization as defined in the SF sense by (25). The first term, on the other hand, may be interpreted as the frequency offset at t that was accumulated since the last syntonization. This interpretation is justified from the fact that starting the fractional integrator from a zero initial state at t = 0 is equivalent to a syntonization of the oscillator.

Thus a comparison between (32) and (34) shows that the logarithmic term found by Greenhall may be interpreted as the part of the TIE that comes from the unknown initial frequency offset when an initial syntonization is not performed. It is clear, under these conditions, that the rms value of the initial frequency offset is a function of t, i.e. of the time elapsed between the last syntonization and the start of the observation interval.

-VI- CONCLUSION

The IC prediction theory developed in this paper shows that when a non-stationary random process, such as a flicker, is simulated by shaping a white noise process by a linear filter that starts at t=0 from an zero initial state then the output of the filter is not the wanted non-stationary process but the error on an optimal prediction of the process. This confirms Greenhall's analysis based on LS prediction. The IC prediction error may be interpreted as the part of the prediction error that comes exclusively from the unknown future of the process. This implies that the trend parameters are perfectly known and include all the relevant information contained in the past behavior of the process. In practice, however, the trend parameters must be estimated against a noise background and part of the prediction error comes from the estimation of the trend parameters as pointed out by Kartaschoff.

The SF prediction theory, on the other hand, includes a nearly optimal estimation of the trend parameters and proves that it is possible to define a linear predictor for a non-stationary process in such a way that the error on the prediction is stationary. The error on the SF prediction of order i is given by the structure function of order i+1. It is shown that a first order prediction corrects the initial time error and the initial frequency offset. The associated first order error is then given by the second order structure function of x(t) which can be expressed in terms of the Allan variance. A second order prediction would also correct the initial frequency drift and so on. Thus, in the general case, the order i that minimizes the structure function of order i+1 for a given prediction interval τ yields the optimal order of prediction.

In the second part of this paper, the general theoretical results are applied to the particular case of the estimation of the TIE. It is found that the conventional definition of the TIE is identical to the error on a first order prediction of x(t). This means that a simulation in which the shaping filter starts from a zero initial state condition, as in Kartaschoff's work, is appropriate for the simulation the TIE produced by a flicker of frequency but not for a simulation of the flicker process itself.

However the conventional definition of the TIE does not specify by which method the initial time and frequency offsets must be estimated. If the LS predictor is selected as the estimation method then the mean quadratic TIE is found to be just a factor of 1.35 larger than the one given by an ideal IC prediction in the case of a frequency flicker. It is shown, finally, that the non-stationary logarithmic term found by

Greenhall belongs to the non-stationary TIE that results from an initial synchronization without syntonization.

-VII- REFERENCES

[1] C.C.I.R. Report 898 (Mod F), Dubrovnik, 1986, p. 1.

[2] C. A. Greenhall, "Initializing a flicker Noise Generator, " *IEEE Trans. on Instrum. and Meas.*, vol. IM-35, no. 2, June 1986, pp. 222-224.

[3] D. W. Allan, "Time and Frequency (Time-Domain) Characterization, Estimation, and Prediction of Precision Clocks and Oscillators, " *IEEE Trans. on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. UFFC-34, no. 6, November 1987, pp.647-654.

[4] D. W. Allan, D. J. Glaze, H. E. Machlan, A. E. Wainwright, H. Helwig, J. A. Barnes, J. E. Gray, "Performance, Modeling, and Simulation of Some Cesium Beam Clocks, " *Proc. 27th Annual Symposium on Frequency Control*, Cherry Hill, 12-14 June 1973, pp. 334-346.

[5] D. W. Allan, J. E. Gray, H. E. Machlan, "The National Bureau of Standards Atomic Time Scale: Generation, Stability, Accuracy and Accessibility, " in Time and Frequency: Theory and Fundamentals, B. E. Blair, Editor, NBS Monograph no. 140, Washington DC, May 1974, chapter 9, pp. 205-231.

[6] P. Kartaschoff, "Computer Simulation of the Conventional Clock Model, " *IEEE Trans. on Instrum. and Meas.*, vol. IM-28, no. 3, September 1979, pp. 193-197.

[7] J. A. Barnes, A. R. Chi, L. S. Cutler, D. J. Healy, D. B. Leeson, T. E. McGunigal, J. A. Mullen, W. L. Smith, R. L. Sydnor, R. Vessot, G. M. R. Winkler, "Characterization of Frequency Stability," *IEEE Trans. on Instrumentation and Measurement*, vol. IM-20, May 1971, pp. 105-120.

[8] J. A. Barnes, S. Jarvis, "Efficient Numerical and Analog Modeling of Flicker Noise Processes, " NBS Technical Note no. 604, June 1971.

[9] M. S. Keshner, "1/f Noise", *IEEE Proceedings*, vol. 70, no. 3, March 1982, pp. 212-218.

[10] A. Papoulis, Probability, Random Variables and Stochastic Processes, McGraw-Hill, New-York, 1965, Sec. 10-3, p.347.

[11] V. Radeka, "1/f Noise in Physical measurements, " *IEEE Trans. on Nuclear Science*, vol. NS-16, October 1969, pp. 17-35.

[12] M. Abramowitz, I. A. Stegun, Handbook of Mathematical Functions, Dover, 1964, Equation 29.3.7.

[13] J. Makhoul, "Linear Prediction: A Tutorial Review, " *IEEE Proceedings*, vol. 63, no. 4, April 1975.

[14] J. J. Gagnepain, J. Groslambert, R. Brendel, "The Fractal Dimension of Phase and Frequency Noises: Another Approach to Oscillator Characterization, " *39th Annual Symposium on Frequency Control*, Philadelphia, 29-31 May 1985, pp. 113-118.

2 ND EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 16 - 17 - 18 MARCH 1988

Session TF-01 Number 2

OSCILLATOR CHARACTERIZATION BY CORRELATION METHOD

Ali MAHMOUD

Laboratoire de Physique et Métrologie des Oscillateurs du CNRS
associé à l'Université de Franche-Comté-Besançon
32, avenue de l'Observatoire - 25000 Besançon - France

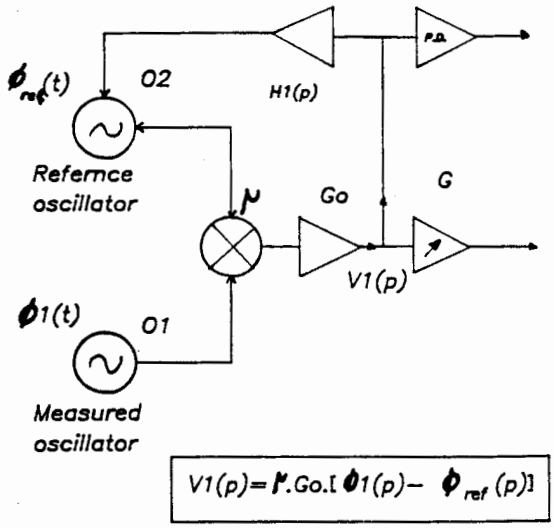
Abstract

The measurement method is based on the calculus of intercorrelation between the relative phase or frequency fluctuations of three oscillators compared two by two using two mixers. Two auxilary oscillators are thus controlled by the main oscillator whose stability is to be measured. The delivered signal of each mixer is proportional with respect to phase fluctuations of both oscillator. One may consider that the three oscillators are independent. Thus, the intercorrelation function of the signals delivered by the two mixers is proportional to the phase fluctuation autocorrelation of the common oscillator. This allows one to characterize the single oscillator independently of the two others. The spectral purity is obtained by fast Fourier transform (FFT) of the intercorrelation function already calculated. One may equally determine Allan variance. This method allows one to reach a resolution of -185 dB.

The device realizes the following functions : analog comparison of phases of the three oscillators two by two, numerical calculation of the intercorrelation function, calculation of the fast Fourier transform (FFT) and visualization of the results.

1. MEASUREMENT METHOD BY PHASE LOCKED LOOP (PLL)

This method needs a reference oscillator which oscillates at the same frequency as the one that is being measured and which can be electronically controlled [1]. The basis scheme is given on figure 1. The error voltage, $V_1(t)$, that holds the necessary informations allows one to reach the phase of frequency fluctuations of the measured oscillator [2].



(1)

fig1: basic scheme of measurement method by phase locked loop

The noise of the low noise amplificators which is -165 dB limits the measurement bench resolution. On the other hand, with this method, we can characterize two oscillators. The measure gives

$$\overline{\Phi_m^2} = \overline{\Phi_1^2} + \overline{\Phi_{ref}^2} \qquad (2)$$

Thus, we have to make some hypothesis on the reference oscillator :

- either it is better than the oscillator which is to measure and in this case : $\overline{\Phi_m^2} = \overline{\Phi_1^2}$
- or it has the same quality and : $\overline{\Phi_1^2} = \overline{\Phi_m^2}/2$.

2. MEASUREMENT PRINCIPE WITH TWO MIXERS

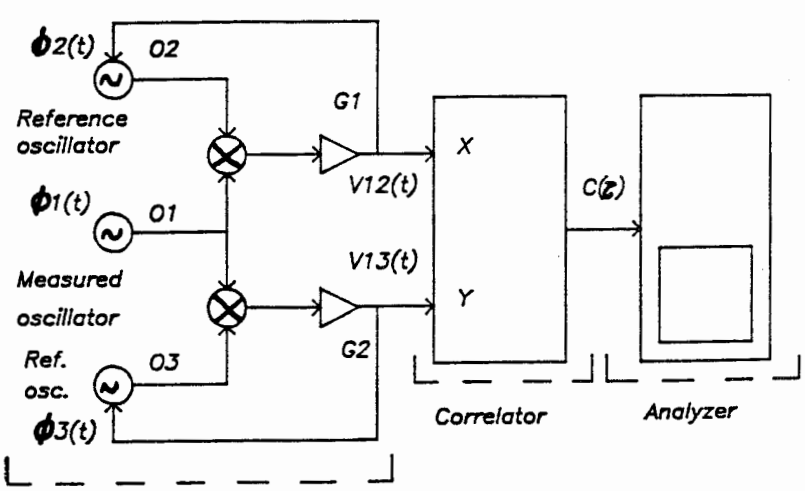


fig2: basic scheme of measurement method with two mixers

In order to increase the resolution (-165 dB) a measurement method by intercorrelation [2] was developed. Two regular benches are this time in parallel.

For measuring oscillator 01 spectral purity, we put under control two other oscillators 02 and 03 on 01. We obtain the two error voltages $V_{12}(t)$ and $V_{13}(t)$ which can be expressed like

$$V_{12}(t) = K_1 \left[(\Phi_1(t) - \Phi_2(t)\right] + b_{12}(t) \tag{3}$$

$$V_{13}(t) = K_2 \left[(\Phi_1(t) - \Phi_3(t)\right] + b_{13}(t) \tag{4}$$

K_1 and K_2 are parameters related with the phase lock loop gains. $b_{12}(t)$, $b_{13}(t)$ represent the additional noises of the two ways of the measurement system.

By calculating the intercorrelation function of the two signals $V_{12}(t)$ and $V_{13}(t)$ we obtain

$$C_{V_1V_2}(\tau) = \lim_{T\to\infty} \frac{1}{T} \int_0^T V_{12}(t)\, V_{13}(t-\tau)\, dt \tag{5}$$

$$\begin{aligned} C_{V_1V_2}(\tau) = {} & K_1K_2 \left[C_{\Phi_1\Phi_1}(\tau) - C_{\Phi_1\Phi_3}(\tau) - C_{\Phi_2\Phi_1}(\tau) + C_{\Phi_2\Phi_3}(\tau)\right] \\ & + K_1 \left[C_{\Phi_1 b_{13}}(\tau) - C_{\Phi_2 b_{13}}(\tau)\right] \\ & + K_2 \left[C_{b_{12}\Phi_1}(\tau) - C_{b_{12}\Phi_3}(\tau)\right] \\ & + C_{b_{12}b_{13}}(\tau) \end{aligned} \tag{6}$$

If the different phase noises $\Phi_i(t)$, and the additional noise $b_{ij}(t)$ are independent, most of the intercorrelation functions are zero

$$C_{\Phi_i\Phi_j}(\tau) = 0 \qquad i \neq j$$

$$C_{\Phi_i b_{ij}}(\tau) = 0$$

$$C_{b_{1i}b_{1j}}(\tau) = 0$$

then

$$\boxed{C_{V_{12}V_{13}}(\tau) \# K_1\, K_2\, C_{\Phi_1\Phi_1}(\tau)} \tag{7}$$

which directly gives the stability of the common oscillator 01.

So by Fourier transform we achieve the spectral purity of the measured oscillator [3]

$$\boxed{S_{V_1 V_2}(f) = K_1 \; K_2 \; S_{\Phi_1}(f)} \qquad (8)$$

Figure 3 shows the regular method resolution and the two mixers bench resolution.

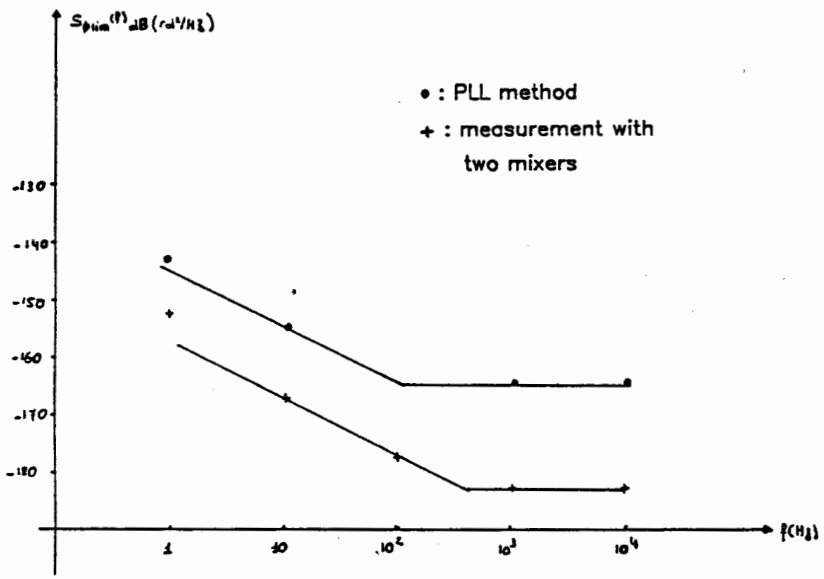


Fig. 3 : Comparison of the resolution of the two methods

The resolution gain obtained with this method is the result of the decorrelation of the noises b_{12} and b_{13}. However this resolution is limited at -183 dB, because there always remains a certain correlation between the different noises [2].

4. THE CORRELATOR

The correlator which was previously relalized [4] is a numeric correlator which allows the calculation of the analog signal autocorrelation function or of the intercorrelation of two different signals. This function is

$$C_{XY}(\tau) = \frac{1}{T} \int_0^T X(t)\; Y(t-\tau)\; dt \qquad (9)$$

T : measure time

and this calculation needs the following process :

- time delay : $y(t) \rightarrow y(t-\tau)$
- product : $X(t) \cdot Y(t-\tau)$
- integration

Figure 4 shows the schematic diagram of the correlator.

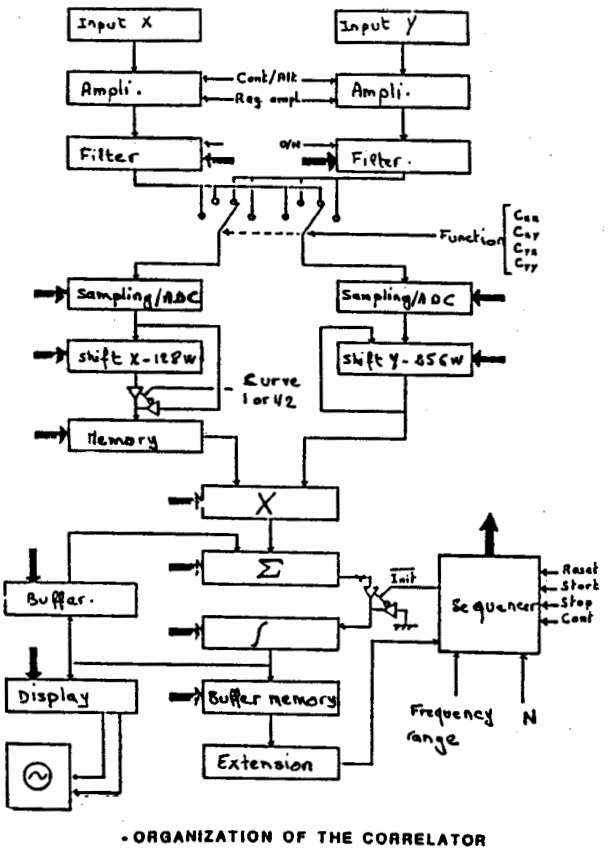


Fig. 4 : Schematic diagram of the correlator

The analog signal is levelled and filtered in order to avoid problems of spectral overlapping. Then it is sampled by a digital-analog converter. We obtain the following connections :

$t \rightarrow K T_e$

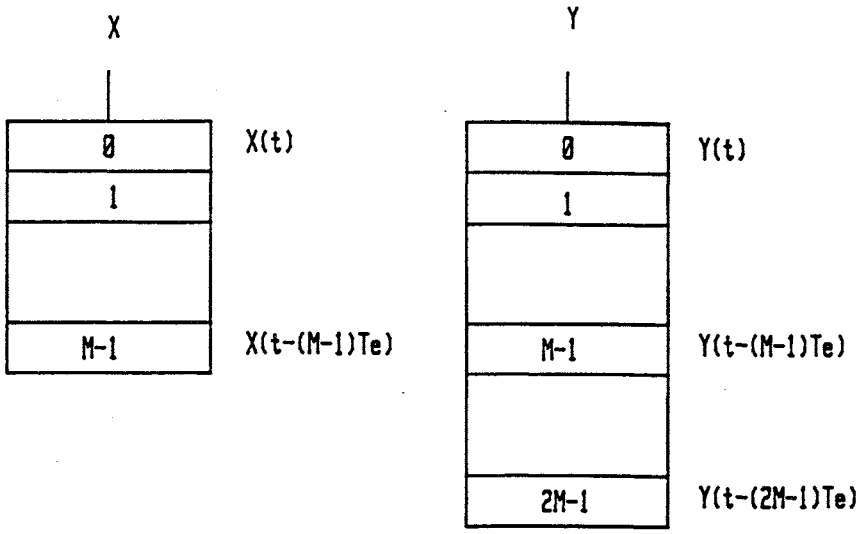
$\tau \rightarrow \ell T_e$

$T \rightarrow N T_e$

T_e : sampling period

$$C_{xy}(\ell) = \frac{1}{N} \sum_{K=0}^{N-1} x(k)\, y(k-\ell) \qquad (10)$$

with numeric signals we can realize the time delays by using a shift register according to the following scheme



$$X(M-1)\ Y(0) \rightarrow \tau = -(M-1)T_e$$
$$\vdots$$
$$X(M-1)\ Y(M-1) \rightarrow \tau = 0$$
$$\vdots$$
$$X(M-1)\ Y(2M-1) \rightarrow \tau = MT_e$$

The time delays can vary between $-(M-1)T_e$ and MT_e.

The correlator has a buffer memory which binds it to an outside calculating device. That will enable us to make the necessary treatment of the correlation function.

The calculation of the correlation function is made in real time for the input signal frequencies up to 5 kHz, which corresponds to a sampling frequency of 12.8 kHz.

The correlator characteristics are the following [4]

- input X and Y impedance = 10 kΩ
- coupling : DC input or AC input with low frequency cut-off at 10 Hz
- sampling frequency at 2.56 B where B is the maximum frequency of the analysis range
- measurement span : 60 dB, limited by weighting window
- average in time : from 1 to 10^{12} measures
- frequency ranges : 1 Hz to 100 kHz with step of : 1 - 2 - 5 ...

5. SPECTRAL ANALYSIS

The last step consists in the calculation of the Fourier transform of the correlation function, which is available in the correlator buffer memory. We use an algorithm of Fast Fourier Transform (FFT) [5] and the calculus is made with 512 points. The principle scheme is given in Figure 5.

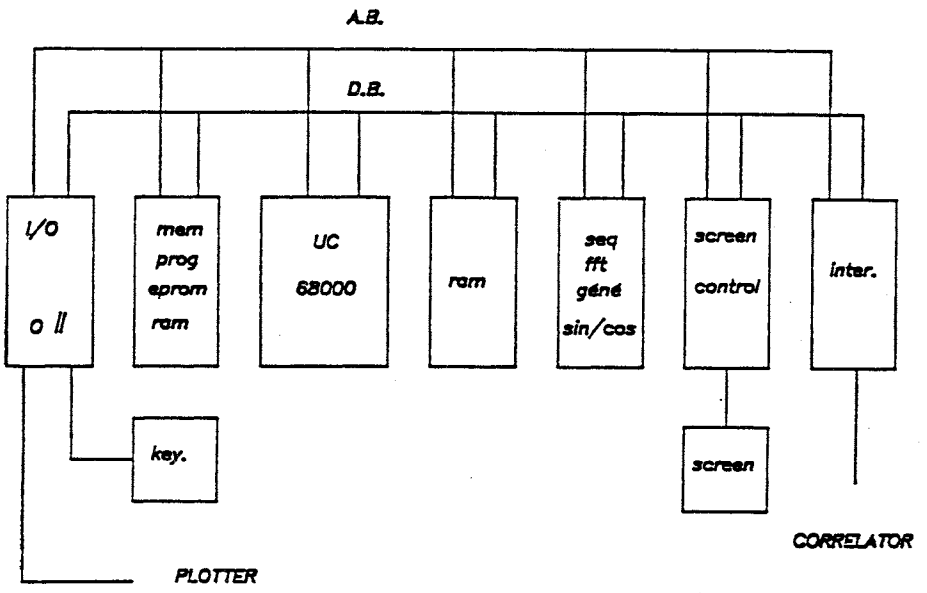


fig5: basic scheme of the analyzer

We used a 68000 microprocessor [6] since the correlation function is available in the buffer memory and there is no need for calculating in real time.

The calculus is realized by the microprocessor associated with a FFT unit which generates the address of the elements necessary for the calculation.

The device has three extensions :

- a parallel interface which establishes the connection with the correlator,
- an interface in series (RS232) which is connected to a plotter in order to record the results
- in addition a second interface in series which could establish a connection with an other device.

The calculations are realized with a floating O.point and with 32 bits that allows a good accuracy of measurement. All the numbers are expressed as $N = M * 2^K$.

M is coded with 16 bits and we obtain a dynamic of 96 dB. This dynamic is sufficient since we are limited by the correlator dynamic which is only 60 dB. The latter can be improved by the truncation of the correlation function which reduces the window effect. The time for the calculation of the Fourier transform is less than 2 sec.

6. RESULTS

The system was tested by means of known functions (sinus, white noise, $1/f^2$ noise ...) and the results are shown in Figures 6 and 7.

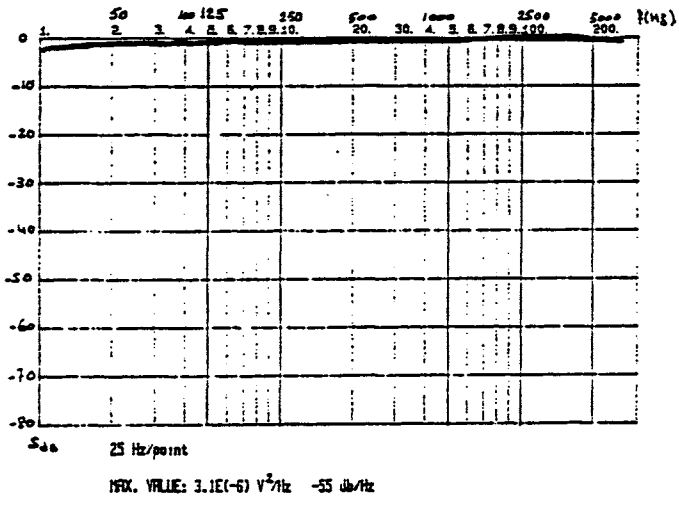


fig6: wite noise spectrum

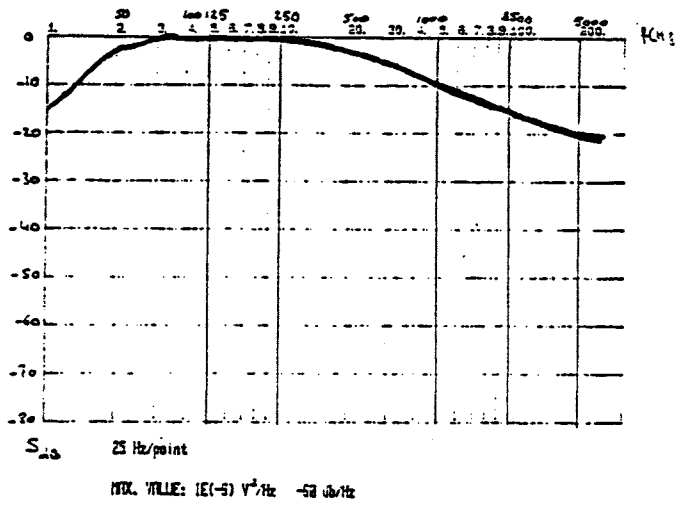


fig7: $1/f^2$ noise spectrum

On the other hand, we observed the effect of the weighting windows on harmonic signals [3],[7]. Thus we have realized the Fourier transform of the correlation function ($C(\tau)$) of a sinusoïdal signal for each of the following cases :

a) $C(\tau)$ is multiplied by the natural window $\Pi_{\tau M}$ (Fig. 8)

$$\Pi_{\tau M} = \begin{cases} 1 & |\tau| < \tau_M \\ 0 & |\tau| > \tau_M \end{cases}$$

b) $C(\tau)$ is multiplied by the Bartlet window $f(\tau)$ (Fig. 9)

$$f(\tau) = \begin{cases} 1 - \dfrac{|\tau|}{\tau_M} & |\tau| < \tau_M \\ 0 & |\tau| > \tau_M \end{cases}$$

c) $C(\tau)$ is multiplied by the Parzen window $g(\tau)$ (Fig. 10)

$$g(\tau) = \begin{cases} 1 - 6\left(\frac{|\tau|}{\tau_M}\right)^2 + 6\left(\frac{|\tau|}{\tau_M}\right)^3 & |\tau| \leq \tau_M/2 \\ 2\left(1 - \frac{|\tau|}{\tau_M}\right)^3 & \tau_M/2 < |\tau| < \tau_M \end{cases}$$

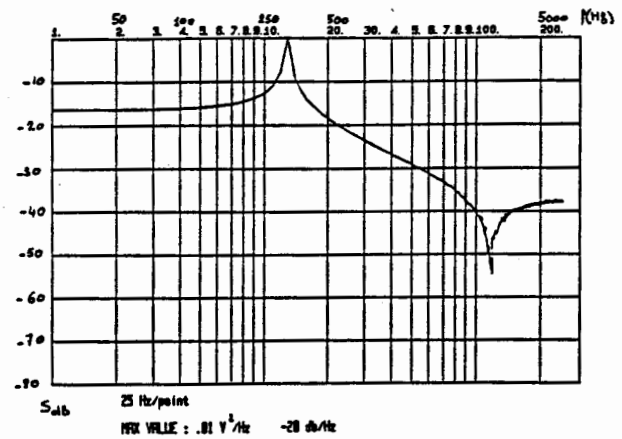


fig8: the correlation function is multiplied by the natural window

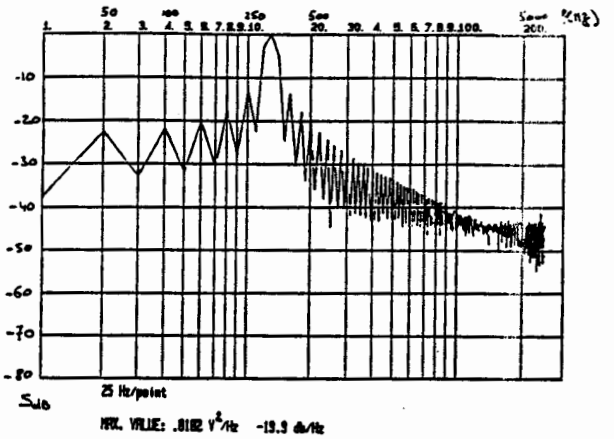


fig9: the correlation function is multiplied by Bartelett window

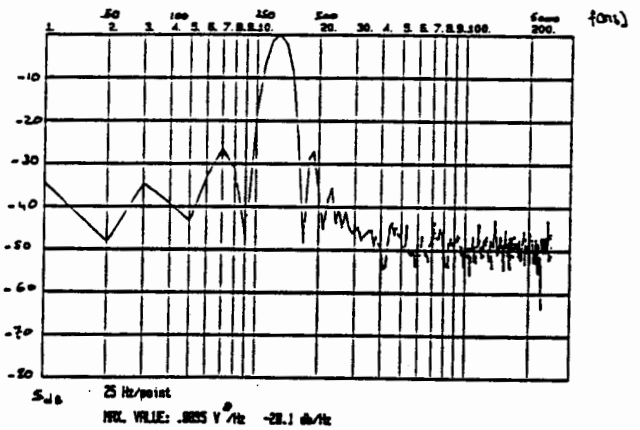


fig10: the correlation function is multiplied by Parzen window

Thus we established that the Parzen window is the more adapted for increasing the dynamic. We obtain a gain of 20 dB referred to the first case.

In order to determine the resolution limit of this measurement bench we use the mounting of Figure 11.

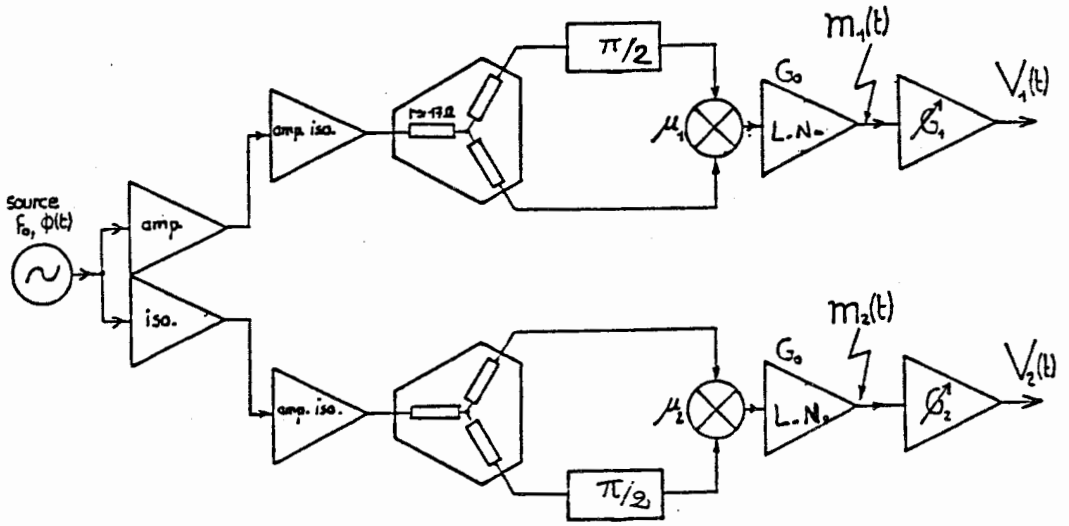


Fig. 11 : Mounting for measuring the resolution limit

The delivered signal by the ultrastable source is applied to both regular benches. The isolation amplifiers permits one to isolate the source from the measured system. The phase low noise of these amplifiers is eliminated by comparison with the inputs of each mixer. The signals $V_1(t)$ and $V_2(t)$ may be written as [2]

$$V_i(t) = G_o\, G_i \left\{\mu_i \left[\Phi_1(t) - \Phi_2(t) + \Psi_i(t)\right] + \frac{e_{n_i}(t)}{G_o}\right\} \qquad (11)$$

$$i = 1,2$$

$\Psi_i(t)$ is the phase noise due to the mixer. $e_{n_i}(t)$ is the additional noise originated from the mixer and the low noise amplifier.

When the source is sufficiently stable we have the following identity

$$\Phi_1(t) \equiv \Phi_2(t) \tag{12}$$

Thus, the bench noise is

$$S_{\Phi_{lim}}(f) = \frac{|S_{m_1 m_2}(f)|}{(\mu G_o)_1 (\mu G_o)_2} \tag{13}$$

where $S_{m_1 m_2}(f)$ is the spectrum of the signals m_1, m_2.

The measurements that we have performed lead to a resolution limit equal to -185 dB (rd^2/Hz). Figure 12 shows the curve $S_{\Phi_{lim}}(f)$.

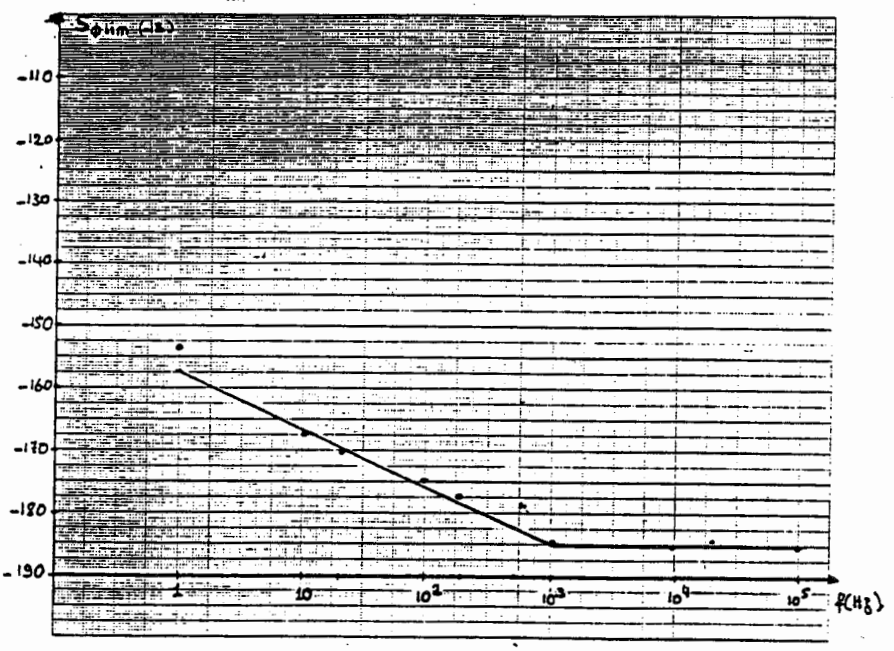


Fig. 12 : Resolution limit of the measurement system

REFERENCES

[1] J. Groslambert, G. Marianneau et M. Olivier, "Méthodes expérimentales de caractérisation des oscillateurs", L'Onde Electrique, vol. 55, n° 2, Février 1975.

[2] D. Fest, "Caractérisation des oscillateurs de très haute stabilité par une méthode de corrélation", Thèse de Docteur-Ingénieur, LPMO, Besançon, 1982.

[3] J. Max, "Méthodes et techniques de traitement du signal", 4e ed., Masson, 1985.

[4] D. Pierrejean, "Etude et réalisation d'un corrélateur pour la mesure de la stabilité et de la pureté spectrale d'oscillateurs ultrastables" Thèse de Docteur de l'Université de Franche-Comté, LPMO, Besançon, 1985.

[5] E.O. Brigram, "The Fast Fourier Transform", Prentice Hall, Inc. 1974.

[6] M. Aumiaux, "Microprocesseurs 16 bits", Masson, 1985.

[7] F.J. Harris, "On the use of windows for harmonic analysis with the Discrete Fourier Transform", Proceedings of the IEEE, vol. 66, n° 1, jan. 1978.

2 ND EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 16 - 17 - 18 MARCH 1988

Session TF-01 Number 3

MEASUREMENT OF PHASE AND FREQUENCY FLUCTUATIONS OF MICROWAVE OSCILLATORS

Jacques GROSLAMBERT and Marcel OLIVIER

Laboratoire de Physique et Métrologie des Oscillateurs du CNRS
associé à l'Université de Franche-Comté-Besançon
32, avenue de l'Observatoire - 25000 Besançon - France

Abstract

This paper is concerned with the measurement of phase- and frequency-spectral density in the 2-18 GHz frequency range. Frequency jitter in a specified offset band from the carrier, e.g. 100 Hz - 12 kHz can also be measured.

The phase locking method is used for determining the close-to-carrier phase spectral density of the oscillator under test. A clean reference source is necessary. This reference consists in a quartz oscillator of high spectral purity operated at 1 GHz. A comb generator gives a combination of harmonics and sub-harmonic frequencies of this reference oscillator. A YIG oscillator, working in the 2 - 6 GHz frequency range is phaselocked on one of these spectral lines, a synthesized IF frequency being inserted in the phase locked loop.

The microwave signal of the oscillator under test is down-converted to 10 MHz by direct or harmonic mixing with the output of the YIG oscillator. A 10 MHz voltage controlled crystal oscillator is used to detect, by means of the phase locking method, the phase and frequency noise of the oscillator under test.

The performances and possibilities of the measurement system will be described in the paper.

INTRODUCTION

At the present time, the requirements for low noise microwave sources in radar and communication systems resulted in the development of good long and short term stability microwave oscillators. The noise sidebands, in the region extending from 1 Hz to several kilohertz off the carrier can be reduced by using a microwave oscillator phase locked on the harmonics of a VHF crystal oscillator. The use of high Q dielectric cavity and low noise electronics gives a very low spectrum noise floor at intermediate modulation frequencies, in the range from 10 kHz to 10 MHz.

The amount of frequency modulation is such that the spectrum analyser display of the signal will show a carrier line and a noise background determined only by the analyser bandpass filter. The analysis of the phase and frequency fluctuations require the demodulation of the signal. The spectral analysis of the amplified demodulated signal gives the spectral density $S_{\Phi}(F)$ or $S_{\Delta f}(F)$ of the phase -or frequency- noise of the oscillator under test.

The demodulation process require either a frequency discriminator or a phase detector. The discriminator method uses a resonant cavity that act as F.M. to A.M. or P.M. converter, but the sensitivity of this method degrades as 1/F as the frequency of the carrier is approached. The advantage of this technique is that it does not require an active reference oscillator, but it is not appropriate for a high sensitivity near carrier measurement.

The phase detector method, using a reference oscillator in addition to the oscillator under test, yields the overall best sensitivity. In the microwave domain, it requires that the two oscillators be phase-locked to maintain phase quadrature. This method, which can be used in conjonction with an heterodyne technique has been used for the realization of a very performant noise measurement system of microwave oscillators which work in the frequency range from 2 to 18 GHz.

The noise measuring equipment consists in two basic sections :

In the first section a reference microwave signal is synthesized. The signal of the oscillator under test is down converted to 10 MHz by direct or harmonic mixing with this reference signal.

In the second section, the phase and frequency fluctuations of the oscillator under test are detected, using a quartz crystal voltage controlled oscillator and the phase locking method.

Each of these sections will be discussed and finally the performances and possibilities of the system will be given.

1. GENERATION OF A REFERENCE SIGNAL

The phase noise test set must be able to analyse the phase -or frequency- fluctuations of microwave oscillators of good spectral quality. The specification of the phase noise spectrum, at 1 GHz, is given in Fig. 1. The phase noise of the system, and particularly the phase noise of the reference source must satisfy this specification.

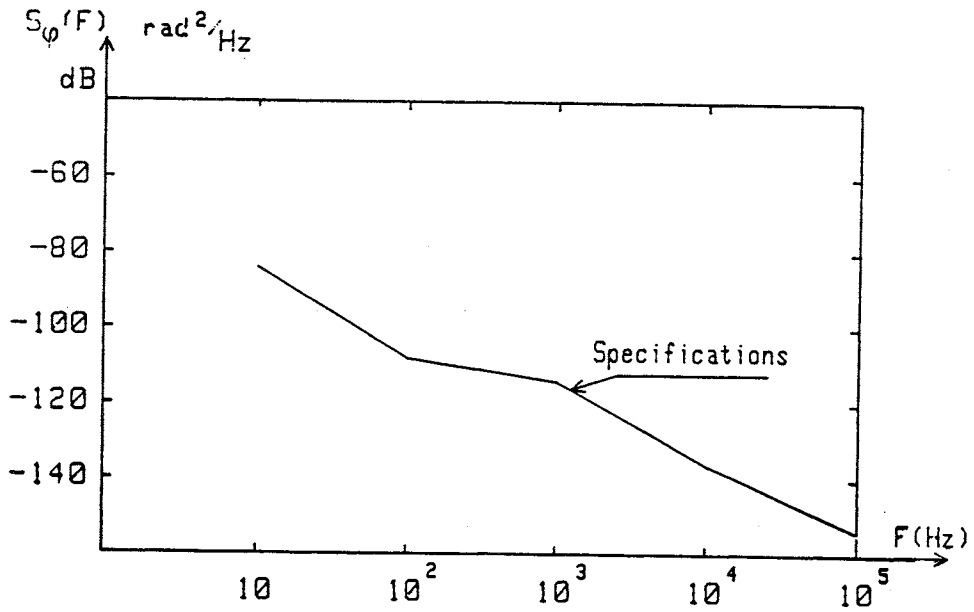


Fig. 1

The basic "idea" of the system is to use a low noise reference signal at 1 GHz to generate a comb of signals ranging beyond 18 GHz. These microwave signals down-convert the microwave source into the radiofrequency domain where the analysis can be done.

To generate directly a low frequency iF beat, we use a comb of signals spaced by 100 MHz.

The output of the comb generator must be filtered : a YIG filter could be used, but the signals to be filtered are so closed spaced that the filtering action of the filter cannot be completely effective. Some phase noise of the narrow bandwidth YIG filter could also be introduced.

A solution to this filtering problem is to use a YIG oscillator phase-locked on the proper harmonic. The filtering effect is more effective. Another advantage of this solution is that a synthesized low frequency signal can be introduced into the phase locked loop, so that the microwave reference signal can be varied by step as low as 1 Hz, and the IF beat frequency signal can be adjusted exactly at 10 MHz. The analysis of this signal is greatly simplified.

We will first describe the different elements of this section, that is, the reference oscillator, the harmonic generation and the locked YIG oscillator.

A/ REFERENCE OSCILLATOR

A microwave signal of high spectral purity can be obtained at 1 GHz, either with an oscillator operating at this frequency, or after the multiplication of the output of a R.F. or V.H.F. oscillator.

The contribution of the external thermal noise on the phase noise of an oscillator can be minimized if the oscillator operates at high frequency and high power. On the other hand, the influence of the internal noise, especially flicker noise, can be reduced, using a high Q R.F. quartz oscillator.

We have choosen a bulk wave, fifth overtone quartz crystal oscillator, operating at 1 GHz, and developed as a prototype by C.E.P.E. [1]. The phase spectral density $S_\Phi(F)$ of this oscillator has been tested, using a R.F. phase noise test set developed at the LPMO. Fig. 2 illustrates the results obtained with two identical oscillators. The specifications are well satisfied for Fourier frequency above 1 kHz.

The spectrum presents a F^{-3} slope, corresponding to F.M. flicker noise, for Fourier frequencies inside the resonator bandwidth. For this range of frequencies, the phase noise spectrum could be improved, using a phase locking of the 1 GHz oscillator, after division, on a 5 MHz high Q quartz crystal oscillator. A classical analysis of the loop gives the phase spectrum S_{Φ_a} of the phase locked oscillator (Fig. 3).

$$S_{\Phi_a} = S_\Phi \frac{1}{|1 + G/N|^2} + S_{\Phi_r} \frac{|G/p|^2}{|1 + G/N|^2}$$

N is the division factor of the frequency divider and G is the open loop gain (Fig. 4).

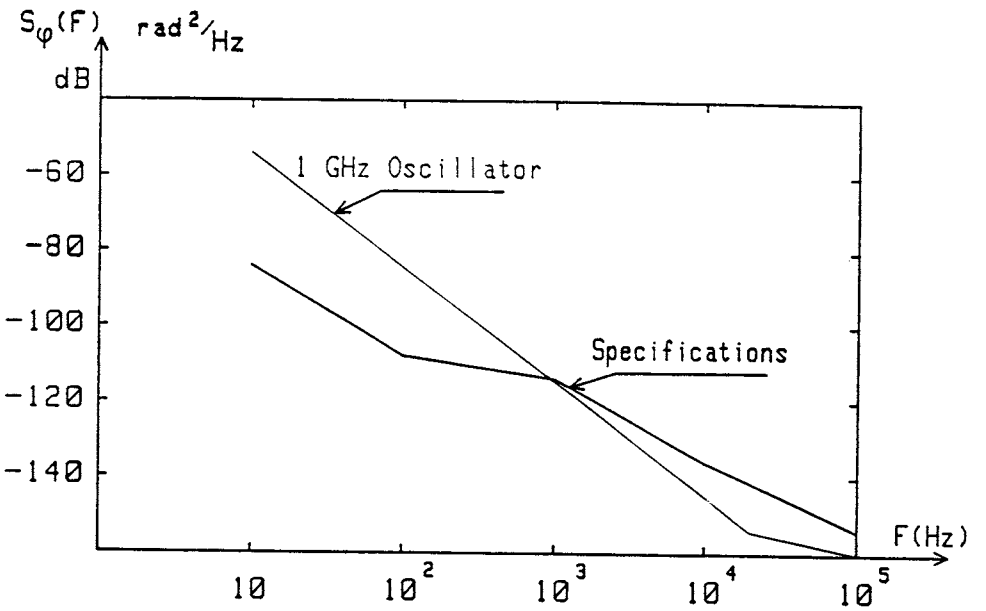


Fig. 2

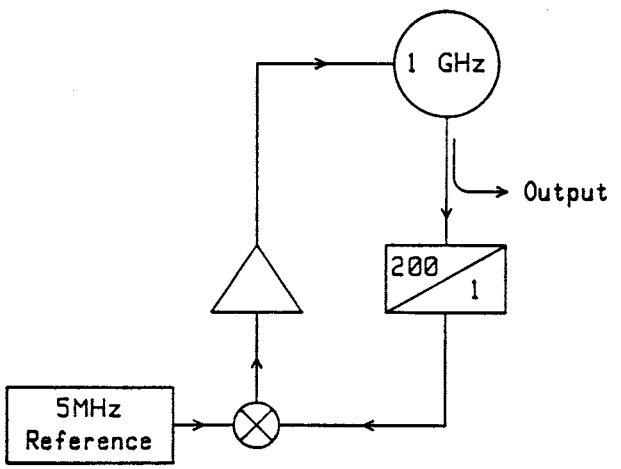


Fig. 3

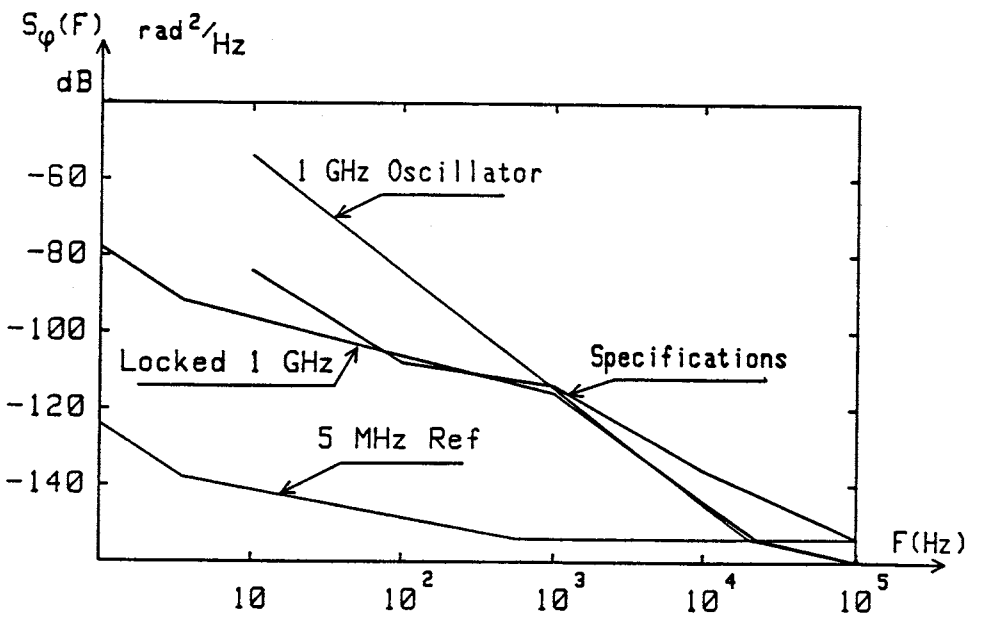


Fig. 4

When the bandwidth of the phase locked servo-loop is adjusted at 1 kHz, the phase noise spectrum of the phase-locked oscillator at 1 GHz satisfy the specifications.

B/ COMB SIGNAL GENERATOR

A S.R.D. comb generator is used to generate harmonics and subharmonics of the 1 GHz reference oscillator. A frequency divider, operating at the input of this reference oscillator, generates a signal at 100 MHz which is added to the reference signal before amplification and multiplication (Fig. 5).

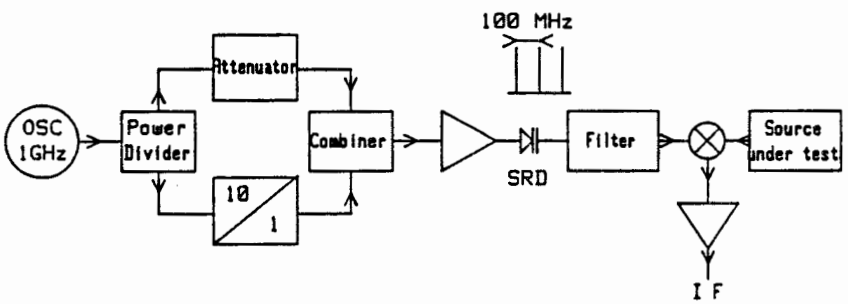


Fig. 5

The phase noise S_{Φ_1} of the 100 MHz signal is given by the relation :

$$S_{\Phi_1} = S_{\Phi} / N^2 + S_{\Phi_d}$$

S_{Φ_d} is the phase noise of the ECL divider, which is much lower than the specification, and thus S_{Φ_1} can be neglected.

The output of the comb generator is filtered, using a phase locked YIG oscillator.

C/ PHASE LOCKED YIG OSCILLATOR

A 2 - 6 GHz YIG oscillator is locked on one of the comb signal, and a synthesized output is introduced into the loop (Fig. 6).

The frequency of the YIG oscillator is predetermined, using the primary winding "TUNE" of the YIG.

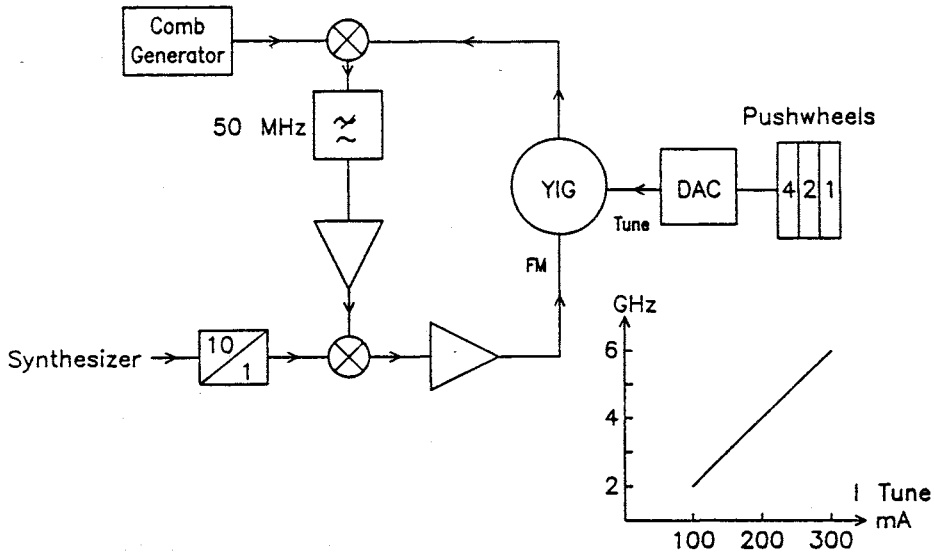


Fig. 6

This predetermination is done, using pushwheels which feed a 12 bits digital-analogic converter. The output signal goes, through a V/I converter, into the Tune winding of the YIG oscillator. The frequency of the YIG can thus be varied by step of 10 MHz ; a fine adjustement of this frequency is also possible.

A classical analysis of the phase locked loop gives the expression of the phase spectral density of the phase locked YIG oscillator ($S_{\Phi_y}(p)$) as a function of the open loop gain G_T.

$$S_{\Phi_y}(p) = \frac{S^{\ell}_{\Phi_y}(p)}{|1 + G_T|^2} + \left|\frac{G_T}{1 + G_T}\right| \left\{\frac{S_{\Phi_s}(p)}{N^2} + S_{\Phi_d} + S_{\Phi_3} + S_{\Phi_{M1}}\right\}$$

$S^{\ell}_{\Phi_y}$ is the YIG phase spectrum, $S_{\Phi_s}(p)$ is the output spectrum of the synthesized signal introduced into the loop after a division by N (N = 10), S_{Φ_r} and S_{Φ_d} are respectively the phase noise spectrum of the comb reference signal and the phase noise of the divider.

The microwave mixer introduces a white phase noise $S_{\Phi_{M1}}$ which should be lower than the specifications. This condition is satisfied if the level if the IF frequency at the output of the mixer is higher than -25 dBm.

We can see from Table I that this requirement is well satisfied in the operating range of the YIG oscillator, that is 2 - 6 GHz.

GIGAHERTZ

MEGAHERTZ

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
0	-6	-9	-12	-7	-10	-12	-12	-15	-18	-19	-15	-17	-19	-20	-20	-25	-25
100	-8	-12	-12	-8	-10	-8	-10	-11	-25	-15	-18	-18	-17	-20	-25	-24	-25
200	-7	-10	-12	-10	-10	-8	-11	-11	-24	-13	-16	-17	-18	-19	-25	-24	-25
300	-8	-10	-12	-14	-8	-10	-13	-16	-19	-15	-16	-16	-17	-17	-25	-25	-26
400	-14	-14	-9	-12	-8	-8	-9	-17	-13	-15	-18	-18	-18	-20	-26	-26	-24
500	-16	-10	-13	-11	-8	-13	-9	-18	-12	-15	-15	-19	-18	-20	-23	-26	-24
600	-14	-9	-15	-11	-7	-12	-10	-15	-15	-18	-14	-18	-18	-21	-20	-26	-24
700	-15	-10	-12	-9	-11	-11	-11	-12	-16	-16	-16	-18	-19	-22	-21	-26	-27
800	-12	-13	-10	-7	-12	-8	-11	-12	-20	-15	-18	-18	-16	-22	-21	-25	-26
900	-9	-17	-12	-8	-12	-8	-12	-12	-20	-15	-18	-18	-18	-20	-26	-25	-26

Table I

The free running YIG phase spectrum and the locked YIG phase spectrum are presented on Fig. 7.

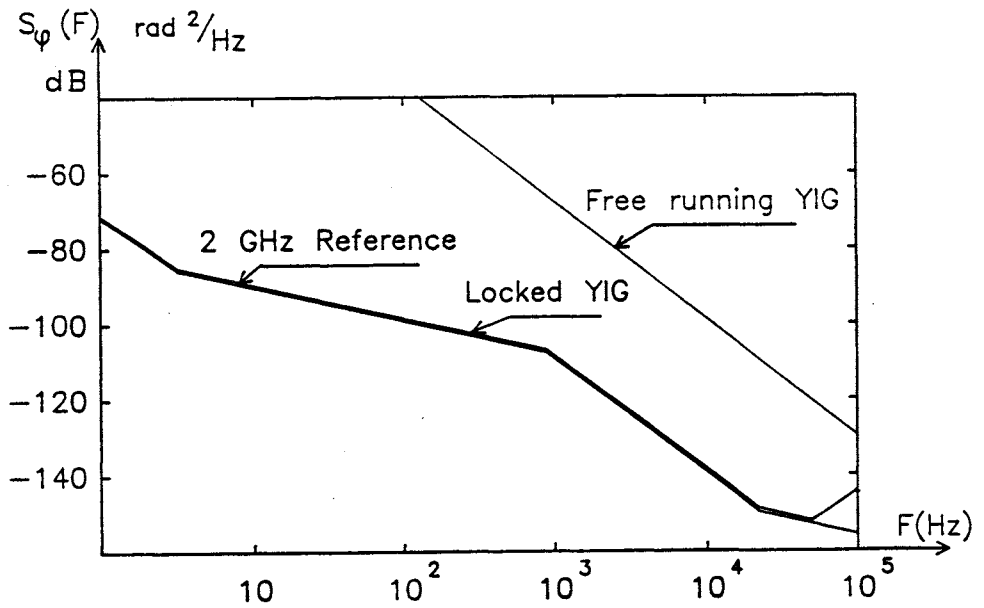


Fig. 7

The phase-locked YIG spectrum is identical to the reference oscillator spectrum inside the bandwidth of the servoloop, in the range from 10 Hz to 300 kHz.

2. SIGNAL ANALYSIS

The signal of the oscillator under test is down converted to 10 MHz by direct or harmonic mixing with the output of the YIG oscillator (Fig. 8).

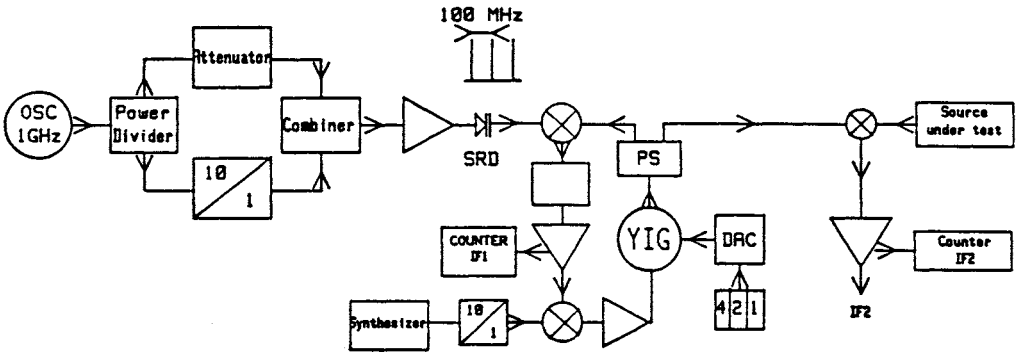


Fig. 8

The demodulation of the phase and frequency fluctuations uses either a passive demodulation when the microwave source is of medium spectral quality, or an active demodulation when the oscillator under test is of high spectral purity.

a) Passive demodulation

A Foster-Seeley discriminator is used to demodulate the frequency fluctuations of the IF signal. It has a bandwidth ranging from 10 Hz to 100 kHz.

b) Active demodulator

A quartz crystal voltage controlled oscillator (PCF3) is phase locked on the 10 MHz IF signal. A classical analysis of the phase locked loop gives the expression of the output signal V_3 :

$$V_3 = \frac{\mu G}{p + \mu K G(p)} [\dot{\Phi}_I - \dot{\Phi}_r]$$

$\dot{\Phi}_I$ represents the frequency fluctuations of the signal to be tested, and

$\dot{\Phi}_r$ the frequency fluctuations of the V.C.O. quartz reference.

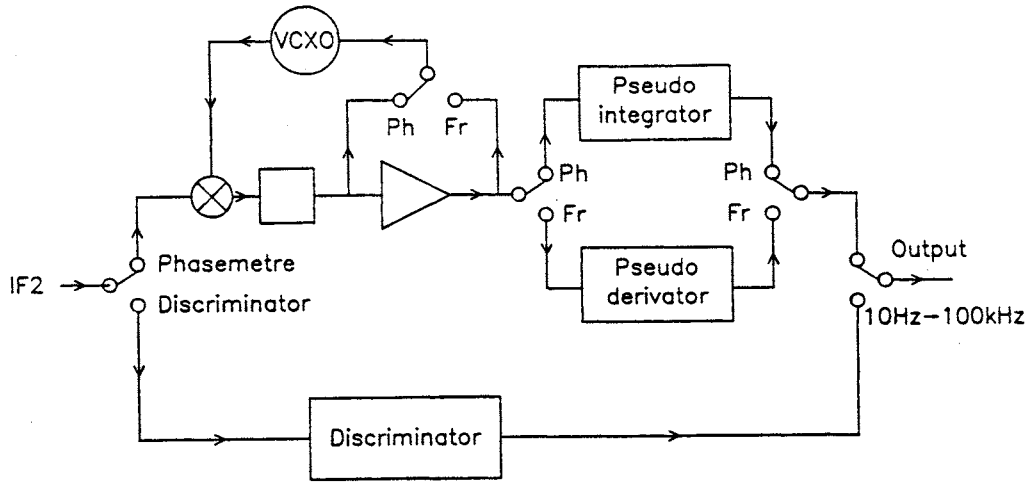


Fig. 9

Two situations must be considered :

a) $\mu KG(p) \gg p$

$$V_3(p) = \frac{\dot{\Phi}_I - \dot{\Phi}_r}{K}$$

The spectral analysis of $V_3(p)$ gives the frequency noise spectrum of the microwave oscillator : the noise of the VCO reference oscillator is lower than the specifications, and it can be neglected.

b) $\mu KG(p) \ll p$

$$V_3(p) = \mu G(p)[\Phi_I - \Phi_r]$$

The output signal is proportional to the phase fluctuations of the microwave oscillator outside the servo-loop bandwidth, which is 50 Hz for $G = 1$.

The conditions a) or b) can be realized using a variation of the open loop gain.

We obtain a noise measurement system which gives a signal proportional, either to the frequency fluctuations or to the phase fluctuations of the oscillator under test. A pseudo-integrator or a pseudo-derivator can be inserted at the output of the servo-loop, and the requirement of phase or frequency proportionality throughout the analysis bandwidth can thus be fulfilled.

A R.F. spectrum analyser must be used to obtain the phase -or frequency- spectrum of the microwave signal.

Phase and frequency fluctuations display

The RMS phase or frequency fluctuations of the oscillator under test, integrated over a predetermined bandwidth are displayed on the front panel of the noise measurement system.

This measurement uses the phase or frequency demodulated signal. It goes through two bandpass filters whose bandwidth are respectively 10 Hz - 100 kHz or 100 Hz - 12 kHz. The filtered signal feeds a RMS detector and the RMS output signal is then displayed (Fig. 10).

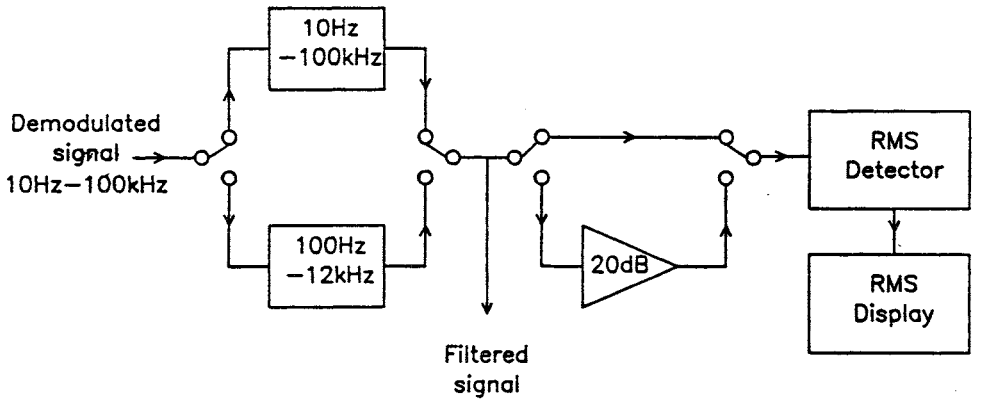


Fig. 10

Calibration and performances

The calibration of the apparatus is realized using a known sinusoïdal frequency modulation of the R.F. synthesizer. The observation of the amplitude of the detected spectral line enables us to scale the output spectrum.

The overall system noise floor of the phase -or frequency- measurement system has been calculated. It is represented on Fig. 11. The absolute noise contribution of the reference signal to the noise floor is predominant.

CONCLUSION

We have developed a phase and frequency noise measurement system which works in the range from 2 to 18 GHz. The demodulation and analysis of the microwave oscillator under test is performed in the R.F. domain, using a

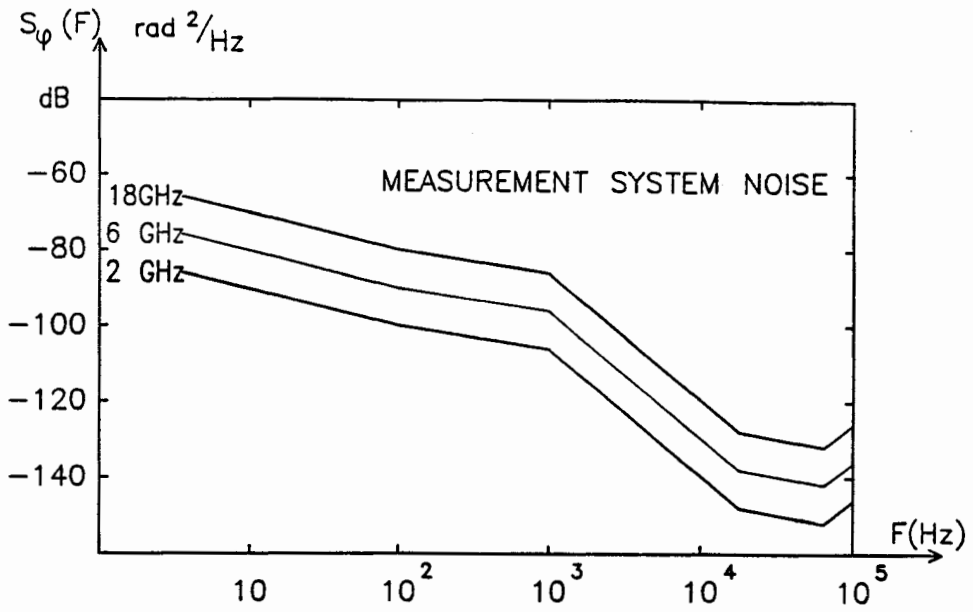


Fig. 11

microwave synthesized signal as a down converter. This system gives the possibility of testing most of the microwave source which are presently available.

Acknowledgements

The authors would like to express their thanks to the C.N.E.S. for providing the financial support ot these research.

REFERENCE

[1] M. Olivier, J. Groslambert (LPMO), J. Chauvin (CEPE), "Oscillateurs à quartz à 1 GHz utilisés en synthèse microonde", 1st European Time and Frequency Forum, Besançon, mars 1987.

Session TF-01 Number 4

Specifications and performances of ultra-stable oscillators in the Doris orbitography and positioning system.

J.B. LAPORTE - CENTRE NATIONAL D'ETUDES SPATIALES - TOULOUSE

The equipment used on board satellites or in stations belonging to the earth network are most frequently transmitters or receivers of electromagnetic waves.

There is a wide range of applications for ultra-stable quartz oscillators in this context. The criteria defining their specifications depend on the space missions (communications, earth observation, orbitography, location and altimetry).

This paper will specifically concentrate on the following two points :

1. Presentation of a space project whose objective is the precise determination of satellite positions and accurate location of beacons on the ground (project DORIS).

2. The ultra-stable oscillators used in onboard and ground equipment : specifications and performance.

ABSTRACT

The equipment used on board satellites or in stations belonging to the earth network are most frequently transmitters or receivers of electromagnetic waves.

There is a wide range of applications for ultra-stable quartz oscillators in this context. The criteria defining their specifications depend on the space missions (communications, earth observation, orbitography, location and altimetry).

This paper will specifically concentrate on the following two points:

1. Presentation of a space project whose objective is the precise determination of satellite positions and accurate location of beacons on the ground (project DORIS).

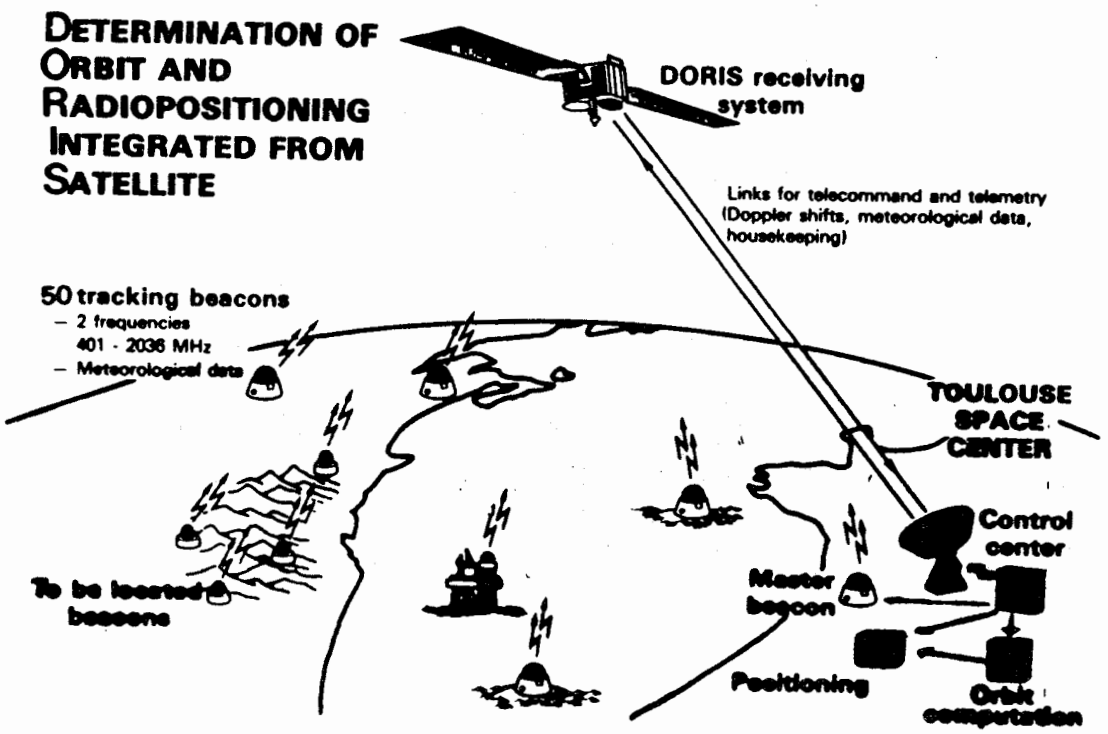
2. The ultra-stable oscillators used in onboard and ground equipment: specifications and performance.

1. PRESENTATION

Satellite-integrated radiopositioning and orbit determination (hereafter referred to as "DORIS").

The DORIS system was designed and developed by the French Space Agency (CNES), the Research Group into Space Geodesy (GRGS), and the National Geographic Institute (IGN) to meet new needs in the field of fine orbit determination (target accuracy of 10 cm) and the accurate location of beacons on the ground.

This system is based on measuring the doppler shift of radioelectric signals emitted by beacons on the ground and received on board the satellite when it passes over these beacons.

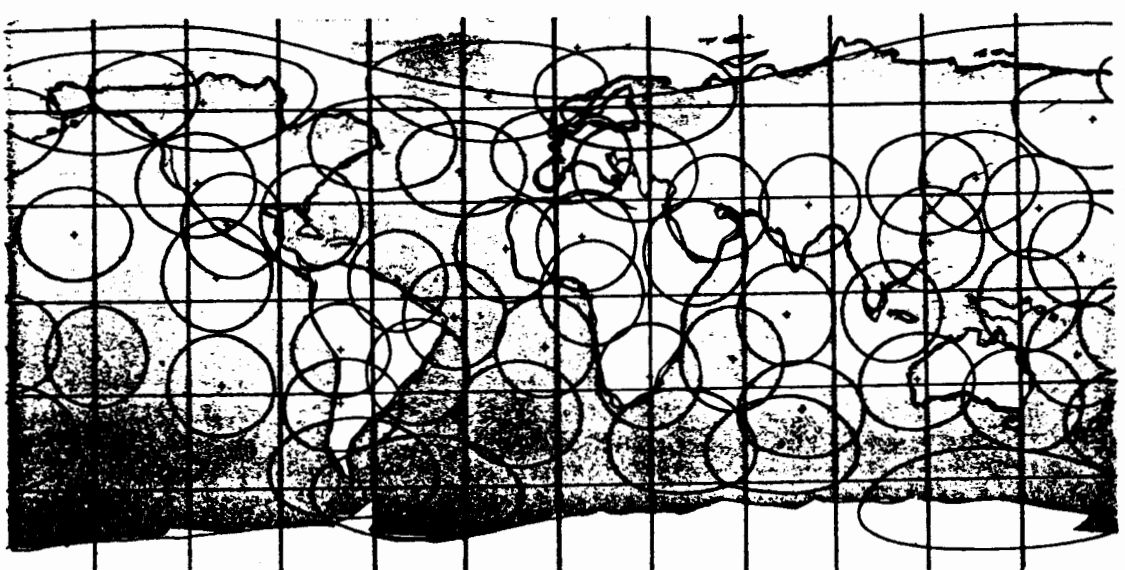


2. USES OF DORIS

2.1. ACCURATE DETERMINATION OF A SATELLITE'S POSITION

By measuring the doppler shift, it is possible to calculate the position of a satellite in space with a precision of approximately 10 cm by combining a network of beacons spread uniformly around the globe and emitting radioelectric signals with precise measurements taken on board the satellite.

Such precision is needed at present to meet various objectives but in particular to determine sea height by radar altimetry. It is for this purpose that the DORIS instrument will be taken on board TOPEX, the American oceanographic satellite (1992).



Carte représentant la distribution des balises d'orbitographie et les zones dans lesquelles les satellites SPOT sont en visibilité.

2.2. ACCURATE LOCATION OF POINTS ON THE GROUND

Once the satellite's orbit is known, it is possible to locate beacons in unknown positions accurately (10 cm). DORIS will thus be particularly suited to monitor natural phenomena:

- measurement of surface deformation in seismic or volcanic areas,

- continental drift,

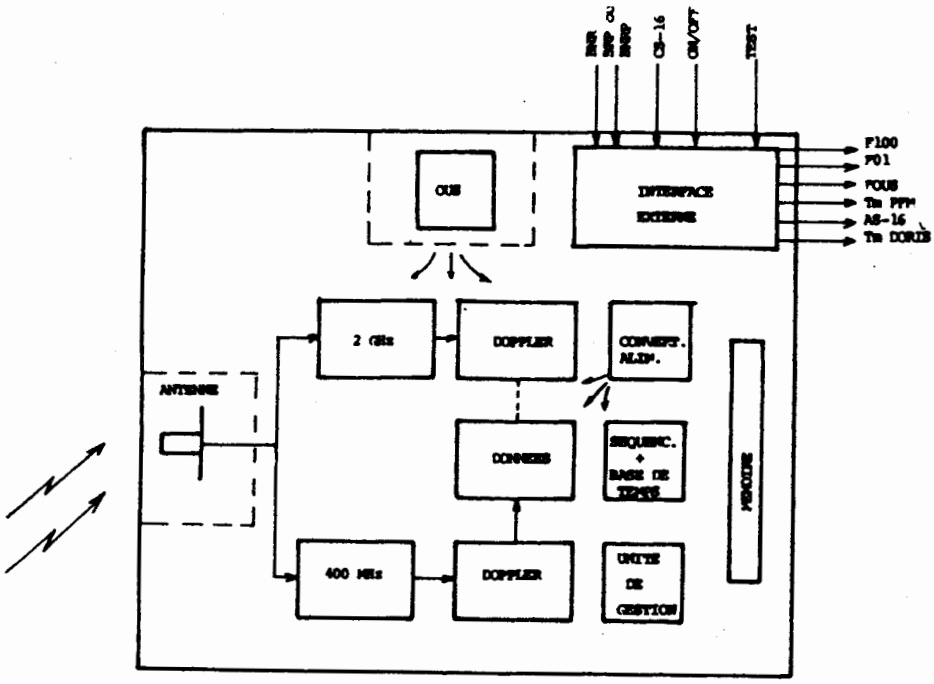
- monitoring sites during civil engineering work (e.g. building dams or tunnels).

The first edition of the DORIS instrument will be taken on board SPOT2, the French satellite due to be launched in March 1989. It is planned to carry later editions onboard satellites SPOT3, SPOT4 and SPOT5.

3. DESCRIPTION OF THE DORIS SYSTEM

3.1. ONBOARD EQUIPMENT

This comprises both a 401.25 MHz and 2036.25 MHz receiver including a doppler counting system controlled by an ultra-stable oscillator (USO) and memory-interface management assembly.



3.2. ORBITOGRAPHY BEACONS

There are 50 orbitography beacons designed to operate in laboratory conditions. A beacon is composed of two emitters, one at 401.25 MHz and the other at 2 036.25 MHz controlled by an USO and:

- atmospheric sensors,
- control unit,
- power supply unit.

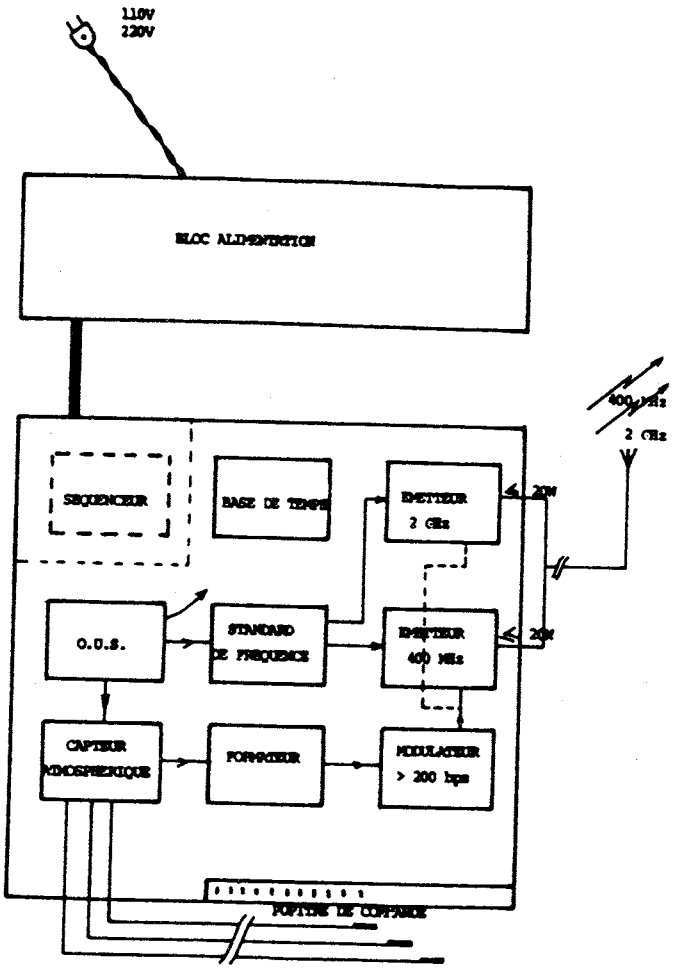


Fig.2 : Balises d'orbitographie

DIAGRAM OF AN ORBITOGRAPHY BEACON

3.3. LAND BEACONS

These beacons operate in the same way as orbitography beacons. Their housing is suited to use in any part of the globe.

3.4. THE MASTER BEACON

Whenever the satellite passes this beacon, CNES's Control Centre in Toulouse can transmit the data needed for its operating programme to the onboard receiver.

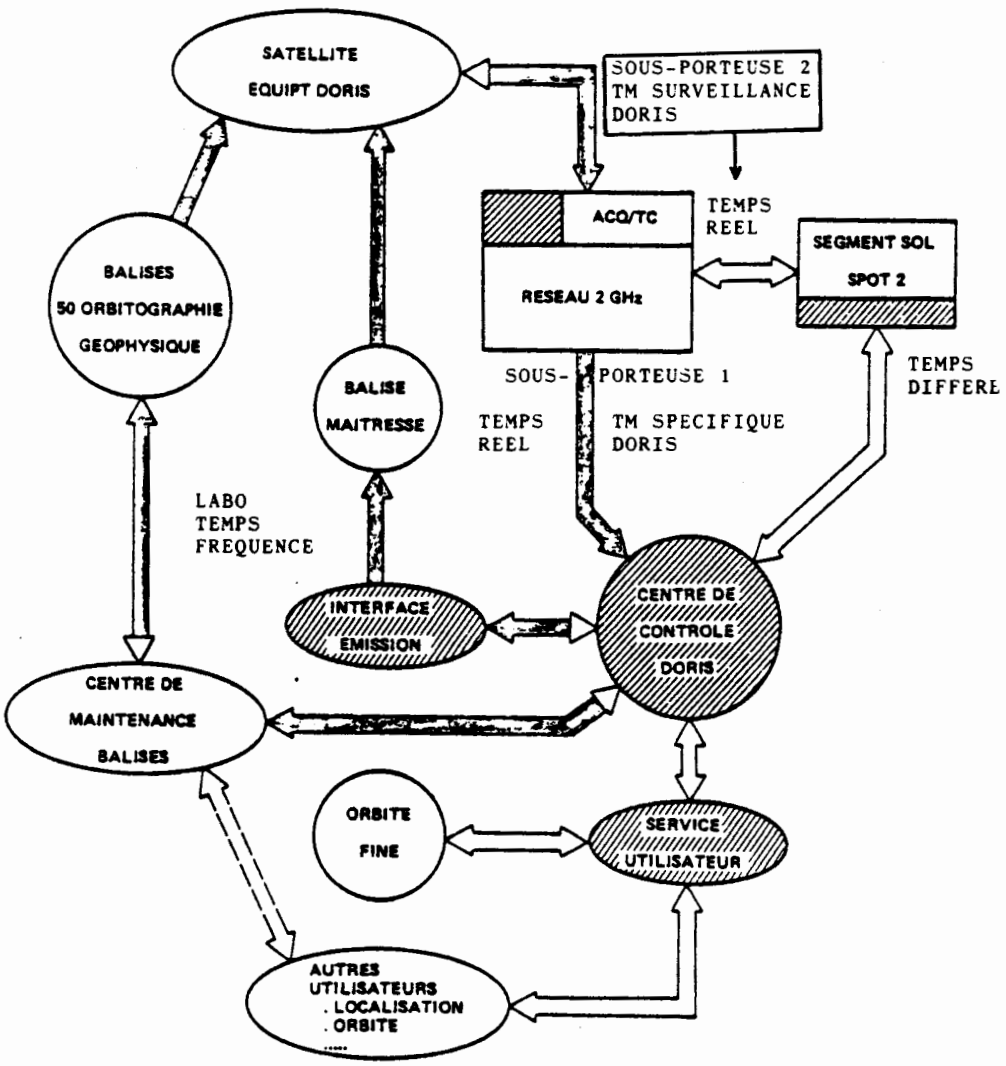
The beacon-satellite links are all synchronized to TAI (International Atomic Time) through the master beacon in Toulouse, itself referenced to the CNES Caesium clocks' atomic time and TAI. It is aimed to know the onboard time to within $\pm$ 10 s.

The master beacon is controlled by a reference frequency from an USO assembly slaved to a Caesium clock designed and made by the Laboratory for the Physics and Metrology of Oscillators in Besançon, France (LPMO).

This control system provides a reference whose performance in the short, medium and long term is compatible with DORIS specifications, using as it does a high time constant.

3.5. THE DORIS CONTROL CENTRE

The DORIS Control Centre in Toulouse receives the measurements taken by the satellite, monitors the onboard receiver and beacons and draws up the telecommands and programmes of work for the instrument.



SYSTEME DORIS / SPOT 2

4. SPECIFICATIONS AND PERFORMANCE OF THE USOs USED IN DORIS

4.1. MAIN SOURCES OF ERRORS SIGNIFICANTLY REDUCING TARGET ACCURACY

These sources are summarized in the table below:

CAUSES OF ERROR	EFFECT ON THE ACCURACY
Onboard USO + ground USOs	8 cm
Tropospheric error (after correction by model)	6 cm
Instrument errors - doppler counter - link noise - electronic circuits - delays	3 cm
Onboard and ground antennae	1 cm
Ionospheric error (after correction by model)	2 cm

Ionospheric errors are largely eliminated by using two frequencies (2 GHz and 400 MHz).
Supposing these errors add up quadratically, a target accuracy of 10 cm would appear reasonable.

The errors noted above for the onboard and ground USOs correspond to the specifications of chapter 4.2.

It may be seen that USOs play a large part in the overall causes of error.

4.2. SPECIFICATIONS FOR ONBOARD AND GROUND USOs

In order to work out a satellite orbit to within 10 cm, it is necessary to be able to measure its radial velocity to within 0.3 mm/second. This implies frequency stability of 2.10^{-3} Hz on the frequency transmitted at 2 GHz. This degree of stability assumes that the performance of onboard and ground USOs are less than or equal to the values shown in the table below.

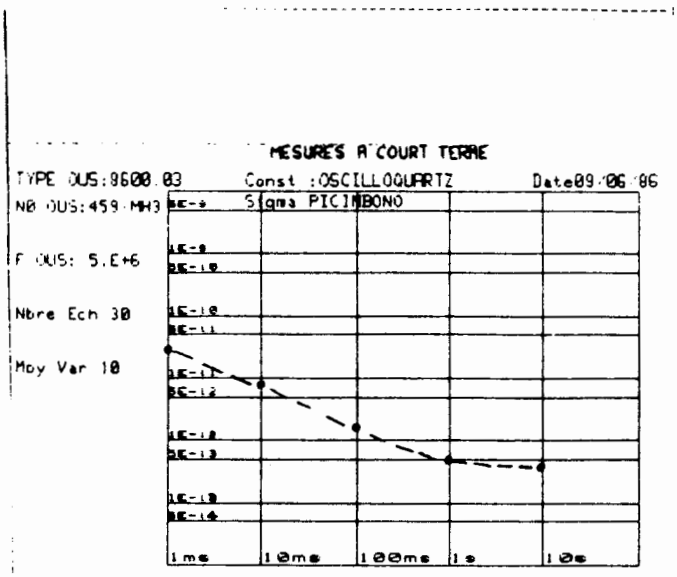
	ONBOARD USO	GROUND USO
FREQUENCY	5 MHz	5 MHz
Short-term stability according to Picinbono (10 s)	$\leqslant 5.10^{-13}$	$\leqslant 5.10^{-13}$
Medium-term stability with a temperature gradient of 5 °C/hour Slope	$\leqslant$ +/- 2.5×10^{-13}/mn	$\leqslant$ +/- 2.5×10^{-13}/mn
Sigma	$\leqslant 1.10^{-12}$	$\leqslant 1.10^{-12}$
Long-term stability	1.10^{-10}/24 hours	1.10^{-10}/24 hours
Consumption	3 W in space vacuum	3 W
Power supply	24 VDC	22 VDC
Temperature	0 to 40 °C	0 to 45 °C
Pressure	10^{-6} mb and 1100mb	1100 mb

4.3. PERFORMANCES MEASURED AT CNES ON ONBOARD AND GROUND USOs

4.3.1. Short-term performance

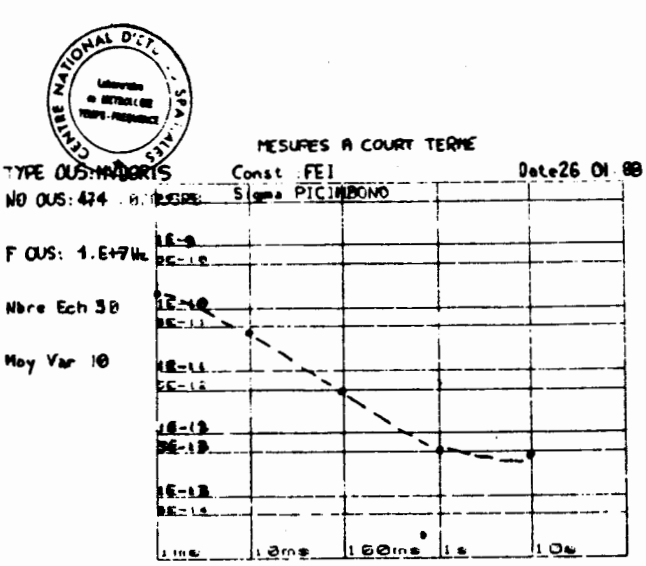
This performance is characterized by the Picinbono variance of 3 consecutive frequency samples for a count time of = 10 seconds.

$$\sigma y_{(\tau)} = \frac{1}{F_0} \sqrt{\frac{\sum_{i=1}^{N} \left(\frac{F_{(i+2)} + F_{(i)}}{2} - F_{(i+1)} \right)^2}{2N}}$$



VARIANCE DE PICINBONO SUR 30 ECHANTILLONS			
T	Minimum	Moyenne	Maximum
1 ms	1.954E-11	2.820E-11	3.672E-11
10 ms	5.419E-12	8.040E-12	1.192E-11
100 ms	9.425E-13	1.668E-12	2.431E-12
1 sec	4.044E-13	5.028E-13	6.220E-13
10 sec	2.820E-13	3.929E-13	5.425E-13

GROUND USO



VARIANCE DE PICINBONO SUR 30 ECHANTILLONS			
T	Minimum	Moyenne	Maximum
1 ms	1.333E-10	2.407E-10	3.248E-10
10 ms	3.018E-11	5.912E-11	5.497E-11
100 ms	5.604E-12	4.61nE-12	5.519E-12
1 sec	5.458E-13	8.324E-13	7.267E-13
10 sec	2.848E-13	4.334E-13	5.548E-13

ONBOARD USO

The 5.10^{-13} specification is clearly met by both onboard and groun oscillators.

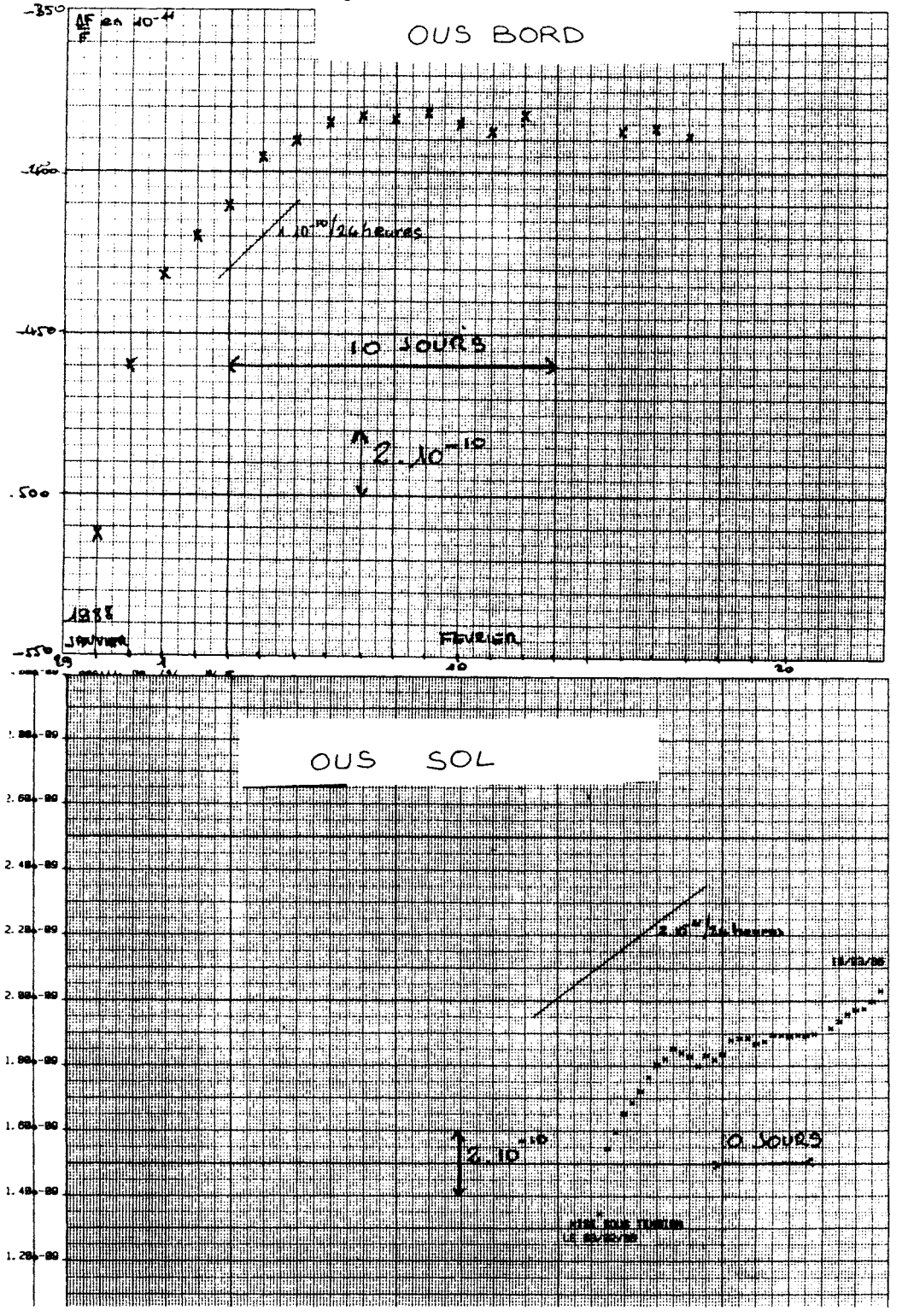
4.3.2. Long-term performance

Long-term deviation does not affect the doppler measurement directly but influences the time-tagging of the beacon message throughout the mission.

The values obtained are as follows:

6.10^{-11}/24 hours for the onboard USO after 1 week.

2.10^{-11}/24 hours for ground USOs.



4.3.3. Medium-term performance

Such performance is defined for the average time during which the satellite is visible to a beacon (15 minutes). The straight line of regression is calculated using the method of least squares of Slope p/minute and Sigma dispersion F/p on tests of 90 frequency measurements counted over 10 seconds.

Slope/mn

$$P/mn = \frac{12 \sum_1^N i \cdot fi - 6(N+1) \sum_1^N fi}{N(N^2-1) \cdot T \cdot F\phi}$$

Sigma dispersion

$$\sigma_{F/P} = \sqrt{\sum_1^N \frac{(fi - fip)^2}{N}}$$

$$fip = T \cdot p \left(i - \frac{N+1}{2} + M \right)$$

$$M = \frac{\sum_1^N fi}{N}$$

T = $\frac{1}{6}$ minute

F = Fnominal USO

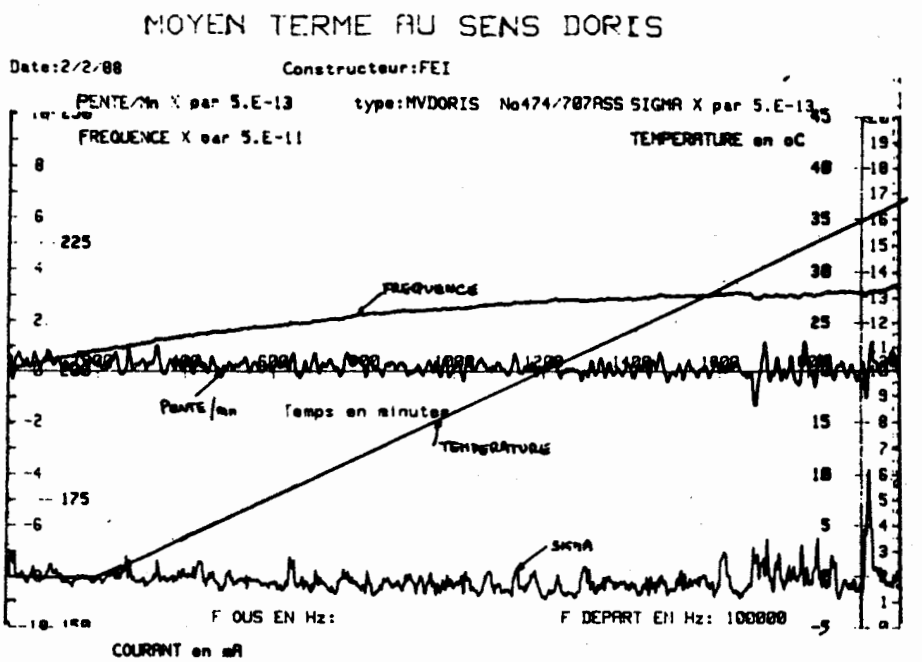
N = 90

Throughout the evaluation process, the oldest samples are rejected over time and new ones entered into the formula.

The evaluation of medium-term stability is checked at a stabilized temperature with a temperature gradient of 1.5 to 5 °C/hour in the 0 to 45 °C range.

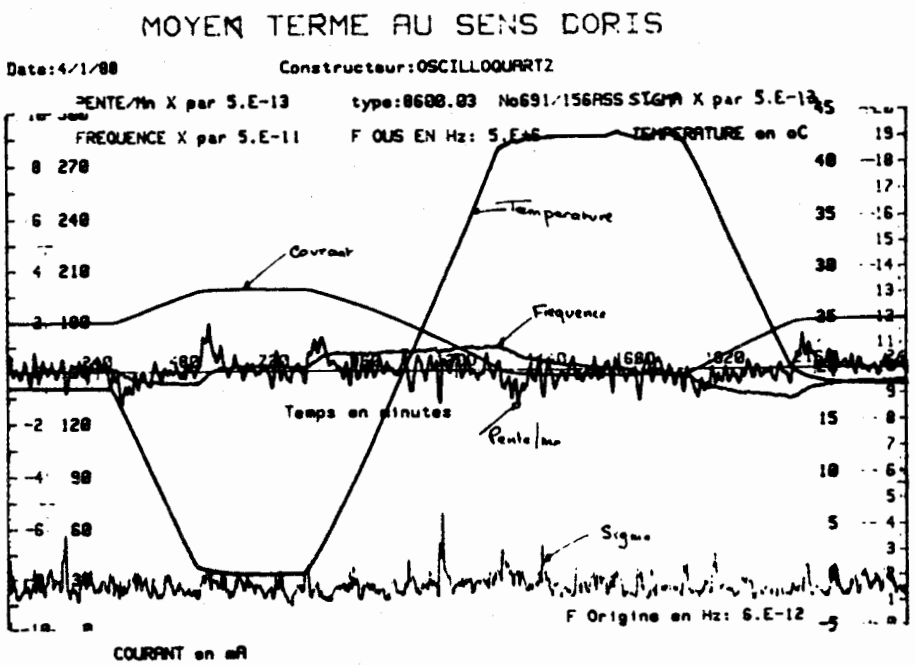
The onboard USO was evaluated in a vacuum (10^{-6} TORR).

Its performance is as shown below.



The ground USOs were evaluated in a ventilated oven.

Their performance is as shown below.

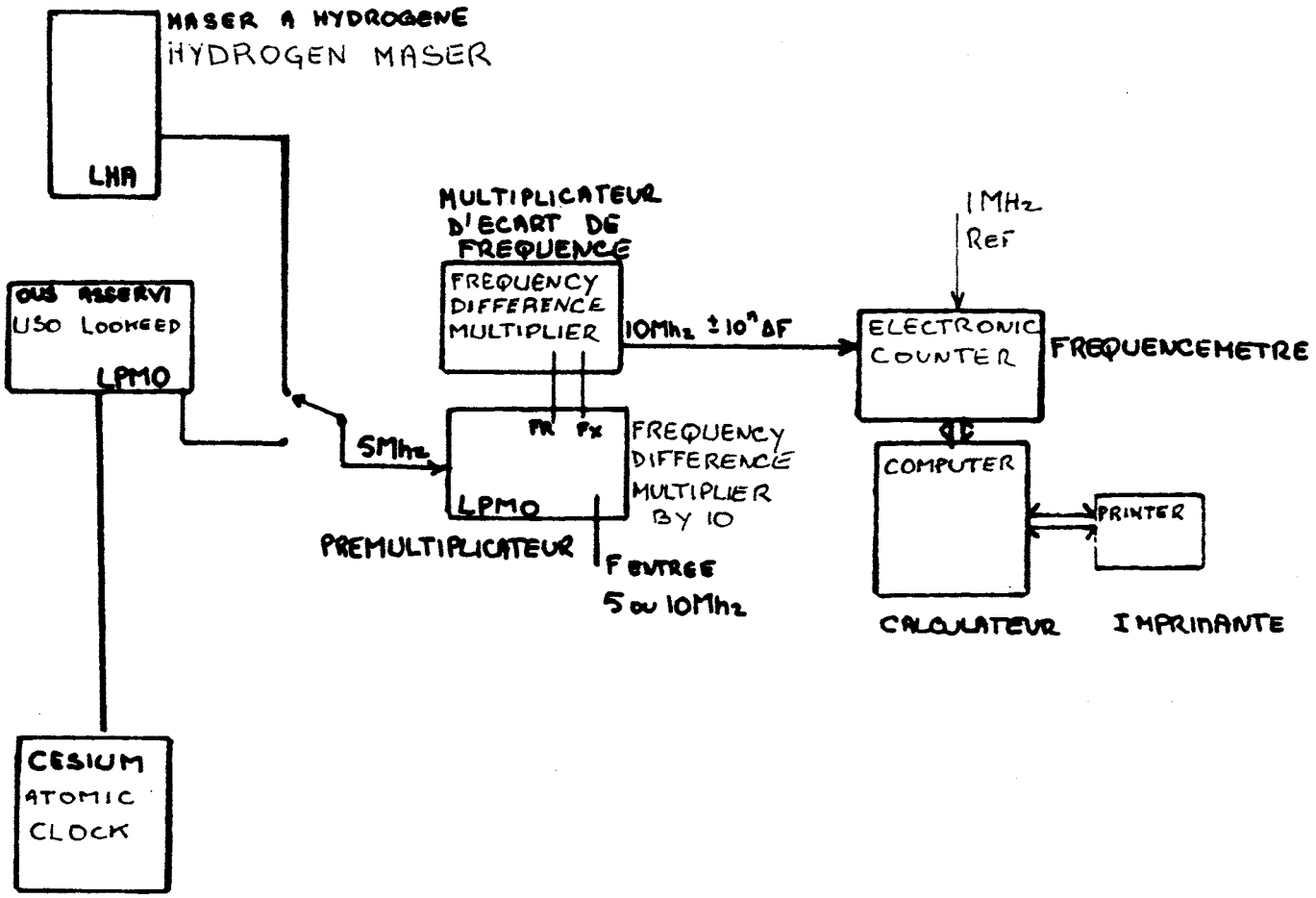


5. MEASURING FACILITIES USED

5.1. REFERENCES

The USOs were evaluated in the CNES "Time-Frequency" laboratory with reference to a hydrogen maser designed and made by the Atomic Clock Laboratory (LHA, Orsay) or to an USO assembly slaved to an atomic standard: assembly designed and made by the LPMO, Besançon.

5.2. DIAGRAM OF THE MEASURING BENCH



This bench can highlight differences in frequency expressed in $\Delta F/F$ of 1.10^{-15} for 5 MHz input frequencies.

6. THE EFFECTS OF A MAGNETIC ENVIRONMENT ON USOs

The onboard USO showed a relatively high level of sensitivity when a magnetic field was applied. This effect appeared during DORIS integration tests for SPOT2. The main cause of these magnetic fields is the satellite's magnetic couplers (used for stabilizing the satellite in space).

Frequency jumps of 2 to 3.10^{-11} were recorded.

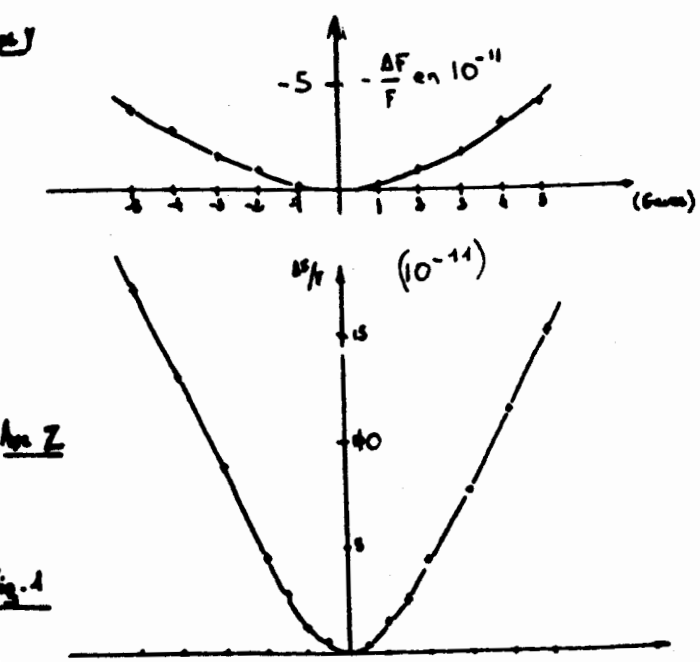
In order to considerably attenuate the magnetic field affecting the USO (40 dB), the oscillator was housed in a double-layer screen made of material with high magnetic permeability.

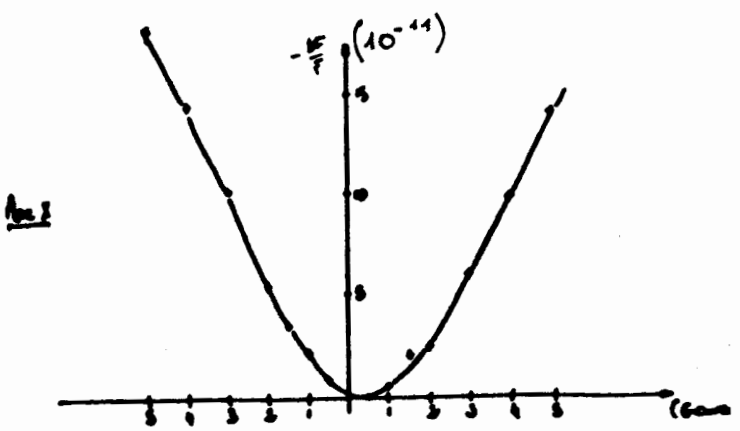
Measures taken in the CNES-INTESPACE non-magnetism laboratory were used to quantify the USO's sensitivity. Such evaluations were made using a compensated magnetic field under earth conditions.

A negative frequency variation occurs whenever a magnetic field is applied along a given axis.

The maximum differences in frequency recorded were -2.9×10^{-11} for a field of 60.000 nT.

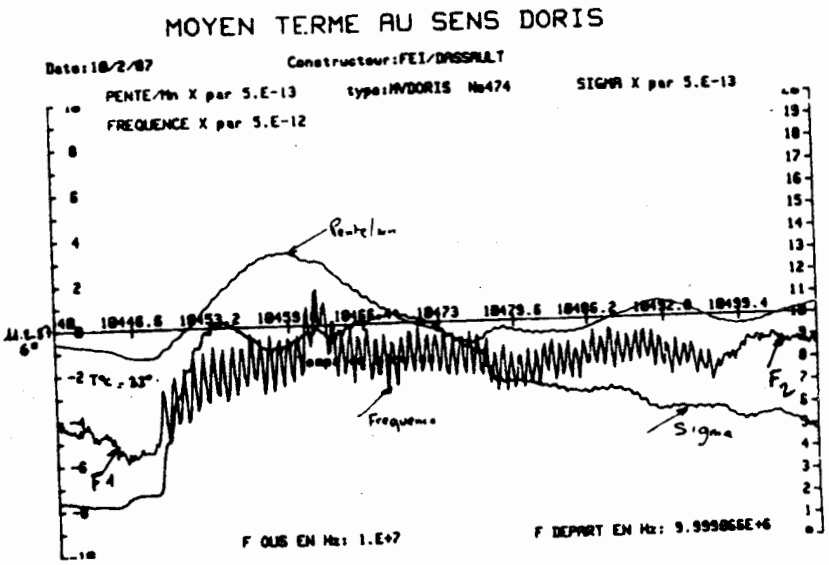
Further measurements of sensitivity were taken using a strong field (max. of 10 Gauss) and gave the following results:





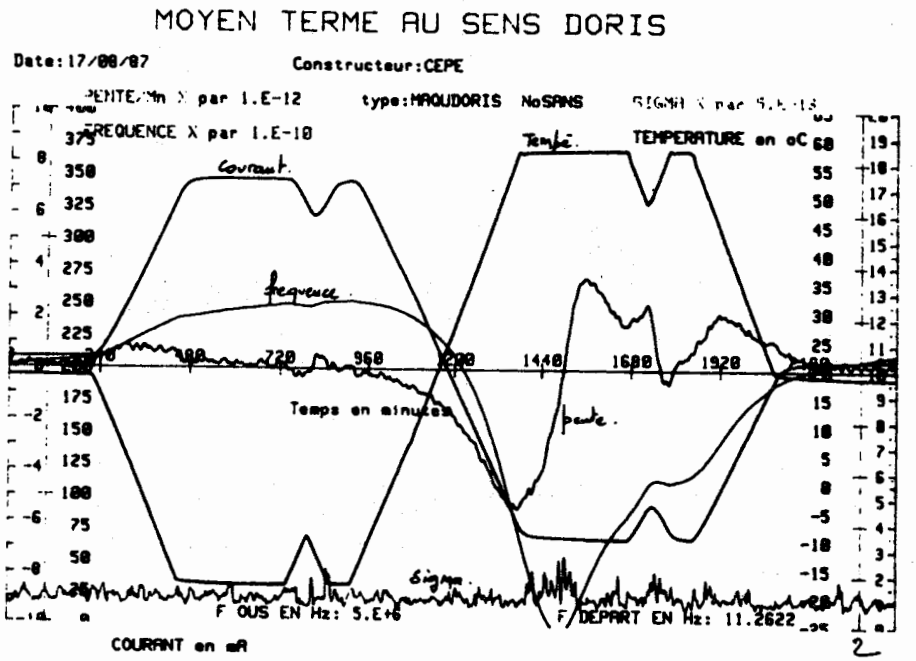
7. THE EFFECTS OF A THERMAL ENVIRONMENT ON USOs

7.1. 1st CASE



This graph shows frequency oscillations which always appear at a similar temperature in the 33 °C to 35 °C range. Points F1 and F2 show that the frequency is stable before and after these oscillations.

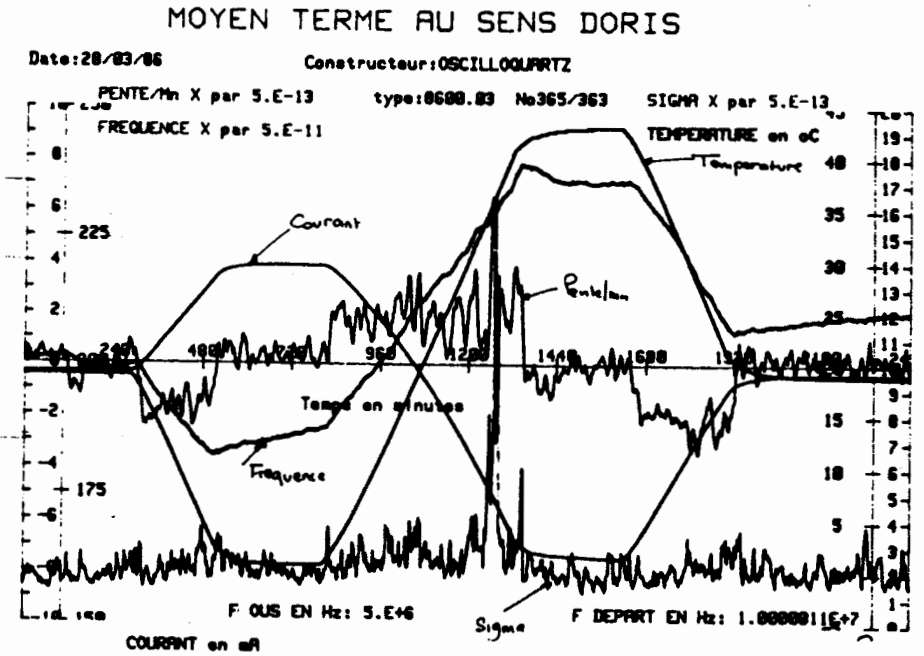
7.2. 2nd CASE



The variation in frequency is of 2.10^{-10} in the - 20 to + 30 °C range and 1.10^{-9} in the + 30 to + 55 °C range.

The temperature gradient applied is 10 °C/hour in the - 20 to + 55 °C range.

7.3. 3rd CASE



The frequency follows the temperature profile.

There is a variation in frequency of 5.10^{-10} in the 0 to 45 °C range.

This type of performance lies outside the tolerances required for the DORIS system.

8. CONCLUSION

Faced with the performance currently required for accurately determining the position of satellites and accurate location on the ground, it is necessary to have extremely effective ultra-stable quartz oscillators.

QAS and BVA, new technologies in quartz, have led to improvements in both short- and long-term performance. However, the gains achieved are now offset by effects of the environment, particularly temperature, magnetism and pressure. Manufacturers must therefore take these criteria into account in order to improve the performance of ultra-stable oscillators.

TF-02

G P S

1. Nanosecond time comparisons in Europe using the GPS
   B. Guinot, W. Lewandowski
   Bureau International des Poids et Mesures, Sèvres - France

2. GPS High Precision Time Transfer Methods and Related Equipment
   R. Gounon
   Sercel, Carquefou - France
   J.N. Lelong
   Osafrance SA, Boulogne - France

3. Operational Highly Precise Time Transfers in France by GPS and by the TV Method
   M. Granveaud, P. Parcelier, G. Freon, R. Tourde
   Laboratoire Primaire du Temps et des Fréquences, Paris - France
   M. Vincent
   Observatoire de Besançon - France
   P. Grudler
   Centre d'Etudes et de Recherches Géodynamiques

4. Improvement of international time comparison of OFM using GPS common view satellite observations
   L. Prost, H. Brandenberger
   Swiss Federal Office of Metrology, Wabern - Switzerland

2 ND EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 16 - 17 - 18 MARCH 1988

Session TF-02 Number 1

Nanosecond time comparisons in Europe using the GPS

by B. Guinot and W. Lewandowski

Bureau International des Poids et Mesures
Pavillon de Breteuil, 92312 Sèvres Cedex

Summary

The time comparisons by the Global Positioning System (GPS) have a resolution of the order of 3 ns for a single measurement over 13 minutes. However systematic differences of 10 ns (one sigma) often appear between data derived from different satellites and also from the same satellite observed at different positions in the sky. Over the usual distances in Europe, the errors due to the uncertainties of the satellite position should not exceed 1 ns; the error of the ionospheric model have also a small influence and cannot explain the biases. The main source of systematic errors lies in the geodetic coordinates of the receivers.

We have derived a method of differential positioning using the data of time comparisons themselves, without requiring extra trackings and extra data transmission. The uncertainties of the coordinates are normally below 50 cm. We have also established a software which takes into account corrections of geodetic coordinates in the current processing of GPS time comparisons.

Applying these methods to the European laboratories, we restore the homogeneity of time comparisons, whichever be the observed satellites. Averaging over short duration, one day, for instance, becomes possible. On one-day means, an accuracy of 1,5 ns is reached (providing a good calibration of instrumental delays).

Introduction

We assume that the reader has some knowledge of the Global Positioning System, GPS (or NAVSTAR). Otherwise, he could find easily descriptions of the system in navigation journals (for instance, Milliken and Zoller, 1980).

The GPS is now widely used for time comparisons. In Western Europe, among 14 national laboratories which participate to the world-wide coordination of time, 11 are equipped with GPS time receivers.

We have found that over distances of about 1000 km or less, which are common in Western Europe, the main source of errors in the GPS time comparisons may be the adoption of wrong geodetic coordinates of the antennae. Even the TRANSIT positioning, with conversion in the WGS84 used by GPS, is not sufficiently accurate. Fortunately the coordinate errors can be derived from the time comparisons themselves.

We will briefly present our method of coordinate correction and its application to all laboratories of Western Europe. This work was based on schedules of satellite tracking which were not optimized for geodetic work. But the new schedule, implemented on 1987 December 15, has been greatly improved in these respects and more accurate results can be expected toward July 1988.

Geometrical errors in the GPS time comparisons by simultaneous tracking

The GPS time comparisons are usually based on simultaneous tracking of the satellites. These "common views" are not necessary, but have the advantage of cancelling the role of the satellite clocks (Allan and Weiss, 1980).

Let A and B be the positions of the GPS antennae of two laboratories (also designated by A and B) and S_k the position of the satellite they track. The unit vectors of AS_k and BS_k are $\vec{a}_k$ and $\vec{b}_k$ (fig. 1). The true value of the clock comparison at the instant of tracking is designated by

$$\text{clock reading of A - clock reading of B} = U_{ab}.$$

It is everywhere assumed in this paper that the relativistic conversion in geocentric coordinate time is made.

If the broadcast ephemeris corresponds to the erroneous position S'_k, the measured time comparison $(U_{ab})_k$ is erroneous, and we define

$$E_k = U_{ab} - (U_{ab})_k. \qquad (1)$$

E_k is given by

$$E_k = - c^{-1} \overrightarrow{S'_k S_k} \cdot (\vec{a}_k - \vec{b}_k), \qquad (2)$$

c being the velocity of light. For a given $\overrightarrow{S'_k S_k}$, E_k is roughly proportional to the chord $|\overrightarrow{AB}|$.

If the adopted coordinates for B correspond, not to B, but to B', the measured time comparison requires a correction F_k defined as in (1)

$$F_k = - c^{-1} \overrightarrow{B'B} \cdot \vec{b}_k, \qquad (3)$$

independent of the location of A.

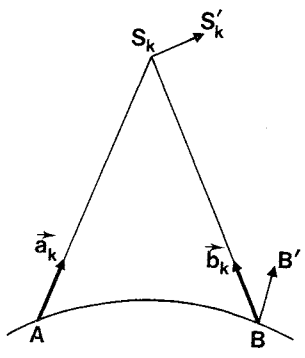


Fig. 1

We can thus get an idea of the errors due to the satellite ephemeris for the short distances of interest in this study. The laboratories USNO and OP, which are 6000 km apart, have a good positioning. When considering the clock differences between USNO and OP, for the same day, as given by different common views (different satellites, or same satellite at different locations in the sky), their standard deviation with respect to the mean decreased from 15 ns in 1985 to 7 ns in 1987. We adopt in the following 10 ns as a typical value.

If we assume that this scattering is entirely due to E-type errors, we get a very conservative estimate of the contribution of the E-type errors at 600 km by dividing by 10. Thus a maximum of this contribution is 1,0 ns (one sigma) on a single common view. As it is customary to observe 10 to 20 common views per day, the E-type errors should contribute by less than σ = 0,3 ns on the daily average of clock comparisons.

Determination of station coordinates

The period of the GPS satellites is half a sidereal day. The tracking schedules for time comparisons repeat themselves a one sidereal day-interval. Thus, each scheduled common view (SCV) identified by index k (k = 1, ..., n) is observed at constant altitude and azimut. According to (3) F_k remains constant, so that for the determination of station coordinates, it is possible to work on averages extended over several months. These averages further decrease the influence of the errors of the satellite ephemerides; they also reduce the errors due to the adopted values for the differential refraction delays.

For the same day, we assume that the $(U_{ab})_k$ are all transfered at the same instant (0 h UTC, for instance). Usually the adopted time comparison is

$$\hat{U}_{ab} = \frac{1}{n} \sum_{k=1}^{n} (U_{ab})_k = U_{ab} - \Delta T, \qquad (4)$$

with the unknown ΔT defined by

$$\Delta T = \frac{1}{n} \sum_{k=1}^{n} F_k. \qquad (5)$$

From (4) and (3), we arrive at the equation

$$c^{-1} \overrightarrow{B'B} \cdot \vec{b}_k + \Delta T = (U_{ab})_k - \hat{U}_{ab}. \qquad (6)$$

The right-hand member of (6) is obtained every day. The time average, for each k, can be computed: we assume henceforward that it has been done. In the left-hand member, the unknowns are ΔT and the components ΔX, ΔY, ΔZ of $\overrightarrow{B'B}$ in the usual coordinate system, O at the geocenter, OZ toward the pole, OX in the prime meridian, OY toward East. More precisely, ΔX, ΔY, ΔZ are corrections to add to the adopted coordinates of B in order to express them in the same frame as the coordinates of A (differential positioning).

The system of k equations (6) is solved by the least square method. The $\vec{b}_k$ are easily expressed in the OXYZ system, using the altitude and azimut of S_k which are included in the standard format for GPS data exchanges.

The method has been applied to the European laboratories shown in the table of distances (Table I) and also to Israel. Some distances exceed largely 1000 km, but the results seem nevertheless interesting.

Over the 45 possible links between pairs of laboratories, 25 have been computed and a global fit gives the corrections to adopted coordinates reported in Table II. We insist on the fact that the positioning is differential. The arbitrary origin has been adjusted on

IEN, OP and VSL which have used TRANSIT positioning. The standard deviations are valid for individual stations, but their combination does not give the standard deviation of relative coordinates for pairs of laboratories — in some cases the latter may be much smaller.

We will not discuss here the geodetic aspect of the work. We only note that our results agree with the TRANSIT absolute positioning, taking into account the uncertainty of the latter (± 1 m), but that they are more precise. Comparisons with laser ranging on LAGEOS and VLBI (± a few centimeters) has not yet been possible.

Use of coordinate corrections for accurate time comparisons

Let us consider again a pair of laboratories. Instead of using ΔT to correct $\hat{U}_{ab}$, it is better to correct the $(U_{ab})_k$ by applying coordinate corrections to, let us say, B, by (3). This is easily done by computation, using the known altitude and azimut of S_k.

Fig. 2 illustrates the improvement brought in the case of differential errors of a few metres. The random uncertainties on daily averages of time comparisons (18 SCV's) are, for instance:

OP-PTB	0,7 ns
PTB-OMSF	1,0 ns

A bias could remain if the satellite positions are broadcast with a systematic shift with respect to the adopted coordinates. It would amount to 1 ns when $\overrightarrow{S'_k S_k} \,//\, \overrightarrow{AB}$ and equal to 7 m, for all satellites, the most unfavorable case. This is unlikely, and the bias should be well below 1 ns.

Another bias is due to the remaining error of the differential coordinates of B with respect to A. It is maximum for $\overrightarrow{B'B} \,//\, \overrightarrow{BS_k}$.

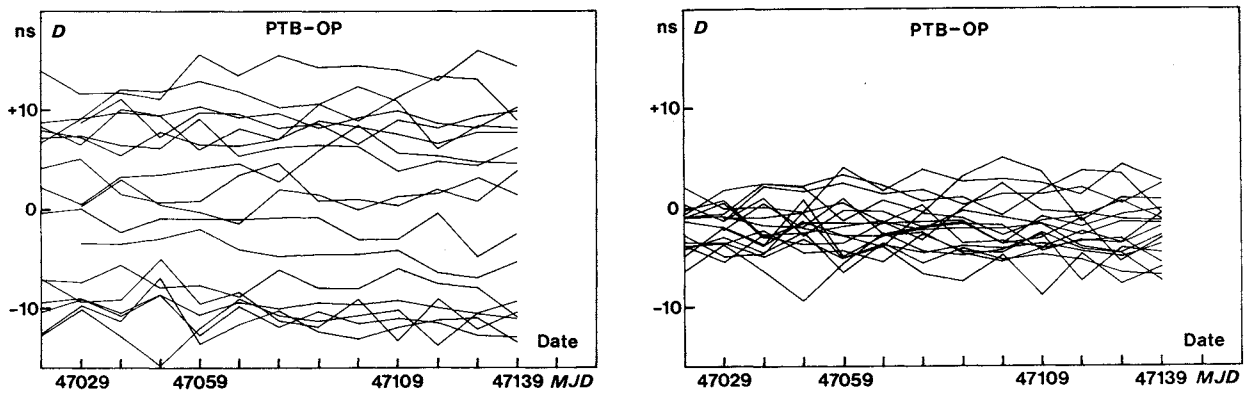


Fig. 2.- D is $(U_{ab})k - \hat{U}_{ab}$ of equation (6), averaged over 10-day intervals. One line is dawn for each k, on the left with the adopted coordinates, on the right after our coordinate correction.

A conservative estimate is $c^{-1}|\overrightarrow{B'B}|/2$, corresponding fo $\overrightarrow{B'B}$ vertical and uniform repartition of S_k on the sky. $\overrightarrow{B'B}$ is evaluated by

$$\overrightarrow{B'B} = (\sigma^2_X + \sigma^2_Y + \sigma^2_Z)^{\frac{1}{2}}$$

where the σ are those of the differential positioning. After quadratic sum with the random uncertainty, we get the following total uncertainty due to positioning:

for OP-PTB : 1,0 ns,
for PTB-OMSF : 2,2 ns (over 2100 km).

Uncertainties of GPS time comparisons

It is not within the scope of this paper to discuss in detail the other sources of uncertainty (see Nard et al., 1987).

The differential receiver delays can be calibrated with uncertainties of the order of 1 ns, or even less (Lewandowski et al., 1986). We will take 1 ns.

In the conditions of this study, short distances, observations spread in time and direction, it seems reasonable to assume that the systematic component of the differential refraction, not corrected by the model, is well below 1 ns; we take 0,5 ns.

Other sources of uncertainty may be multipath propagation, anisotropy of the antenna delay, receiver delay depending on Doppler effect, etc. In some cases, we have found clear indication of systematic errors well above the random uncertainties. However, as these errors seem to be negligible with good instrument properly operated, they are not taken into account here.

Combining the different uncertainties, on the daily means of time comparisons, with good instrumentation, the total uncertainty is thus

$$\sigma = 1,5 \text{ ns}.$$

Conclusions

1. Over distances up to about 1000 km, the station coordinate uncertainty may be the main source of uncertainty in the GPS time comparisons. The TRANSIT doppler positioning (± 1 m) is not sufficient. The GPS positioning using the time receiver positioning softwares (± 5 m) is worse.

2. It is possible to reduce the errors due to the positioning to about 1 ns by analysis of the time comparison data, without extra-tracking and extra-data transmission. This analysis requires that changes of location of antennae be accurately measured (± 5 cm), and be reported to cooperating laboratories, together with the changes of adopted coordinates.

3. Corrections due to improved adopted coordinates can be easily performed by software, when computing the time comparisons (formulae available from BIPM). We therefore recommend to avoid frequent changes of adopted coordinates. After check of our results (stability, comparison with other

methods of positioning), BIPM could recommend a global coordinate improvement at some agreed date.

4. Accurate local geodetic ties should be measured (± 5 cm) between GPS antennae and sites of space geodesy (doppler, laser ranging) and VLBI positioning.

5. It is important to shield the GPS antennae from multipath propagation, and also to investigate other causes of biases of the GPS time receivers.

Acknowledgements

We thank the colleagues of the laboratories mentioned in the text for their cooperation and especially for their help in providing the necessary data on their coordinates.

References

Allan, D.W. and Weiss, M., 1980, "Accurate Time and Frequency during Common Views of a GPS Satellite", Proc. 34th Annual Symposium on Frequency Control, Fort Mounmouth, NY, p. 334-346.

Lewandowski, W., Weiss, M.A. and Davis, D., 1986, "A calibration of GPS equipment at time and frequency standard laboratories in the USA and Europe", Proc. of the 18th PTTI meeting p. 265-279, [also published in Metrologia, 24, p. 181-186 (1987)].

Milliken, R.J. and Zoller, C.J., 1980, "Principles of operation of NAVSTAR and System Characteristics", in "Global Positioning System", Papers published in Navigation, Institute of Navigation, p. 3-14.

Nard, G., Rabian, J., Gounon R., 1987, "Utilisation des signaux du GPS en mode differentiel instantané pour des applications temps-fréquence de haute précision", Actes, 1er Forum Européen Temps-Fréquence, p. 237-243.

Annex - Acronyms used in the text

IEN	Istituto Elettrotecnico Nazionale Galileo Ferraris, Torino, Italia.
INPL	National Physical Laboratory, Jerusalem, Israel.
NPL	National Physical Laboratory, Teddington, United Kingdom.
OMSF	Real Instituto y Observatorio de la Armada, San Fernando, Cadiz, Espana.
OP	Observatoire de Paris (Laboratoire Primaire du Temps et des Fréquences), Paris, France.
ORB	Observatoire Royal de Belgique, Bruxelles, Belgique.
PTB	Physikalisch-Technische Bundesanstalt, Braunschweig, Bundesrepublik Deutschland.
STA	Swedish Telecommunications Administration, Stockholm, Sweden.
TUG	Technische Universität, Graz, Österreich.
USNO	US Naval Observatory, Washington D.C., USA.
VSL	Van Swinden Laboratorium, Den Haag, Nederland.

Table I - Distances between the laboratories considered in this study, in thousands of kilometres

	IEN	INPL	NPL	OMSF	OP	ORB	PTB	STA	TUG
INPL	2,8								
NPL	0,9	3,6							
OMSF	1,4	3,7	1,6						
OP	0,6	3,3	0,4	1,5					
ORB	0,7	3,3	0,3	1,7	0,3				
PTB	0,9	3,0	0,7	2,1	0,7	0,5			
STA	1,7	3,3	1,4	3,0	1,5	1,3	0,9		
TUG	0,7	2,4	1,2	2,1	1,0	0,9	0,7	1,4	
VSL	0,8	3,3	0,3	1,9	0,4	0,1	0,4	1,2	1,0

Table II - Corrections to the adopted coordinates, in a frame of reference defined by OP, IEN, VSL. The standard deviations are SX, SY, SZ (see the text). Unit : 1 metre.

Lab	ΔX	SX	ΔY	SY	ΔZ	SZ
IEN	0,29	0,13	- 0,11	0,09	- 1,14	0,29
INPL	- 5,28	0,44	- 0,17	0,28	- 8,97	0,82
NPL	2,85	0,32	- 3,50	0,23	- 4,66	0,90
OMSF	- 0,74	0,32	3,24	0,18	- 6,26	0,65
OP	0,31	0,14	1,22	0,09	0,05	0,32
ORB	2,26	0,32	- 1,96	0,20	1,35	0,70
PTB	- 0,18	0,13	- 2,95	0,09	- 0,34	0,31
STA	-12,02	0,50	1,88	0,27	- 6,65	0,64
TUG	- 0,39	0,24	- 0,04	0,19	- 3,16	1,16
VSL	- 0,61	0,18	- 1,10	0,13	1,08	0,45

2 ND EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 16 - 17 - 18 MARCH 1988

Session TF-02 Number 2

G.P.S. HIGH PRECISION TIME TRANSFER METHODS AND RELATED EQUIPMENT

René GOUNON
SERCEL
B.P. 64 - 44471 CARQUEFOU CEDEX - FRANCE

Jean-Noël LELONG
OSAFRANCE S.A.
36, Rue de Silly - 92100 BOULOGNE - FRANCE

1 - INTRODUCTION

The GPS system has now been acknowledged for several years as the most accurate means of measuring time deviations between remote clocks ; an accuracy of about 10 ns can fairly easily be achieved.

Furthermore, the quality of transmitted signals as well as the high degree of space correlation of most errors relating to the system suggest that it is still possible to improve the system's performance, in order to achieve an accuracy of 1 to 3 ns owing to relative time data.

Such an accuracy, however, cannot be contemplated unless an equipment with adequate structure and high technical capabilities is used, together with strict methods for performing the measurements and working out the results.

This paper briefly deals with those subjects which seem most important, on the basis of results from a record campaign conducted in late 1987, initiated by the Bureau International des Poids et Mesures, Section du temps (BIPM/ST), with the assistance of the departments of the Laboratoire Primaire du Temps et des Fréquences from Paris Observatory and the Time & Frequency department from the Centre National d'Etudes Spatiales space research centre in Toulouse, France. We would like to thank the heads of those laboratories, who allowed us to achieve those investigations, as well as all the persons who contributed to that experiment.

2 - THE RECORD CAMPAIGN

The record campaign was conducted over about ten days in october, 1987. A SERCEL NRT1 single frequency, multi-channel receiver was installed in the LPTF laboratory in PARIS, another like receiver was installed in TOULOUSE. The two receivers were therefore 600 km apart. (See FIGURE 1).

The GPS constellation available at that time is shown in FIGURE 2.

Starting form that situation, we devised an acquisition scheme for the signals from the satellites (FIGURE 3).
A major characteristic in that scheme can already be pointed out : the receiver in use can simultaneously handle the signals from several satellites via four independent channels. We therefore could :

- assign channel No. 1 to acquisition according to the standard scheme defined by the B.I.P.M., for its time comparisons between laboratories in the European area.

- assign the other channels — when the constellation pattern allowed so — to observing 1, 2 or 3 more satellites simultaneously.

- Check the reliability of the measurements, twice a day, by temporarily assigning all four channels to the signal from a single satellite (periodic check on the real signal of interchannel calibration).

Note that the SERCEL NRT1 receiver is characterized by an absolute calibration of about 1 nanosecond, and a relative stability, between independent channels, of typically 0.2 ns.

Of course, the record scheme is the same on both sites (common-view acquisition) and it was replicated in the same way every day over the trial period.

3 - EVALUATION PRINCIPLES

In order to get the closest evaluation of the relative importance of the various factors liable to affect the quality of results, we heeded the following procedure :

- Each laboratory records the relative time deviation between its own clock and the GPS time, according to the scheme described above. Each record session produces time data resulting from separate measurements on the signals from several satellites. An example of such results is shown in FIGURE 4.

- Those data are then grouped with a view to calculate the deviations between the LPTF time and the CNES time, for each satellite separately. An example of that differential calculation, consistent with the raw measurements shown in FIGURE 4, is shown in FIGURE 5.

- Then the dispersion of those results is evaluated by subtracting those differential time measurements — simultaneously performed from several satellites — two by two.

- The quality criterion adopted for that evaluation is an analysis of the distribution of the deviations collected throughout the trial period, meaning that close to 200 samples are available.

4 - RESULTS

4.1. Antenna positions

Because we are dealing with time transfer between two distant sites, the relative positions of antennas have to be known accurately, in the same WGS 84 geodetic system as that used by the GPS.

In the experiment described in this paper, we had an opportunity to install the antennas close to reference geodetic points with coordinates, in the WGS 84 system, known with an accuracy better than 1 metre for each site (coordinates provided by the National Geographical Institute).

The results from the processing of the time measurements performed according to the method described above are shown in FIGURE 6. The standard deviation of the discrepancies between simultaneous, differential time measurements in PARIS and TOULOUSE, two by two, is 3.5 ns.

The NRT1 receiver, owing to its multichannel structure, is perfectly suited for the radiopositioning function, so we made an attempt at sharpening the relative positioning by performing a geodetic measurement before processing time measurements.

Only a few hours were necessary for us to determine the relative position with an uncertainty reckoned below 0.4 m on each dimension. The uncertainty was reckoned by analyzing the coherence of the results from 9 sessions of about 20 to 45 min, involving different constellation patterns.

The data resulting from processing the same time measurements as those shown in FIGURE 6, but, this time, using the position data directly computed by the receivers, are shown in FIGURE 7. We can notice that the standard deviation is still further reduced (from 3.5 to 2.8 ns).

It can be pointed out that the benefit would have been still more substantial if we had not been able to take advantage of the vicinity of reference geodetic points with as good an accuracy as those provided by the National Geographical Institute. As a matter of fact, it is often possible to get accurate reference points, but the problem is they are not expressed in the WGS 84 system. Then tricky geodetic conversion is required, often leading to an accuracy no better than one to a few metres. Never forget that a 0.3 metre error in the direction of a satellite involves a 1 nanosecond error on the time measured from this satellite.

4.2. Need for protecting the antenna against multiple paths

The most limiting factors for accurate time transfers are those factors completely decorrelated at both sites. The most disturbing factor is due to unwanted waves reflected by obstacles close to the antenna.

We had the opportunity to emphasize this phenomenon at a second site in TOULOUSE where the installed antenna was purposely unprotected against wave reflections.

The results are provided in FIGURE 8 : the standard deviation on the distribution rises up to 3.2 ns. More especially, the distribution pattern deteriorates and the 95 % value changes from 5.7 to over 10 ns.

4.3. Processing mode

The most commonly used processing mode is the following : during 13 minutes, a tracked satellite provides 52 individual time measurements (one measurement every 15 seconds). At the end of this observation period, we perform a linear regression over the 52 measurements. From this, we extract :

- the measurement of the time at the beginning of the 13-min period
- the regression slope, providing the local clock drift with respect to GPS time.

This procedure would fit perfectly if the signals from GPS were also perfect, that is, if they were not suffering from effects due to multiple paths. Unfortunately, and although using the most protected antennas, residual effects remain and the individual time measurements are as shown in FIGURE 9. This curve shows that applying a linear regression, and more particularly choosing the time measurement corresponding to the beginning of the observation period, reveals a noise, other than that produced by the clock drift. This noise depicts disturbances proceeding from multipath residuals.

To eliminate this disturbance, and taking into account the high-stability clocks used by the great laboratories, we thought that it could be interesting to replace that traditional calculation just by the calculation of an average and operating the time measurement at half-time in the observation period.

The drawback of this method, however, is that it cannot provide an estimation of the drift. In fact, as shown on FIGURE 9, the regression slope is more representative of the effects of the multiple paths than of the local clock drift. As the NRT1 receiver can acquire the phase of the GPS signals, we preferred to take advantage of this feature and to calculate the drift by simply observing the phase, which as everybody knows, is less sensitive to multiple paths and receiver noise than the code measurement.

In order to appraise the advantages of this method, we indicate (FIGURE 10) the results provided by the traditional processing, in which the linear regression is dated at the beginning of the observation period. Those results should be compared with those obtained by the average calculation, FIGURE 7. The standard deviation decreases from 4.8 to 2.8 ns, the 95 % value from 12 5.7 ns, that is a gain of about 2.

4.4. Increasing the observation periods

Multiple paths result in disturbances that may last beyond 13 min when the reflection point is close to the antenna. For this reason, we thought it interesting to analyze the possible gain an observation period longer than 13 min could afford.

To do so and when we had the opportunity, we calculated for each satellite the average of the times provided by two successive 13-min periods prior to comparing the results between the satellites. In this way, we obtained the same results as a 26-min observation period would have

provided. Unfortunately, only a few of these periods were possible, the SV acquisition program being not initially intended for this purpose.

Nevertheless, we nearly got 50 measurements of that type. Their distribution is illustrated in FIGURE 11. Strictly speaking, this result should not be compared to that in FIGURE 7. — It derives from all the measurements over 13 minutes — but to that proceeding from the twofold observation periods (FIGURE 12). The similarity between the statistic result provided by the incomplete data (FIGURE 12) and that provided by the complete data (FIGURE 7) proves that making a selection among the 13-min periods does not significantly modify the final result.

Thanks to this method, the standard deviation on the distribution drops below 2 ns and more especially, the value at 95 % changes from 5.3 ns to 3.2 ns with no value exceeding 4 ns. This proves that inter-satellite time measurements are coherent on condition that they are properly processed.

5 - PROPAGATION

When dealing with GPS precision, one frequently mentions the errors due to ionospheric and tropospheric propagation.

The former are most correlated at both sites — even far distant sites. In our survey, the calculations we did using the ionospheric parameters transmitted from the satellites revealed a deviation less than 1 ns between PARIS and TOULOUSE for a satellite with an elevation of about 30 degrees. In fact, this deviation is primarily due to the difference — a few degrees — between the elevations seen from the two sites at a given time.

At first view, the tropospheric corrections should be correlated to a lower extent, since they depend on the local meteorological conditions. Yet they do not vary much with the meteorological parameters. (See FIGURE 13). Here again, the difference of about 2 ns observed between PARIS and TOULOUSE for the same satellite — elevation about 30 degrees — is chiefly due to the fact that the elevations seen from both sites are slightly different.

In both cases, we are convinced that using a good model of correction versus elevation is most important.

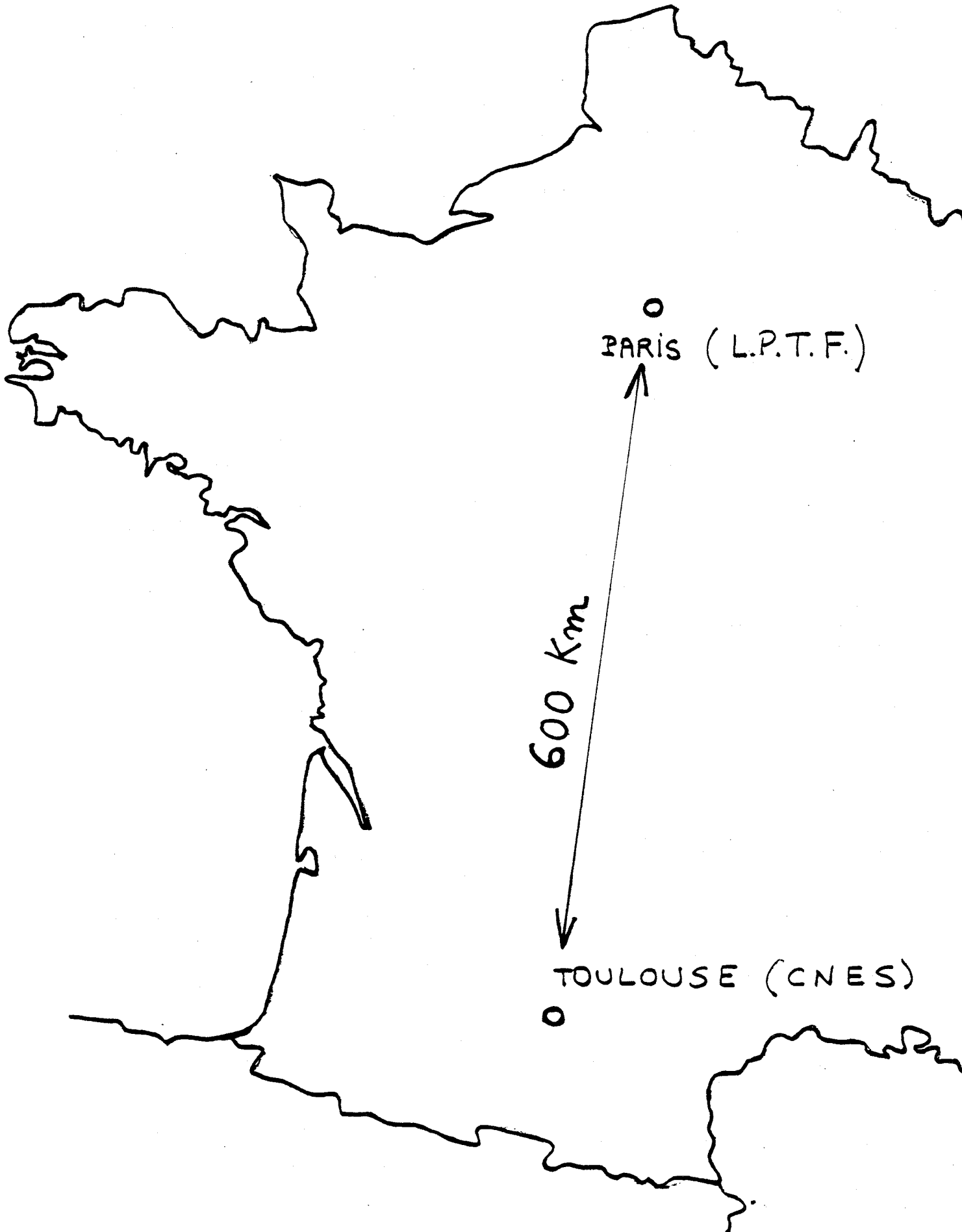
6 - CONCLUSIONS

Through our survey, we proved that a precision better than 2 ns is possible in time transfer applications (common view, 600 Km base). However, such a precision can be reached only if both an appropriate equiment and strict processing methods are implemented.

The main cause of error is linked with multiple path propagation and is therefore independent of the distance between the sites. The effect of multiple path propagation can be nearly eliminated using both a well-protected antenna and appropriate processing methods. This experiment allowed a significant improvement of these techniques.

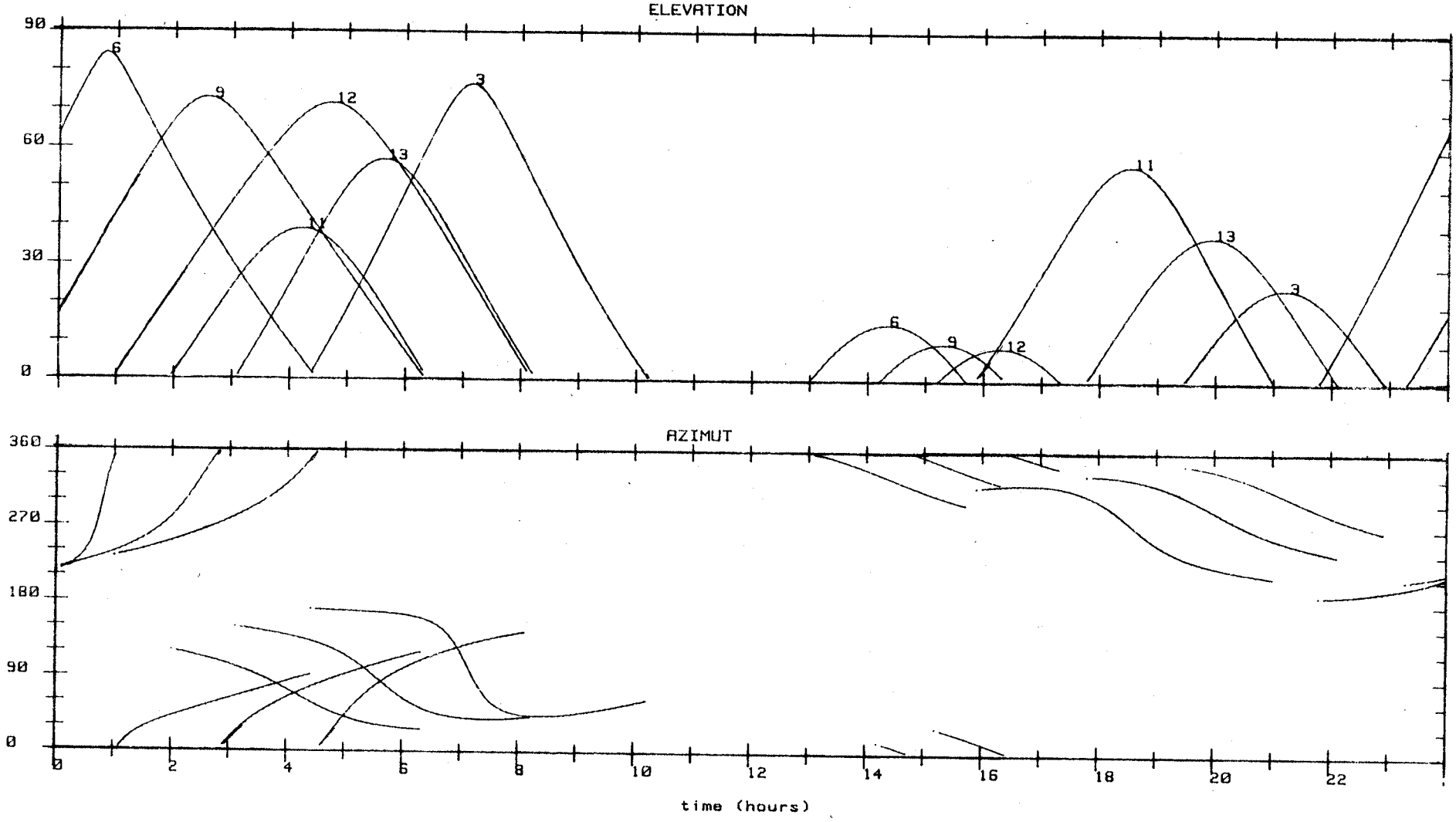
In that domain of application, and considering the system's very high capability for space-correlation of the ionospheric errors, the dual-frequency technique is not necessary to reach the best precision. Nevertheless other sophisticated methods for measuring ionospheric delays are possible and will be investigated.

On the contrary, the multi-channel structure for the receivers is necessary : only this type of equipment can provide, first, a quick and most accurate means to determine the relative positions of the antennas, and second, to guarantee the quality of the results by offering an efficient means to appraise the consistency of the measurements.



1

Week 405 oct 17,1987 at 48°50′ 9.543"N 2°20′ 4.828"E 140.95m (PARIS OBSE)
Week 414 Almanachs; Ev min: 0



2

GPS TIME

Time : 09:45:30
Date : 03-02-88
MJD:7222 Day:3

MENU

Schedule select
Auto Acq-Process
Multi C. Acquis
Multi C. Process
Schedule set-up
Path predictions
NRT.1 Tool-Box

SELECT

SERCEL
TEST
ESSAI
CNESBR
CNESAG
* LPTF
LABO-RN
New Sche

SYSTEM CONDITIONS

NRT 3 - Cal.:402+150- 0
4732 Ko free - Acq.file 150

SELECTED SCHEDULE

48°50′ 9.543"N 2°20′ 4.828"E 140.95 m
X: 4202784.868m Y: 171348.898m Z: 4778678.72
Date: oct 17,1987 MJD:7085 Smpl: 15.0 s

n	CV	Time	C1	C2	C3	C4	n	CV	Time	C1	C2	C3	C
1	lp	0h38	6	9			9	lp	18h53	13	11		
2	lp	2h 6	9	6	12		0	lp	19h42	13	11		
3	lp	2h41	6	9	12		1	lp	19h58	3	11		
4	lp	2h57	9	6	12		2	lp	20h55	3	13		
5	lp	3h21	6	9	12								
6	lp	3h37	12	9	11								
7	lp	3h54	11	12	9	13							
8	lp	4h18	9	12	11	13							
9	lp	4h34	11	13	12	9							
0	lp	4h58	12	13									
1	lp	5h19	13	12									
2	lp	5h39	3	13	12								
3	lp	6h17	12	13	3								
4	lp	7h 4	13	12	3								
5	lp	8h14	3	3	3	3							
6	lp	8h44	3										
7	lp	16h41	11										
8	lp	18h37	11	11	11	11							

4 CHANNELS

3

UTC-OP (LPTF) - GPS

JRT1 : 11

SV #	MJD	TIME HH MM SS	LNTH SEC	UA RE	EL DG	AZM DEG	IONO .1NS	TROP .1NS	IO DE	REF-GPS Ø.1NS	SLOPE *E-13	RMS .1NS
.1	7Ø84	16:47:15	78Ø	8	22	322	1Ø3	211	7	-56651	13	61
.1	7Ø84	18:43:15	78Ø	8	56	267	58	98	11	-56694	-1	4Ø
.3	7Ø84	18:59:15	78Ø	12	27	323	93	176	13	-566Ø3	8	73
.1	7Ø84	18:59:15	78Ø	8	53	254	61	1Ø2	13	-56711	-1	36
.3	7Ø84	19:48:15	78Ø	8	38	3Ø0	76	132	6	-56699	39	46
.1	7Ø84	19:48:15	78Ø	8	34	228	82	145	15	-56783	-7	91
3	7Ø84	2Ø:Ø4:15	78Ø	8	12	34Ø	13Ø	382	1Ø	-56876	-2	53
.1	7Ø84	2Ø:Ø4:15	78Ø	8	26	224	95	181	15	-56736	26	92
3	7Ø84	21:Ø1:15	78Ø	8	24	318	1ØØ	197	12	-56837	14	94
3	7Ø84	21:Ø1:15	78Ø	8	26	26Ø	96	186	7	-56663	11	73
9	7Ø85	Ø2:Ø8:15	78Ø	8	7Ø	291	52	87	6	-56764	-15	39
6	7Ø85	Ø2:Ø8:15	78Ø	8	51	44	62	1Ø4	12	-56691	-4	49
2	7Ø85	Ø2:Ø8:15	78Ø	8	27	254	93	176	15	-56592	2	89
6	7Ø85	Ø2:43:15	78Ø	8	37	57	78	136	12	-56695	13	92
9	7Ø85	Ø2:43:15	78Ø	8	72	35Ø	52	85	6	-56777	-9	51
2	7Ø85	Ø2:43:15	78Ø	8	4Ø	269	73	126	4	-56673	-2Ø	53
9	7Ø85	Ø2:59:15	78Ø	8	69	14	52	87	8	-56794	-14	58
6	7Ø85	Ø2:59:15	78Ø	8	3Ø	63	88	159	14	-56655	23	81
2	7Ø85	Ø2:59:15	78Ø	8	46	277	67	113	6	-56673	-8	43
6	7Ø85	Ø3:23:15	78Ø	8	22	71	1Ø5	218	14	-5667Ø	23	68
9	7Ø85	Ø3:23:15	78Ø	8	62	39	55	92	8	-56816	1	59
2	7Ø85	Ø3:23:15	78Ø	8	54	291	6Ø	1ØØ	6	-56697	11	49
2	7Ø85	Ø3:39:15	78Ø	8	59	3Ø2	56	94	6	-5666Ø	2	34
9	7Ø85	Ø3:39:15	78Ø	8	57	52	58	97	8	-56824	17	65
1	7Ø85	Ø3:39:15	78Ø	8	35	84	8Ø	14Ø	7	-566Ø7	1Ø	34
2	7Ø85	Ø3:56:15	78Ø	8	64	317	54	9Ø	7	-56685	-13	53
9	7Ø85	Ø3:56:15	78Ø	8	51	62	62	1Ø5	1Ø	-56844	12	61
9	7Ø85	Ø4:2Ø:15	78Ø	8	42	75	71	121	1Ø	-56815	15	55
2	7Ø85	Ø4:2Ø:15	78Ø	8	7Ø	344	52	87	7	-56658	27	53
3	7Ø85	Ø4:2Ø:15	78Ø	8	34	132	81	144	9	-56759	-5	5Ø
2	7Ø85	Ø4:36:15	78Ø	8	71	8	52	86	7	-56693	Ø	43
9	7Ø85	Ø4:36:15	78Ø	8	36	82	78	137	1Ø	-56774	28	36
2	7Ø85	Ø5:ØØ:15	78Ø	8	7Ø	44	52	86	9	-56676	-3	4Ø
3	7Ø85	Ø5:ØØ:15	78Ø	8	5Ø	113	63	1Ø5	11	-56788	8	4Ø
2	7Ø85	Ø5:21:15	78Ø	8	65	69	54	89	9	-56687	-15	64
2	7Ø85	Ø5:41:15	78Ø	8	59	87	57	95	9	-56652	-4	46
2	7Ø85	Ø6:19:15	78Ø	8	45	11Ø	67	114	11	-56653	-29	33
3	7Ø85	Ø6:19:15	78Ø	8	48	51	65	1Ø9	13	-5677Ø	-19	48
3	7Ø85	Ø7:Ø6:15	78Ø	12	29	39	91	169	15	-56786	-15	8Ø
2	7Ø85	Ø7:Ø6:15	78Ø	8	26	128	95	182	13	-56648	-2	72
3	7Ø85	Ø7:Ø6:15	78Ø	8	76	1Ø3	51	84	9	-56616	-7	45
3	7Ø85	Ø8:16:15	78Ø	8	5Ø	43	63	1Ø7	11	-56737	-29	38
3	7Ø85	Ø8:16:15	78Ø	8	5Ø	43	63	1Ø7	11	-56737	-31	38
3	7Ø85	Ø8:16:15	78Ø	8	5Ø	43	63	1Ø7	11	-56738	-29	38
3	7Ø85	Ø8:16:15	78Ø	8	5Ø	43	63	1Ø7	11	-56737	-29	38
3	7Ø85	Ø8:46:15	765	8	35	44	8Ø	14Ø	11	-56791	-28	73
1	7Ø85	16:43:3Ø	78Ø	8	22	322	1Ø3	211	8	-566ØØ	16	84
1	7Ø85	18:39:3Ø	78Ø	8	56	267	58	98	11	-56569	5	28
3	7Ø85	18:55:3Ø	78Ø	8	27	323	93	175	14	-56663	17	76
1	7Ø85	18:55:3Ø	78Ø	8	53	253	61	1Ø2	13	-56553	7	38
3	7Ø85	19:44:3Ø	78Ø	8	38	3Ø0	76	132	4	-56653	1Ø	46
1	7Ø85	19:44:3Ø	78Ø	8	34	228	82	146	13	-56519	-4	87
1	7Ø85	2Ø:ØØ:3Ø	78Ø	8	26	224	95	182	15	-56499	8	89
3	7Ø85	2Ø:57:3Ø	78Ø	8	24	317	1ØØ	197	12	-56684	8	87
3	7Ø85	2Ø:57:3Ø	78Ø	8	26	26Ø	96	186	7	-56653	3	83
9	7Ø86	Ø2:Ø4:3Ø	78Ø	8	7Ø	291	52	87	7	-56722	-12	44
6	7Ø86	Ø2:Ø4:3Ø	78Ø	8	51	45	62	1Ø4	13	-56789	-1	57
2	7Ø86	Ø2:Ø4:3Ø	78Ø	8	27	254	93	176	15	-56645	-7	8Ø

3 SATELLITES SIMULTANEOUS ACQUISITIO

CALIBRATION CONTROL

4

LPTF - CNES(1) DIFFERENCES

NRT1 : 11

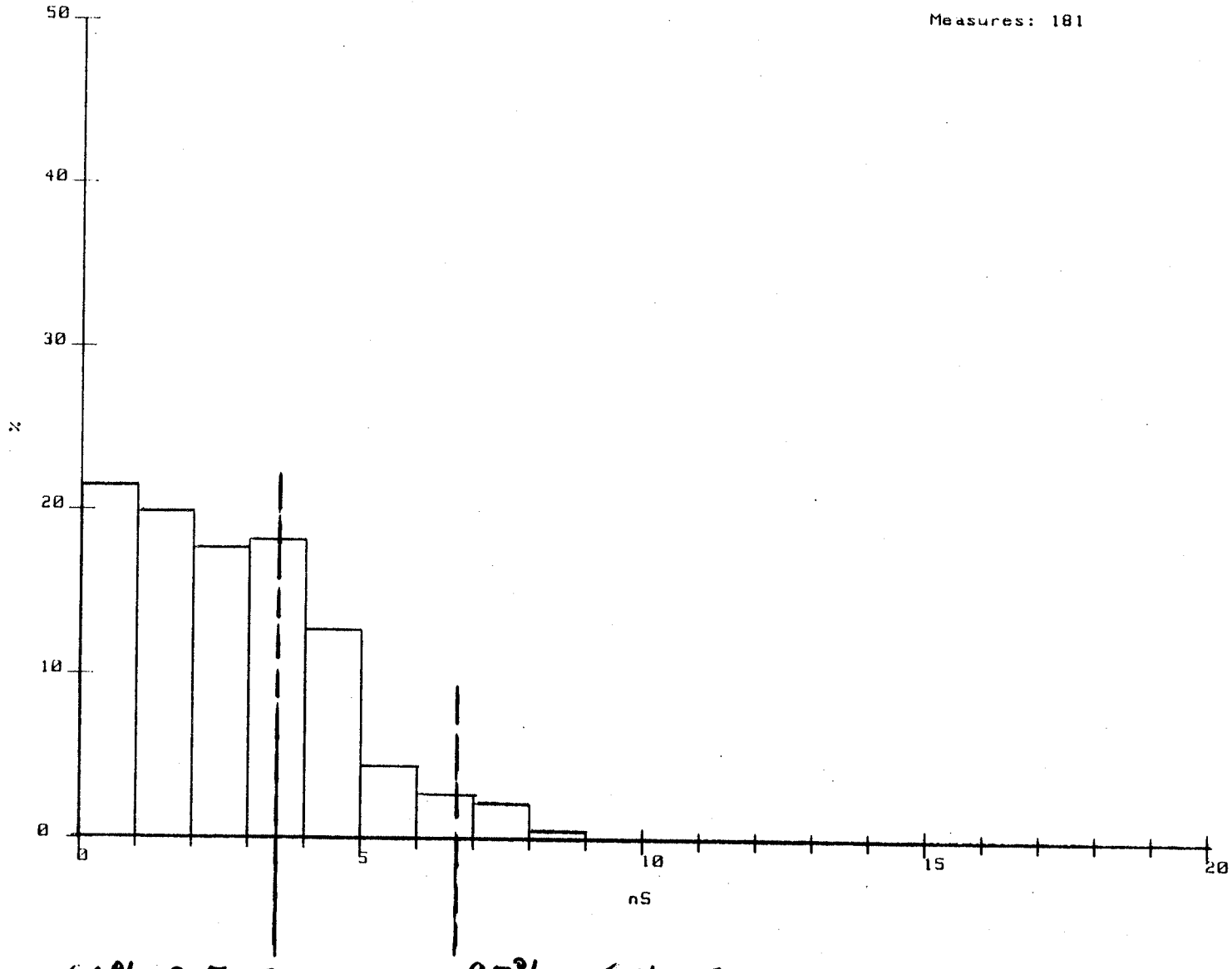
SV #	MJD	TIME HH MM SS	LNTH SEC	AG HR	ELEVATION LPTF	ELEVATION CNES	IONO .1NS	TROP .1NS	LPTF-CNES Ø.1NS	SLOPE PS/S	RMS .1NS
11	7Ø84	18:43:15			56	56			-314813		
11	7Ø84	19:48:15			34	38			-314Ø5		
13	7Ø84	21:Ø1:15			26	27			-314899		
9	7Ø85	Ø2:Ø8:45			7Ø	67			-314942		
6	7Ø85	Ø2:Ø8:45			51	46			-314946		
12	7Ø85	Ø2:Ø8:45			27	3Ø			-314982		
6	7Ø85	Ø2:43:45			37	32			-314993		
6	7Ø85	Ø2:59:15			3Ø	27			-314973		
12	7Ø85	Ø2:59:45			46	46			-314983		
6	7Ø85	Ø3:23:45			22	19			-315Ø54		
12	7Ø85	Ø3:39:15			59	56			-314993		
12	7Ø85	Ø3:56:15			64	6Ø			-315Ø1		
12	7Ø85	Ø4:2Ø:15			7Ø	63			-314999		
13	7Ø85	Ø4:2Ø:45			34	38			-315Ø37		
12	7Ø85	Ø4:36:45			71	65			-315Ø55		
9	7Ø85	Ø4:36:45			36	34			-315Ø33		
12	7Ø85	Ø5:Ø0:45			7Ø	64			-315Ø25		
13	7Ø85	Ø5:Ø0:45			5Ø	52			-315Ø45		
12	7Ø85	Ø5:21:45			65	61			-315Ø79		
12	7Ø85	Ø5:41:45			59	57			-315Ø51		
12	7Ø85	Ø6:19:45			45	46			-315Ø92		
13	7Ø85	Ø6:19:45			48	43			-315Ø89		
3	7Ø85	Ø8:16:15			5Ø	44			-315220		
3	7Ø85	Ø8:16:45			5Ø	44			-315219		
3	7Ø85	Ø8:16:45			5Ø	44			-315223		
3	7Ø85	Ø8:16:45			5Ø	44			-315222		
3	7Ø85	Ø8:46:45			35	3Ø			-315254		
11	7Ø85	18:39:3Ø			56	56			-315455		
13	7Ø85	19:44:3Ø			38	35			-315437		
11	7Ø85	19:44:45			34	38			-315473		
11	7Ø85	2Ø:Ø0:45			26	31			-315442		
6	7Ø86	Ø2:Ø4:3Ø			51	46			-315586		
12	7Ø86	Ø2:Ø4:45			27	3Ø			-315Ø6		
6	7Ø86	Ø2:39:45			37	32			-315653		
6	7Ø86	Ø2:55:3Ø			3Ø	27			-315632		
12	7Ø86	Ø2:55:45			46	46			-315628		
6	7Ø86	Ø3:19:45			22	19			-315663		
12	7Ø86	Ø3:35:3Ø			59	56			-315622		
12	7Ø86	Ø3:52:3Ø			64	6Ø			-315632		
12	7Ø86	Ø4:16:3Ø			7Ø	63			-315648		
13	7Ø86	Ø4:16:45			34	38			-315673		
12	7Ø86	Ø4:56:3Ø			7Ø	64			-315682Ø		
13	7Ø86	Ø4:56:45			5Ø	52			-315682		
12	7Ø86	Ø5:17:45			65	61			-315725		
12	7Ø86	Ø5:37:45			59	57			-315685		
12	7Ø86	Ø6:15:45			45	46			-315733		
13	7Ø86	Ø6:15:45			48	43			-315720		

BETWEEN SATELLITE DIFFERENCES

5

ALL 13mn MEASURES (MEANS)

"IGN" FIXES

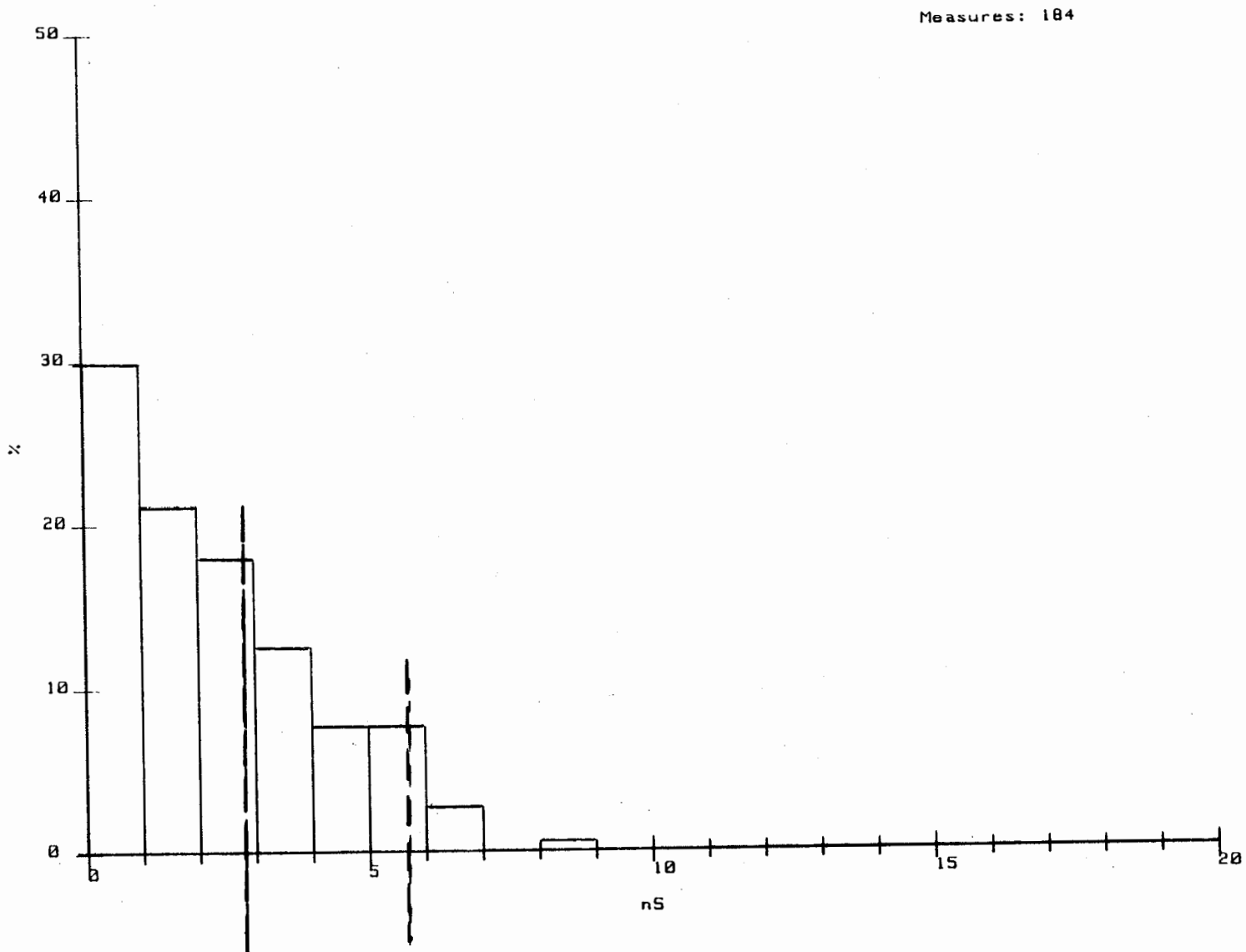


68% = 3.5 nS

95% = 6.7 nS

6

ALL 13mn MEASURES (MEANS)
DIFF. GPS FIXES

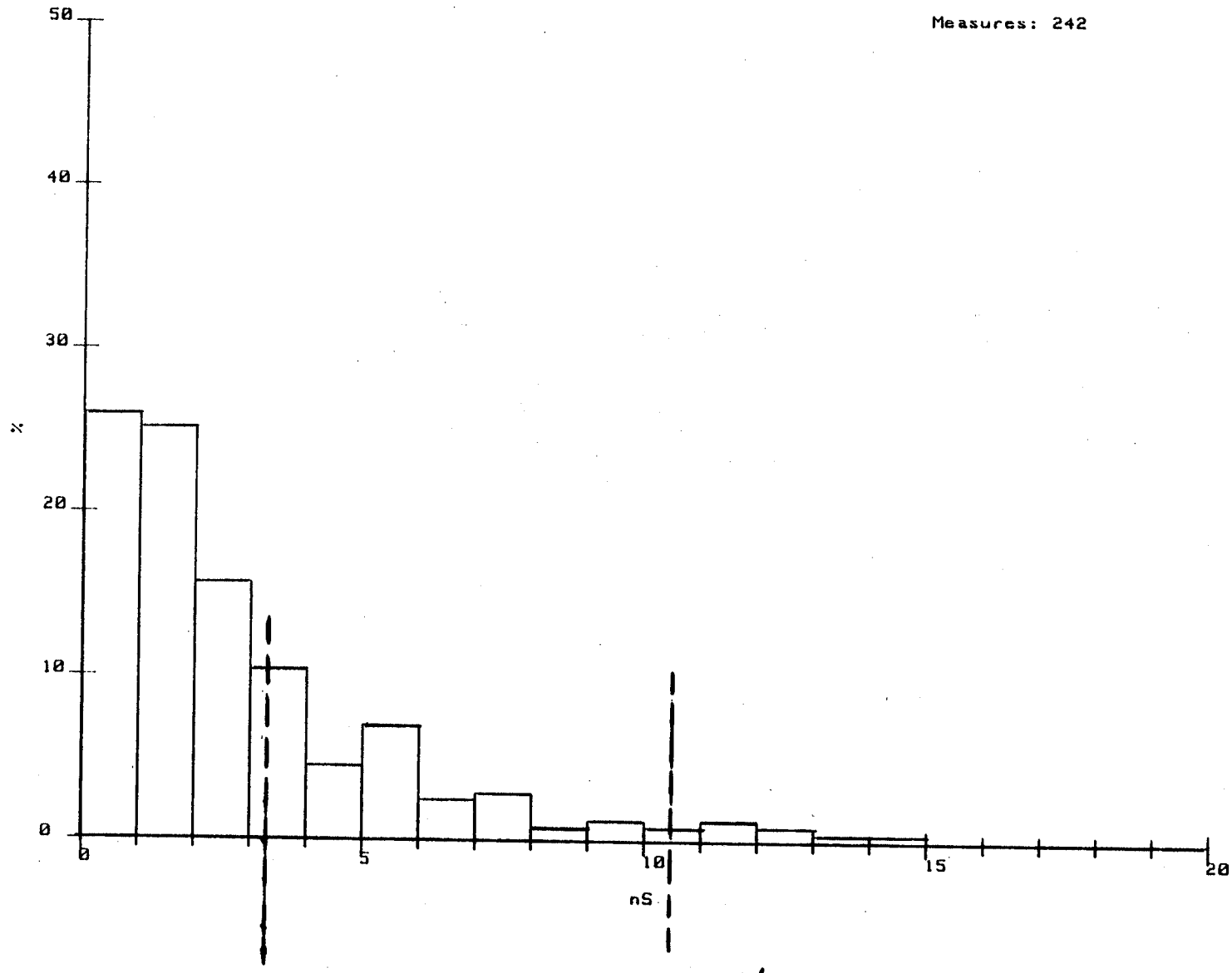


68% : 2.8nS

95% : 5.7nS

ANTENNA WITHOUT PROTECTION

LPTF-CNES All 13mn measures (Antenna without protection on CNES site)

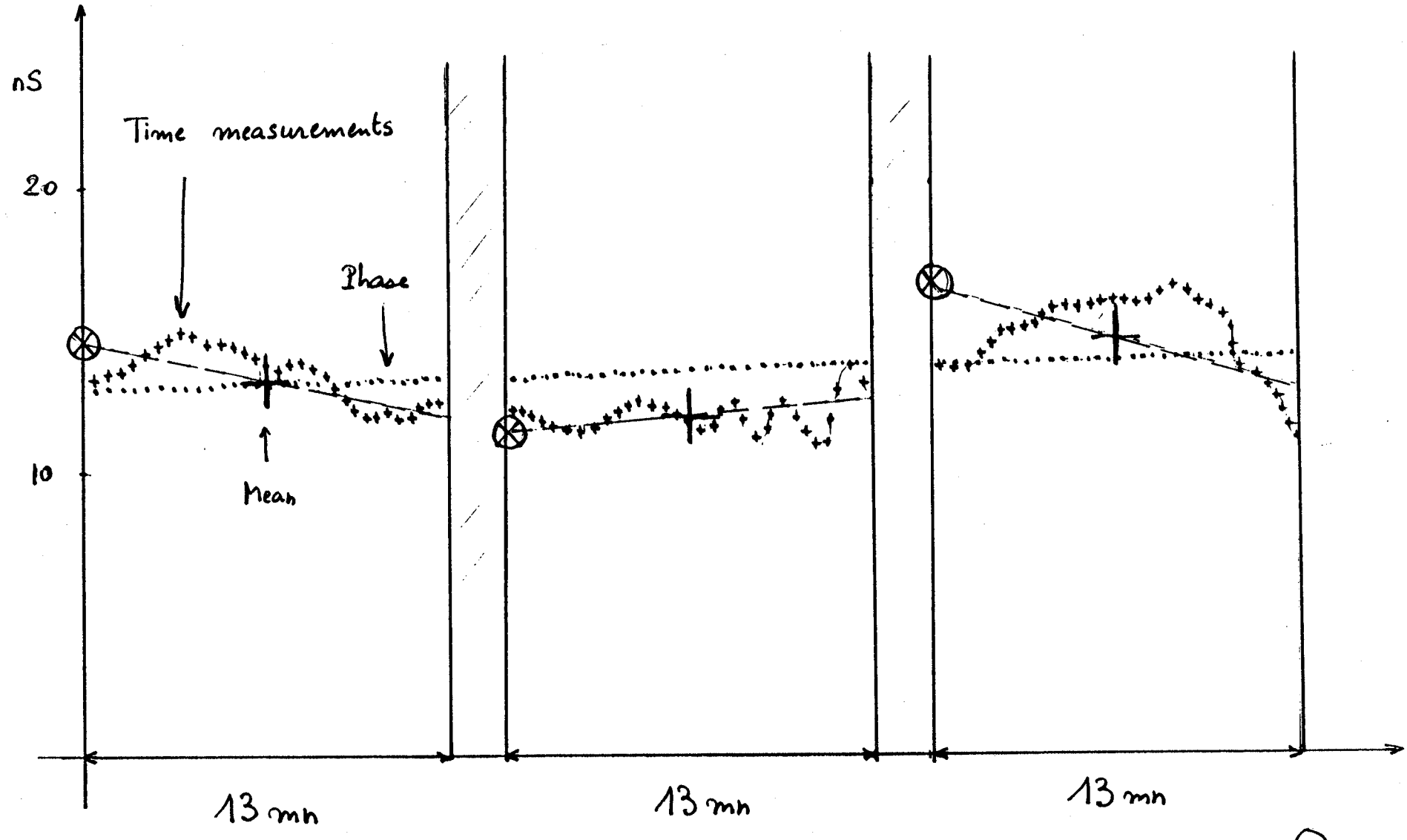


68% : 3.2nS

95% : 10.5nS

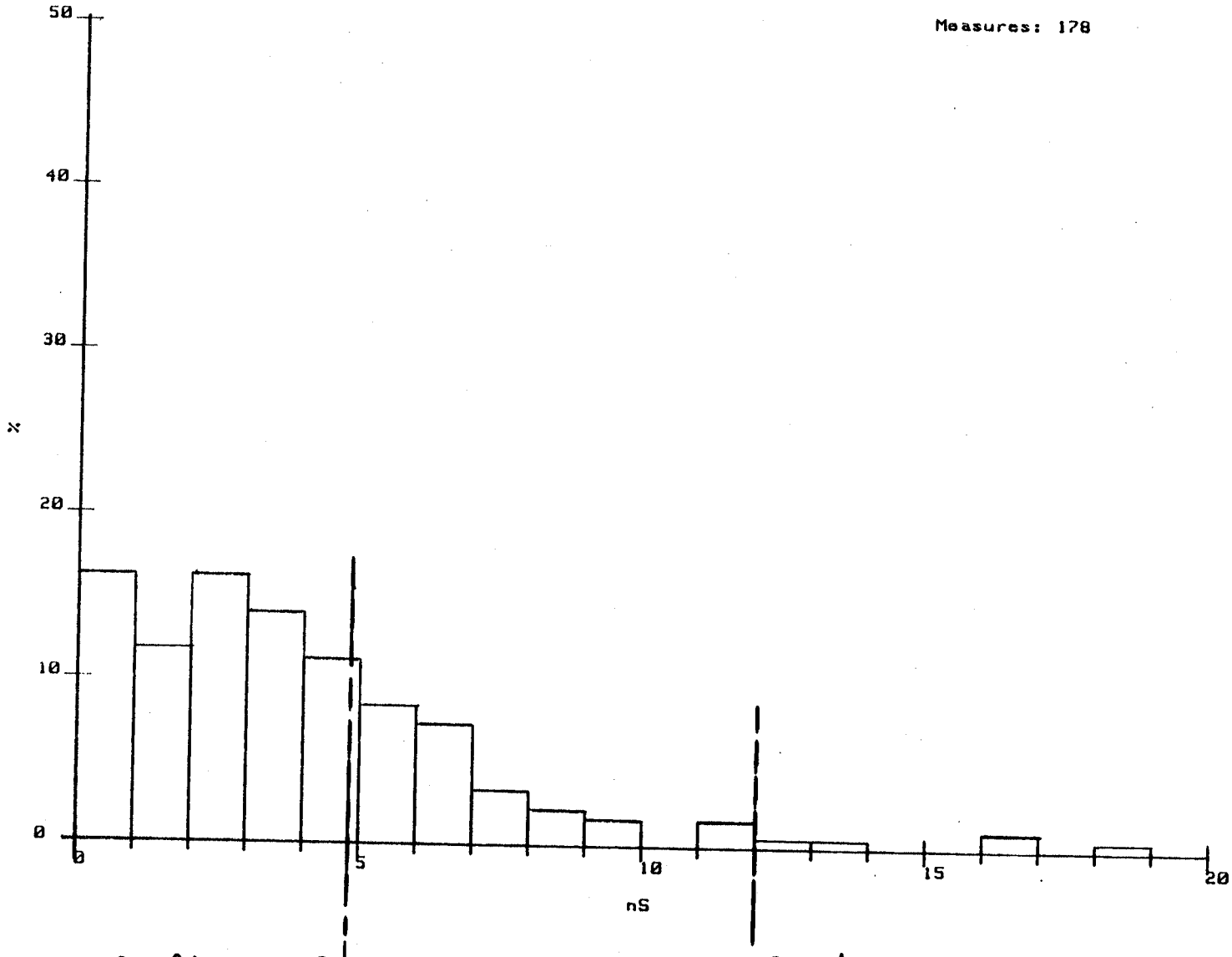
8

TIME AND PHASE MEASUREMENTS



LINEAR REGRESSION

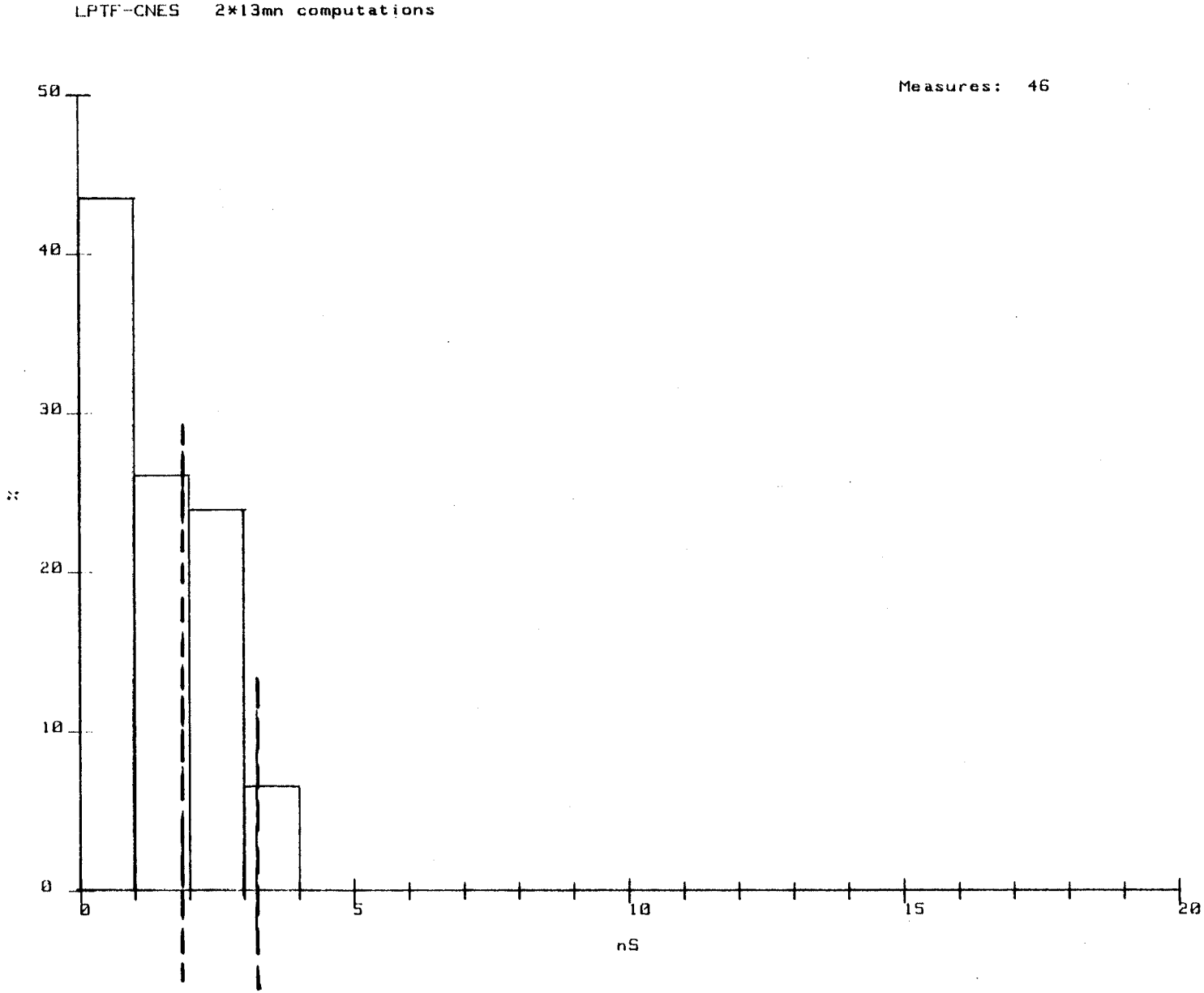
LPTF-CNES All 13mn measures (Linear least squares estimation)



68% : 4.8 nS

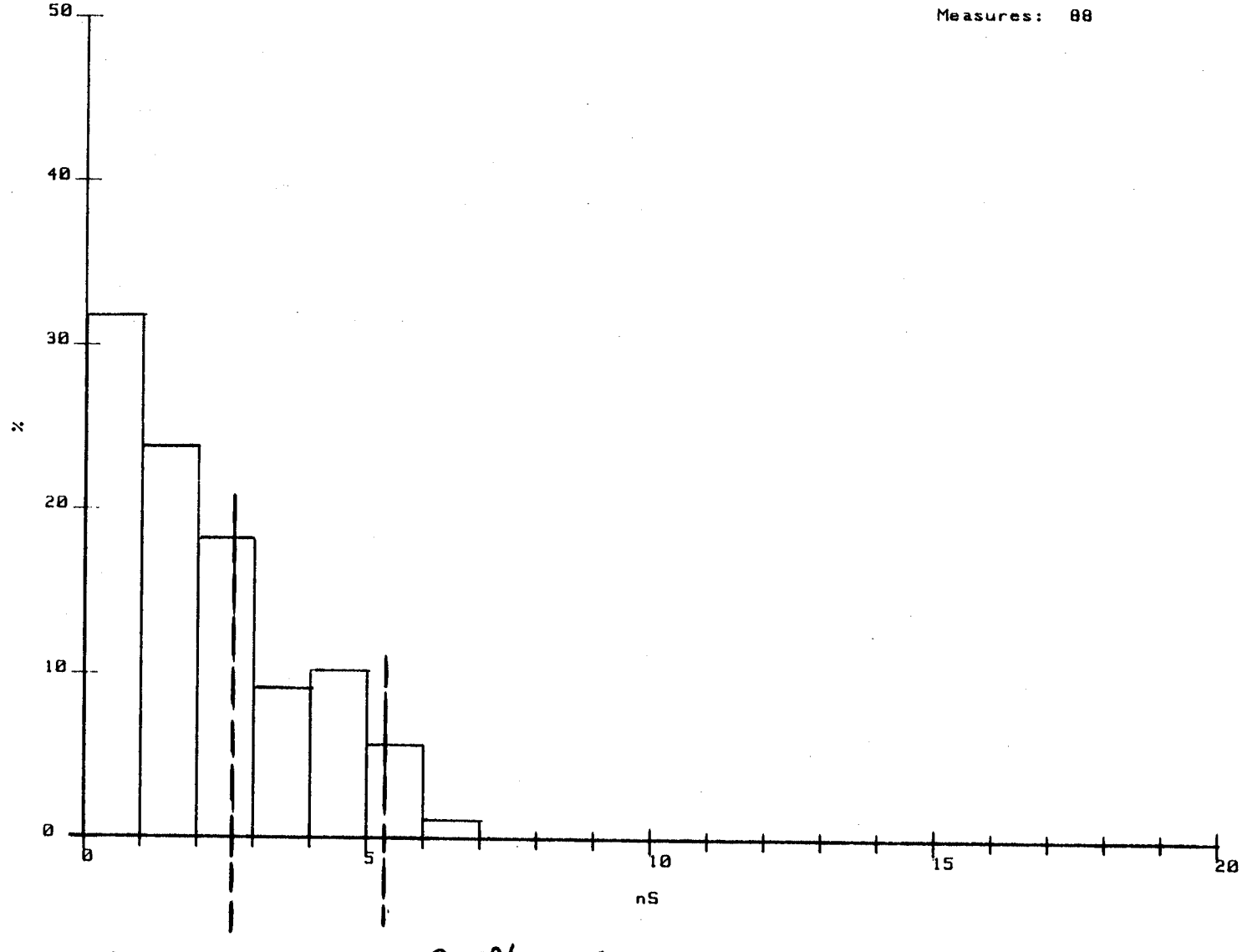
95% : 12 nS

2 x 13 mn COMPUTATIONS



68% : 1.9nS 95% : 3.2nS

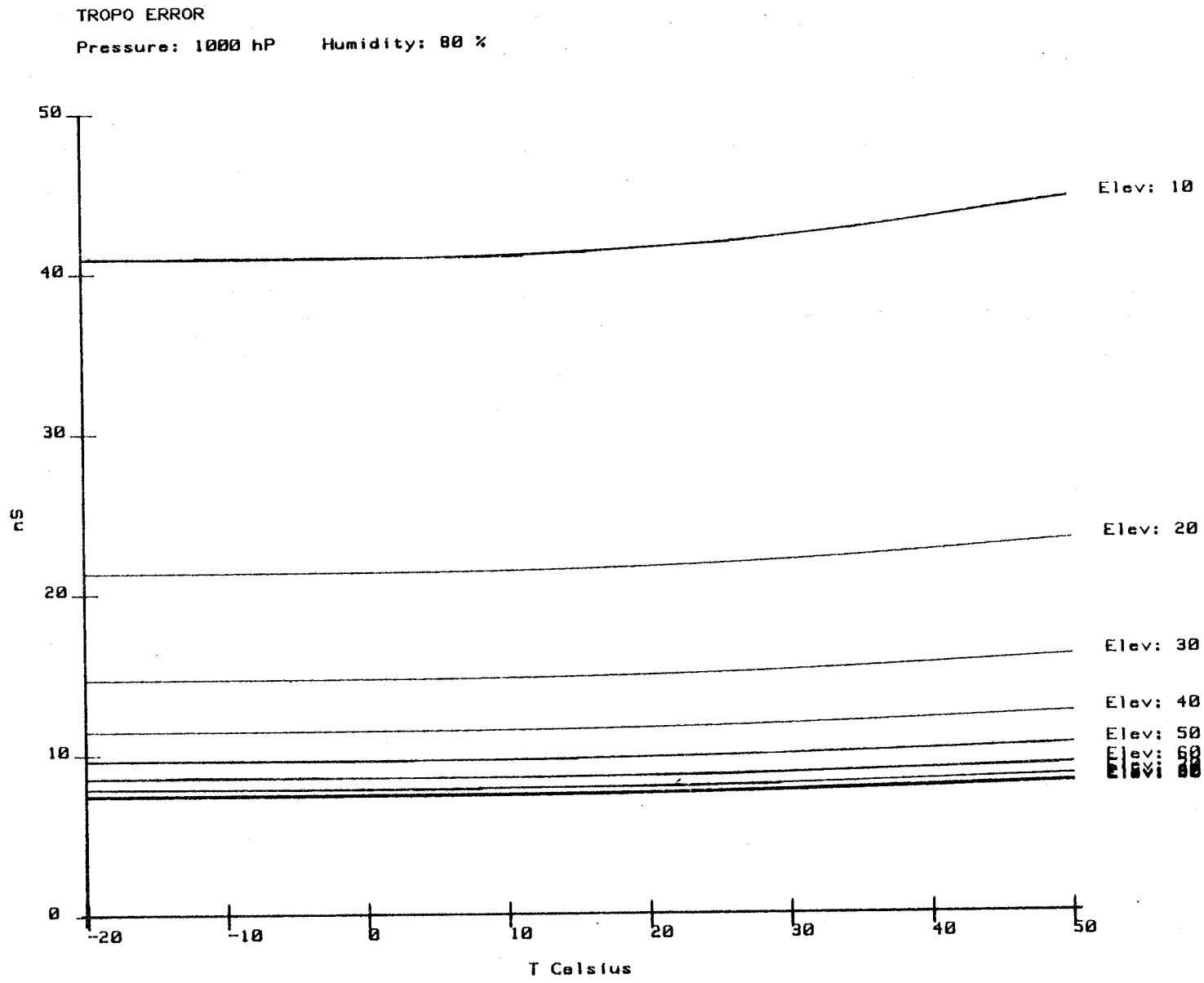
LPTF-CNES 13mn measures used in 2*13mn computations



68% = 2.7nS 95% = 5.3nS

12

TROPO CORRECTION SENSITIVITY



2 ND EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 16 - 17 - 18 MARCH 1988

Session TF-02 Number 3

OPERATIONAL HIGHLY PRECISE TIME TRANSFERS IN FRANCE BY GPS AND BY THE TV METHOD

M. GRANVEAUD, P. PARCELIER, G. FREON, R. TOURDE
Laboratoire Primaire du Temps et des Fréquences,
Observatoire de Paris
M. VINCENT, Observatoire de Besançon
P. GRUDLER, Centre d'Etudes et de Recherches Géodynamiques

Abstract

The french atomic time reference has been computed from cesium clocks data running in various centers. It has been decided to take advantage of the GPS possibilities to link these clocks except for short distances where the TV method is equally performant.

The time/ frequency metrology is based upon the continuous running of atomic clocks and on the fast continuous comparison of these devices between themselves. So the quality of a time reference is strongly dependent on the time precision and time accuracy of the clocks comparisons. In case of local time references where all the participating clocks are located near each other, there is no difficulty to ensure time intercomparisons at the level of or better than 1 nanosecond. As far as international or, in some cases, national time references are concerned, the quality of the clock comparisons is questioned. Another important point is the availability of the clock comparison results. It can be of interest to get them definitive in quasi real time.

In this general context, the french atomic time reference TA(F) is currently obtained from readings of an ensemble of about 20 commercial cesium clocks which run in 9 laboratories. Until 1987, these devices were daily intercompared through common reception of a television signal. In order to improve the general performances of the system, it has been decided to take advantage of the possibilities offered by GPS. It is worth noting that the GPS as well as the TV method and the LORAN are passive time intercomparisons methods ;the time signal (GPS, TV, LORAN) is received by 2 (or more than 2) stations A and B and dated by each one with respect to its own Master Clock. The time synchronization between the Master Clocks of A and B is simply computed as the difference of the two datations, taking into account the necessary time corrections. The best conditions of operating these 3 methods are generally obtained by receiving quasi simultaneously the same transmitted signal (it is named common view in GPS). In fact, it is a necessity of operating in such a way if the time signal is not monitored by a very stable source, like a cesium standard.

This work has support of the Bureau National de Metrologie - France

In the following of this paper, it is reported results obtained through TV and GPS in the framework of the french atomic time scale.

1. Possibilities of the local terrestial TV method

The passive TV method is 20 years old [1] ; it has been largely operated since 1967, especially in Europe [2]. Improvements of the method has been realized by choosing a well defined pulse of the TV frame . In France, it has been chosen the test line LT1 of the 17th line ; its rise time is 100 nanoseconds high and 45 microseconds wide. Results concerning the daily time comparisons of two french laboratories in the Paris area are shown ; these ones are 5 and 8 km away from the transmitting station and the distance between them is 5 km. Precisions of the results as given by the standard deviation of 31 measurements obtained in 30 minutes are of the order of a few nanoseconds (typically 3 ns).

Figure 1 shows the square root of the two sample variance of the daily frequencies computed from the time differences clock (lab. 1) - clock (lab. 2). These results are representative of the stability of the two cesium clocks and of the TV time link. Furthermore the variance relative to each clock has been computed by the three corner hat method. The variance referring to the TV method is deduced by substracting the variances of the two cesium clocks. This time intercomparison method, as used in France, introduces a white frequency noise exhibiting $\sigma_y(2,\tau) = 6\ 10^{-14}\ \tau^{-1/2}$ with τ in days. In the specific case of figure 2, the TV link noise is lower than the cesium clocks one for $\tau \geq 3$ days.

2. Preliminary results through GPS measurements

Since February 1988, cesium clocks at the Paris Observatory and at the Besançon Observatory have been intercompared through GPS common wiew measurements. Both GPS receivers are of the NBS type. Before this compaign, the two receivers were located near each other at the Paris site ; the coordinates of their antenna were different essentially in altitude by 1.5m. This situation is worth of interest with respect to two points of view : first, it allows to appreciate the local GPS conditions, especially the environmental conditions ; second it corresponds to the same situation as a GPS receiver calibrating transportation.

A. Colocation of two GPS receiver

Both receivers measure the time synchronization UTC(OP) - GPS via the international time comparison schedule, i.e., 24 passes per day since mid December 1987. Results of interest are the differences [UTC(OP) - GPS] Rec 1 - [UTC(OP) - GPS] Rec 2 which are denoted D(t). Let us focus on a series of D(t) obtained from 1987, December 22 (MJD = 47151) to 1988, January 27 (MJD = 47187), i.e., 37 days. The D(t) can be characterized from two points of wiew : first, the D(t) belong to 24 different time series, each one corresponding to one pass with one elevation and one azimut of specific satellite. Table 1 shows the average and the standard deviation obtained for each 24 series. After disregarding the passes number 3,13 and 15 which have got a bad local geometry (elevation and/or azimut), the extreme values are the following ones ; for the average : 9.22 and 16.68 ns ; for the standard deviation : 2.35 and 7.28 ns. Figure 2 shows the D(t) relative to the pass 15 and clearly there is no stationarity of the first order. The D(t) corresponding to the above extreme averages are presented by figure 3 ; the D(t) values referring to 9.22 are practically all below these ones showing an average of 16.68, during the 37 day period. Finally, figure 4 shows the D(t) values for the passes 2 and 11 which have got the extreme standard deviation values. These results are an indication of the degradation coming from the Parisian laboratory environment.

Computation of the two sample variance of the 24 D(t) series gives insight into the noise of the two GPS acquisition systems. It is found a general law in τ^{-1} (figure 5) for the square root of the variance with a $\sigma_y(2,\tau)$ value which is generally smaller than 10^{-14} for $\tau \geq 8$ days.

The D(t) values can also be considered on a purely statistical point of view without taking into account the pass involved in the measurement. Such an approach is generally done in case of calibration by GPS receiver transportation as the calibration result is obtained by the simple average of the differential measurements carried out during one or two days.
Table 2 lists the averages and the standard deviations of the D(t) corresponding to one day (generally 24 values).

PASS N°	Sat	Time on MJD = 47151	El (°)	Az (°)	m (ns)	σ (ns)
1	12	0 h 06	71	31	15.0	3.6
2	11	0 h 22	32	41	14.5	2.4
3	03	1 h 10	18	101	12.9	5.1
4	09	0 h 38	57	84	11.6	4.4
5	13	0 h 54	45	162	12.0	3.9
6	12	2 h 04	33	121	13.6	4.8
7	13	2 h 24	29	39	13.7	2.7
8	03	4 h 18	28	46	13.6	3.8
9	11	12 h 16	26	322	13.7	4.8
10	11	13 h 49	56	281	13.7	2.8
11	13	14 h 24	30	320	15.4	7.3
12	13	15 h 22	38	288	14.2	4.1
13	03	16 h 22	25	313	17.1	6.6
14	06	18 h 22	36	199	9.2	4.1
15	09	19 h 19	22	219	5.1	10.7
16	06	20 h 02	84	294	13.0	3.6
17	09	21 h 19	71	294	12.4	3.5
18	06	21 h 39	45	50	14.8	2.8
19	12	21 h 59	42	270	11.3	3.6
20	06	22 h 31	24	68	16.1	3.5
21	11	22 h 47	33	91	16.7	4.6
22	12	23 h 03	64	312	13.2	3.8
23	09	23 h 30	40	74	15.0	3.9
24	13	23 h 46	39	128	15.5	4.8

TABLE 1. Characteristics relative to the 24 series of GPS differential measurements carried out every day from 1987 December 22 to 1988 January 27 (37 days).

DATE	MJD	m (ns)	σ (ns)
1987 déc. 22	47151	15.4	5.4
23	47152	12.0	3.6
24	47153	11.4	4.4
25	47154	14.4	5.9
26	47155	14.5	5.1
27	47156	14.1	4.0
28	47157	13.6	5.3
29	47158	11.5	4.8
30	47159	10.5	4.5
31	47160	14.1	5.1
1988 Jan. 01	47161	12.9	4.6
02	47162	12.0	4.3
03	47163	14.5	4.4
04	47164	10.7	4.6
05	47165	13.0	3.4
06	47166	12.3	5.1
07	47167	13.3	5.1
08	47168	12.8	6.9
09	47169	13.7	5.7
10	47170	13.8	6.0
11	47171	14.3	5.8
12	47172	12.9	5.1
13	47173	14.0	6.2
14	47174	12.3	5.6
15	47175	13.2	5.7
16	47176	12.3	5.1
17	47177	14.0	5.6
18	47178	14.5	4.7
19	47179	13.9	4.4
20	47180	15.6	4.5
21	47181	15.5	3.6
22	47182	15.0	4.6
23	47183	14.7	4.8
24	47184	16.3	4.8
25	47185	13.9	5.6
26	47186	13.7	4.2
27	47187	13.9	5.0

TABLE 2. **Daily Characteristics of GPS differential measurements.** (24 measurements per day)

The averages ranges from 10.5 up to 16.3 ns and the standard deviations from 3.4 to 6.9 ns. Both of them show less dispersion than the above results of table 1. Nevertheless a one day calibration result will differ by 5.8 ns if it is performed on 1987, December 30 (MJD = 47159) or on 1988, January 24 (MJD = 47184). Figure 6 presents the repartition rates relative to the 24 D(t) measurements carried out on 1987, December 30 and on 1988, January 24. Considering the two 24 elements ensembles, it looks like a general translation of the values by 6 ns. A detailed examination of the values show that fifty per cent have changed by more than 6 ns between the two dates.

If we consider the D(t) values obtained during the whole 37 days span period, the average is 13.5 ns and the standard deviation 5.2 ns. Besides, it is found that the D(t) population under consideration satisfies a test of gaussian distribution. It can be questioned if the same characteristics are obtained with measurements performed a few months before. Table 3 presents the results obtained with other ensembles.

D(t) ensemble	Nb of data	m ns	σ ns	gaussian distribut.	remarks
Dec. 1987 Jan. 1988	892 37 days	13.5	5.2	well	New schedule
Jul. 1987 Aug. 1987	591 32 days	13.0	7.7	well	New location antenna receiver2
Apr. 1987 May 1987	625 37 days	9.9	12.7	pretty well	First location antenna receiver2

Table 3 - Characteristics relative to 3 ensembles of differential measurements between two GPS receivers located at Paris Observatory.

Clearly the Dec-Jan ensemble is more satisfactory than the Apr-May one, the sigma value being divided by 2.5. The last location of the antenna of receiver 2 is better than the previous one. A change of the standard deviation is observed from the Jul-Aug ensemble to the Dec-Jan one ; the only intentional change was the updating of the schedule mid December which has introduced 25 % more passes than before. It can be thought that the current schedule is more adapted to the Paris Observatory environment than the previous one.

A selection of the passes has been operated on the three ensembles of measurements with respect to the elevation and the azimut (Table 4). It can be noted an improved value of the standard deviation with the first ensemble (Dec 1987-Jan 1988) after selection of 67 % of the passes.

In conclusion, in order to improve the calibration results of a GPS receiver transportation, it is important to take into account the local environment conditions and to get data over several days (at least 3 days) ; consistency of the results from one day to the other should be tested in order to eliminate the outliers. After selection and cancellation of the non consistent data, the final result would be given by a simple average.

B. Colocation of three NBS type GPS receivers

From 1987, December 30 to 1988, January 10, (i.e. 12 days) a third NBS type GPS receiver was set up at the T/F laboratory of the Paris Observatory. It gave the opportunity of testing the calibration results of the previous BIPM campaign. The antenna was fixed up on a dedicated pole and the coordinates were deduced from these ones of the other antenna. During a 11 days period, the same passes were schedule on the 3 NBS type GPS receivers. Applying the three corner hat method to the measurements carried out through 24 passes, it has been obtained an estimate (resulting of time and ensemble averages) of the noise level specific to each receiver ; for τ = 1 day, it is found σ_y (2,τ = 1 day) = 6 10^{-14} with a one sigma uncertainty of 4 x 10^{-14}. The estimate of σ_y (2,τ = 2 days) is 3 x 10^{-14} in agreement with a τ^{-1} law for the noise.

If we look at the daily time differences, their mean values are in the range 0.1 ; 3.4 ns. with standard deviations of 5.1 up to 7.1 ns. It confirms the necessity of operating a GPS receiver transportation on a duration superior to 1 day.

The ensemble of measurements carried out over 12 days has got a mean value of 1.8 ns with a standard deviation of 6.2 ns. Selection of passes (elevation $\geq$ 30° and azimut out of 250-330° for elevations in the range 30-35°) and cancellation of 2 outliers lead to a mean value of 1.7 ns with an improved standard deviation of 4.6 ns.

D(t) ensemble	Nb of data	m (ns)	σ (ns)	gaussian distribut.	remarks
Dec 1987	892 (37 d)	13.5	5.2	Well	new schedule
Jan 1988	483 el ≥35°	13.2	4.0	Pretty well asymmetry	
	596 el ≥30°	13.6	4.1	Well	el ≥ 35° if 250°≤az≥330°
Jul 1987	591 (32d)	13.0	7.7	Well	new location antenna receiver 2
Aug 1987	303 el ≥35°	14.6	6.9	Well	
	367 el ≥30°	13.3	7.1	Well	el ≥ 35° if 250°≤az≤330°
Apr 1987	625 (37d)	9.9	12.7	Pretty well	first locati. antenna receiver 2
May 1987	475 el ≥35°	12.2	12.4	Well	
	540 el ≥30°	10.7	12.9	Well asymmetry	el ≥35° if 250°≤az≤330°

Table 4. Statistical characteritics relative to the 3 ensembles without and with selection of the passes.

C. GPS time comparisons between Paris and Besançon

The GPS Receiver 2 was set up at the Besançon Observatory on 1988 February 2. Coordinates of the antenna as obtained from the National Geographical Institute were introduced (it is also the case at Paris). Results of 35 days of GPS comon view intercomparisons have been collected. A least square adjustement of daily measurements leads to sigma values in the range 5.9 ns. The square roots of the two sample variances have been computed from the time data UTC (OP) - UTC (Besançon) obtained from specific passes ; figure 7 presents the mean value out of 16 passes. The noise of the Besançon cesium clock is relatively high and so after two days, the GPS noise in τ^{-1} becomes negligible before the clocks one.

Differential navigation solutions have been obtained by comparing the Besançon and Paris coordinates measurements. It has been found :

$$\Delta X = - \ 0.2 \pm 1.7 \text{ m}$$

$$\Delta Y = - \ 10.5 \pm 0.9 \text{ m}$$

$$\Delta Z = - \ 8.4 \pm 1.6 \text{ m}$$

showing a significant incoherence between the GPS coordinates and these ones introduced. Importance of the antenna coordinates has been highlighted by the BIPM team [3]. It is planned to perform specific tests with respect to the coordinates and to the environment in order to improve the Besançon-Paris time link.

Conclusion

Two time intercomparison systems will be used by the T/F laboratory of the Paris Observatory to establish the french atomic time scale. One is the passive terrestrial TV method whose possibilities (precision, frequency noise) are worth of interest in case of laboratories distant by a few kilometers from the transmitter. For long distances (a few hundreds kilometers), GPS time intercomparisons will be carried out. Noise characteristics of these two systems are homogeneous with respect to the noise of the commercial cesium standards : after 2-4 days, the time link noise becomes lower than the cesium clocks one.

References

[1] TOLMAN, J., PTACEK, V.,SOUCEK, A., STECHER, R. : Microsecond clock comparison by means of TV synchronizing pulses, Trans IEEE, IM-16, 247 (1967).

[2] PARCELIER, P., FREON, G. : Utilisation de la télévision dans la métrologie temps-fréquence, Comptes Rendus Colloque Int. URSI mesures Télécom 264 (1970).

[3] GUINOT, B., LEWANDOWSKI, W. : Accurate time comparisons and positioning by GPS over distances up to 1000 km, to be published.

Figure captions

Figure 1. Stability of :
(a) two cesium clocks plus TV time link
(b) two cesium clocks
(c) TV time link - Results obtained from 242 time measurements.

Figure 2. Time values D(t) obtained through pass N° 15- see Table 1.

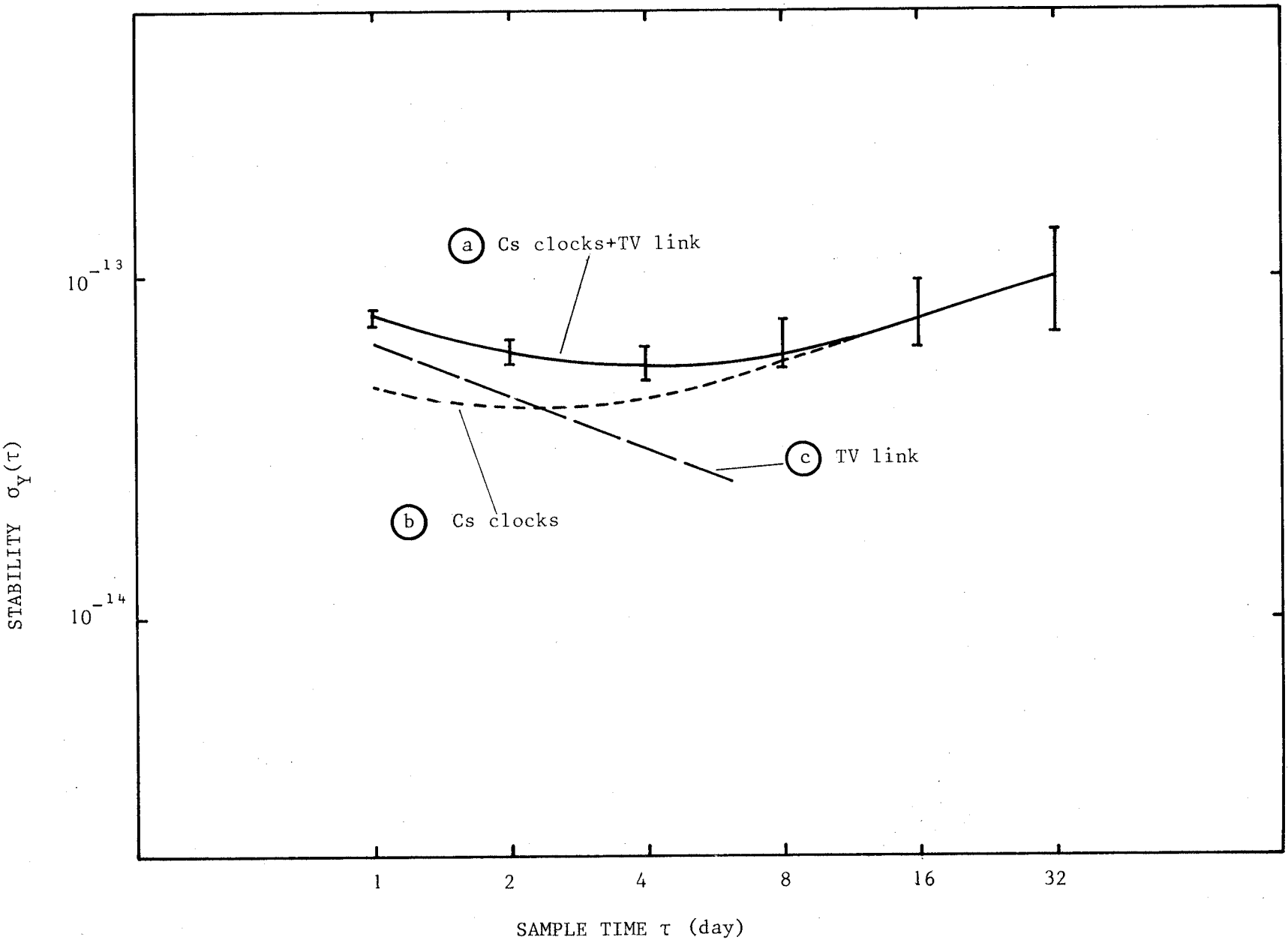
Figure 3. Time values D(t) obtained through passes N° 14 (a) and 21 (b) - see Table 1.

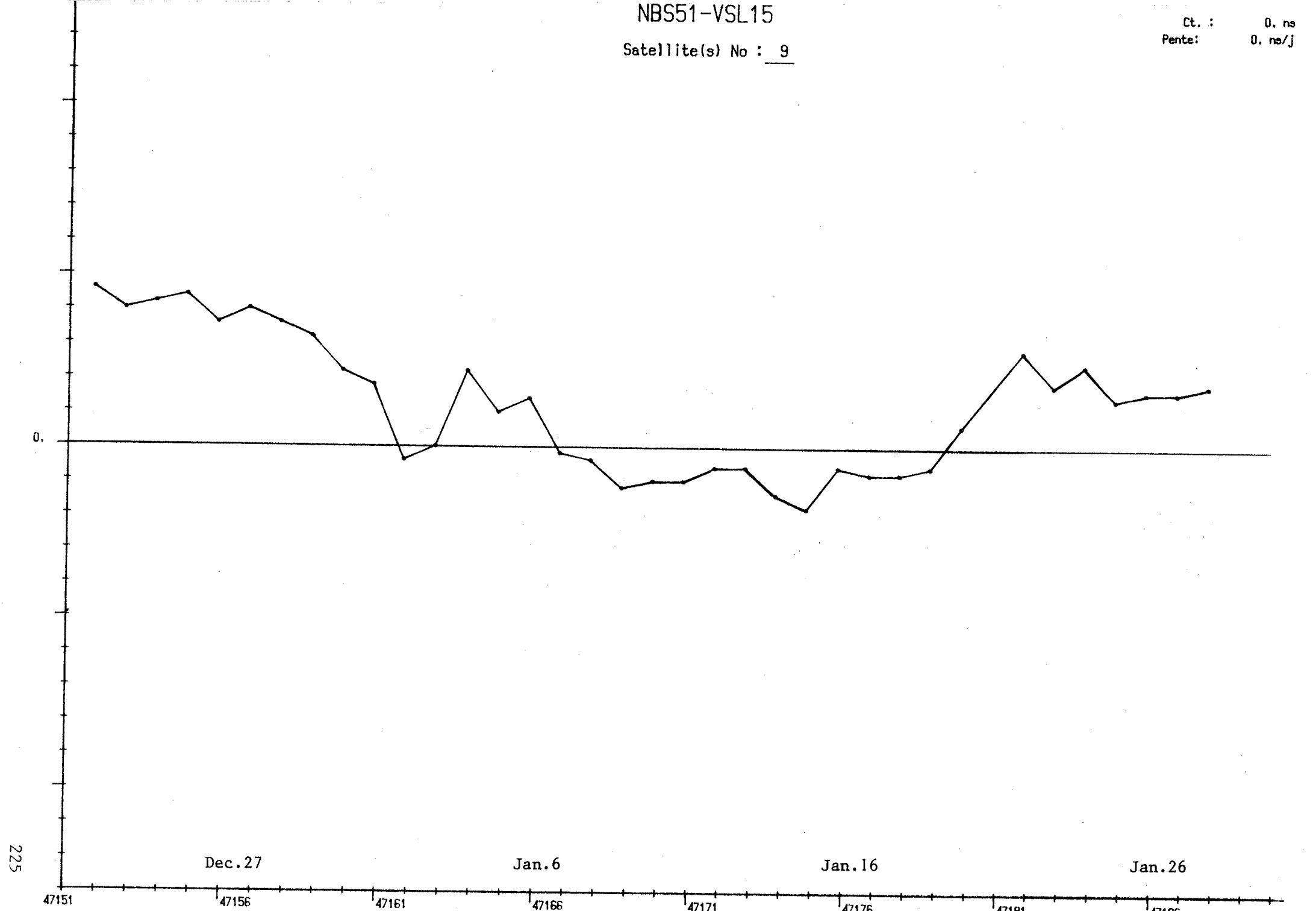
Figure 4. Time values D(t) obtained through passes N°2 (a) and 11 (b) - see Table 1.

Figure 5. Stability of the daily frequencies obtained from the D(t) measurements. Total dispersion is indicated ; results referring to 5 specific passes are showed ; they are obtained from 37 time measurements.

Figure 6. Repartition rate of the 24 D(t) measurements of December 30 (a) and of January 24 (b).

Figure 7. Frequency stability of the time differences UTC(OP) - UTC (Besançon) through GPS (35 days, 16 satellite passes).

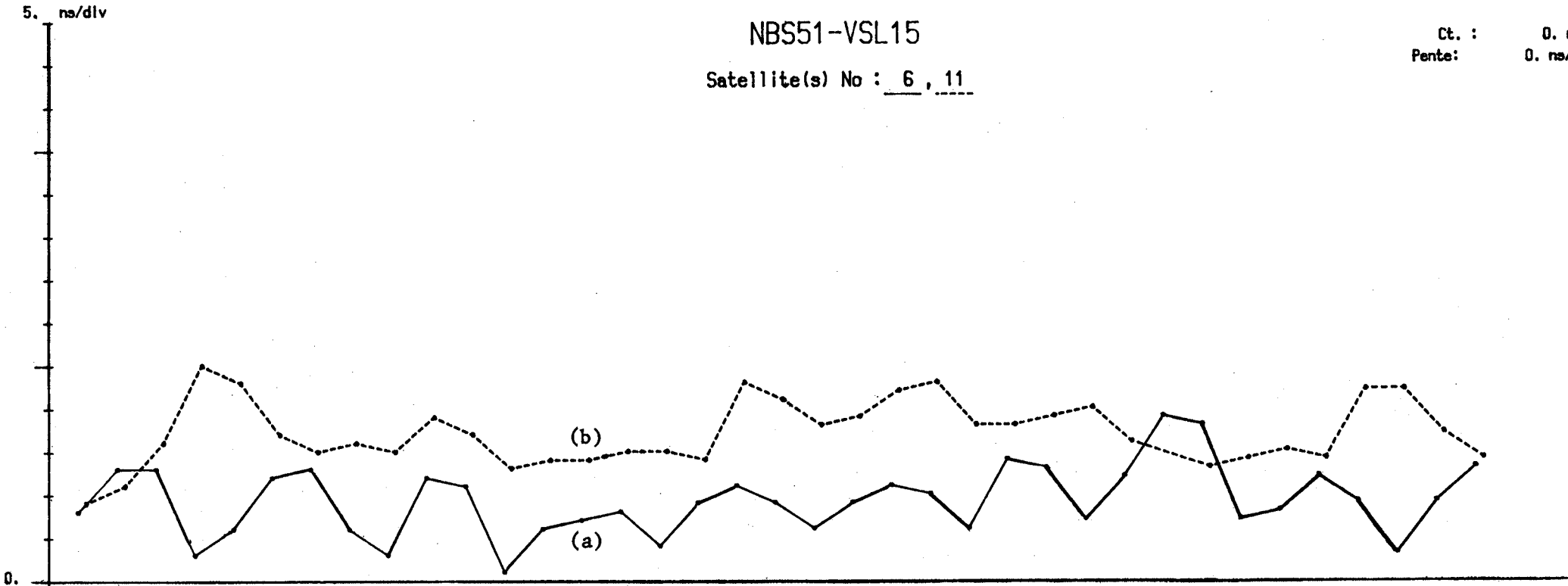

a Cs clocks+TV link
c TV link
b Cs clocks
10^{-13}
10^{-14}
STABILITY $\sigma_y(\tau)$
1
2
4
8
16
32
SAMPLE TIME τ (day)


NBS51-VSL15
Satellite(s) No : 9
Ct. : 0. ns
Pente: 0. ns/j
0.
Dec.27
Jan.6
Jan.16
Jan.26
47151
47156
47161
47166
47171
47176
47181

NBS51-VSL15

Satellite(s) No : 6 , 11

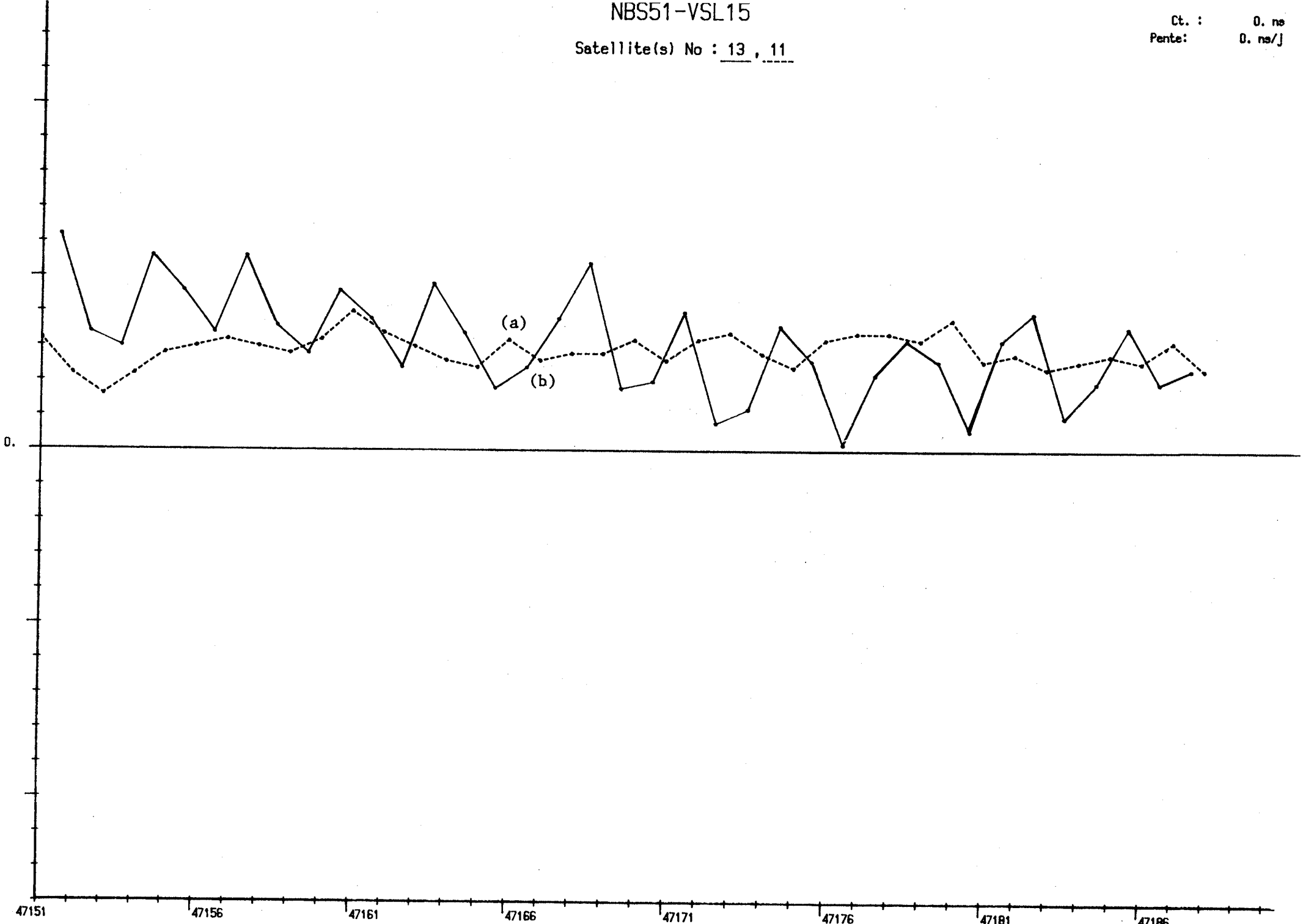
Ct. : 0. ns
Pente: 0. ns/J

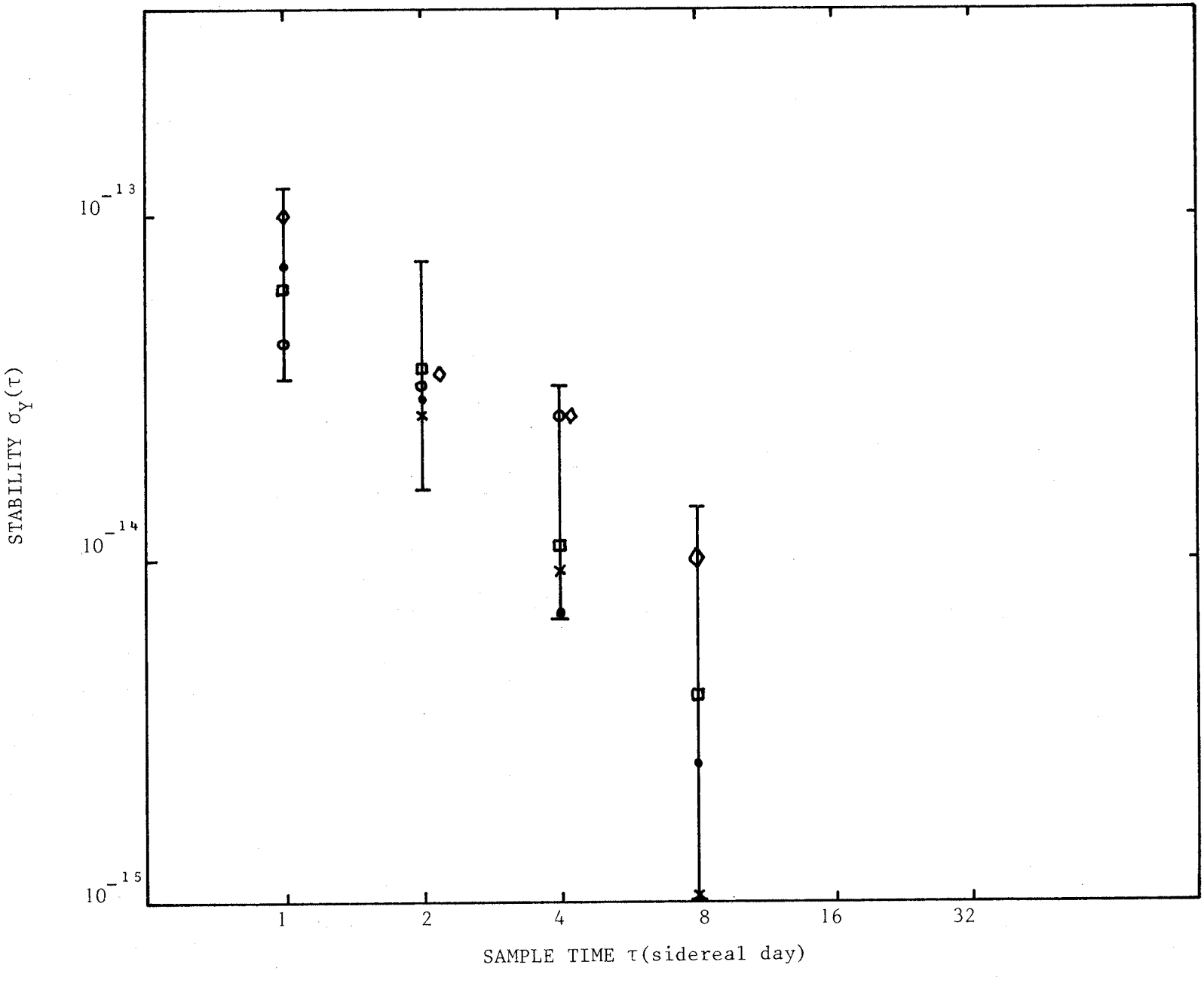


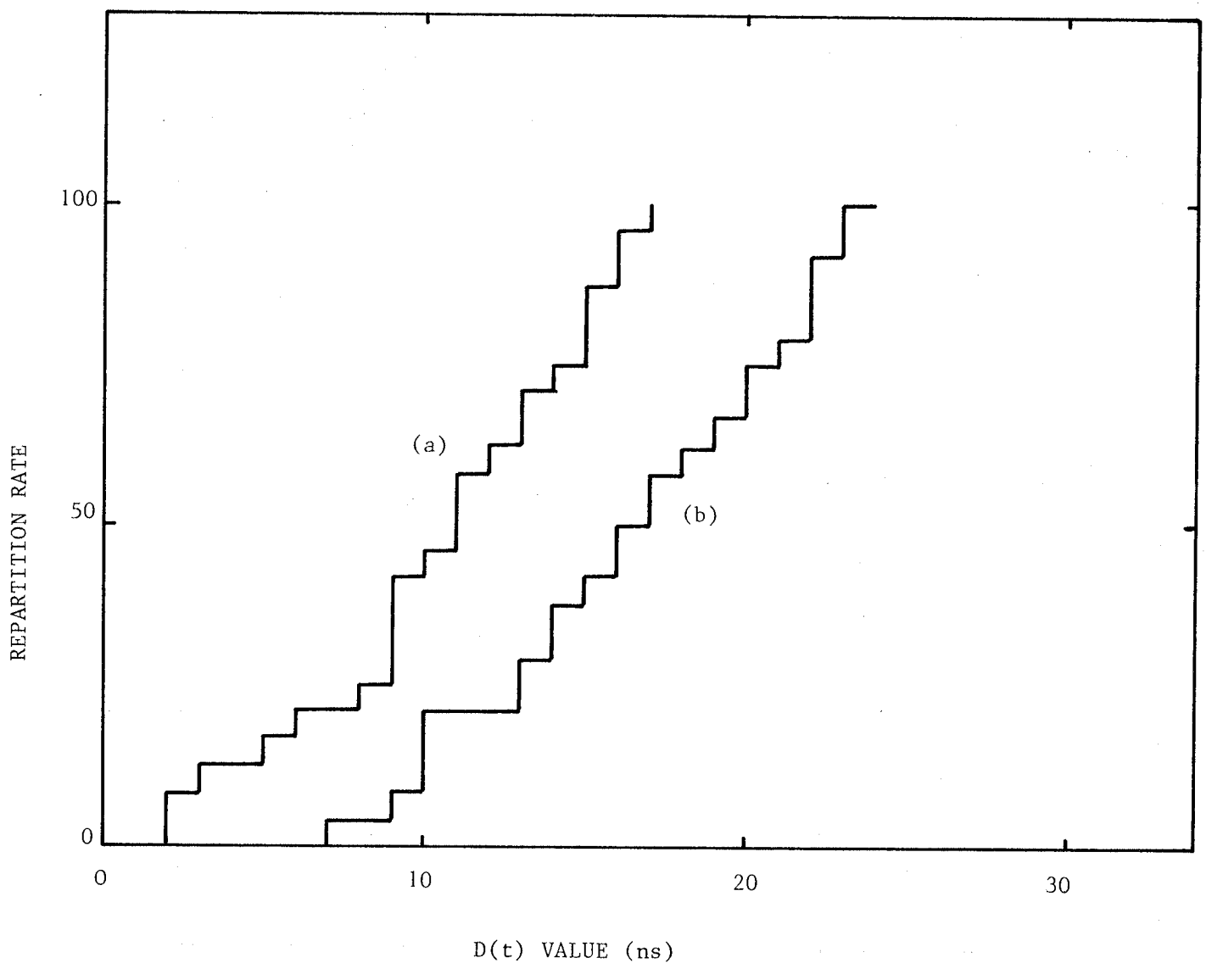
NBS51-VSL15

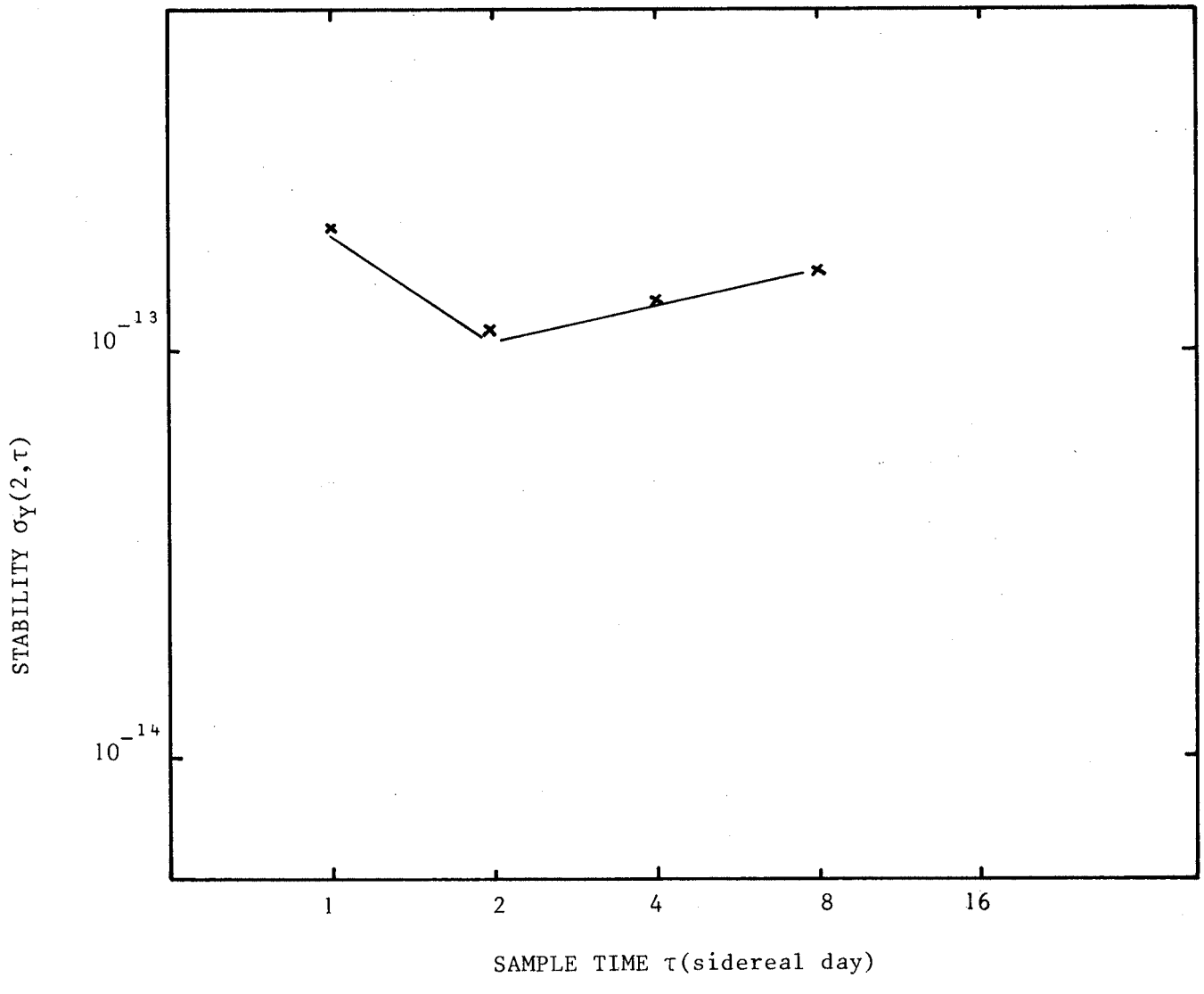
Satellite(s) No : 13 , 11

Ct. : 0. ns
Pente: 0. ns/J




STABILITY $\sigma_y(\tau)$
10^{-13}
10^{-14}
10^{-15}
1
2
4
8
16
32
SAMPLE TIME τ(sidereal day)


100
50
0
REPARTITION RATE
(a)
(b)
0
10
20
30
D(t) VALUE (ns)


STABILITY $\sigma_Y(2,\tau)$
10^{-13}
10^{-14}
1
2
4
8
16
SAMPLE TIME τ(sidereal day)

2 ND EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 16 - 17 - 18 MARCH 1988

Session TF-02 Number 4

Improvement of international time comparison at OFM using GPS common view satellite observations

L. Prost, H. Brandenberger

Swiss Federal Office of Metrology, Wabern, Switzerland

SUMMARY

In Switzerland the atomic time scale TA(CH) is elaborated by comparisons of 15 clocks situated at different sites and linked to OFM by LORAN–C or TV signals. Until February 1987 TA(CH) was transferred to TAI by LORAN–C as well. Since then, a GPS link is operating and the results of common view observations with the Physikalisch–Technische Bundesanstalt Braunschweig (PTB) are used to transfer TA(CH) to TAI. A calibration of the link was achieved by a clock transportation OFM–PTB–OFM. The results available from March 87 to December 87 show that an improvement of a factor of 10 in the stability of TAI–TA(CH) may be expected.

1. Introduction

Until 1982 Switzerland contributed to the establishment of TAI and UTC through several independant laboratories namely the Observatoire de Neuchâtel (ON), the Swiss PTT (PTCH) and the horlogical industry (ASUA). However, in 1982 these last frequency standards were transferred to the Swiss Federal Office of Metrology (OFM) and 2 time scales TA(OFM) and UTC(OFM) were published by BIH. In 1984 it was decided that all these contributions should be coordinated and managed by OFM which is also responsible for the link to the foreign laboratories. An free running atomic time scale TA(CH) and a

time scale UTC(CH) were created, based on the contributions of the three laboratories ON, PTCH and OFM.

In the following, we present the elaboration of time scales in Switzerland. Further both the results of the comparison of the local Swiss atomic time scale and the international time scale are shown, covering a period of nearly three years. During this period of time, part of the comparison was done by LORAN–C and the rest by GPS. Using this particular comparison method, an improvement of nearly an order of recognition in the stability of the time scale had thus been achieved.

2) Elaboration of the time scales TA(CH) resp. UTC(CH)

The clocks contributing to the time scales are located at three different sites. As the OFM is responsible for the computation of the time scale, the clocks of the two other laboratories are linked by LORAN–C and TV to OFM, as shown on figure 1.

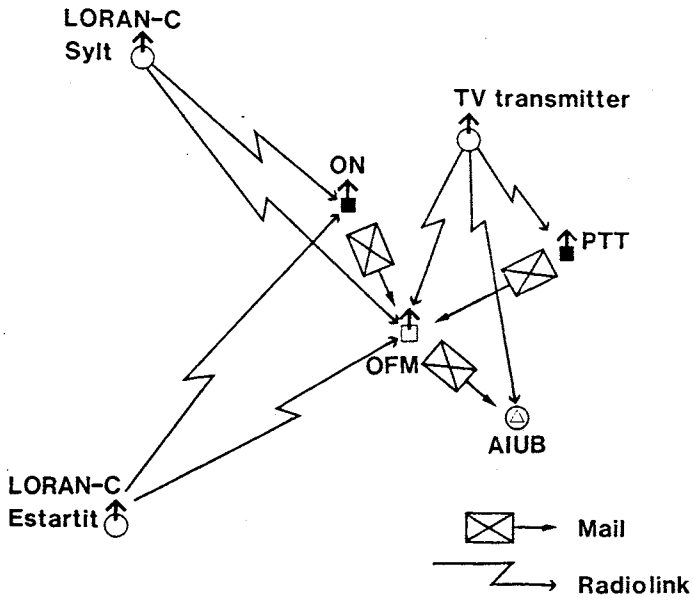


Fig. 1: Clock links of ON and PTT to OFM

In Switzerland, LORAN–C signals from two different chains, namely Norwegian Sea (Sylt) and Mediterranean Sea (Estartit) may be received. Advantage is taken from this fact to link ON to OFM. Up to now, the results of the comparison at ON and PTT are communicated to OFM by mail. As a consequence of these comparison methods, the external clocks are affected with a rather high transmission noise level which lowers the stability of these clocks and for which improvement is seeked.

The stabilities of an OFM internal clock and an external clock versus TA(CH) are shown on figure 2.

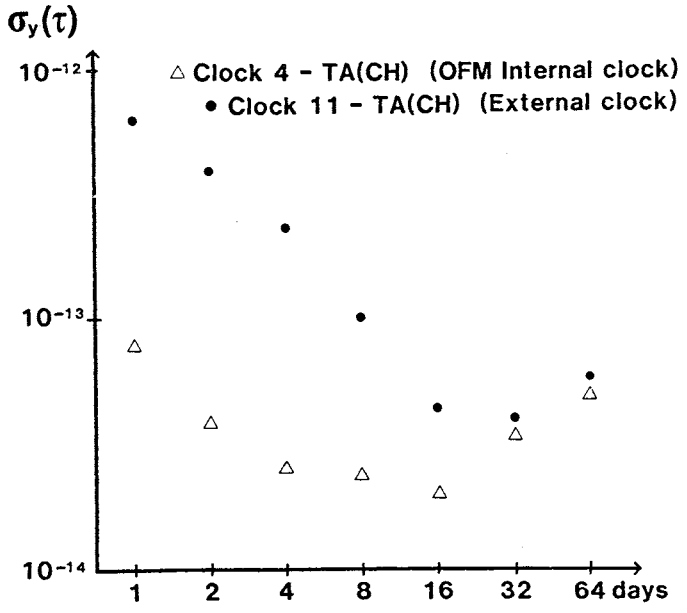


Fig. 2: Stabilities of an internal and an external clock versus TA(CH)

The clocks at OFM are located in a room about 8 m below ground. The environment has a temperature stability of ± 0,3 ^{0} and the stability of the relative humidity is ± 5 %.

The stability of an atomic time scale is function of the stabilities of individual clocks. The knowledge of clocks parameters is an essential requirement to compute a good time scale.

Before a new clock is accepted in the pool, it is first observed for about two months and its parameters namely the rate, the white frequency noise coefficient h_0, and the flicker of frequency coefficient h_{-1}, are determined. Knowing h_0 and h_{-1}, an optimum filter constant for the clock rate prediction is calculated. The time scale is calculated using the algorithms developed at NBS [1]. The prediction interval is 1 day. Two time scales, TA(CH) and UTC(CH) are calculated. TA(CH) is a completely free running. UTC(CH) differs from TA(CH) only in the rate.

TAI is derived from EAL (Echelle atomique libre) by adding a linear function of time with a convenient slope, which was obtained by evaluating the EAL scale interval duration with reference to the primary frequency standards. In the same way UTC(CH) is derived from TA(CH) in such a manner that UTC(CH) – UTC deviations remain small. The rate of TA(CH) is compared with TAI and then the necessary rate correction is applied to obtain UTC(CH). The frequency corrections are small and are introduced only at greater time intervals (loose steering).

No clock steering or physical modifications on the clocks (i. e. C–Field) are applied except for repair, of course. TA(CH) and UTC(CH) are pure paper clocks and are not physically realised.

3) International comparison of CH time scales before the introduction of GPS measurement

Until February 1987, the international comparison was also done using the 2 LORAN–C paths. The daily comparison UTC(CH) – LORAN–C(Sylt) and UTC(CH) – LORAN–C(Estartit) were communicated to the BIH as well as the ten days mean on the modified Julian day (MJD) ending with 9.

At the same rythm, the results of the different clocks in the pool versus UTC(CH) were communicated to the BIH. The stability of TAI–TA(CH) and the residuals of a linear least square fit TAI–TA(CH) for the period January 85 to February 87 are given in figures 3 and 4. The results used in these and the following figures are those published in BIH Circular D.

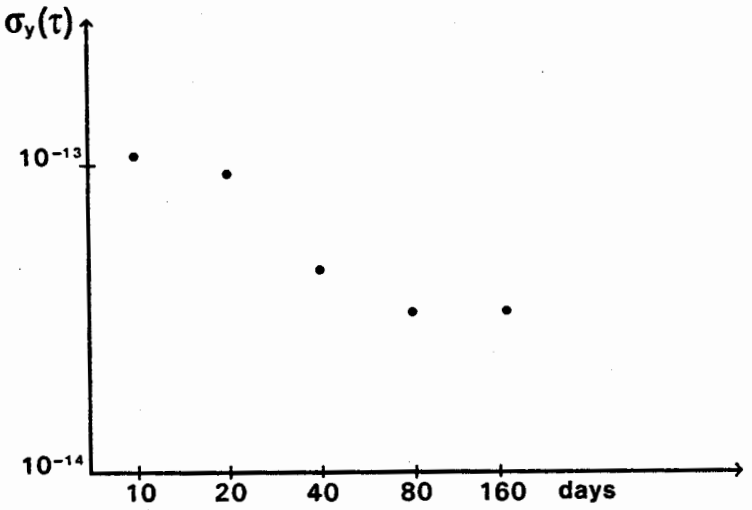


Fig. 3: Stability of TAI–TA(CH), international comparison by LORAN–C

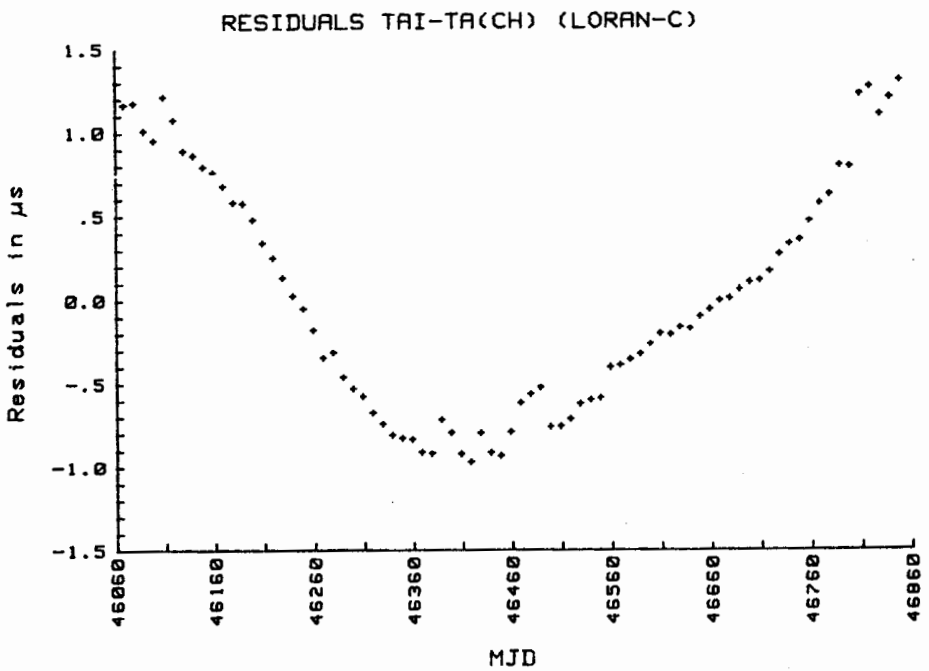


Fig. 4: Residuals of linear fit TAI–TA(CH)

The rather high transmission noise of the LORAN–C links (figure 2) affects of course also the stability of the local time scale compared to the international time scale. With the GPS system having been in operation for several years, a much more reproductible method of time comparison was made available. Therefore, in fall 1986, OFM acquired a GPS receiver. After a small period of getting acquainted with the instrument, we decided to use common view observations with PTB to compare UTC(CH) – UTC(PTB). PTB was chosen because this laboratory operates two primary standards and as that contributes to the steering of TAI. Meanwhile, OFM had also installed a link to Mark III and GPS information from PTB was made available to us. The position of the antenna had been determined by executing several navigation solutions during summer 1986 with a receiver lended by Oszilloquartz. The results of averaged coordinates given by that receiver and the definitive one were the same within the 1σ confidence interval. A first analysis of the results showed that an absolute calibration of the link OFM–PTB by a clock transportation would be useful to take full advantage of the GPS time comparison.

It is worth to mention that PTB switched from WGS 72 to WGS 84 system on Jan. 6th 1987 and OFM only on Jan. 28th. During that period, systematic errors were introduced and the dispersion in the comparison results raised significantly.

4) Calibration of the GPS link by clock transportation

From Febr. 16th to Febr. 24th 1987, in collaboration with Oscilloquartz, a clock transportation was performed between OFM and PTB. The difference of the clock transportation comparison UTC(OFM) – UTC(PTB) and GPS comparison UTC(OFM) – UTC(PTB) was –64 ns on Febr. 16th and –73 ns on Febr. 24th. A mean value of –68 ns was adopted to correct the results delivered by the GPS comparison.

The GPS comparison itself is done in the following manner: As OFM does not dispose of a physical realization of UTC(CH), a reference clock is used as 1 PPS generator feeding the GPS receiver. For every day, the difference

$$(\text{Ref(OFM)–GPS}) - (\text{UTC(PTB)–GPS})$$

is calculated for each common view observation, and a linear least square fit is done on the 20 daily observations. After the linear least square fit,

UTC(CH) – Ref(OFM) + (Ref(OFM) – GPS) – (UTC(PTB) – GPS)

is calculated for that particular date and the correction for the link is applied. This result is communicated to BIH. Typical values of the dispersion from the last square fit are about 4–6 ns. The non–disponibility of a physical realization of UTC(CH) is the reason that no results of the GPS measurements were communicated to Mark III, but if other laboratories are interested, the results of the GPS measurement (Ref Clock–GPS) and the daily results of the paper clock (Ref Clock–UTC(CH)) and (Ref Clock–TA(CH)) could be transferred to March III.

5) International comparison of CH time scales after introduction of GPS measurments

Since February 1987, the daily comparisons UTC(CH) – UTC(PTB) are also communicated to BIH in 10 days rythm. The residuals of a least linear square fit and the stability of TAI–TA(CH) for the period of time from February to December 1987 are represented on figures 5 and 6. The vertical scales are the same as in figures 3 and 4.

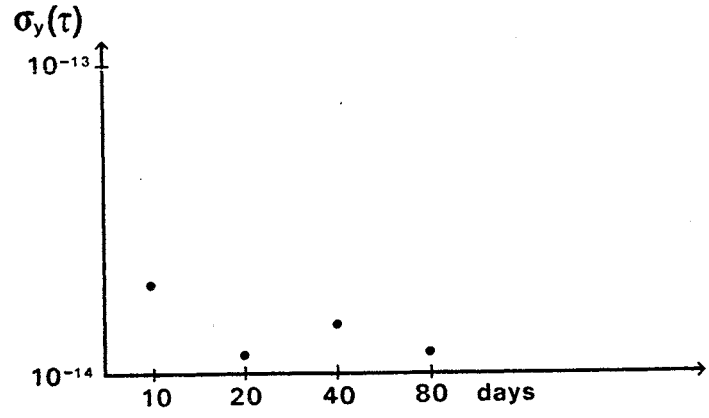


Fig. 5: Stability of TAI–TA(CH), international comparison by GPS

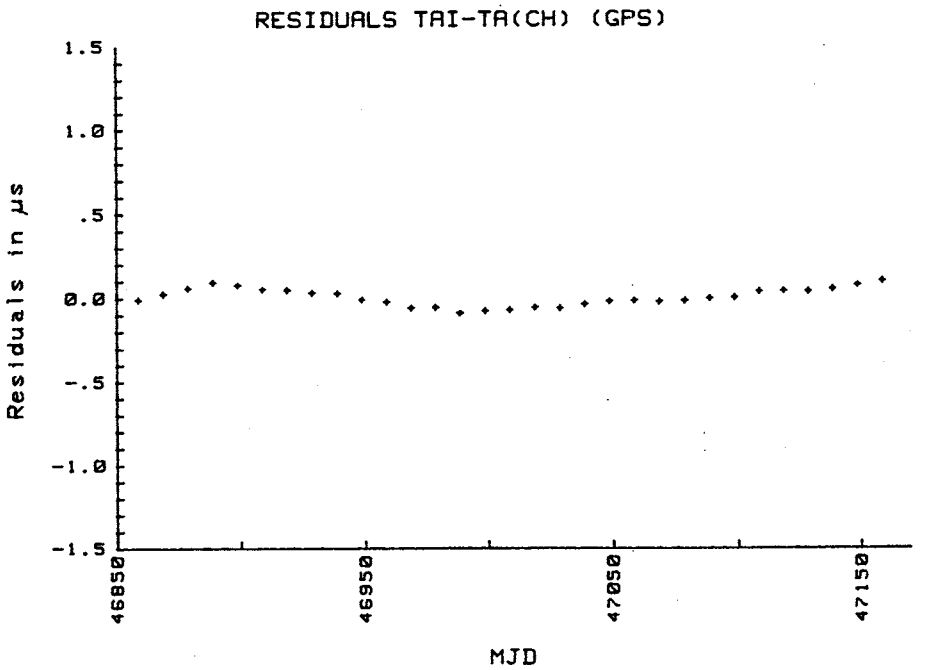


Fig. 6: Residuals of a linear fit

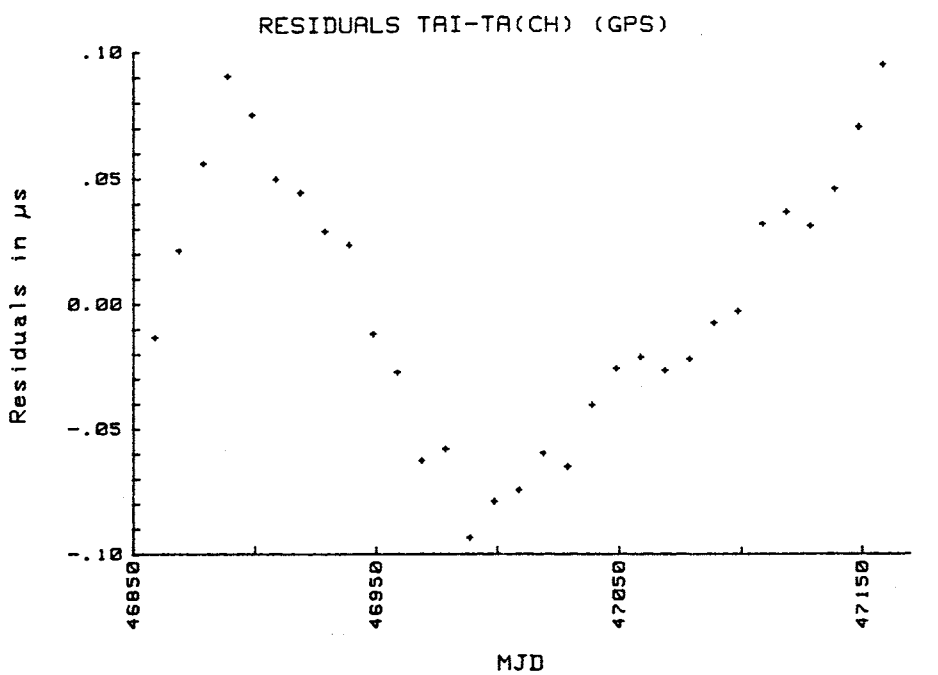


Fig. 7: Residuals of a linear fit (higher resolution).

The comparison TAI–TA(CH) over the entire time interval MJD 46059–47159 is shown on figure 8. The arrow indicates the effective start of GPS measurements in the time comparison. Figure 9 shows the residuals of the comparison with a higher resolution.

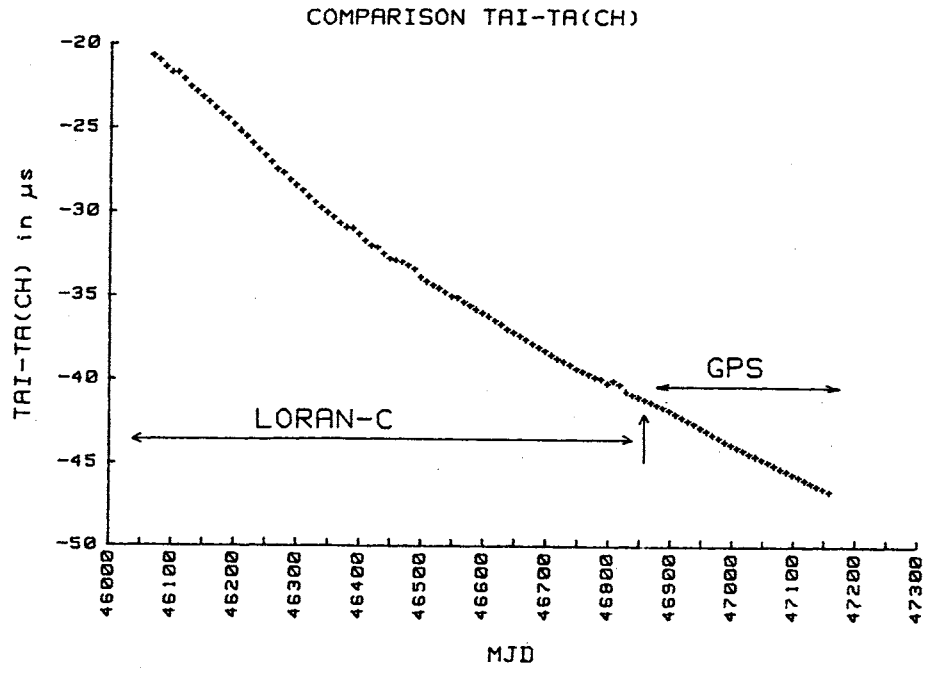


Fig. 8 Comparison TAI–TA(CH) over the entire time interval

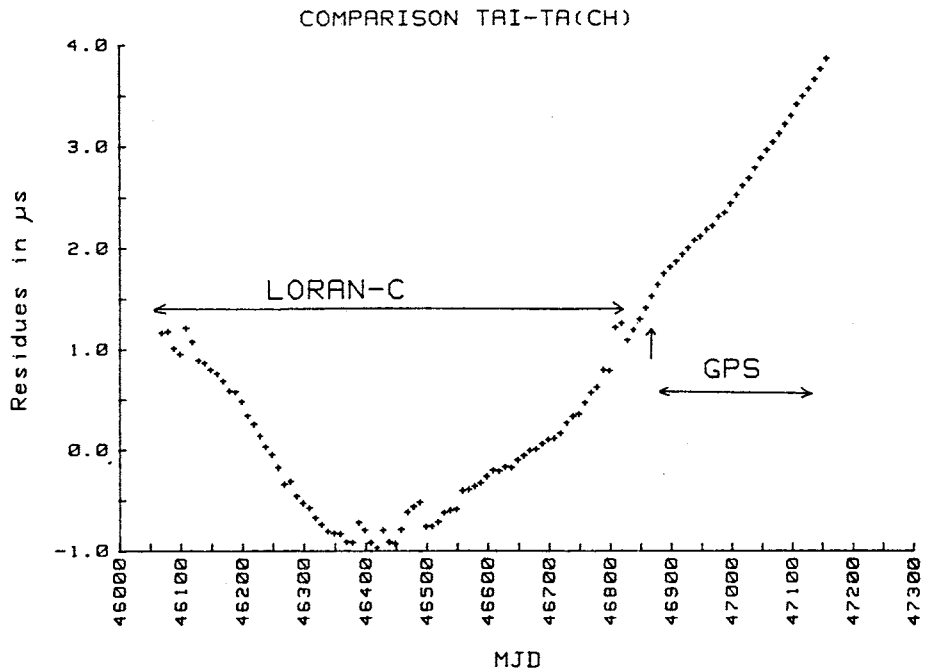


Fig. 9: Residuals of the comparison TAI–TA(CH) over the entire time interval.

6) Conclusion

It has been shown that the introduction of GPS time comparison techniques resulted in an improvement of the stability of the time scale TA(CH) versus TAI. On the other hand, the external clocks are still affected by a higher instability due to the transmission noise for time intervalls shorter than 20 days. The quality of these clocks data could significantly be improved if the corresponding laboratories could be equipped by a GPS receiver to compare their results to OFM.

REFERENCES

[1] B. E. Blair, Editor, Time and Frequency: Theory and Fundamentals, NBS–Monogr. 140, 1974.

TF-03

Timing and Dissemination

1. Present Status of UTC (NPLI)
   B.S. Mathur, P.N. Taneja, G.K. Goel, A. Sen Gupta, P. Banerjee, A.K. Hanjura
   National Physical Laboratory, New-Delhi - India

2. Some sources of timing for communication systems*
   J.L. Jespersen, D.W. Hanson
   National Bureau of Standards, Boulder - USA

3. Measurement of Delay Variation in Digital Communication Networks
   J.P. Mellana, K. Hilty
   Swiss Post Office, Bern - Switzerland

4. Short Distances Frequency Distribution Systems
   E. Rubiola
   Politecnico di Torino, Torino - Italy

5. Slave Time Code Generators*
   D. Ceppi
   CIR, Gals - Switzerland

Session TF-03 Number 1

PRESENT STATUS OF UTC (NPLI)

B.S.Mathur, P.N.Taneja,G.K.Goel,A.Sen Gupta,
P.Banerjee and A.K.Hanjura
National Physical Laboratory,
Dr.K.S.Krishnan Road,
New Delhi-110012,
INDIA

ABSTRACT

The UTC (NPLI) is based on a Hewlett-Packard commercial Cesium clock out of a group of four such standards. Linkage of this time scale was started in 1979 using VLF signal reception, periodically confirmed using portable clock trips. VLF measurements were continued till middle of 1984. In the meanwhile since 1983, the UTC linkage was supplemented using measurement on the NNSS signal. Using these measurements and similar data obtained from USNO it was possible to obtain a link UTC (NPLI)- UTC (USNO). The results were in good agreement with the VLF measurements except for a discripancy of 15 μ sec. The NNSS was discontinued in June 1985 as the USNO data was no longer available.

In August 1986, a GPS receiver was acquired and after an initial process of testing and commissioning, regular data taking on an observation schedule prescribed by BIH was started in February, 1987. Using the data in commonview mode with PTB, computations are being made of UTC (NPLI)-UTC(PTB). The GPS data is also currently being communicated to BIH.

INTRODUCTION

The time and frequency section of NPL, India (NPLI) is endowed with the responsibility of generating and maintaining the Indian standards of Time and Frequency. This has been achieved since 1974 with the help of one out of a group of commercial cesium clocks. The total number of clocks in the group has varied from time to time between two and six. Also

the choice of the master clock whose output was considered as the UTC (NPLI) has varied depending on the individual clock performances. The procedure generally is that all clocks in the group are kept in close synchronisation and are regularly intercompared. The clock designated as the master is replaced only when its main functions or performance deteriorates. Over the years our experience has been that the master clock has continued operating over stretches of two to three years. At present we have a HP 5061B with option 004 Cesium clock as the source of UTC (NPLI) since May 1987.

One of major problems being faced by us at NPL before the advent of GPS was that of a regular intercomparison link between UTC (NPLI) and those of other international timekeeping laboratories. This was because the Indian subcontnent not being in the ground wave range of any LORAN-C transmitter is prohibited the use of this internationally accepted technique for submicrosecond time intercomparison. Various alternative techniques that have been employed over the years are

a) a two way time transfer experiment with PTB .,F.R.G. using satellite symphonie in May-June 1979.

b) a VLF phase time transfer using 16 KHz GBR transmissions between June 1979 and July 1985[2,3].

c) a NNSS satellite time transfer link between July 1982 to July 1985[3] and

d) portable clock trips from USNO in Sept. 1980 and from New Zealand in Feb. 1981.

There was a period of about two years between July 1985 and May 1987 when we had unfortunately no time intercomparison link and finally since May 1987 upto present we have made use of GPS satellite system in commonview mode. The present intercomparison results are of internationally acceptable quality and the UTC (NPLI) data is being included in the BIH clock ensemble. In the present we describe briefly the results of the various international intercomparisons of UTC (NPLI) starting with the VLF link which was the first regular intercomparison technique employed at NPL.

2. VLF RESULTS

We shall not go into details of the VLF technique of intercomparison. This is adequately described earlier[2,3,4] and here it suffices to say that measurements are made only once at midday and seasonal variation of phase due to solar

zenith angle changes are taken account of. Fig.1 shows the variation of UTC (NPLI) - UTC (USNO) over the period July 1979 to June 1985. Choice of UTC (USNO) was made due to availability of VLF data from USNO regularly over this period. Link could be easily established using these results with any other laboratory A by making use of the published B.I.H. Bulletins of UTC (USNO) - UTC (A). There are two gaps in the measurements during October 1981 to February 1982 and again during October 1982 to March 1983. We observe the following features in the VLF results.

a) The jitter in the day to day VLF measurements is of the order of 2-3 microseconds during summer months and increases to about 10-12 microseconds during winter months. This is due to greater instability of the ionospheric D-region during winter[5].

b) The accuracy of the VLF results is within 5 microseconds was verified at least on two occassions by portable trips from USNO and New Zealand.

The rate of the UTC (NPLI) was relatively stable having a value around 7 to 10 X 10^{-13} relative to UTC (USNO) during the period from July 1979 to June 1983. After this the UTC (NPLI) had large changes in its rate reversing its direction during 1983-84 and again reattaing a value around $5X10^{-13}$ in 1984-85. From December 1985 onwards there is again large changes in rate of UTC (NPLI). All these measurements are of course not verified by any other technique.

3. NNSS RESULTS

The Navy Navigational Satellite System (NNSS) has currently 5 satellites in polar orbit of about 1000 km altitude. These satellites although basically meant for navigation can also be used for precise time link with USNO[6]. Results of UTC (NPLI) - UTC (USNO) are also shown in Fig.1 over the period July 1982 to June 1985. The points are the weekly averaged value for all satellite passes and the r.m.s. jitter is shown with vertical error bars. It is apparent from these observations that the general trend of the UTC (NPLI) graph obtained by NNSS is in agreement with VLF results except for the period February to June 1985. We observe however that the values obtained using NNSS are lower than those using VLF by about 15-20 microseconds. This discrepancy could not be resolved and we feel that it could be because of an unaccounted equipment delay in our NNSS receiving set up.

4. GPS COMMONVIEW RESULTS

The VLF and NNSS links were discontinued from July 1985 onwards due to equipment malfunctioning and other difficulties. In August 1986 a GPS receiver was purchased and regular data taking on an observation schedule prescribed by BIH was begun in February 1987. We do not intend to give a description of the GPS system or its capabilities in regard to time transfer as this is adequately described in the literature[6,7,8]. We have used our GPS observations in a commonview mode[8] with PTB. The PTB data is available to us from published bulletins from PTB. The observations schedule is such that we get nine commonview passes of 15 minutes duration each day from 5 GPS satellites. The results of this GPS link are shown in Fig.2. Plotted in this figure are one point each day which is an average of all nine commonview comparisons. The r.m.s. jitter in the nine values each day varies over 10-15 nanoseconds. With such high precision of the comparison we observe that not only the rate of UTC (NPLI) can be very accurately determined but also its short term small fluctuations can be ascertained. During the initial period from 1st June to 15th July'87 the relative frequency offset between UTC (NPLI) and UTC (PTB) was -0.72×10^{-13}. Around July 15,1987 there was a large change in the frequency of UTC (NPLI) for some inexplicable reason. Thereafter the UTC (NPLI) has remained fairly stable in its rate of $\pm 4.16 \times 10^{-13}$ over the remaining period upto the present. There have been short term fluctuations of about $\pm 1 \times 10^{-13}$ over this frequency offset.

From July 1987 onwards the GPS observations of GPS-UTC (NPLI) are being communicated by us to BIH and the computation by BIH of UTC (BIH) -UTC (NPLI) at ten day intervals are available in circular D published by BIPM. These values are plotted in Fig.3 with sign inverted i.e., we have plotted UTC (NPLI)-UTC(BIH). This graph reflects the same behaviour as in Fig.2. This is because UTC (PTB) is extremely stable and has a very small frequency offset relative to UTC (BIH).

5. REFERENCES

1. Mathur, B.S., Banerjee. P.,Saxena,M.,Nand Kumar and Suri, A.K., Journ. of Inst. Electron & Telecom. Engg. (India), Vol. 27, No.10,Oct. 1981.

2. Sen Gupta, A., Goel, G.K., and Mathur, B.S., Journ. of Inst. Electron & Telecom Engg (India), Vol. 27, No.10, Oct. 1981.

3. Sen Gupta, A., Goel, G.K., Taneja, P.N., Hanjura, A.K. and Mathur, B.S., Journ. Inst. Electron & Telecom Engg. (India), Vol.30, 1984.

4. Becker, G., Fisher, B., Kramer, G., Proc. Colloque International de Chronometric Paris, 1969.

5. Belrose,J.S. - AGARD Radio Wave Propagation -AGARD Lecture series XXIX July, 1968.

6. Putkowich, K., Journ. Instt.Electron & Telecom Engg. (India), Vol. 27, Oct. 1981.

7. Allan, D.W. and Weiss, M.A., Proc. 34th Ann. Freq.Control Symp. May, 1980.

8. Allan, D.W. et al, IEEE Trans. on I & M Vol. IM 34, June, 1985.

6. FIGURE CAPTIONS

Fig. 1 :Values of UTC (NPLI) - UTC (USNO) over the period July 1979 to June 1986 using (a) 16 KHz VLF link (b) NNSS link and (c) Portable clock.

Fig. 2 : Values of UTC (NPLI) - UTC (PTB) using GPS commonview technique.

Fig. 3 : Values of UTC (NPLI) - UTC (BIH) as obtained from BIPM circulars.

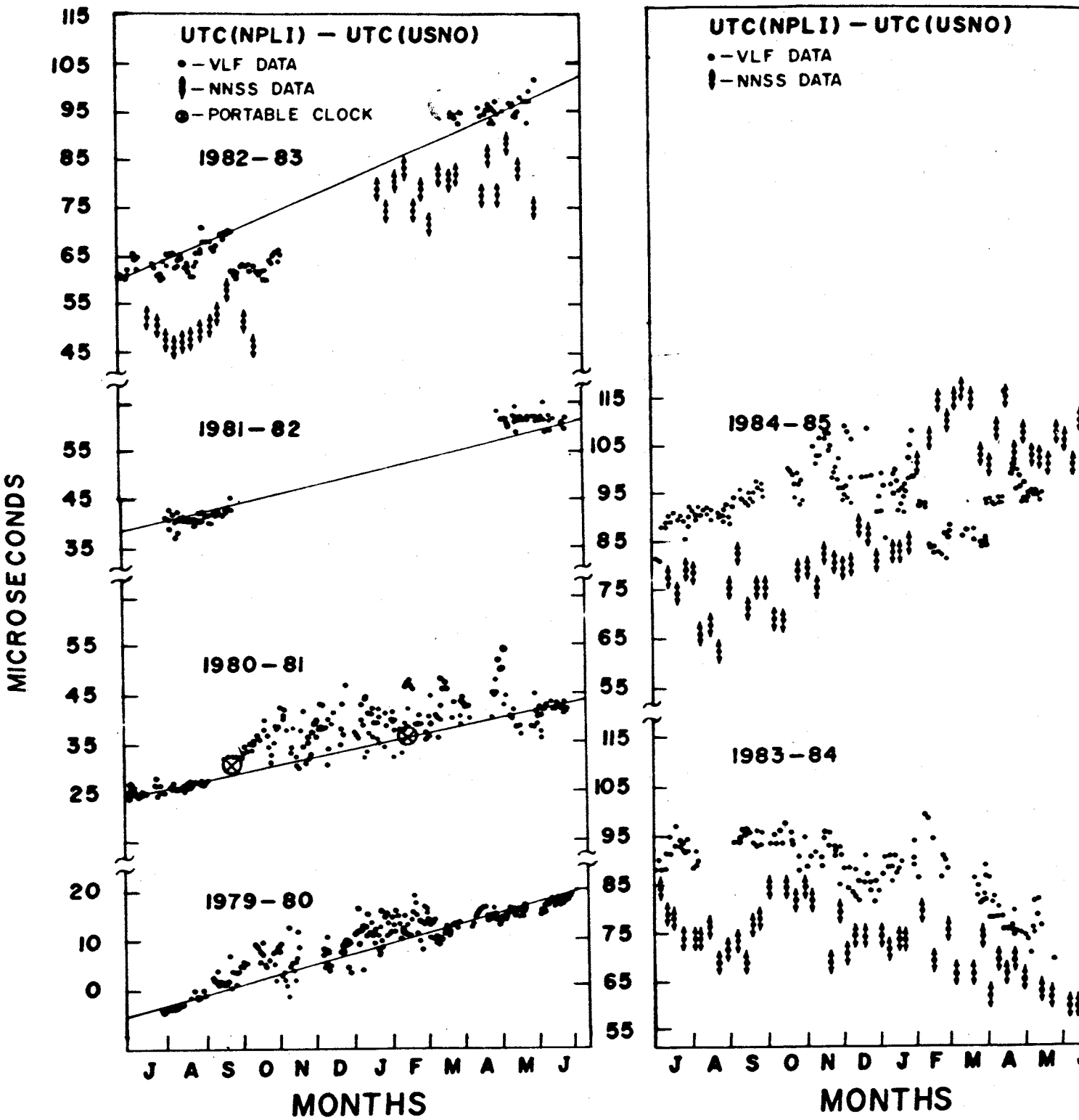


Fig. 1 : Values of UTC (NPLI) - UTC (USNO) over the period July 1979 to June 1986 using (a) 16 KHz VLF link (b) NNSS link and (c) Portable clock.

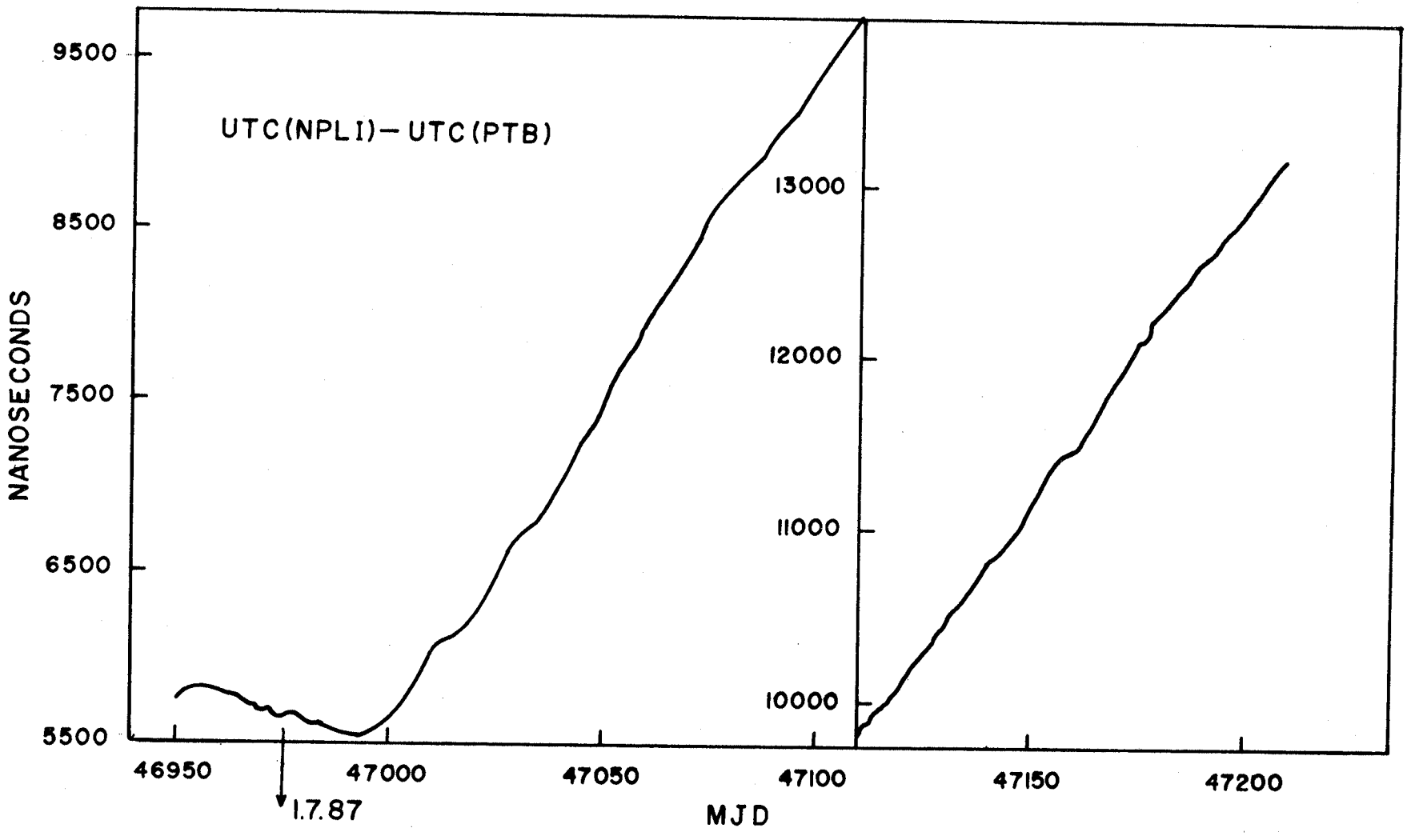


Fig. 2 : Values of UTC (NPLI) - UTC (PTB) using GPS commonview technique.

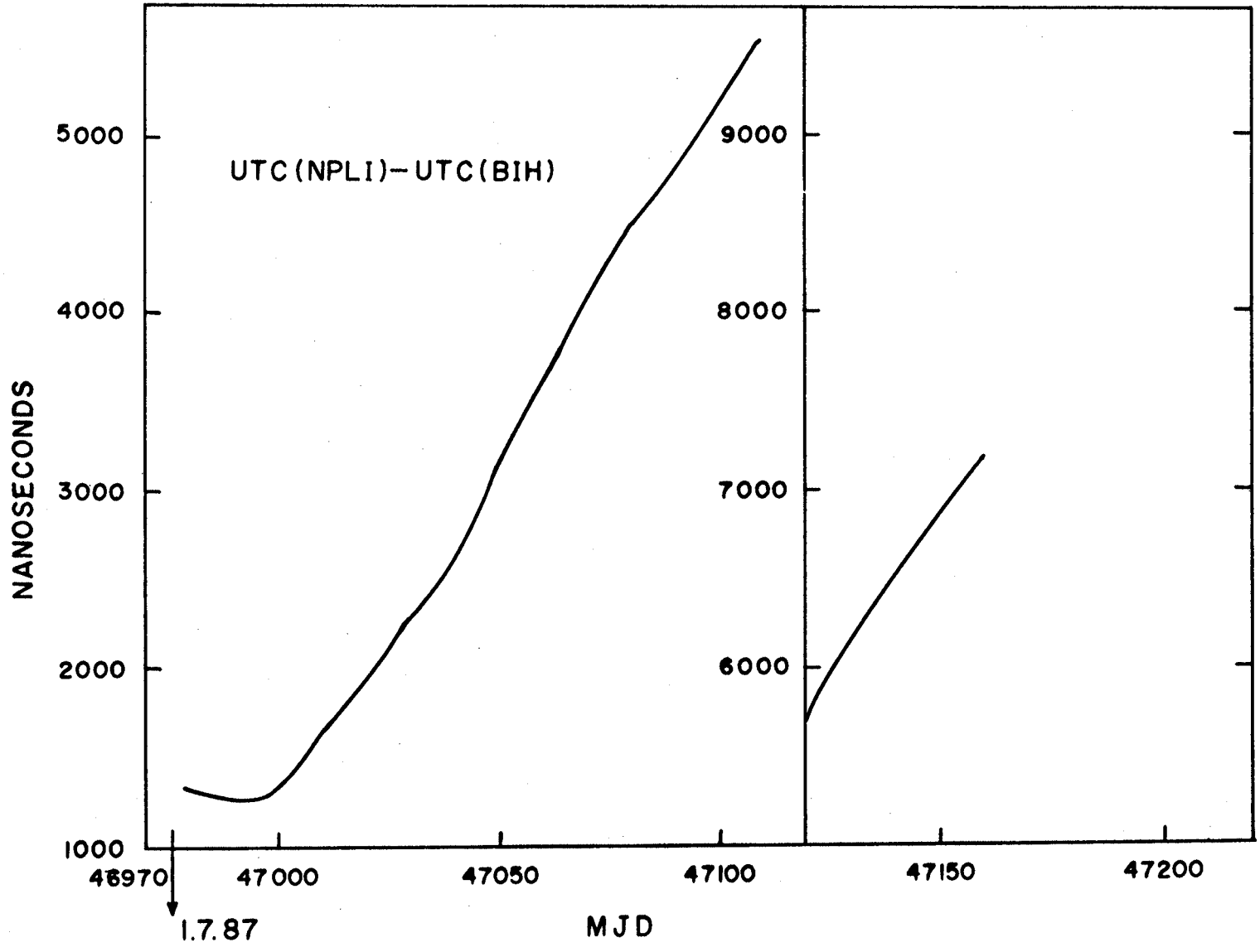


Fig. 3 : Values of UTC (NPLI) - UTC (BIH) as obtained from BIPM circulars.

2 ND EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 16 - 17 - 18 MARCH 1988

Session TF 03 Number 2

SOME SOURCES OF TIMING FOR COMMUNICATION SYSTEMS

J.L. JESPERSEN and D.W. HANSON
NBS, Boulder Colorado 80303 USA

One of the important applications for accurate time is in digital communication systems. Here the applications range from code alignment in code division multiple access systems--spread spectrum modulation is an example--to decoding information that has been stream enciphered. An important problem in all these applications is how does one initially establish the necessary timing throughout a communication network, and, once it is established, how can it be maintained?

Generally speaking, two approaches are possible. In the first the communication network itself is used to establish and maintain network timing. In the case of television, for example, timing is provided by synchronization pulses which are incorporated into the TV signal itself. Thus in the "international" approach part of the communication capacity of the TV system itself is reserved for network timing. In the other approach, the "external approach", network timing is provided by some time source external to the communication network. For example, one could use LORAN-C to provide the necessary time information at major nodes in a communication network. In many cases there is a mix of both internal and external sources of time information.

Much of the network that is now carrying wideband digital communication traffic was in fact built and used earlier for analog communication. Here the communication system itself, where the emphasis was on frequency coordination, syntonization, rather than synchronization, was sufficient to distribute the necessary timing information. But with the more stringent timing requirements of digital communication, the essentially analog system is often not a viable option for providing the necessary timing.

NBS has for some time now been participating with various industry wide telecommunication associations in their efforts at evaluating and solving the timing needs of future and evolving communication systems. Both internal and external time sources are being considered. In particular external sources such as GPS and two-way satellite time coordination techniques are being examined for possible use in future operational systems.

2 ND EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 16 - 17 - 18 MARCH 1988

Session TF-03 Number 3

MEASUREMENT OF DELAY VARIATION IN DIGITAL COMMUNICATIONS NETWORKS

Jean Philippe Mellana and Kurt Hilty
Swiss Post Office, Telecommunication Research and Development Div.
CH 3000 Bern 29, Switzerland

Abstract

A method to measure the slow delay variations (wander) in digital transmission networks, while normal traffic is in progress, has been developed. The measurement system uses the TV-broadcast system as a common reference time base. There is no need to loop back the lines nor to prepare them in a special way for the measurements. The system is observing the traffic on the transmission lines in a passive mode. Measurements has been carried out on an optical 140 Mbit/s point-to-point transmission system of 50 km length.

1. Introduction

The importance of timing and synchronization in integrated digital communications networks has been a subject of research, development and sometimes controversy for many years. Reference (1) describes a network timing concept for the Swiss digital network, based on the hierarchical master-slave principle and using triple redundant timing generators at each major node. The design principles of the frequency control loops linking a slave node to its master over a 2 Mbit/s PCM line have been discussed in reference (2). It is stressed there that the optimum design requires knowledge on

a) the characteristics of the slave clocks
b) the delay variations expressed e.g. by their spectral density.

Methodes for testing clocks under various operational conditions have been proposed (3). However, reliable data on the path delay variations are more difficult to obtain, since it ist not easy to organize long-term path delay measurements on operational transmission systems. Only jitter measurements on the lowest hierarchy PCM links are known (4). Determining the path delay variations of an operational system, without looping back the signals, requires an adequate reference line. The well known technique for clock comparison (5), (6) and our many years experience therewith encouraged us to use the TV-broadcast system as a common reference time base.

2. Television System

In the television system the synchronisation pulse is generated by the studio which produces the program. Generally it has been derived from a quartz - or a rubidium oscillator and then distributed by means of microwave links to the different TV-broadcast stations. Therefore a relative timing information is available throughout the country, without making any modification on the TV-system. In an ordinary TV-receiver there is neither an adequate amplitude stabilization

nor an adequate level stabilization, and the picture information affects the synchronization pulse. More sophisticated TV-receivers (Block diagram Fig. 1) with a special clamping circuit and a line selector have been developed. The radiofrequency section is composed of commercially available modules including automatic frequency control and automatic gain control facilities. After the separation of the sound and picture signals, a video limiter eliminates the picture information. A special clamping circuit is used to stabilize the black level immediately following the synchronisation pulse. Then, an electronic switch selects the synchronisation pulse of the arbitrarily chosen line no. 16 (a line without picture information) (7).

Two of these TV-receivers have been tested over a long period using the signals from the same broadcast station in a setup shown in Fig. 2. The measurements show variations of less than 3 ns a day, using an averaging time of 100 s. The variations from day to day are less than 8 ns. These variations are due to daily adjustements of the synchronization and video level at the TV-studio. This effect requires a good match (symmetry) of both TV-receivers. The measured temperature sensitivity in the ordinary temperature range from 10°C to 30°C is less than 0.4 ns/°C. Similar receivers have also been successfully in use for many years for clock comparison within Switzerland and with the Bureau International de l'heure in Paris.

In order to test the time stability of the TV-distribution system the two receivers were tuned to different UHF-channels (No 40 and No 50) distributing the same TV-program supplied over different microwave links (Fig. 3). Channel 40 is supplied by a path via Zürich (dashed line) and Channel 50 via Geneva (solid line). The total path difference is 670 km and incorporates 12 microwave links. Measurements over 3 months showed total time variations smaller than 20 ns. Simultaneously at three critical points (Pélerin, La Dôle and Chasseral) the automatic gain control of the microwave links was recorded, but no correlation to the delay variations could be detected. This experiment demonstrated the excellent time stability of the TV-distribution system.

3. Principle of measurement

The principles for measuring delay variations of digital transmission lines using the TV-distribution system as a time reference are shown in Fig. 4. At site 1 the frame pulse of the digital transmission (2.048 Mbit/s hierarchy) is extracted and divided to obtain a pulse repetition period of 40 ms. This pulse represents the clock of the transmitting PCM equipment at site 1. Let's suppose that a TV-synchronisation pulse is transmitted at time T_S from the TV-station and the path delay to site 1 is τ_1. Then the TV-sync pulse arrives at receiver 1 at time

$$T_{E1} = T_S + \tau_1 \qquad (\tau_1 = \text{path + receiver delay})$$

A time interval counter measures the time difference T_{d1} between the TV-sync pulse and the frame pulse T_{clock},

$$T_{d1} = T_S + \tau_1 - T_{clock}$$

At site 2 the frame pulse is extracted from the incoming data stream (2.048 Mbit/s). The time of arrival corresponds to the clock time T_{clock} delayed by the transmission path τ_L. For site 2 the following equations are valid:

$$T_{E2} = T_S + \tau_2 \qquad (\tau_2 = \text{path + receiver delay})$$

and the time interval counter reads

$$T_{d2} = T_S + \tau_2 - T_{clock} - \tau_L$$

Taking the time interval measurements at the same time instant (i.e. the same TV-sync pulse) the difference between the readings of the two counters is:

$$\Delta T_{1-2} = (\tau_1 - \tau_2) + \tau_L$$

This difference is proportional to the path delay τ_L and its variation, as long as the TV path delay and the receiver delay remain stable, as shown in section 2, therefore:

$$\Delta T_{1-2} = \text{const} + \tau_L$$

4. Measurement system

In order not to disturb the PCM transmission, special demultiplexers are used, which are connected parallel to the transmitting and to the receiving line (Fig. 5). These demultiplexers contain circuits for dividing the frame repetition rate (500 Hz) by a factor of 20 to 25 Hz (corresponding to the 40 ms sync pulse period). A computer initiates the demultiplexers and controls the time interval counters so that they measure at the same time instant at both measurement stations. To improve the measurement resolution, a time interval averaging mode with a measuring time of 100 s is used. After a measurement the computer reads the data from both counters, starts a new measurement, checks the data for plausibility, calculates the time difference and stores the data on a disc for further availability. The first measurements were carried out with both measurement-stations in the laboratory, using a looped back 2.048 Mbit/s symmetrical pair of about 10 km length. The initial delay on the PCM-path was 38'172 ns and diminished during the observation time of one month for 100 ns (Fig. 6). Assuming a temperature coefficient of delay of 3 ns/km°C (2) for a symmetrical pair with paper isolation, the measured delay variation corresponds to a temperature change of about 3°C which appears to be plausible.

5. Field measurements between distant sites

When measuring a transmission path in the field, both measurement systems are located at different sites (Fig. 5) and they have to be remotely controlled by the computer using modems and leased telephone lines. Special attention has to be paid to the initialization and synchronization problems because there are some undefined delays (larger than 100 ns) in the interfaces and the modems used for triggering the measurements. The distance between the two sites is about 50 km and the same TV-station could be used as a common time reference. The measurements were carried out on a 140 Mbit/s optical fibre transmission system (at the 2 Mbit/s level), which was operating during this time for commercial data transmission. Up to now measurements on a period of more than 10 months have been collected and various effects could be observed. There are days (75-80% of measuring time) where short-term delay variations are as small as ± 5 ns (Fig. 7) and other days where delay variations up to ± 50 ns can be observed (Fig. 8). On a few days delay variations exceeded 400 ns (Fig. 9), while the graphic was clipped to ± 100 ns.

Pending further investigation we suppose that these variations originate in the waiting time jitter of the multiplex systems using positive justification (bit stuffing).

The long-term delay variations (wander) of the transmission path mentioned have not yet been significantly determined, (Fig. 10) as an optical fibre shows a very small temperature coefficient of 0.035 ns/km°C (2). The diagramm is corrected for a few discrete steps (at: "A" = + 50 ns, "B" = - 100 ns, "C" = + 150 ns) probably due to modifications of the network. During the small periods at D, E, F the program has to be stopped for revisions.

6. Conclusion

We have described a time delay measurement system which allows long-term observation of the path delay and its fluctuation on digital transmission lines carrying actual traffic. The main advantage of this method is that it can be implemented by means of picking up the signal at both ends and thus without interferring with the network operation. Other methods known to us up to now did require some interferring action in the operation, such as connection of special clocks or loop back of the circuits. The first experience has shown that the delay measurements can be used as a highly sensitive diagnostic tool. Some of the observed anomalous delay variations might indicate independent failures of some equipment along the line, which can not be detected by means of the usual bit error rate measurements.

The authors gratefully acknowledge the many helpful discussions with Dr. P. Kartaschoff and his support.

References

(1) P. Kartaschoff — A Network Timing Concept for Switzerland
Proc. 17th PTTI Conf pp 287 - 300, Dec 1985

(2) P. Kartaschoff — Reference Clock Parameters for Digital Communications Systems Applications
Proc. 12th PTTI Conf. pp 515 - 548 Nasa Conf. Publ. 2175; Dec. 1980
Goddard Space Flight Center Greenbelt, Maryland

(3) E.L. Varma — Standard Clock Testing Methodology
Contribution to CCITT Com XVIII-E Question 15/VXIII, June 1987

(4) A. Kaeser — Quelques mesures de taux d'erreur et de gique en ligne sur les systèmes à 2 Mbit/s sur câbles
Bull. Tech. PTT (Suisse), Vol 54, No 11, pp 428 - 430 (Nov. 1976)

(5) CCIR Study Groups — Question 3/7 Doc. 7/6-E 1975 (Japan)
Problems on Clock Comparisons using Television Signals

(6) CCIR Study Groups — Question 3/7 Doc. 7/125 1977 (Switzerland)
Improved Time Measurement Technique using the Television Synchronisation Pulses

(7) J.P. Mellana — Procédé de mesure permettant de déterminé le temps de trajet absolu et ses variations à long terme dans un système de transmission.
Internal Report SWISS PTT Bern VD 22.110, 6.4.1983

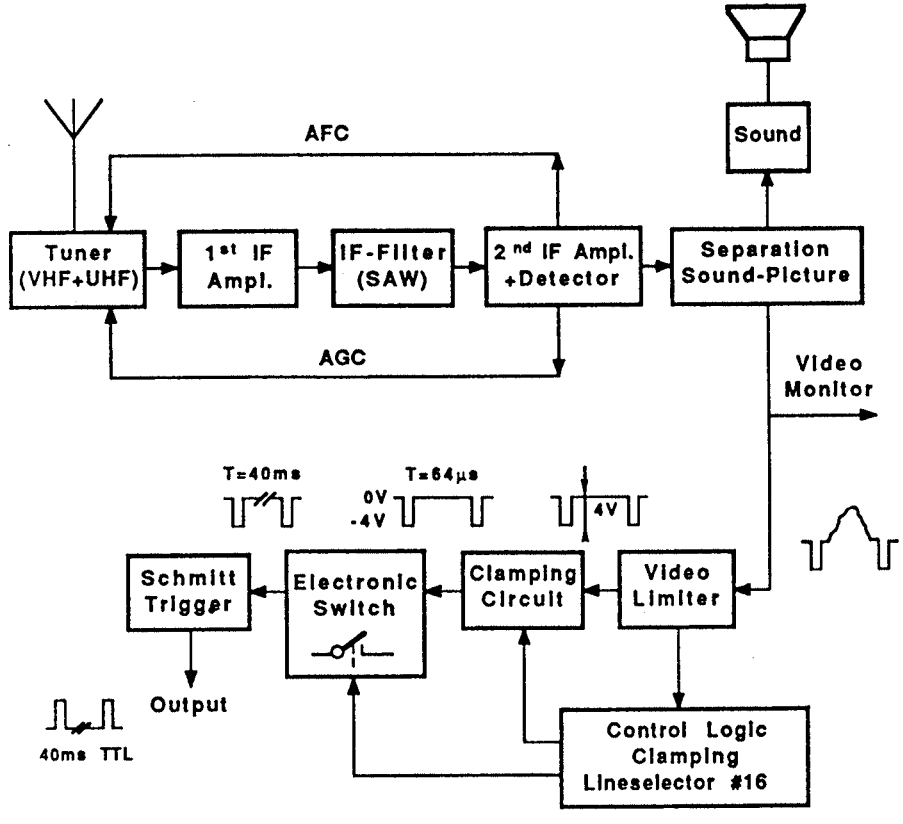


Fig.1 Schematic of TV-Receiver

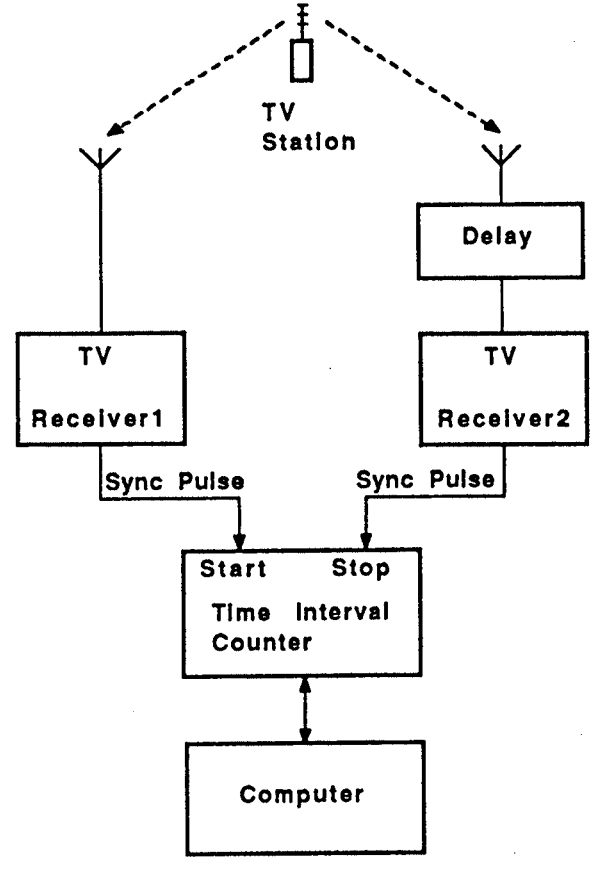


Fig.2 Setup for the Stability measurements of the TV-Receivers

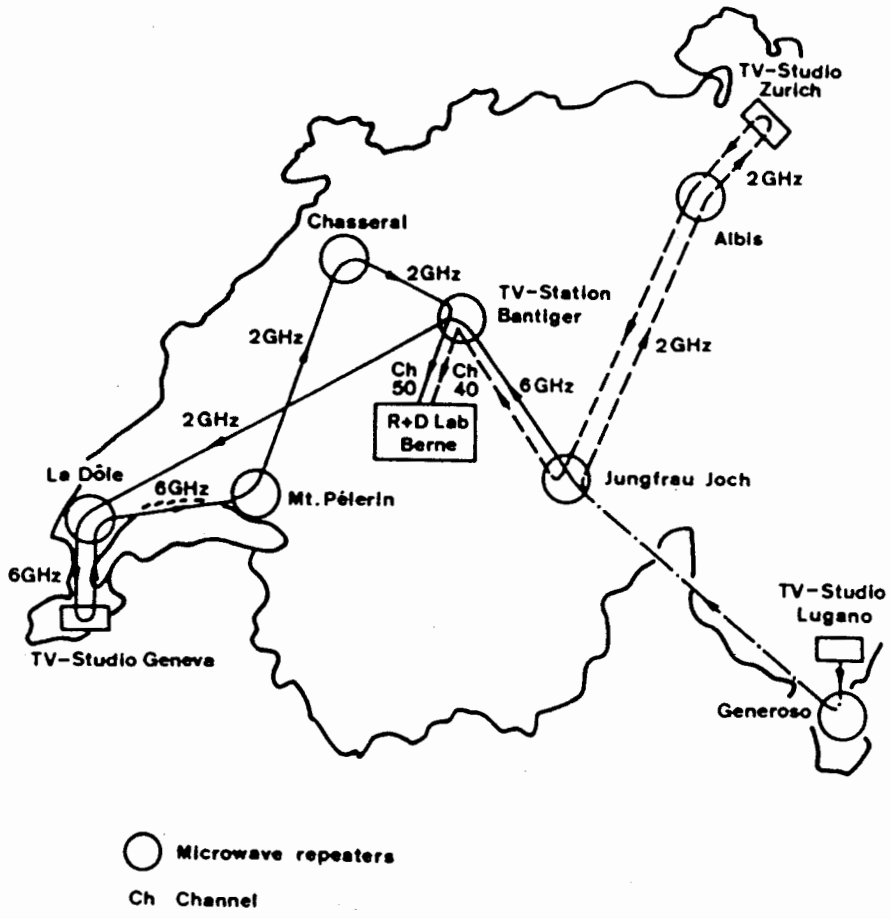


Fig.3 Microwave path for testing the time stability of the TV-distribution system

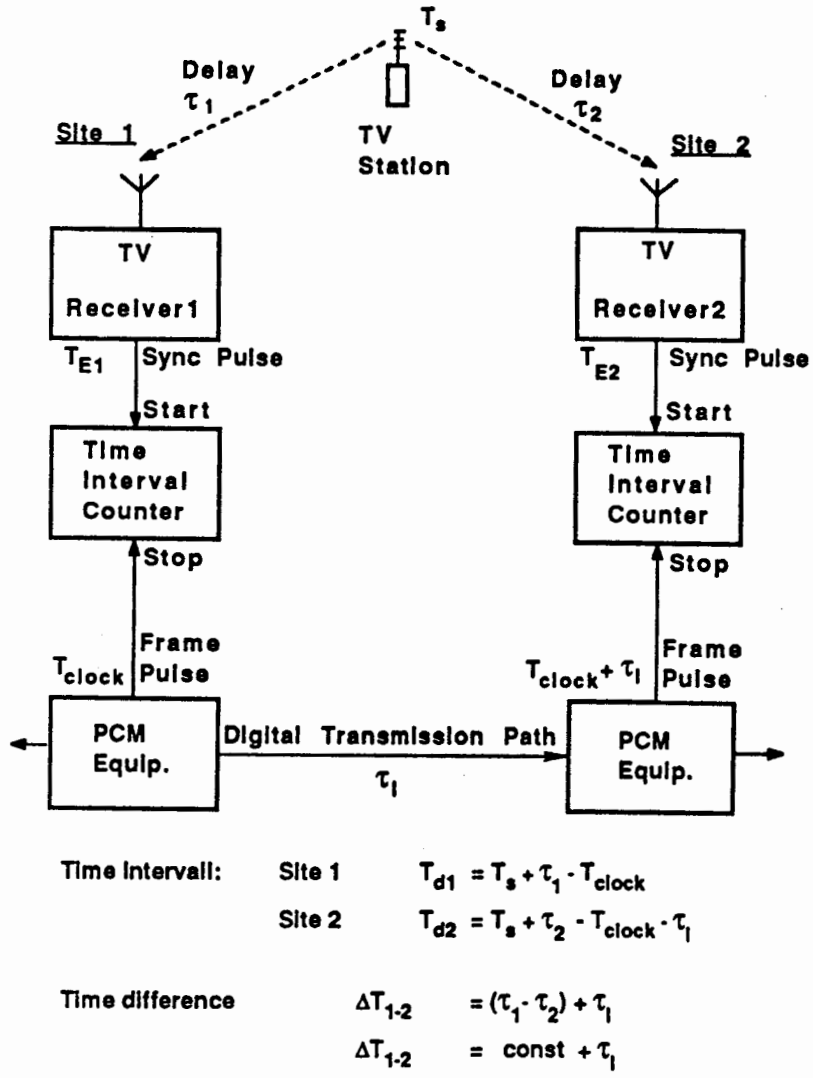


Fig.4 Principle of Measuring Delay Variations of Digital Transmission Pathes

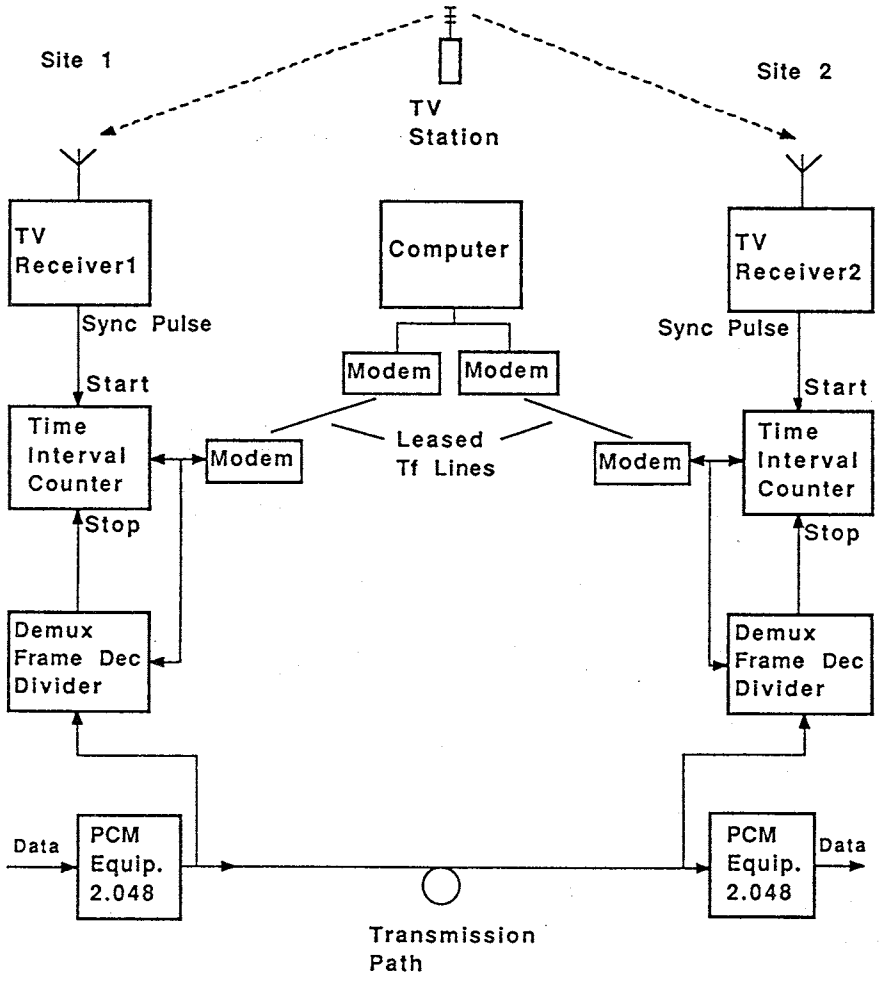


Fig. 5 Setup for Field measurements

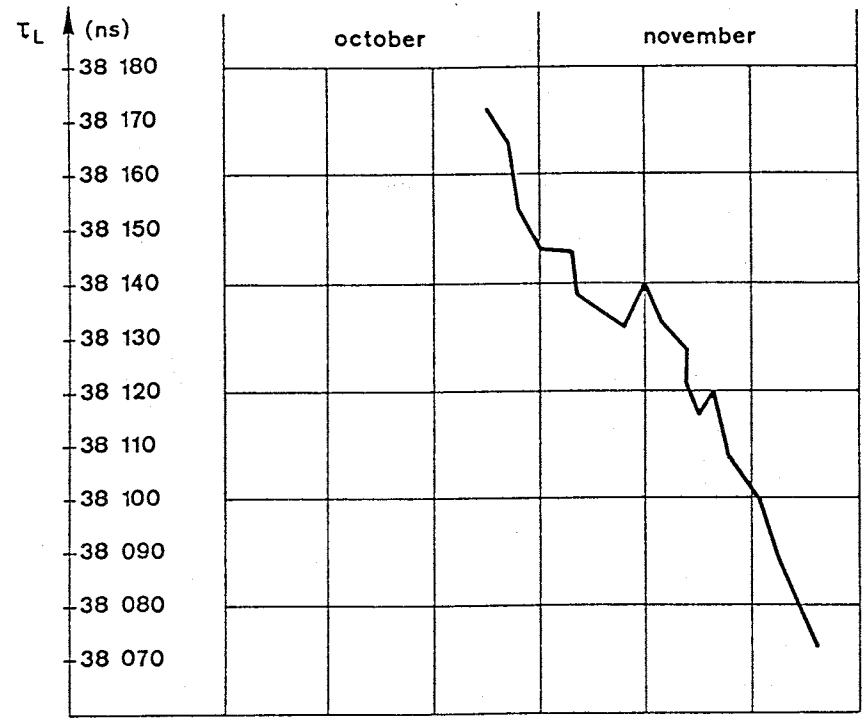


Fig. 6 Long-term Variations on a 2Mbit/s PCM line (symmetrical pair with paper isolation)

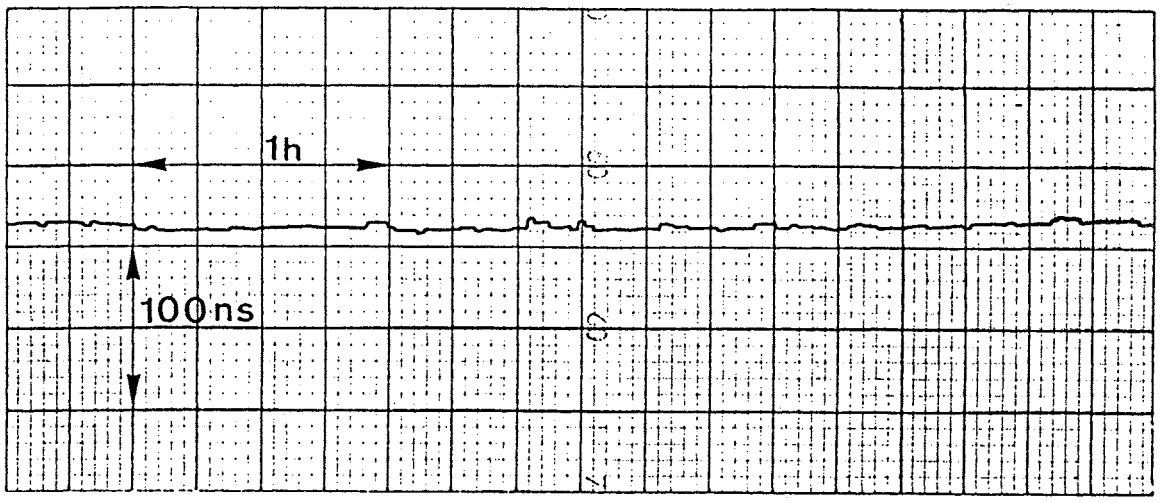


Fig. 7 Short-term Delay Variations on a 140Mbit/s Optical Fibre Transmission Line, quiet period (1.11.87)

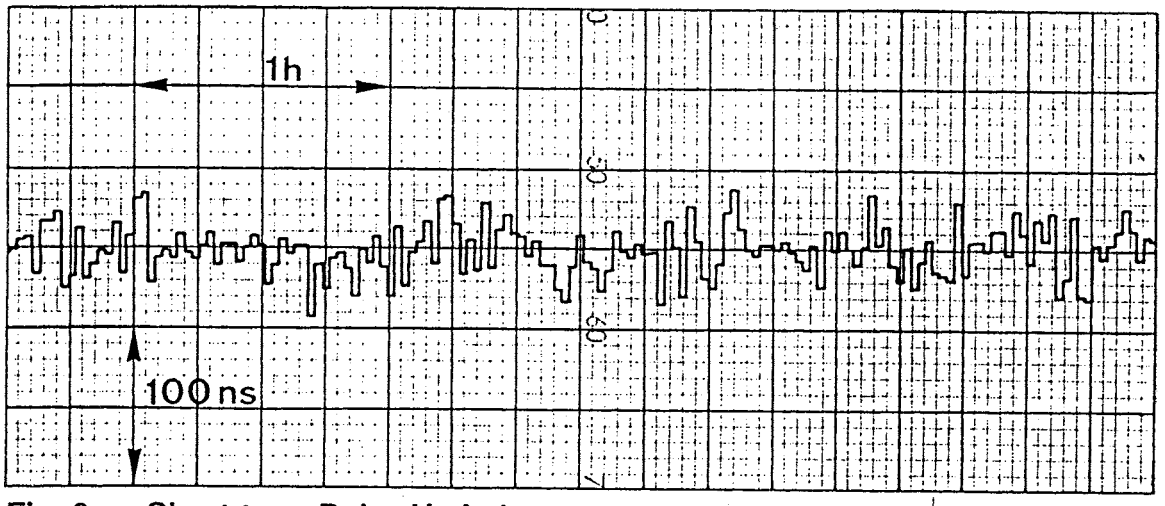


Fig. 8 Short-term Delay Variations on a 140Mbit/s Optical Fibre Transmission Line, disturbed period (14.10.87

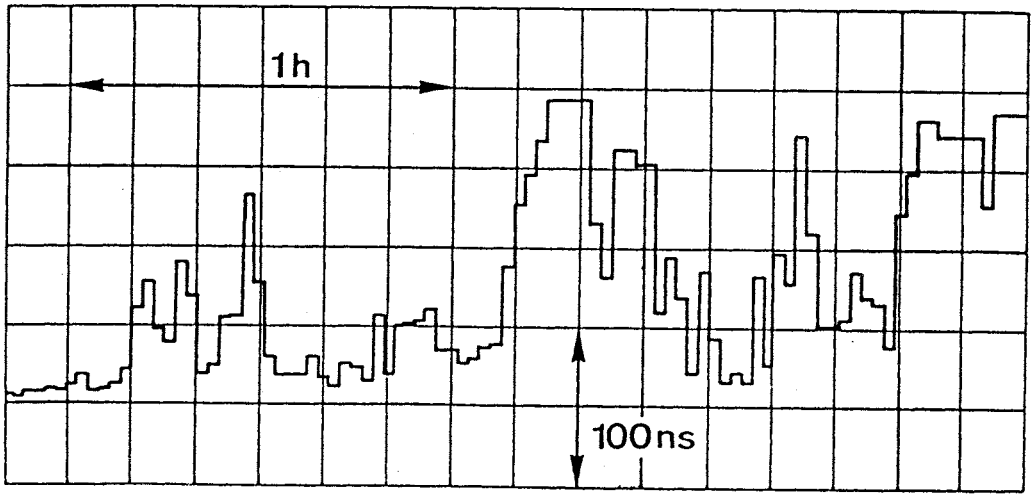


Fig. 9 Short-term Variations on a 140Mbit/s Optical Fibre Transmission Line, disturbed period (6.2.88)

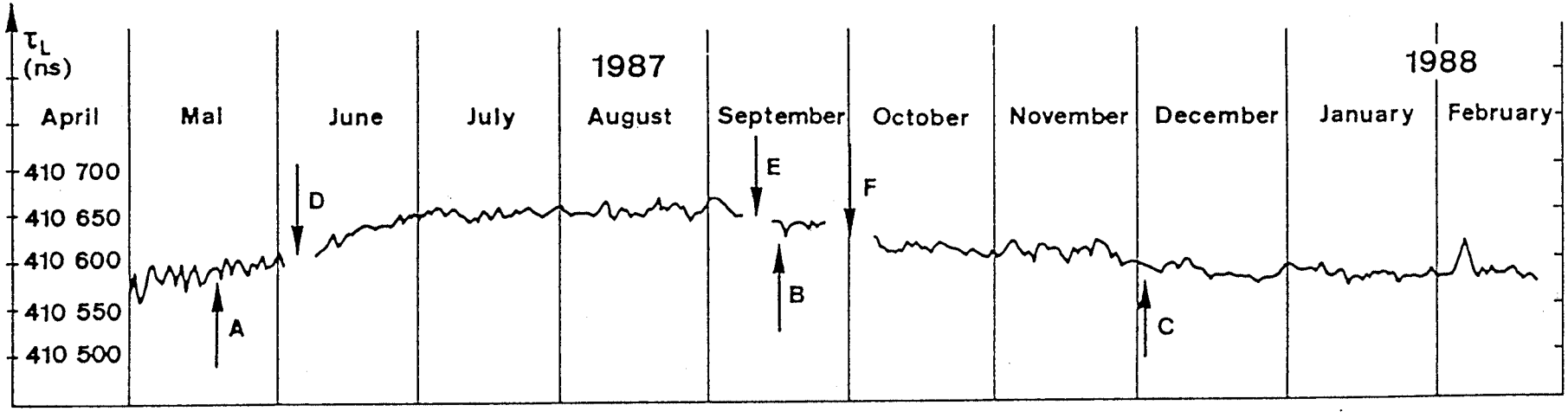


Fig. 10 Long term Delay Variations on a 140Mbit/s optical Fibre Transmission Line (april 1987 to february 1988)

2 ND EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 16 - 17 - 18 MARCH 1988

Session TF-03 Number 4

Short Distances Frequency Distribution Systems

Enrico Rubiola
Politecnico di Torino
Dipartimento di Elettronica

Abstract

In some cases a frequency reference must be established with a minimum of expenses and man power. If a frequency standard is available in another laboratory in the same area, some techniques easier and cheaper than classical ones can be used.

In this paper some tests performed on two different techniques are presented. The first one is a radio link operating in band 8 that has shown accuracy, averaged in a few minutes, near 10^{-11} using as local reference an inexpensive crystal oscillator without any temperature compensation. The second technique is based on a telephone twisted pair that allows frequency transfer with uncertainty around 10^{-13} (one sigma) in 1 day measurements.

Introduction

In some cases, a frequency reference must be established with minimum resources, with given accuracy and a long term stability requirements that cannot be provided by a crystal frequency standard alone. In these cases the use of an atomic frequency standard with the relevant ancillary equipments, such as fail safe power systems and comparison equipments, is not justified. Other methods, as Loran-C or GPS, used to steer a frequency standard, are too much expensive and don't fit man power requirements.

The aim of this paper is to present some tests performed on two different frequency distribution systems which can be used at distances up to about *50 km* when in the area an accurate time and frequency standard is available. The two techniques are the transmission of a low frequency reference tone along a telephone twisted pair and the emission of a frequency stabilized carrier on band 8. Both signals originate at the I.E.N., Torino, where the national time scale UTC(IEN) is kept; all the measurements have been done at Politecnico di Torino, whose distance from IEN is about *6 km*.

Band 8 Radio Link

In this section the performances of a band 8 radio link as frequency dissemination system are discussed.

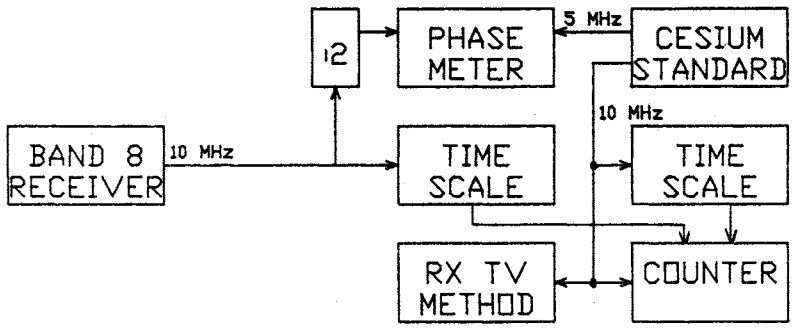


Figure .1: Experimental set-up adopted in band 8 measurements

The system considered is based on a non modulated high stability carrier at *160 MHz* transmitted by the IEN with a power of a few watts; a field intensity of *235* $\pm$*15* $\mu V/m$ is available. The frequency reference source is a caesium beam standard, the same from which of UTC(IEN) time scale is derived.

The long term propagation stability over a path as small as a few tenth of microseconds is expected to be in subnanosecond region; the investigation on this aspect of the problem is out of the scope of this work. Even if a two way link would have been available, separating effects of the various causes would be a serious problem; the system should be treated "as-it-is", considering only overall performances.

Receiver

In a first instance a direct division receiver, consisting in an ECL divider preceded by a sharp cutoff bandpass crystal filter and an amplifier, was adopted; it has shown poor performances because the frequency divider has tendency to overcount carrier cycles in presence of noise. Despite of this, another similar receiver, installed out of the town, where S/N ratio is better, was not affected by this problem.

For the present experiment another receiver has been designed, with a PLL based on a low-cost low-performances crystal oscillator. The servo time constant has been chosen to be *80 ms*, large enough to limit the click phenomenon but reasonably small in order to get experimental results that depend only on the coming RF signal when the measures are averaged on *1 s* or more; an averaging time less than *1 s* is considered useless for frequency dissemination purposes.

The receiver above described has long term stability versus environmental parameters (i.e. ambient temperature, that is not stabilized) that cannot be measured with a *zero-uncertainty* of *1 ns*. During measurements some clicks (phase jumps of *6.25 ns* or multiple) occurred; their probability is around one per day.

Measurement Methods

The long term stability (one week) of the whole system has been measured with television method, taking advantage from common view of the same transmitter. Data fit exactly in the measurement uncertainty ($1\sigma = 15ns$), but this test is considered meaningless: the direct radio link is thought to be significantly better than television system. In effect at the output of the television receiver there is a noise of *10 ns* (*one sigma,* comprehending internal noise and noise associate to the incoming signal) in *60 s* of averaging time, while the output of the receiver has a noise below *1 ns* in the same conditions.

The whole instrumental set-up is shown in fig. 1.

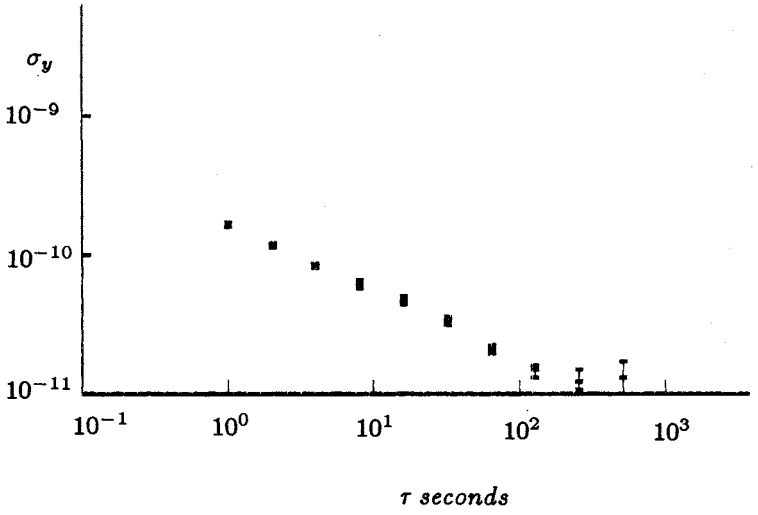


Figure .2: Allan variance measured at the output of the receiver

The phase of the signal available at the output of the receiver has been compared with caesium beam frequency standard. This measurement required the calibration of the standard to fit its frequency offset to UTC(IEN).

Results

The output signal of the receiver, compared with caesium beam standard in more than two weeks of continuous measurements, is stable within *4 ns peak-to-peak* during 10^4 *s*, or *10 ns* day-to-day. The corresponding values of relative frequency differences, 1.1×10^{-13} to 4×10^{-13} are due to the combined effects of propagation, instrumental drifts and differences from local and remote clocks. The effect of some clicks occurred during the measurement has been removed to get these results.

As regards the short term stability, the Allan variance of the *10 MHz* signal obtained from the receiver is shown in fig. 2.

Telephone Link

A telephone twisted pair represents an attractive medium for reference frequency transfer because its reliability compared with cost. A unique bibliographic reference [1] on this subject has been found, with methods and results quite different from these here reported.

A telephone not-switched line has been used for the transmission of a frequency reference at *10 kHz*. The aim of this section is to discuss environmental conditions, used methods and results.

The cable length is estimated to be *10 km* and the attenuation is *40 dB*; there is no amplifier along the line. Despite of the short distance, the situation should be considered less favorable than it seems because the whole path is inside a town with a high volume of telephone traffic; after rejecting frequencies below to *700 Hz* in order to avoid power line interferences, the signal to noise ratio is *12 dB* in *30 kHz* bandwidth.

A band pass filter *8–23 kHz* has been used in some tests, but its thermal drift affect strongly long term results: its thermal stability is around $0.015°/°C$ for laboratory conditions. Phase rotations of *40 ns* peak-to-peak were observed from day to night testing the filter alone. It is preferable to measure the signal without filtering; despite of this, it is necessary to reject *50 Hz* power network interferences.

Three methods of phase detecting have been considered: zero crossing detection with an high performances digital counter, zero crossing detection based on numerical analysis of sampled data and direct phase measurement on frequency multiplied signal. The measurement performed with a

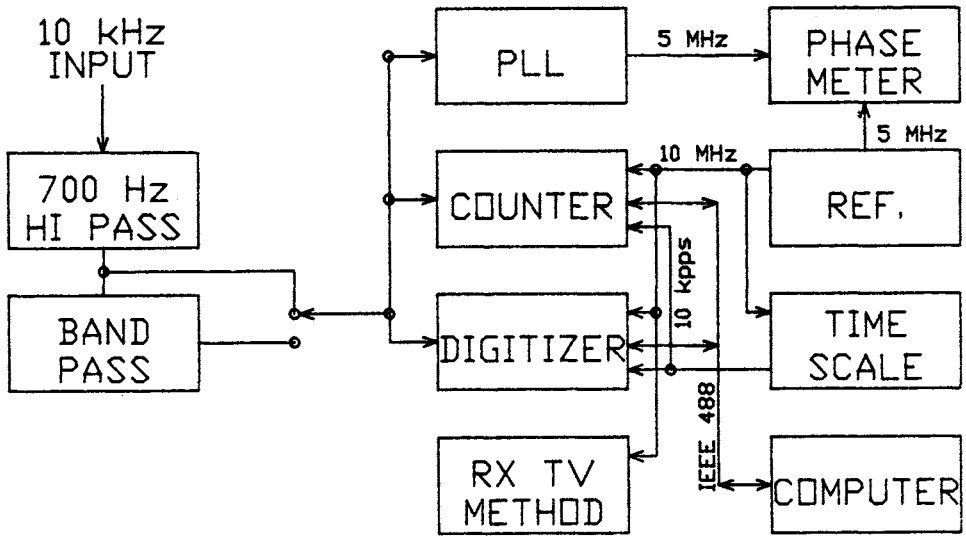


Figure .3: Experimental set-up adopted in telephone cable measurements

phase-meter or a vector voltmeter operating at carrier frequency has not been considered because of the poor resolution: being the period $T = 100\,\mu s$ a resolution of $0.5\,\mu s$ is expected at most from a chart recorder.

The whole instrumental set up is shown in fig. 3. The reference adopted is derived from the band 8 receiver or from a caesium standard.

Digital Counter

One of the simplest way to compare the telephone signal with a local reference is the use of a digital counter. Problems arise from the intrinsically poor triggering characteristics of the counter if frequency is low and noise is added to signal.

The Allan-variance has been measured with two methods: in the first, one value of y_k is the relative frequency averaged on *1 s*, in the second, one value of y_k is derived from average time difference between signal and local time scale using blocks of 10^4 samples.

The first method is essentially limited by instrumental noise due to the slope of the signal.

The second method allows the average on the whole information available and would probably give results two order of magnitude better but the counter used needs *8 s* to take and average *10000* values corresponding to the information available in *1 s*. Only these data are considered as results.

Accordingly with spectrum analyzer response, most of noise power is due to white frequency noise and Allan-variance σ_y goes down with $1/\sqrt{\tau}$ slope.

Digitizer

The second method considered is based on numerical analysis performed on sampled signal. A digitizer has been used, but, considering the operating parameters, a digital low-cost storage oscilloscope would give the same result. The experimental set up does not need especially built instruments, but only a small amount of software for interfacing hardware and detecting zero crossings.

The idea is to extract as information as it is possible from the signal. Acquisition rate is slower than rate that can be obtained with a counter, but many contiguous values are available around a

zero crossing, so an algorithm of fitting can strongly reduce the noise.

The algorithm has been studied to perform noise reduction and cancellation of systematic effects. This is done in three steps:

1. M periods of signal, sampled at frequency f_s, with a dead time between one another due to bus communications, are stored in memory and averaged;
2. zero crossing are searched for and evaluated with a best fit on a small fraction of period;
3. the final phase value is obtained by averaging *positive slope* and *negative slope* partial results.

In step 1 the noise is reduced by a factor $1/\sqrt{M}$; if noise is greater than 1 bit voltage level, quantization noise is reduced by the same way.

In step 2 zero crossing is measured with a best fit on a window of w samples that is very small compared with the period $T = W$ samples. This action corresponds to a large bandwidth filter ($B_w = f_c/w$) so only quantization noise is reduced. Some classical technique of least-squares data fitting has been tested: linear, Gram-polynomial and trigonometric. The trigonometric one gives the best practical results; moreover, it is more robust than other ones with respect of positioning error of the central sampling point.

With the averaging performed in step 3 measured values are independent by dc drift of the instrument because, if ground level is varying, the two measures are shifted of symmetrical amounts.

It should be noted that:

- the bandwidth of the instrument has been limited to *20 MHz*, that is three orders of magnitude more than signal frequency, so variations of instrumental delay can be neglected,
- the filtering action of the best fit is controlled by a quartz crystal oscillator, so the phase delay due to the equivalent bandwidth can be considered constant,
- the positive-slope and the negative-slope zero crossings are searched for independently; for this reason the two observation windows do not cause interference.

The practical values adopted are: size of averaging blocks $M = 256$, sampling frequency $f_c = 20\ MHz$, sampling window $W = 21$ to 101 samples (1 to $5\ \mu s$), without appreciable differences.

Phase Locked Loop

The last technique considered consists in direct phase measurement, but the incoming frequency has been multiplied with a PLL in order to improve the resolution. The use of phase meter and a chart recording device limits this technique to medium and long term measurements.

The multiplying factor, 500, has been chosen for experimental reasons: with the available instruments the resolution is *1 ns*, that corresponds to a relative value of 1.1^{-14} in one day continuous measurement.

Keeping in mind simplicity and cost requirements, the most difficult task is the design of the phase locked loop. The kernel of the instrument is the phase measurement device: only stability is required, but its performances limit the performances of the whole system. The frequency used, *10 kHz*, is too high to be easily handled by a phase detector based on an high accuracy analog multiplier because the reference "must" be a digital signal; the incoming signal is too noisy to be suitable for a sequential detector. More, drift of delays, due to environmental parameters, such as temperature, give serious problems in nanosecond region.

A somewhat unusual phase meter has been developed in a PLL with a servo time constant of *1 s* (fig. 4). The instrument is based on the following ideas:

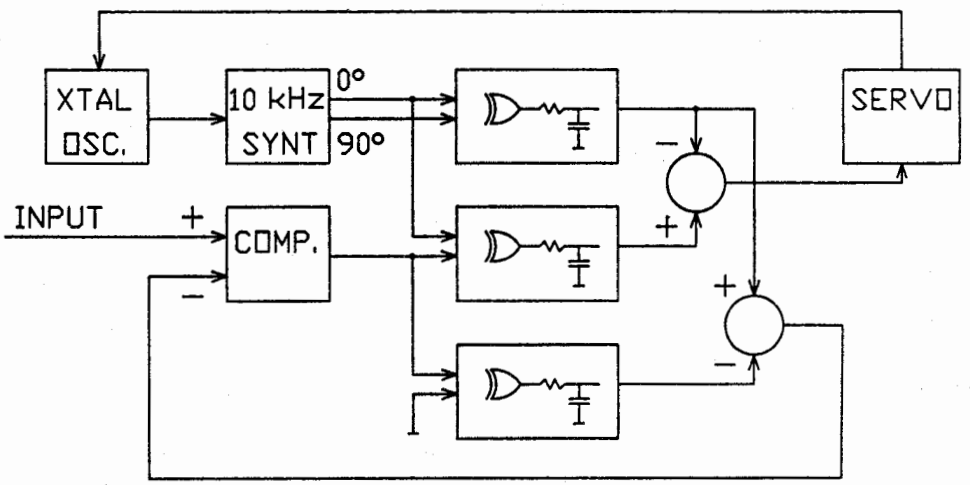


Figure .4: PLL used as frequency multiplier

- from a crystal oscillator working at *5* or *10 MHz* the synthesis of a digital *10 kHz* reference is easy and free of drift problems. Consequently the input signal should be squared and compared with the local reference using the EX-OR logical function. An high speed comparator (*12 ns*), whose delay drift is negligible, can be used;
- with digital synchronous circuits, 0° and 90° references can be synthetised with 50% duty cycle that are accurate within *1 ns* and stable in subnanosecond region;
- three identical circuits are used to perform measurement and compensation: the first gives 90° reference *dc* voltage, the second measures input signal phase and the last, measuring only duty cycle, allows the compensation of comparator offset. All these circuits are in the same chip; providing chip selection with gate delays and output resistances matching criteria, the environmental effects on output average levels can assumed to be equal.

Despite of the use of a low cost quartz oscillator without any temperature compensation or stabilization, the instrument above described has shown thermal variations below 8 *ns peak-to-peak* in a laboratory whose temperature is not stabilized (*8–10 °C* from day to night).

Results

Using the first two methods, the Allan variance of the incoming signal has been measured. The results, compared with the limits of the measuring system are shown in fig. 5. The limits of the systems have been obtained measuring a signal derived from the same frequency reference that drive the instruments.

The PLL system has been tested only with a phase meter and a chart recorder, during more than two weeks. Considering only nighttime values, the long term stability of the whole system is *12 ns* peak-to-peak (one day average value), comprehensive of cable, difference of local end remote clock and instrumental drift. The accuracy of measurement is limited by the noise; in fig. 6 the peak-to-peak noise is shown, versus daytime.

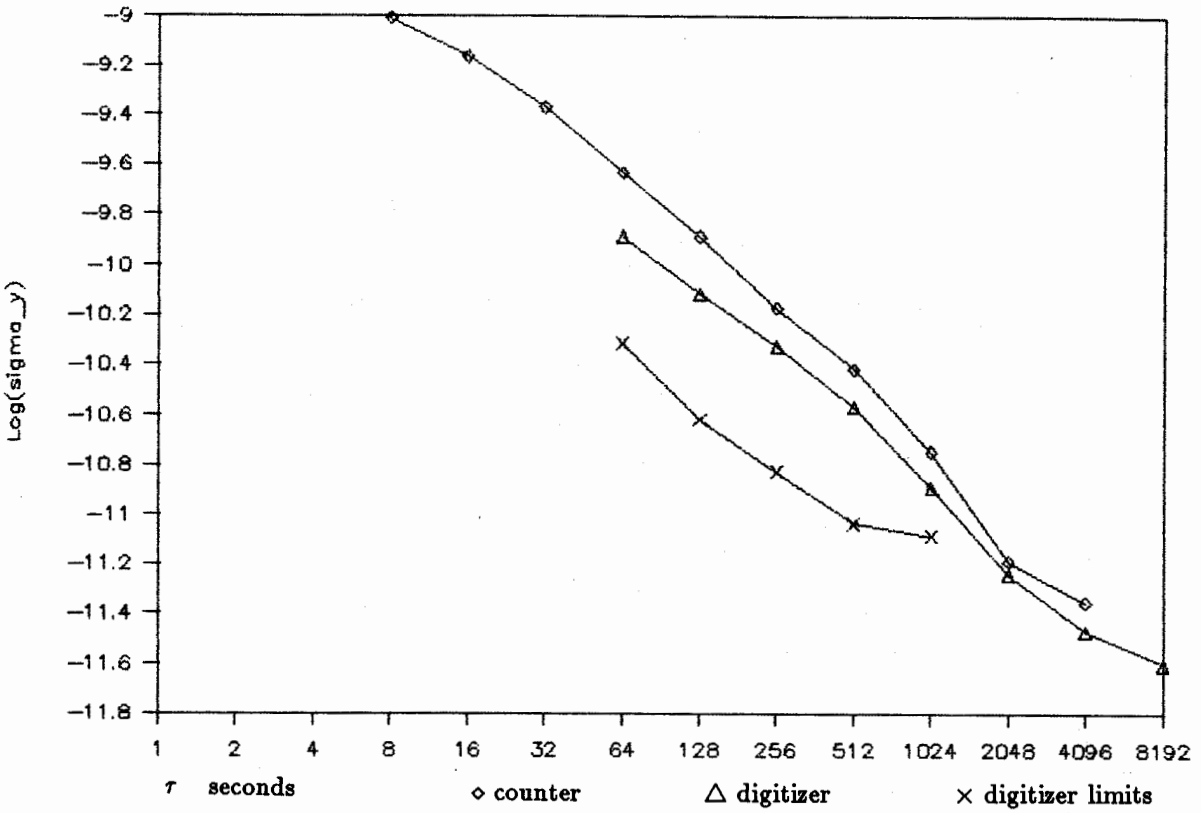


Figure .5: Allan variances of the of the signal measured with counter and digitizer

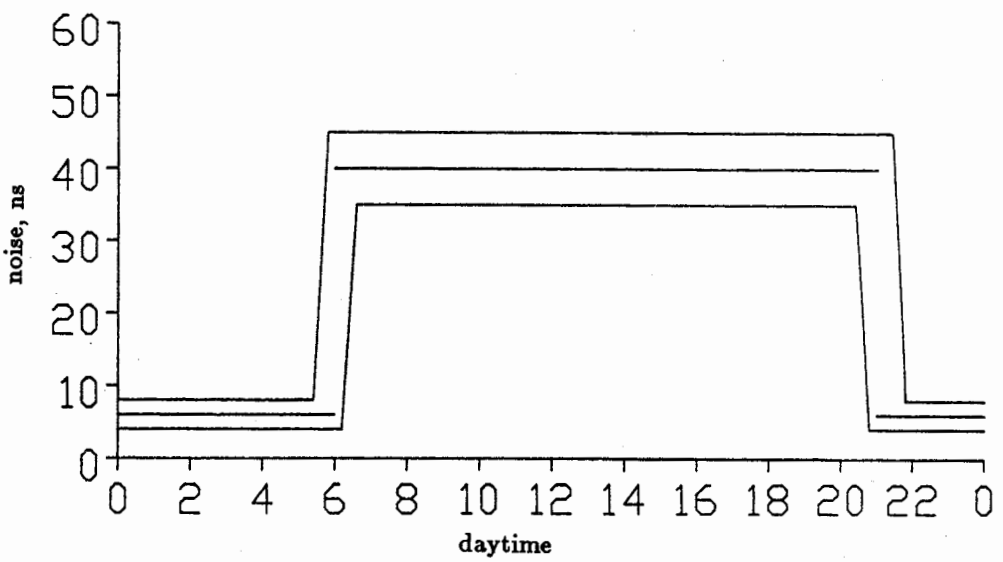


Figure .6: Noise peak-to-peak versus daytime

Final Remarks

The systems considered are quite different as regards accuracy and methods, so it is difficult to decide which is better.

The band 8 system requires a non-standard receiver that must be built and the availability of a radio channel of suitable frequency. It is less reliable than a wire link and the present some problems due to click phenomenon. A longer servo time constant, that could give lower probability of click, had not been experimented because the attention was focused on the signal. The receiver can offer an high quality output signal at *5* or *10 MHz* with stability better than 10^{-10} at one second, near 10^{-13} at one day, that can drive directly a time scale.

The telephone link is not as suitable as the radio link to originate a *5* or *10 MHz* reference signal because of the difficulty of frequency multiplying with adequate stability; the short term stability is not as good as that can be obtained from the other system. The PLL was built only to investigate on the limit of the system; actually it does not meet the man power requirements mentioned in the introduction.

Frequency comparisons via telephone cable can be performed without special instruments. The most attractive methods seems to be the second one, using a low cost digital storage oscilloscope instead of a digitizer. As regards the short term stability, the results seems to be strongly affected by the limits of the instruments.

Bibliography

1. Méthodes de transfert de précision de signaux horaires a courte distance. CCIR rapport 897, Dubrovnik 1986.

2 ND EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 16 - 17 - 18 MARCH 1988

Session TF 03 Number 5

SLAVE TIME CODE GENERATORS

D. CEPPI

COMPAGNIE INDUSTRIELLE RADIOELECTRIQUE

2076 GALS (SWITZ.)

Whenever several pieces of equipment are separated at distances of several kilometers between one another, on a rocket launch site for instance, the necessity of synchronisation becomes evident when a precision of a few microseconds is required. This synchronisation has to be performed by using an IRIG B standard time code on a 1 KHz frequency carrier transmitted by the central time station.

Additionally each piece of equipment must remain synchronised in time with a sustained accuracy of +/- 10 microseconds during a minimum of 24 hours should an interruption occur at the central time station.

To fulfil these requirements, we have developed slave time code generators equiped with BVA type crystal oscillators automatically corrected by digital phase and frequency system. The performance obtained allows to maintain the synchronisation of the slaved generators within +/- 5 microseconds of the central time station whereas the oscillator precision alone is maintained within +/- 5. 10-11.

The paper presented describes the concept and realisation of our slave time code generators.

TF-04

Various Satellite Timing Experiments

1. Timing Data Reduction and Processing for the NASA - Goddard Laser Tracking Network
   E. Detoma, A. Clinton, W. Hanrahan, P. Klushmeider, R. Price, P. Dachel
   Bendix Field Engineering Corp., Columbia - USA

2. Synchronization of Ground Station Atomic Clocks Using Weighted Transportable Cesium Time Standards in the Space Shuttle NAVEX Experiment
   H. Schild
   DFVLR, Wessling, FRG

3. Time Scales and their Estimated Uncertainties in the Space Shuttle Experiment NAVEX
   S. Starker, E. Sappl
   DFVLR, Weissling - FRG

4. Loran-C GPS Hybridization in Time Keeping
   M. D'Amico
   Techniphone, Le Puy Sainte-Reparade - France

2 ND EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 16 - 17 - 18 MARCH 1988

Session TF-04 Number 1

TIMING DATA REDUCTION AND PROCESSING FOR THE NASA - GODDARD LASER TRACKING NETWORK

E.Detoma, A.Clinton, W.Hanrahan
P.Kushmeider, R.Price, P.Dachel

Bendix Field Engineering Corporation
Columbia, Md. - 21045 (U.S.A.)

SUMMARY

The NASA Crustal Dynamics Project has deployed Satellite Laser Tracking Stations and Very Long Baseline Interferometry Systems to measure global and regional crustal motions and Earth rotational parameters. The data obtained from these systems can be used to determine satellite orbital parameters, to monitor the Earth polar motion, the tidal parameters and the length of the baselines between VLBI or Laser ranging sites.

As a part of this effort, the Bendix Field Engineering Corporation (BFEC) is currently providing engineering and technical support in the operations, data collection and reduction for the NASA Crustal Dynamics project.

The accuracy of Laser ranging is critically dependent upon the high stability of the reference oscillator at the tracking stations, and upon precise synchronization between remote stations. Current requirements are to maintain synchronization between Laser Tracking Stations, and to reference them back to UTC(USNO), to within +/- 0.5 microseconds. However, with the improvement of Laser equipment and the demand for more accurate ranging, the need for even more precise timing and higher frequency stability is foreseen for the future.

The Precision Timing Section (PTS) at BFEC has been responsible for maintaining and reporting time positions for the NASA Goddard Laser Tracking Network (GLTN) for the past twelve years. Recognizing that conventional means of timekeeping using LORAN-C, HF and VLF radio transmissions could not practically achieve the required global synchronization for the GLTN, the PTS has used techniques such as Cesium portable clocks in the past and GPS timing receivers in the present, to provide synchronization to the Network.

This paper will present the techniques, methodology and procedures which are used to meet the aforementioned requirements, and to process and reduce the timing data.

New timing equipment and data reduction methods have been developed and are currently under study at BFEC, to increase our capabilities in the generation and distribution of precise time and frequency, in close cooperation with NASA and other U.S. Government agencies.

IMPACT OF TIMING ON RANGING PERFORMANCES

Precise Laser ranging to an Earth satellite imposes severe requirements for the reference time and frequency standard oscillator; in particular:

* the short term stability of the oscillator affects the ranging precision and ranging accuracy is critically dependent upon its frequency accuracy;

* time synchronization is very important and directly affects the error in the determination of the baseline length between ranging stations; in fact, the correlation between measurements taken at different sites with the satellite position is based on the time tags of the range measurements performed at each site.

The accuracy limitations to the two-way ranging of an artificial satellite due to instabilities of the reference oscillator at the tracking station are well known, and this problem has been extensively discussed in the literature [ref. 1].

The range measurement is commonly performed by transmitting a short pulse of light to the satellite, where corner cube reflectors reflect the signal back to the ground station; the station is able to measure the round-trip delay T, from which the distance D, travelled by the pulse, can be easily computed when the speed of propagation c is known:

$$D = cT \tag{1}$$

The time delay T is measured using as a reference the frequency of the station local oscillator, and any offset or instability in the frequency of this oscillator translates into an error on D. Since the frequency fluctuation Δf around the nominal frequency f results in a ΔT fluctuation over an arbitrary time interval T:

$$\frac{\Delta f}{f} = \frac{\Delta T}{T} \tag{2}$$

then the uncertainty Δr on the range measurement r (where $r = D/2$) can be written as:

$$\Delta r = \frac{cT}{2} \frac{\Delta f}{f} \tag{3}$$

The nominal frequency accuracy for commercial Cs beam frequency standards [see ref. 2] is:

Oscillator type	Accuracy
HP 5061B, opt. 004	$+/- 7 \cdot 10^{-12}$
HP 5061B	$+/- 1 \cdot 10^{-11}$

and, for round trip (propagation) times T between 10 and 100 ms, the effect of the absolute accuracy of the oscillator frequency can be neglected.

However, over short time intervals (in the order of the round trip propagation time), the oscillator is more accurate than pure: that is, the noise affecting the output signal is greater than the intrinsic accuracy of the oscillator.

The short term instability of the oscillator translates directly into a contribution to the observed scatter of the ranging data around a mean value (ranging noise): this has been estimated, using the Allan deviation $\sigma_y(\tau)$ as a measure of the frequency stability of the oscillator, and the results are presented in fig. 1 (ref. 3). We are considering, and testing, other types of oscillators, such as BVA XTAL oscillators or disciplined standards, in order to provide a stable frequency source for ranging applications.

But other requirements must be taken into account. To correlate ranging measurements taken at different stations, the data is time-tagged at each site with reference to the local time scale. Each station's time scale is referenced to a common time scale, that is UTC(USNO), by using different time transfer techniques.

Any error in time synchronization between two sites translates directly into an error in time-tagging the data, and this produces an error when correlating measurements taken at different sites. Ranging measurements are not affected directly by this error, but the ranging data processing for computing the baselines is affected.

The current GLTN synchronization requirements limit the time position error versus UTC(USNO) to less than +/- 0.5 microseconds, to achieve sub-centimeter accuracy in the computation of the baselines, but we are working to improve the present performance to meet more stringent requirements for the full network in the near future.

SYNCHRONIZATION

Two techniques have been used in the past and are currently used within the GLTN to precisely synchronize remote clocks:

* *Portable clock trips*, providing a time transfer accuracy estimated to be within 50 and 100 ns;

* *Synchronization via GPS* timing receivers, currently providing global time transfer to better than a few hundred nanoseconds, but with an expected accuracy between 20 and 100 ns when used in the "common view" mode. This mode is not currently used in the network because of operational constraints, but we are working to introduce it, to match the future requirements for the GLTN.

Older LORAN-C timing receivers, providing synchronization to within 1 or 10 microseconds, are being phased out from the GLTN stations; they do not provide the required accuracy, especially when used in remote areas where they have to rely on the skywave mode of propagation.

DATA COLLECTION AND PROCESSING

GPS and LORAN-C time position measurements are made daily at the GLTN stations. GPS data is taken once a day for each visible satellite; the raw time position reading is recorded during the 10-minute long satellite observation pass, and it is usually taken as the time position fit for the total pass computed at the end of the observation. LORAN-C data is collected once a day only at those GLTN stations not equipped with GPS timing receivers.

The measurements (see fig. 5) are then transmitted (via TWX or satellite links, where available) to the Bendix LTN (Laser Tracking Network) Communication Center, and are stored in the daily Laser Operations Report (LOR). In addition, the LOR holds information concerning station time steps, frequency changes, power outages, equipment problems, etc.

At BFEC, all the LTN timing data, i.e. LORAN-C and GPS, is processed for the determination of the time position of the individual clocks, using the PTS Automated Time Position (ATP) Program. The function of the ATP Program is:

* provide daily time position determination;

* calculate long term frequency offsets for the clocks;

* predict when the station time position will exceed 0.5 microseconds [for the present requirements] with reference to UTC(USNO);

* evaluate the validity of the data and data analysis methods.

The system is coded in FORTRAN and runs on the Bendix VAX computer system.

Least squares fitting is the mathematical method used to determine the time position of the station timing system. Both observed and computed time positions are provided as the output of the data analysis program. Normally, a 92-day long time frame provides the working database length required to determine the time position. Reported frequency changes during this period cause the program to break the analysis into smaller segments.

WHY FITTING THE TIMING DATA

The fitting of the timing data is performed to:

* reduce the measurement noise;

* estimate the relative parameters of the clocks involved (modelling);

* detect anomalies, in the measurement data and in the behaviour of the clocks;

* interpolate the timing data;

* predict the future relative position of the clocks.

REDUCTION OF THE MEASUREMENT NOISE BY LEAST SQUARES FITTING OF THE DATA

To reduce the noise associated with the measurements, the fitting process averages several measurements taken at different times using a specified model for the clock (eq. 4) in order to correlate the said measurements. As a consequence, the following implicit assumptions are made:

a) - the noise processes affecting the data are normally distributed, uncorrelated and characterized by zero mean and a finite, non zero standard deviation (see ref. 8);

b) - the clocks will not contribute any significant (non-white) noise of their own to the level of precision to which the uncertainty of the smoothed values of the measurement data is reduced;

c) - as a consequence of the assumption (b), we implicitly assume the use of the clock as an ideal "flywheel" over the averaging interval (to be able to interrelate measurements taken at different times), and we assume that the noise processes affecting the measurements are stationary (not time dependent). This is equivalent to stating that, over the averaging interval, the clock behaves like an "ideal" clock, that is no systematic change in the clock parameters (phase

offset, phase drift and frequency drift) is allowed, to avoid the introduction of a polarization (or bias) in the estimate.

Since we know that the clocks are affected by systematics and by noise processes that do not average out (especially over long time intervals), then the reduction of the measurement noise through the fitting process shall only be carried on at the level at which the contribution of the clock noise to the total uncertainty can be neglected. *This means selecting the length of the data set (span) in order to have the correct amount and length of data to meet the above condition*.

Since the data is taken in the time domain and the fitting process behaves like a low pass filter in the frequency domain, then the equivalent time constant of the fitting procedure is directly related to the span (in time) of the data base and the number of points within the selected span (assuming that all the data points have equal weight).

If a moving fit is computed to limit the increase in the time constant as new measurements are taken, and N is the number of data points taken over a time span S, then the response time of the output of the estimator to a change in the input (the last measurement taken) is a function of S and N: for a fixed N, S determines the rate of acquisition of new data, and each data point has a weight 1/N in the fitting process. Then, it is evident that *the choice of S and N affects the response time of the estimator in reacting to a transient condition in the input*.

Selection of the parameters for the fitting process: *The selection of the number of data points and the length of the data span depends on the measurement noise*. Since, over extended periods of time, the clocks cannot be considered any longer as an "ideal flywheel" to average the measurement data, and the clock instability takes over the measurement noise, then the selection of the length of the data span S and the number of points N used in the fit is very important to obtain the best results.

MODELLING THROUGH THE FITTING PROCESS

A discussion on modelling of standard clocks can be found in ref. 8 and ref. 3, sect. 5.1. The typical model used to characterize the behaviour of a standard oscillator in the time domain, against some reference time scale, is:

$$\Delta T(t) = \Delta T(t_0) + \dot{\Delta T}\,(t - t_0) + \frac{1}{2}\,\ddot{\Delta T}\,(t - t_0)^2 + e(t) \qquad (4)$$

where: $\Delta T(t)$ is the time (phase) error of the oscillator relative to the reference standard at the time t;
$\dot{\Delta T}$ is the relative time (phase) drift;
$\ddot{\Delta T}$ is the relative aging;
$e(t)$ is the random part.

For Cesium beam frequency standards a linear fit is usually applied. Since the random part e(t) of eq. (4) is not part of the linear model when the fit is computed, care should be exercised to apply the fit only at the level at which the contribution to the total uncertainty due to the instability of the clock can be neglected. This has been discussed before: if proper care is not taken, then the fitting process will translate the random part of eq. (4) into the estimation of the parameters of the model, i.e., the instability of the clock will *contribute* to the determination of the value of the phase offset, drift and aging.

The fitting process, when applied properly, will allow characterization of the relative clock performance (it is always in terms of one clock versus another clock) through an estimation of the clock parameters:

phase offset

phase drift ---> frequency offset

phase aging ---> frequency drift

DETECTION OF ANOMALIES

A remote clock, and in general any clock, is only as good as our capability to read it, i.e., to know its time position and its behaviour. Since the clock can be characterized by a model (eq. 4), and the parameters of the model can be estimated, then any significant change that is detected in the parameters when carrying on a new estimate is an indication of a departure from the previous behaviour of the clock, i.e., it is an indication of an anomaly affecting the clock in such a way that the model cannot be applied since the moment in which the anomaly occurs.

Since the model (4) implicitly contains the measurement process, the anomaly can either affect the measurements only (i.e., the clock is unaffected) or it may affect the phase and/or frequency of the clock. We can classify the anomalies in three main categories:

a) - a sudden change in the relative phase of the clock (phase or time jump);

b) - a gradual or sudden change in the relative frequency offset of the clock;

c) - outliers, that usually are bad data points produced by random phenomena (noise spikes, etc.) affecting the measure of a given physical quantity at the moment at which the measurement is taken. When these values do not reflect the true value of the quantity being measured, they are usually discarded from the data set.

If the clock must be maintained to a very strict tolerance as far as the knowledge of its time position is concerned, then the data collected at the sites shall be returned in the LORs as soon as possible, and analyzed in real time.

In fact, at the moment in which an anomaly occurs and it is detected, all the previous knowledge that we had of the clock through its model (4) is lost (not even its time position is saved, since the data taken can be an outlier and the uncertainty of the measurement cannot be assessed from a single measurement), and the clock behaviour is completely unknown until it is again characterized by a new model or by an estimate of the new parameters based on new data.

Since the detection of an anomaly is a critical period in maintaining a time scale, the time during which the knowledge is lost shall be kept to a minimum, by detecting the anomaly as soon as possible and then acquiring as much data as possible to build a new model for the clock as fast as we can. In this way, outliers can be detected quickly, new measurements can be performed to replace those rejected and, in the occurrence of jumps in the phase or changes in the frequency of the clock, corrective actions can be taken rapidly.

INTERPOLATION OF THE DATA

The fitted model can be used to estimate the time position of the clock at any time within two adjacent measurements, by using eq. (4) instead of performing an interpolation between adjacent points. As a consequence, any point within the data span S can be interpolated using eq. (4), with the benefit of reducing the number of data required to estimate the time position of the clock (from the full measurement data set to the few coefficients of the model).

If a linear model (as in the case of a Cesium clock) is used, of the type:

$$\Delta T(t) = c_0 + c_1 t \qquad (5)$$

where: ΔT is the relative time position between the two clocks at the time t;
c_0 is the phase offset, and
c_1 is the phase drift;

then the estimate of ΔT at an arbitrary time t' (*within the data range* S over which the fit was computed), can be written as (ref. 5):

$$\Delta T(t') = \overline{\Delta T} + c_1 (t' - \bar{t}) +/- E(t') \qquad (6)$$

where: $\overline{\Delta T}$ and $\bar{t}$ are the average values (arithmetic mean values) of the ΔT_i and t_i ;

$E(t')$ is the uncertainty of the estimate at the time t':

$$E(t') = t_{1-a/2} \; s_T \left[\frac{1}{N} + \frac{(t' - \bar{t})^2}{S_t}\right]^{0.5} \qquad (7) \quad \text{(ref. 5)}$$

where: $t_{1-a/2}$ is the value of the Student t-distribution for N-2 degrees of freedom and (1-a) confidence interval (usually we use a 95% confidence in the estimate of the error);

s_T is the standard deviation of the fit;

N is the number of data points used in the fit;

S_t is defined as:

$$S_t = \sum_{i=1}^{N} t_i^2 - \frac{\left[\sum_{i=1}^{N} t_i\right]^2}{N} \qquad (8)$$

PREDICTION OF FUTURE DATA POINTS

Once the parameters of the clock have been determined at the time t, the model (5) allows the prediction of the clock position at a *future point* t'; eq. (6) still holds, but now E(t') shall be computed (ref. 5) as:

$$E(t') = t_{1-a/2} \; s_T \left[1 + \frac{1}{N} + \frac{(t' - \bar{t})^2}{S_t}\right]^{0.5} \qquad (9)$$

The prediction can be used to estimate the validity of a new measurement based on the past data.

FILTERING OF THE MEASUREMENT DATA

Filters can be applied to the data to identify and reject outliers. A simple technique that has been implemented in our programs uses the following filtering algorithm:

a) - apply a least squares fit to the data;

b) - compute the standard deviation for the fit;

c) - reject all the data points whose residuals are greater than 1, 2 or 3 times the standard deviation of the fit, according to the level of confidence that has been selected for the filter;

d) - if no points are rejected, exit the filter;
if one or more points are rejected, then go back to step (a) and re-apply the fit to the remaining data.

A more sophisticated filter can be devised, to correct also for the number of data points that are actually used in the fit. In this case, the algorithm will be:

a) - apply a least squares fit to the data;

b) - compute the standard deviation of the fit; if N data points are used at this step of the iteration, then compute the value F(t) of the Student t-distribution for the number of degrees of freedom and the selected confidence level. If the degree of the polynomial is n, then the number of degrees of freedom m will be: $m = N - n - 1$;

c) - compute a tolerance factor T, as the product of the standard deviation of the fit and F(t);

d) - reject all data points whose residuals are greater than T;

e) - if no points are rejected, exit the filter;
if one or more points are rejected, then go back to step (a) and re-apply the fit to the remaining data.

This second filter has the advantage of maintaining constant the associated level of confidence as the data set size is reduced by the effect of filtering data points.

ANALYSIS OF THE VARIANCE VERSUS TIME

This procedure can be used to assess the behaviour of the data in relation to their statistical properties, so that systematic phenomena affecting the data can be detected and identified by a sudden change of the mean values of the statistical parameters of the fit (i.e., a sudden increase in the variance and standard deviation, or a decrease of the correlation coefficient).

The analysis of the statistical parameters (variance and standard deviation of the fit, variance of the coefficients and correlation coefficient) versus time is carried on using the following algorithm:

a) - given a data set consisting of N points, choose a suitable size of n contiguous data point, where $n < N$;

b) - compute a moving fit over the full data set (see fig. 2) using n

points and the following technique:

b.1 - select the first n data points [j = 1]:

$$[x_1, x_2, \ldots, x_n];$$

b.2 - fit the subset defined in (b.1) and compute the variance s^2, the standard deviation, the variances of the coefficients and the coefficient of correlation r^2; save these values in vectors, as the j-th element of each vector;

b.3 - shift the subset over 1 data point, and use the new subset; for j=2 it will be:

$$[x_2, x_3, \ldots, x_{n+1}];$$

As a general rule, for each step j the subset will be:

$$[x_j, x_{j+1}, \ldots, x_{n+j-1}]$$

for n < N this will result in j = 1, 2, ..., N-n+1.

b.4 - go to step (b.2) and loop until j = N-n+1. Then go to step (c);

c) - display/plot the data.

An example of this technique is presented in fig. 2. The values of s_i^2 are computed using the algorithm shown above, and each s_i^2 is plotted at the time t_{n+j-1} [this is because at the time t_j we obtain a new data point and at that time we are able to compute a new estimate of s_j^2]. Changes in the frequency offset of the clock are shown as peaks in the plot of $s_j^2(t)$. Because of the way the $s_j^2(t)$ are plotted, there is a delay of about n/2 data points (for equispaced measurements) between the event (change in the frequency) and the corresponding peak: this is only a consequence of the way the $s_j^2(t)$ are referenced to a t_j in the plot.

A visual analysis of the current data set (fig. 2) reveals immediately the major frequency change occurring near t = 857, but two smaller

frequency changes are not noticeable; in the plot of $s_j^2(t)$, the two smaller peaks mark the two minor changes in the relative frequency of the clock (see also fig. 4, where the actual rates are computed and displayed).

By varying n (the size of the subset) it is possible to investigate the behaviour of the statistical properties of the data in relation to the size of the sample taken (fig. 3). The average values of the statistical parameters (the variance of the slope in fig. 3) are computed for various n, and then plotted as a function of n. This information may be used to determine the optimum size of the data set to be used in the fitting process.

Since the uncertainty W in the estimate of the generic parameter w is inversely proportional to the square root of the number of data points used:

$$W \sim s_w [\, n \,]^{-0.5} \quad ,$$

where s_w is the standard deviation for the baricenter of the data set (see eq. 7 or refs. 5 and 7), then the quantity:

$$s_w^2 / n$$

can be used as an indicator of the improvement in the uncertainty of the estimate as a function of the number of data points used in the fit.

REFERENCES

1. V.S.Reinhardt, FREQUENCY STABILITY MEASUREMENTS FOR TWO-WAY RANGE RATE TRACKING, NASA/GSFC Document X-814-75-315 (December 1975)

2. Hewlett Packard, 1987 PRODUCT CATALOG, pp. 338-340

3. E.Detoma, FINAL REPORT FOR THE LASER TRACKING NETWORK: TIME AND FREQUENCY STUDY, Bendix Field Engineering Corporation Report, NASA Contract NAS5-28027 (October 1987)

4. A.Clinton, E.Detoma, W.Hanrahan, P.Kushmeider, TIMING OPERATIONS AND DATA PROCESSING IN THE GODDARD LASER TRACKING NETWORK, Proceedings of the 19th PTTI Meeting, NASA/DoD, Redondo Beach, Ca. (December 1-3, 1987)

5. M.G.Natrella, EXPERIMENTAL STATISTICS, NBS Handbook 91 (August 1963)

   CRC, STANDARD MATHEMATICAL TABLES (14th edition, 1965)

6. E.Detoma, LINEFIT.BAS: A COMPUTER PROGRAM FOR THE LINEAR LEAST SQUARES DATA ANALYSIS, Bendix Field Engineering Corporation, Internal Report, (October 1987)

7. C.Daniel, F.Wood, FITTING EQUATIONS TO DATA, John Wiley & Sons (1980)

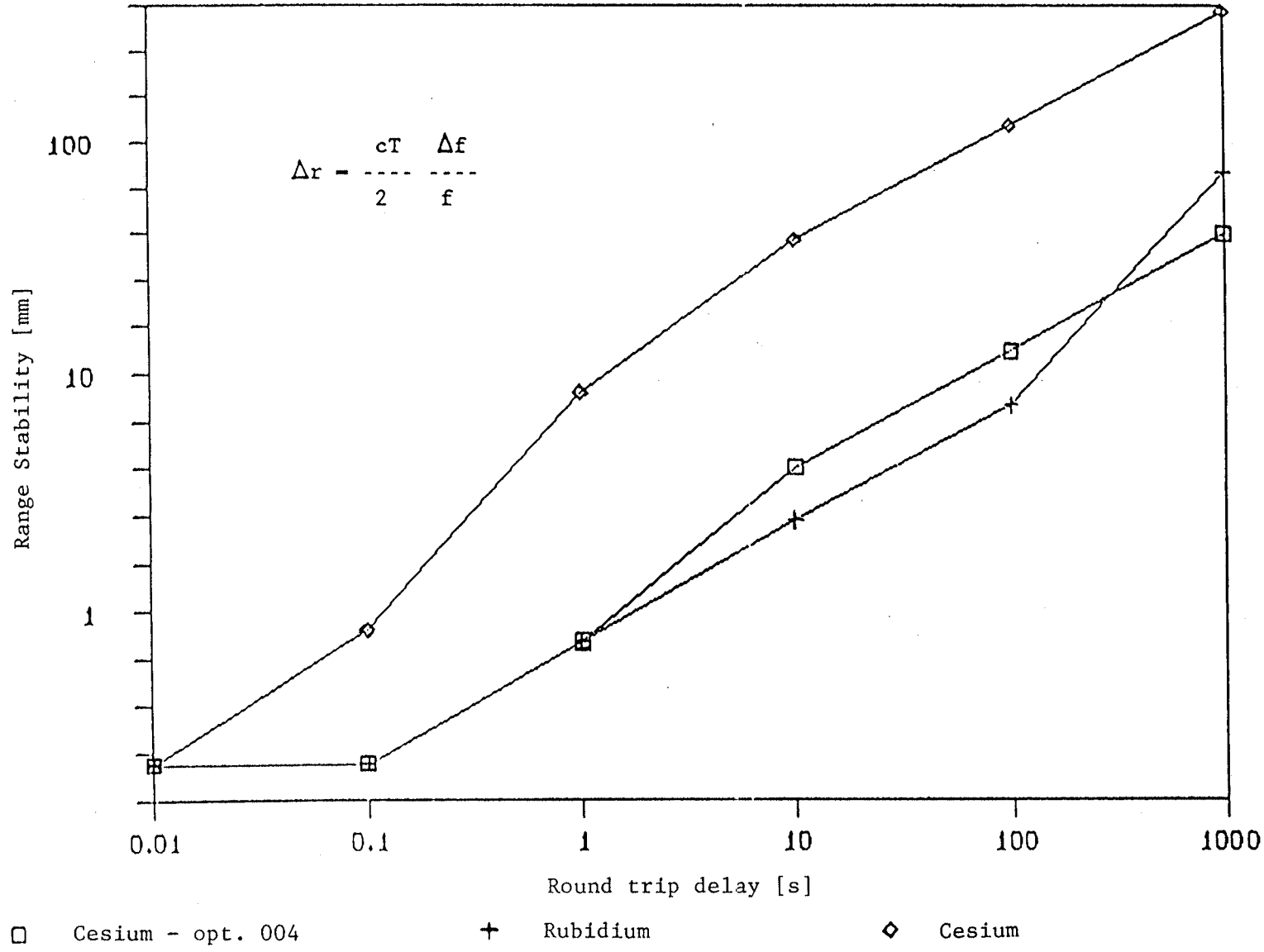
8. J.A.Barnes, THE MEASUREMENT OF LINEAR FREQUENCY DRIFT IN OSCILLATORS, in "NBS Seminar on Frequency Stability and its Measurement", (July 1984)

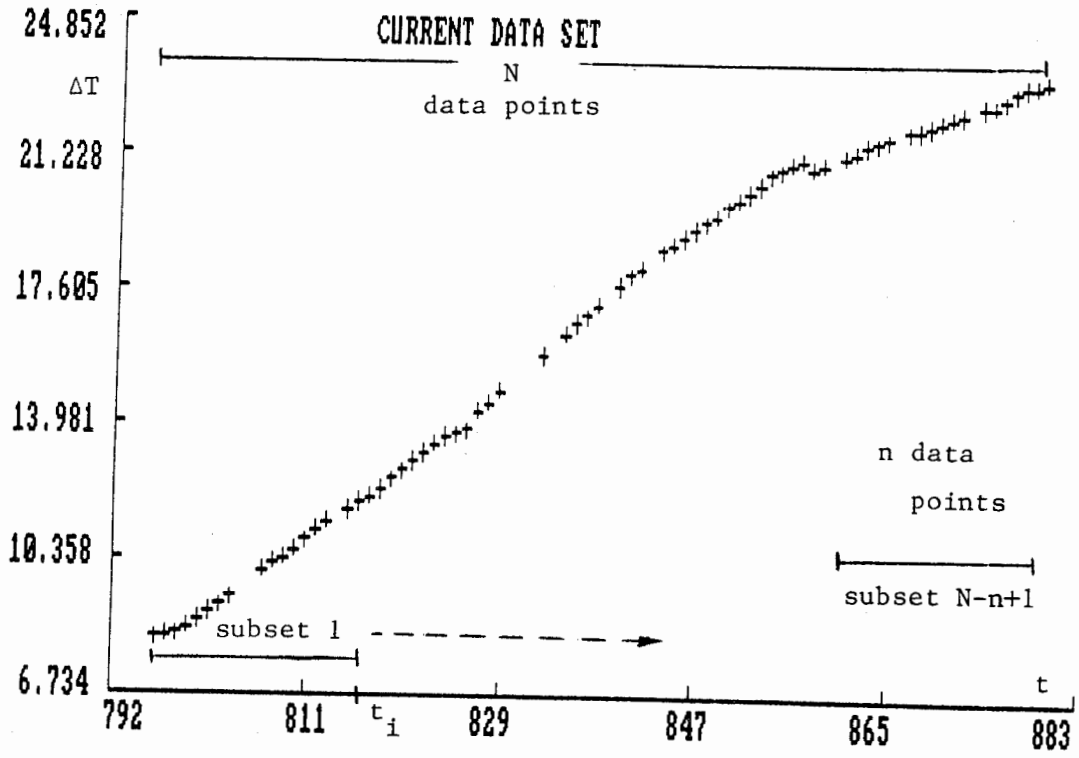
9. J.A.Barnes, DATA ANALYSIS AND CONFIDENCE LIMITS, in "NBS Seminar on Frequency Stability and its Measurement", (July 1984)

Figure 1

RANGE STABILITY

versus CLOCK TYPE





From the subset i [$1 \leq i \leq N-n+1$]

σ_i
s_i^2
....
r_i^2

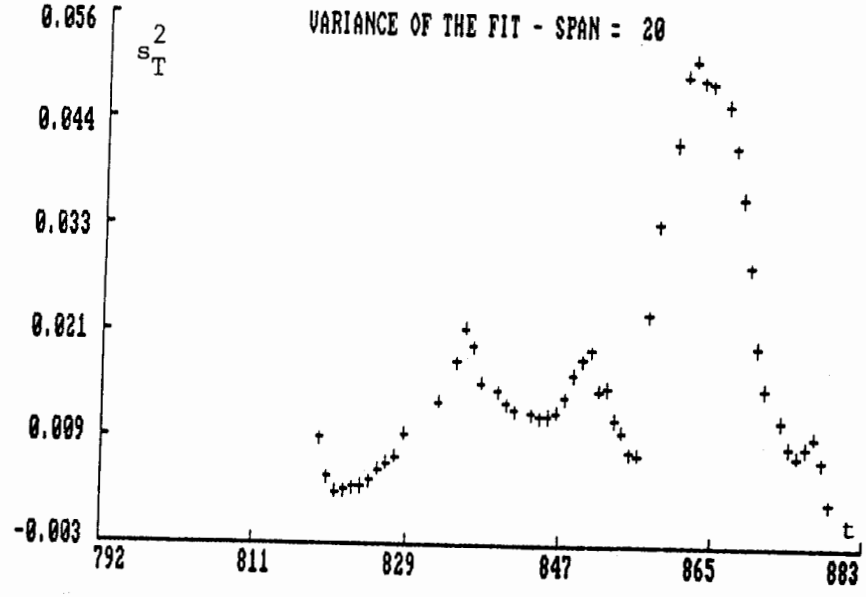
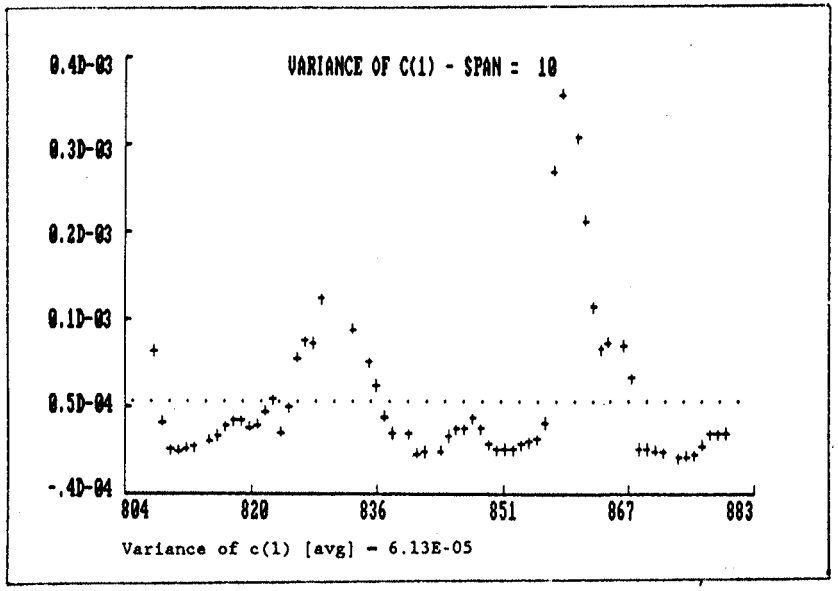
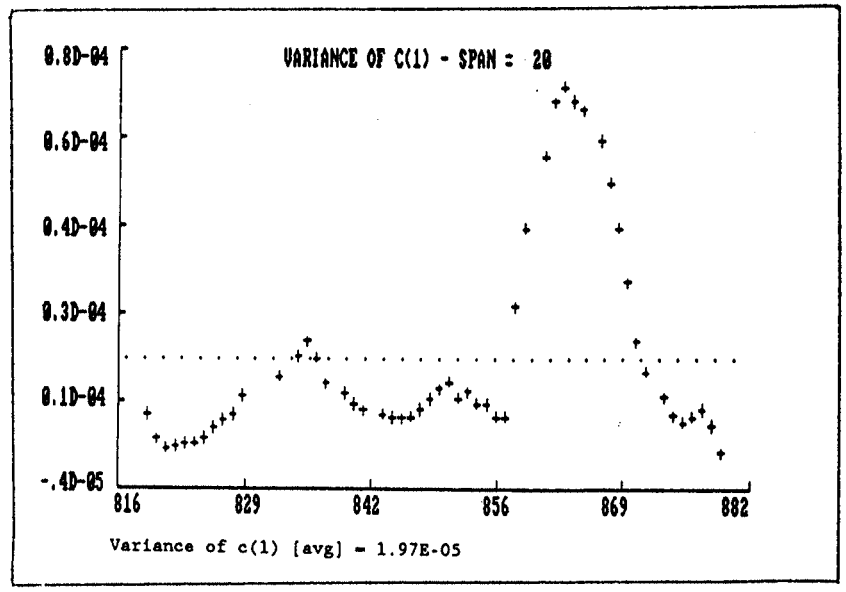


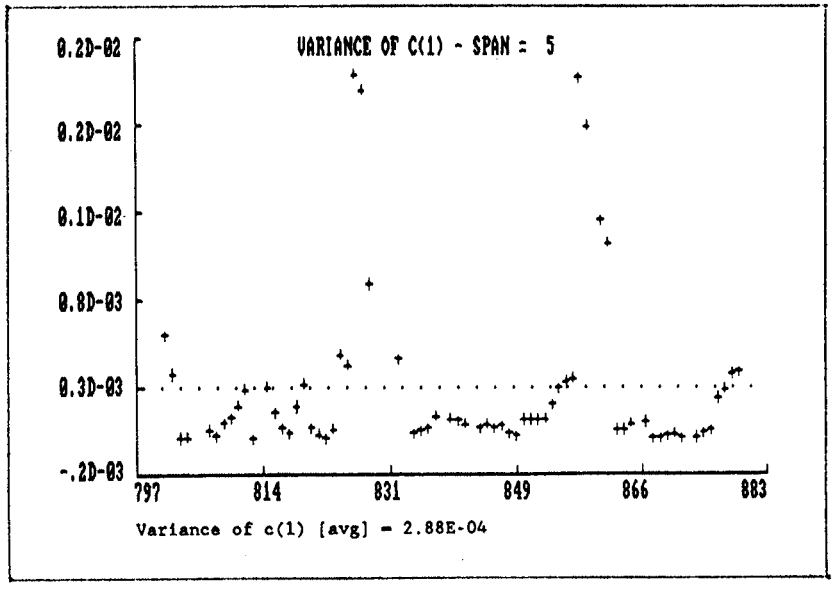
Fig. 2 - PLOT OF THE VARIANCE

plot units:

* vertical scale: microseconds
* horizontal scale: days







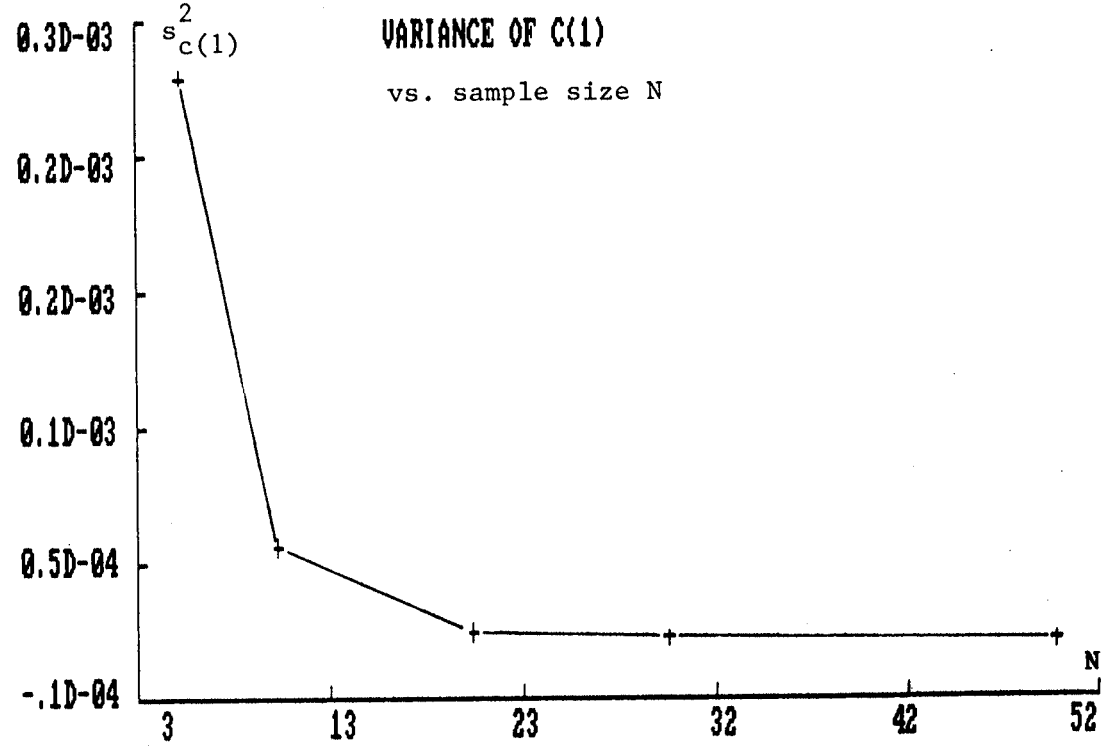
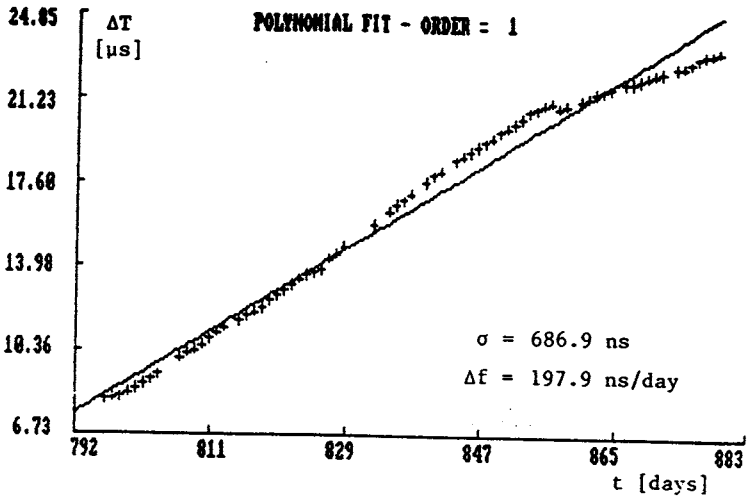
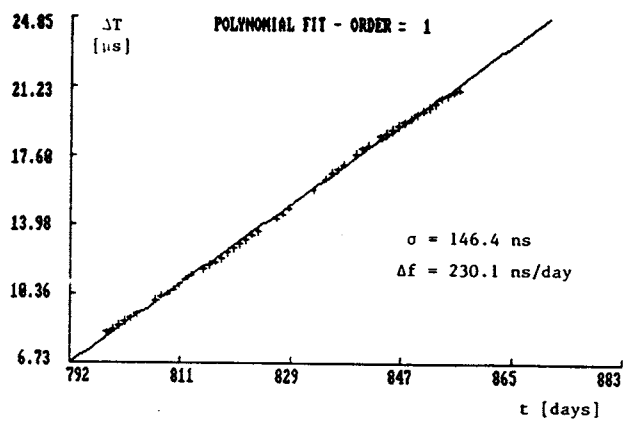
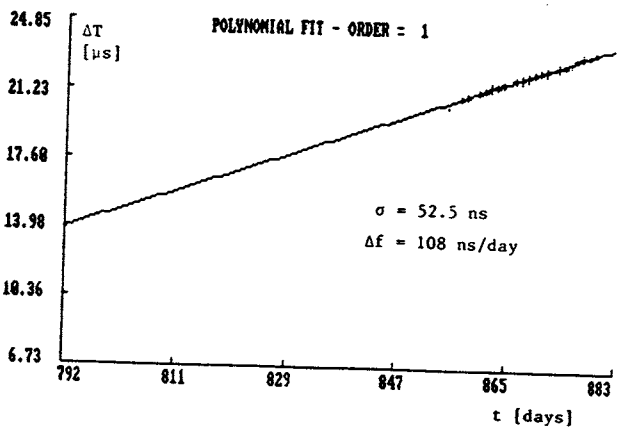
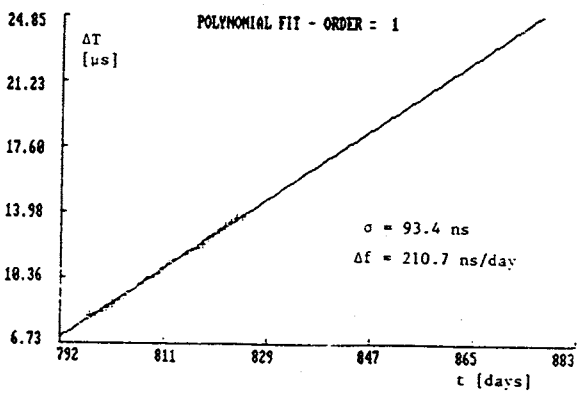


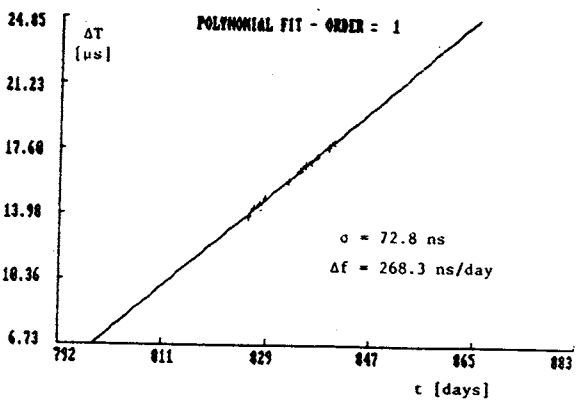
Fig. 3 - CHARACTERIZATION OF VARIANCE VS. SAMPLE SIZE.











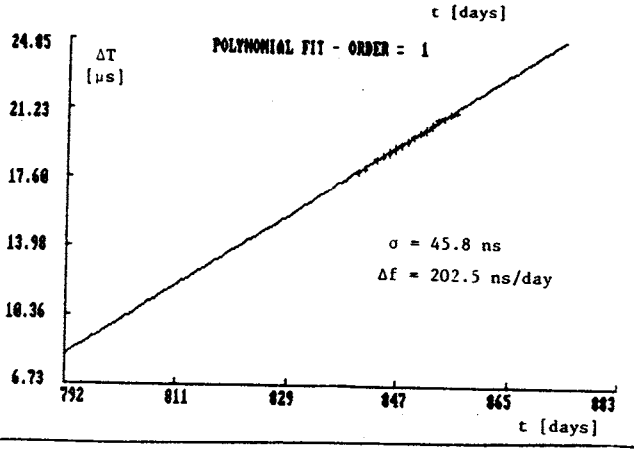


Fig. 4 - DATA ANALYSIS:

AN EXAMPLE

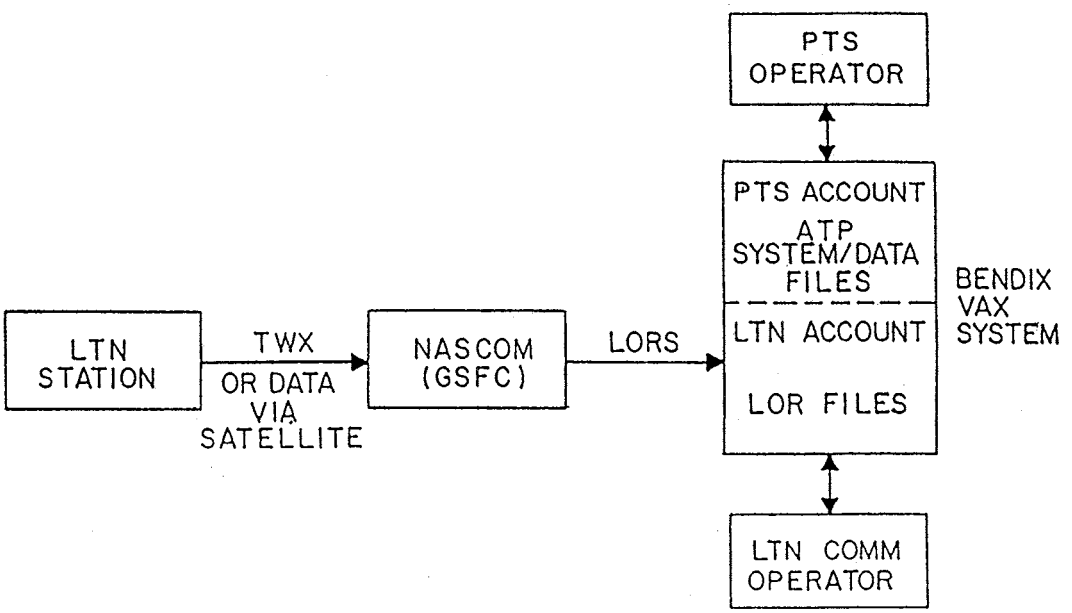


Figure 5 - DATA FLOW WITHIN THE GODDARD LASER TRACKING NETWORK

Session TF-04 Number 2

Synchronization of Ground Station Atomic Clocks Using Weighted Transportable Cesium Time Standards in the Space Shuttle NAVEX Experiment.

Herbert Schild
DFVLR, NE-HF/13
D-8031 Weßling, West Germany
Telephone 08153/28388

Summary

In the navigation experiment called NAVEX and performed during the spacelab mission D1 in autumn 1985 time signals were transmitted from on-board the spacelab to three German ground stations located in Oberpfaffenhofen near Munich, in Braunschweig and Stuttgart. These stations were equipped with a cesium time standard each, which enabled the exact time of arrival of the time signals to be recorded. Some goals of the NAVEX experiment required the ground station clocks to be synchronized.

The paper describes two two-way comparisons involving transportation of clocks to enable synchronization. On these clock travels three transportable cesium clocks were employed enabling averaging. Averaging was performed by using weighting factors for each clock. This paper shows the method of deriving the weighting factors from the clock rate at rest as well as from the clock rate when the clocks were in motion. UTC(PTB) was used as a reference time scale for converting the reading differences between the ground station clocks on the one hand and the transportable clocks on the other hand into reading differences between the ground station clocks and the UTC(PTB).

Introduction

One of the more than 70 experiments performed during the spacelab mission D1 in autumn 1985 was the navigation experiment NAVEX. Goal of this experiment has been to prove different future satellite borne navigation and time dissemination techniques as well as to test special equipment developped for them. In this experiment time marks derived from an atomic clock on-board the spacelab were transmitted to three ground stations located in Braunschweig (G1), Stuttgart (G2) and Oberpfaffenhofen (G3) near Munich. These three stations, the master station of which was G3, have

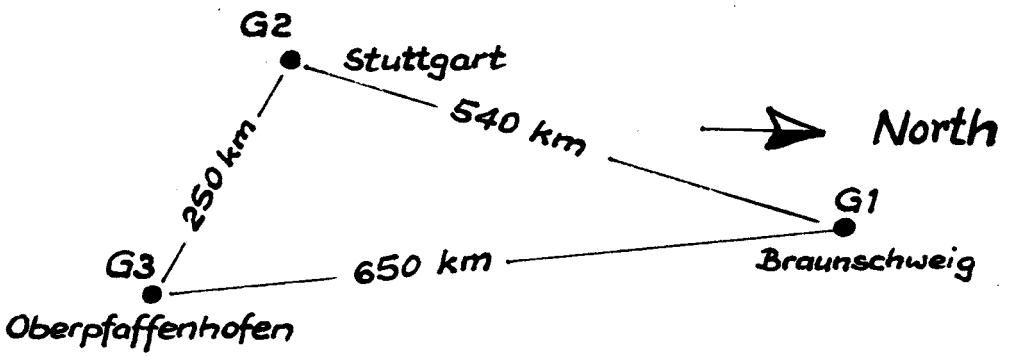


Figure 1. Site plan of the NAVEX ground stations

been equipped with a cesium time standard each, C1, C2, and C3, by means of which the epochs of the arrival of the time marks were measured. Besides the data were stored. The time marks have been one information in a pseudo noise coded L band signal. Moreover in the master station G3 the received time marks were transmitted back on a L band uplink to the shuttle, where the epochs of the arriving signal were measured by means of the on-board cesium clock. The data were stored, too.

One of the measuring techniques under test has been a method of determining the position of the shuttle and another one of determining the position of any receiving station participating in this experiment. Both techniques required all ground stations to be synchronized, in other words, the reading differences between the three ground station cesium standards had to be known. By means of two clock transportations, the first of which was performed in the week before and the second one in the week after the D1 mission, the fitting clock comparisons at each ground station were made to meet these requirements.

Definitions

At first some definitions are given. A reading difference is represented by the expression

$$T[\text{clock1} - \text{clock2}] = \text{a numerical value.}$$

T, in our calculations in microseconds with three decimal places behind the point, represents the time interval from the positive-going 1 pps of the clock 1 to the same pulse of the clock 2.

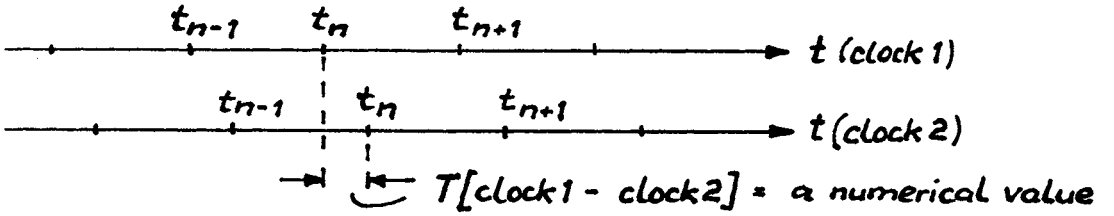


Figure 2. Reading difference shown as the difference of the 1 pps signals of two clocks.

This time interval can be measured with a high precision time interval counter, which is started with the impuls No. n of the clock 1 and stopped with the impuls No. n of the clock 2. The figure in the display of the counter is valued positive if measured in this way. The epochs of our measurements have been taken and rounded up to the full minute (in UT). The variation of the reading difference with time is called reading difference rate or in a word rate. If we know both these terms of two clocks, we can coordinate all events which have been timed with one of these clocks to the time scale of the other clock, assuming a linear rate. If one clock is far distant to the other, a third clock - a transportable one - is used for clock transportations between the clocks in question. By means of a multiple clock transportation back and forth and back again all these parameters mentioned above can be measured. Therefore all events having happened during this long tour can be coordinated to each of these three clocks. This is the principle of the two way clock transportation.

Provisions for the clock transportations

In order to be able to calculate the reading differences between the three ground station cesium clocks and their rates as precisely as possible three steps were taken.

1. Three transportable cesium standards were used, namely the clocks T1, T2, and T3.

2. The influences on the averaged results of the clock transportations by nonlinear rates on the way have been reduced by introducing different weighting factors for the transportable clocks. Therefore and because of a supplementary check the transportable clocks were compared additionally with each other at all ground stations.

3. The UTC(PTB) has been used as a reference time scale for the transportation results, i.e.: all measured reading differences have been related to the UTC(PTB). The additional measurements, which were necessary due to this step did not cause any special expenditure, since the ground station G1 was housed in that building of the Physikalisch-Technische Bundesanstalt in which the UTC(PTB) is created.

Consequently each two-way clock transportation had to begin and to end at the clock laboratories of the PTB. Here at the start of the transports all transportable clocks as well as the cesium clock C1 of the ground

station G1 were compared against the UTC(PTB) and after that C1 against the three transportable clocks. Then the cesium clock C2 in the ground station G2 and the cesium clock C3 in the master station G3 were visited in turns by car. In conclusion the car returned to the ground station G1. At all three ground stations measurements of reading differences of the corresponding station clock against the transportable clocks T1, T2 and T3 were performed. Finally at the station G1 the transportable clocks and the station clock were compared additionally against the UTC(PTB). Each measurement was taken down on record.

In the sequel the first clock transportation will be discussed in detail and later on the essential results of the second transportation will be completed, because the method of weighting in principle is equal in both transports. Moreover the clock C1 of the ground station G1 won't be mentioned any more, because C1 was always compared immediately against the reference time scale UTC(PTB) and because these measurements were in no case influenced by the weighted transportable clocks. The measurements mentioned above between C1 and the transportable clocks had only redundant quality for checking.

Results of the clock transportations

The main goal is to obtain precise reading differences of the UTC(PTB) against the station clocks C2 and C3. Therefore after the first transportation the linear rates of the transportable clocks compared with the UTC(PTB) during the transportation were calculated first. The results are shown in the following table 1.

Clocks under comparison	Reading differences in µs before the transport	Reading differences in µs after the transport	Duration of the transport	Rate $\Delta T/\Delta t$ in ns/d
T[UTC(PTB)-T1] Date	408.249 22.10.85/8:20 UT	408.168 24.10.85/15:42 UT	2.306945 d	- 35.11
T[UTC(PTB)-T2] Date	224.179 22.10.85/8:22 UT	223.986 24.10.85/15:42 UT	2.305556 d	- 83.71
T[UTC(PTB)-T3] Date	93.627 22.10.85/8:21 UT	93.550 24.10.85/15:42 UT	2.306250 d	- 33.39

Table 1

The rate listed in the last column is the quotient ΔT divided by Δt, where ΔT is the difference of the reading differences of the columns 3 and 2 and Δt the duration of the travel in days.

The reading differences measured with the transportable clocks in the ground stations G2 and G3 against the station clocks C2 and C3 are listed in the tables 2 and 3 in the first line. In the second line we find the reading differences against the UTC(PTB) which have been calculated for the same moments by means of the rates of the table 1. Finally in the third line we see the required reading differences of the station clock C2 and C3, respectively, against the UTC(PTB). We find out, that the three

Nº		Clock comparisons and calculations for the station G2 by means of transportable clock T1		T2		T3	
1	comparison on 22.10.85 at 15:19 UT	T[C2 - T1]	314.747 µs	T[C2 - T2]	130.670 µs	T[C2 - T3]	0.094 µs
2	calculated by measured rate for 22.10.85, 15:19 UT	T[UTC(PTB) - T1]	408.239 µs	T[UTC(PTB) - T2]	224.155 µs	T[UTC(PTB) - T3]	93.617 µs
3	difference 2 - 1	T[UTC(PTB) - C2]	93.492 µs	T[UTC(PTB) - C2]	93.485 µs	T[UTC(PTB) - C2]	93.523 µs

Table 2

Nº		Clock comparisons and calculations for the station G3 by means of transportable clock T1		T2		T3	
1	comparison on 23.10.85 at 13:44 UT	T[C3 - T1]	383.604 µs	T[C3 - T2]	199.492 µs	T[C3 - T3]	68.953 µs
2	calculated by measured rate for 23.10.85, 13:44 UT	T[UTC(PTB) - T1]	408.206 µs	T[UTC(PTB) - T2]	224.076 µs	T[UTC(PTB) - T3]	93.586 µs
3	difference 2 - 1	T[UTC(PTB) - C3]	24.602 µs	T[UTC(PTB) - C3]	24.584 µs	T[UTC(PTB) - C3]	24.633 µs

Table 3

results differ from one another. When considering the ground station G2, we find a difference of 38 nanoseconds for the same term and this difference amounts even to 49 nanoseconds at the ground station G3.

We have to assume, that the rates of the transportable clocks have been more or less nonlinear on the travel. We can see it, too, when we look at the figure 3 showing us the different rates of the travelling clocks. We detect, that the total interval of the transport can be divided into two characteristic periods when considering the rates. The period 1 lasts from the beginning of the transport on October 22nd 1985, 8:10 UT till 21:17 UT. The period 2 lasts from October 22nd, 21:17 till October 24th, 1985, 15:42 UT, that is the end of the transport. Moreover we detect at the first glance, that the rate of the clock T3 presents obviously the greatest deviations, for in both graphs, which contain the rate of the clock T3, a significant peak appears at the same point. On the other hand the rate of the clock T1 against the clock T2 takes a fairly linear course. Therefore we can estimate the rate of the clock T3 against the UTC(PTB) in both periods by means of the rates of these last-named clocks T1 and T2 against the UTC(PTB), which we know from the table 1, using the following two statements, the numerical results of which we will average later on.

$$\frac{\Delta T[UTC(PTB) - T3]}{\Delta t} = \frac{\Delta T[UTC(PTB) - T2]}{\Delta t} - \frac{\Delta T[T3 - T2]}{\Delta t} ,$$

$$\frac{\Delta T[UTC(PTB) - T3]}{\Delta t} = \frac{\Delta T[UTC(PTB) - T1]}{\Delta t} - \frac{\Delta T[T3 - T1]}{\Delta t} .$$

But in these equations the rates of the clock T3 against the clock T2 as well as of the clock T3 against the clock T1 are required. They can be derived from reading differences measured during the transport. We used them already for the graphs.

Date	T[T3-T2]	Difference	$\frac{\Delta T[T3-T2]}{\Delta t}$	Date	T[T3-T1]	Difference	$\frac{\Delta T[T3-T1]}{\Delta t}$
22.10.85 8:10 UT	130.552 µs			22.10.85 8:22 UT	314.621 µs		
Period 1 =0.546528 d		50 ns	91.49 $\frac{ns}{d}$	Period 1 =0.541667 d		56 ns	103.38 $\frac{ns}{d}$
22.10.85 21:17 UT	130.602 µs			22.10.85 21:22 UT	314.677 µs		
Period 2 1.767361 d		-165 ns	-93.36 $\frac{ns}{d}$	Period 2 1.763889 d		-58 ns	-32.88 $\frac{ns}{d}$
24.10.85 15:42 UT	130.437 µs			24.10.85 15:42 UT	314.619 µs		

Table 4

In the table 4 we find all the values. In the column 2 there are the reading differences between the clocks T3 and T2 measured on the travel at the ground station G1, G3 and finally G1 again. At the station G3 the measurement was made at the moment, which shows the peak in the graph. In the column 4 there are the rates of the periods 1 and 2. For instance one

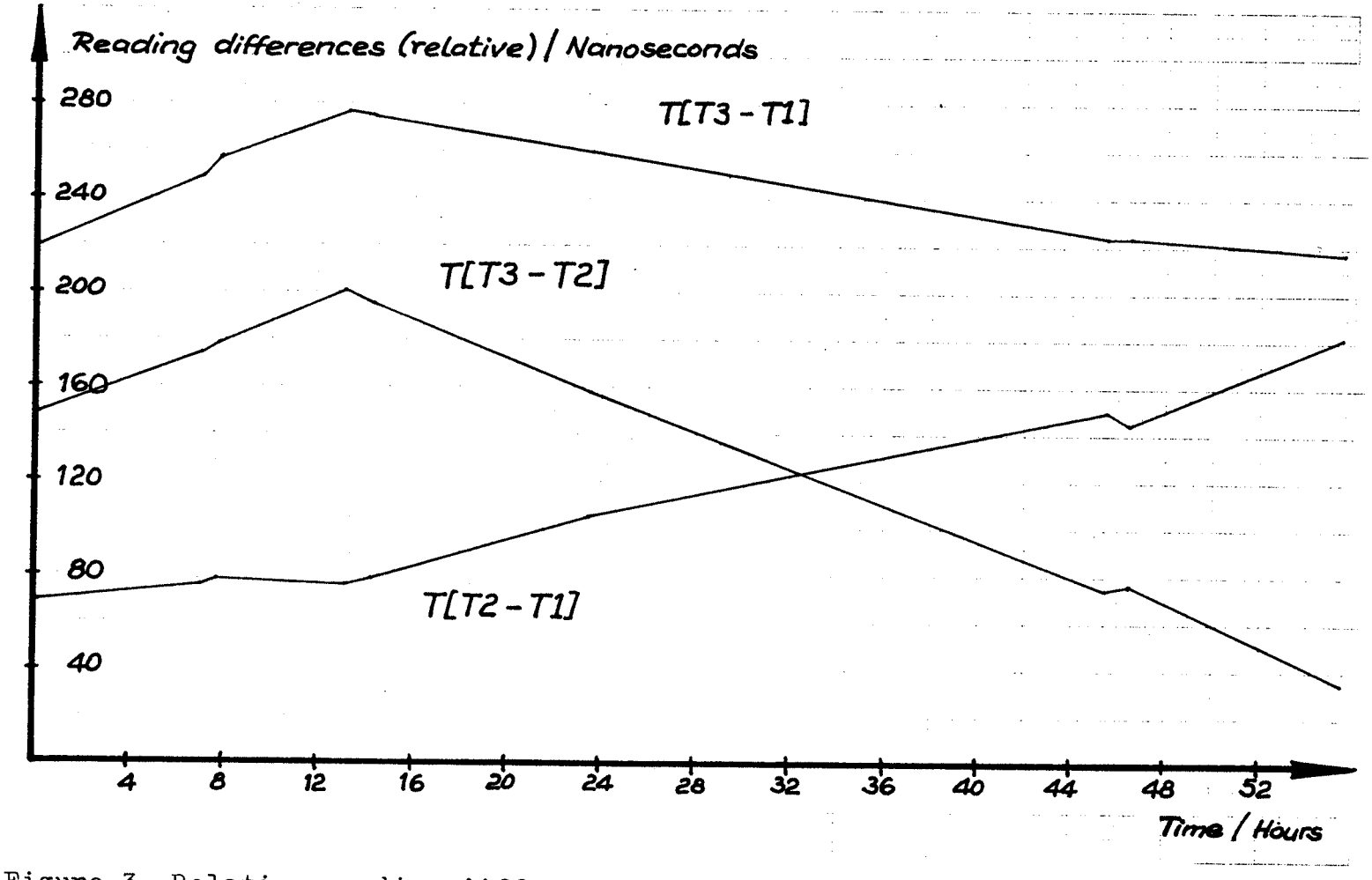


Figure 3. Relative reading differences of the transportable clocks against one another during the 1st clock transportation.

obtains the rate 91.49 ns/d as follows: The subtraction of 130.552 µs in the column 2 from 130.602 µs leads to 50 ns, which are divided by the duration of the period 1, 0.546528 days. In an analogous way with that all rates in this table are derived. Using the statements above with the corresponding rates of the tables 1 and 4 one finds the results in the table 5.

	Estimated rate $\frac{\Delta T[UTC(PTB)-T3]}{\Delta t}$		
	using transportable clock T1	using transportable clock T2	averaged
Period 1	$-138.49 \frac{ns}{d}$	$-175.2 \frac{ns}{d}$	$-144.56 \frac{ns}{d}$
Period 2	$-2.23 \frac{ns}{d}$	$+9.65 \frac{ns}{d}$	$-0.26 \frac{ns}{d}$

Table 5

These results were different, because we can use either the clock T1 or the clock T2 in the formulas above. This corresponds to the graph, in which the rate of the clock T2 against the clock T1 is not fully linear. By adequate averaging the result will be optimized. For this purpose we look at the averaged rates against the UTC(PTB) of both of these transportable clocks, which they had during the two weeks around the transport, when they were in the clock laboratory. On this occasion it must be mentioned, that two of the transportable clocks have been lent from the ensemble of the Physikalisch-Technische Bundesanstalt in Braunschweig. The rates in rest of these clocks have been measured continuously. The 3rd transportable clock was compared periodically against the UTC(PTB) in the DFVLR, so that its rate in rest was known, too. Now we compare all rates in rest with the rates against the UTC(PTB) the clocks had during the transport and we regard the value of the deviation as a figure of reliability.

	Rate in rest in $\frac{ns}{d}$	Rate on transport in $\frac{ns}{d}$	Deviation in $\frac{ns}{d}$	Weighting factor
$\frac{\Delta T[UTC(PTB)-T1]}{\Delta t}$	-30.6	-35.11	-4.5	5.04
$\frac{\Delta T[UTC(PTB)-T2]}{\Delta t}$	-61	-83.71	-22.7	1

Table 6

In the table 6 we find the numerical values. The rate of the clock T1 deviates from its rate in rest by -4.5 ns/d, that of the clock T2 by -22.7 ns/d. Because the clock T1 is five times better, we rate the weighting factor of this clock T1 for averaging at 5.04, of the clock T2 at 1. In the table 5 the last column shows the averaged numerical values of the rate of the clock T3 against the UTC(PTB).

Final weighting factors

Now the initial numerical values of all three transportable clocks are present, not only the rates in rest but also the deviations during the transport. We see them in the table 7.

	Period 1			Period 2		
	Rate in rest in $\frac{ns}{d}$	Deviation in $\frac{ns}{d}$	Weight. fact. α_n	Rate in rest in $\frac{ns}{d}$	Deviation in $\frac{ns}{d}$	Weight. fact. β_n
$\frac{\Delta T[UTC(PTB) - T1]}{\Delta t}$	-30.6	-4.5	25.76	-30.6	-4.5	6.31
$\frac{\Delta T[UTC(PTB) - T2]}{\Delta t}$	-61.0	-22.7	5.11	-61.0	-22.7	1.25
$\frac{\Delta T[UTC(PTB) - T3]}{\Delta t}$	-28.65	-115.91	1	-28.65	+28.39	1

Table 7

The numerical value of the rate in rest of the clock T3 in the last line comes out of a bulletin of the Physikalisch-Technische Bundesanstalt in Braunschweig. The deviation of -115.91 ns/d is the difference between this PTB figure and the averaged number of the table 5. As we have done just now we form the ratios of the deviations of the clocks T1 and T3 as well as of the clocks T2 and T3, for the clock T3 shows the greatest deviation, so we will find the following numbers. In the period 1 we see, that the clock T1 was 25.76 times better than the clock T3; consequently the clock T1 has been valued at 25.76. The clock T2 was 5.11 times better than the clock T3. So the clock T2 got the weighting factor 5.11. In the period 2 the corresponding numerical values are 6.31 and 1.25.

Calculation of the required final reading differences

By means of the weighting factors the reading differences listed in the tables 2 and 3 can be averaged. For calculating the value T[UTC(PTB) - C2], that is the reading difference of the UTC(PTB) and the station clock C2, we form the sum out of the three products weighting factor times reading difference and then we divide this sum by the sum of the weighting factors according to the formula

$$T[UTC(PTB) - C2] = \frac{\Sigma(\alpha_n \cdot R_{G2n})}{\Sigma \alpha_n},$$

where α_n is the weighting factor of the transportable clock T_n and R_{G2n} the reading difference of the ground station G2 obtained by the clock T_n.

For calculating the term T[UTC(PTB) - C3], that is the reading difference of the UTC(PTB) and the station clock C3, there are two ways of averaging. Because we have this term at the peak of the graph we can calculate it with the weighting factors of the period 1 as well as with those of the period 2, so to speak from behind. Because the period 2 is three times greater than the period 1 and consequently any deviation of the rate becomes three times greater, we can only take into account one third of the value of the period 2, when we form the averaged value. Now we use the formulas

$$T^*[UTC(PTB) - C3] = \frac{\Sigma(\alpha_n \cdot R_{G3n})}{\Sigma \alpha_n}$$

and

$$T^{**}[UTC(PTB) - C3] = \frac{\Sigma(\beta_n \cdot R_{G3n})}{\Sigma \beta_n}$$

The characters in the formulas are to be interpreted as above. The calculated values are listed in the table 8.

Reading difference	Numerical value	Date
T[UTC(PTB) - C2]	93.492 µs	22.10.85, 15:19 UT
T*[UTC(PTB) - C3]	24.6001 µs	
T**[UTC(PTB) - C3]	24.6029 µs	
T[UTC(PTB) - C3]	24.601 µs	23.10.85, 13:44 UT

Table 8

In the 2nd line we see the averaged numerical value of the reading difference of the clock C2 against the UTC(PTB). This value was valid for October 22nd, 1985, 15:19 UT in the ground station G2. In the 5th line finally we see the numerical value of the reading difference of the clock C3 against the UTC(PTB). This is the averaged value of both numbers in the lines 3 and 4. It was valid for October 23rd, 1985, 13:44 UT in the master station G3.

The 2nd clock transportation

In the 2nd clock transportation, which was performed soon after the D1 mission, in principle the same method of weighting was used. But this transport was to be divided into 3 periods, as we can see in the graphs of the figure 4, in which we find the reading differences of the transportable clocks. There the reading differences of the transportable clocks against one another as well as against the UTC(PTB) are shown. By means of

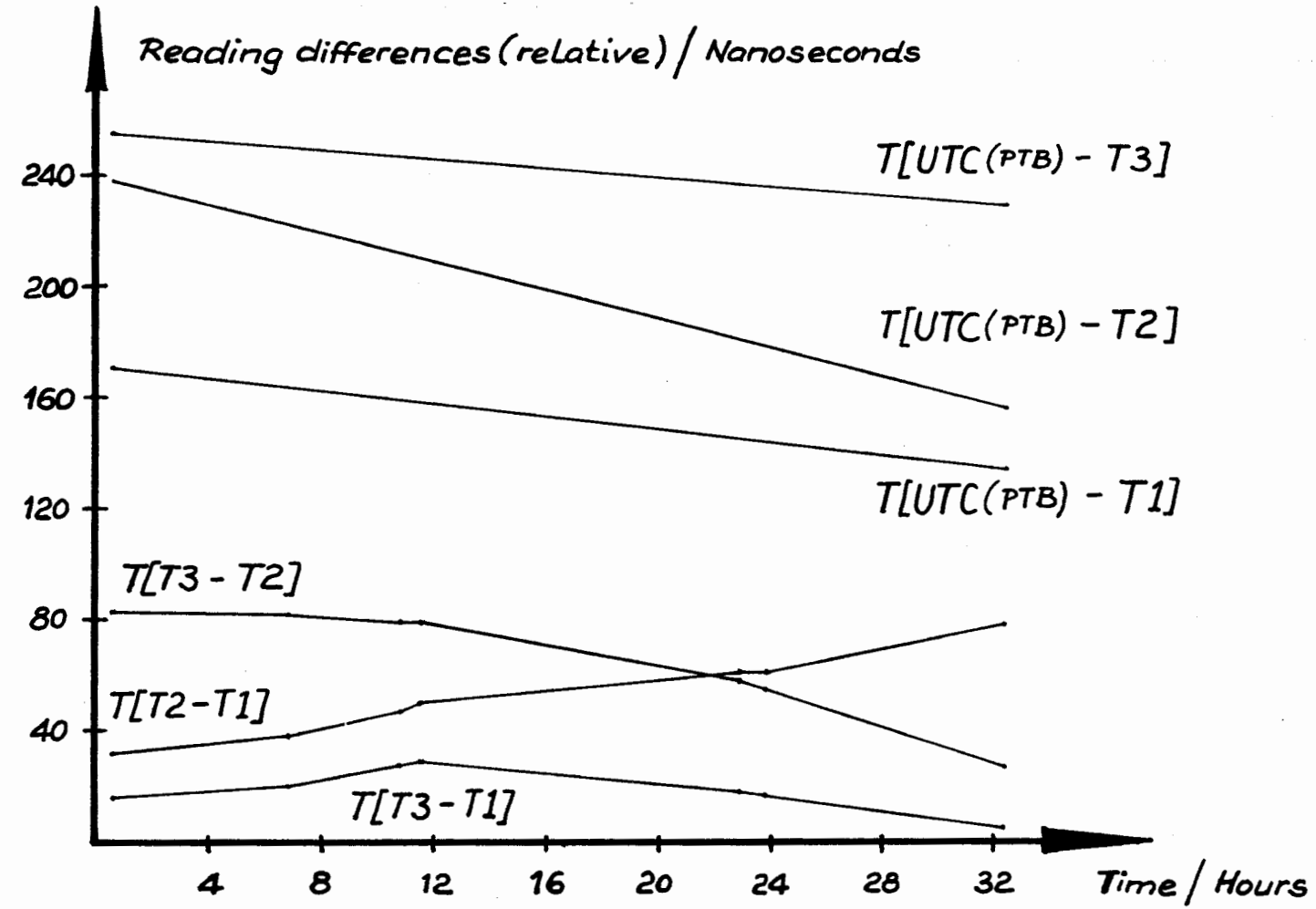


Figure 4. Relative reading differences of the transportable clocks against the UTC(PTB) as well as against one another during the 2nd clock transportation.

similar considerations as in the first clock transportation we obtain weighting factors for averaging the results of each transportable clock. These factors are listed in the table 9.

Transportable clock	Period 1	Period 2	Period 3
T1	5.72	1.49	5.2
T2	8.84	2.31	8.05
T3	1	1	1

Table 9

In this transport the clock T3 has got the least weighting factor again, but in contrast with the first transport the clock T2 obtained the highest weighting factor. The table 10 shows the results obtained with the single clocks as well as the averaged final result.

Transportable clock	T[UTC(PTB) - C2]	T[UTC(PTB) - C3]
T1	91.666 µs	23.543 µs
T2	91.663 µs	23.541 µs
T3	91.671 µs	23.554 µs
Averaged value	91.665 µs	23.543 µs
Date	12.11.85, 13:48 UT	13.11.85, 00:10 UT

Table 10

Review

In a final review we summarize. The reading differences of the NAVEX ground station clocks had to be gained with high accuracy. The common-view technique of GPS has not been applicable, because no GPS receiver was available at the station G2 in Stuttgart. Therefore the method of clock transportation remained only. High expenditure in magnetic shielding and in the stabilizing of the temperature as known from other clock transportations has not been applicable, because of the expensive special equipment. Instead of that three weighted transportable clocks have been used. For this weighting the rate behaviour of the clocks on transport has been considered important. Instead of over-all-statistics only the weeks around the transportation time have been taken in consideration. The question, how precisely the UTC(PTB) has been brought to the master station G3 at Oberpfaffenhofen can be answered with two common-view GPS results. In both transports the GPS reference clock C4 in Oberpfaffenhofen, which is standing only 500 meters off C3, was compared, too. In the 2nd column of the table 11 we see the numerical values of the reading differences of the clock C4 against the UTC(PTB) derived from the transport, in the 3rd column we find the same reading differences, derived from GPS common-view measurements and in the 4th column there are the differences of both results in nanoseconds.

Date	T[C4 - UTC(PTB)]		Difference
	transportable clocks	GPS common view	
23.10.85, 13:56 UT	114.678 µs	114.681 µs	- 3 ns
13.11.85, 00:10 UT	116.395 µs	116.397 µs	- 2 ns

Table 11

Session TF-04 Number 3

TIME SCALES AND THEIR ESTIMATED UNCERTAINTIES IN THE SPACE SHUTTLE EXPERIMENT NAVEX

Siegfried Starker, Dipl.-Ing.; Emil Sappl, Dr.-Ing.
Deutsche Forschungs- und Versuchsanstalt für Luft- und Raumfahrt (DFVLR)
Institut für Hochfrequenztechnik
D-8031 Wessling
Fed. Republic Germany

ABSTRACT

During the Space Shuttle Experiment NAVEX several methods of time transfer have been investigated with respect to future applications in satellite navigation systems. A main point was the investigation of a two-way microwave time transfer between a Cs-clock on-board the Shuttle and a reference clock at the ground control station. Pseudo-noise code signals with correlation receiver techniques and spread spectrum data transfer were applied. The measurements were performed with accuracies of $\sigma \approx \pm 5{,}8$ ns for the clock readings and $\sigma_y \approx \pm 0{,}16 \cdot 10^{-13}$ for frequency determination during 80 hours measurement time. The relativistic frequency shift of the flying Cs-clock was determined with an uncertainty of $\sigma_{\bar{y}} = 1{,}5 \cdot 10^{-13}$, which is equivalent to 0,5 % relative accuracy.

INTRODUCTION

To provide satellites with high accurate time scales will be a basic task in future time distribution and navigation systems. A well known application of such satellite borne timescales are one-way-ranging systems, as for instance GPS-NAVSTAR and GLONASS. In these systems the control of satellite clocks is performed by one way measurements only. In the Space Shuttle Experiment NAVEX [1] the additional use of two way measurements for a time transfer between on-board and ground based clocks was investigated. The advantage of this modification is that a very accurate time comparison between a single ground station and a satellite clock can be performed, eliminating the uncertainty of satellite position and ground station. Byond this the number of necessary ground stations for satellite position and time control can be reduced from 4 to 3.

THE NAVEX EXPERIMENT

Fig. 1 shows the configuration of the experiment NAVEX. The two-way clock comparison was performed between the Space Shuttle and the ground control station at DFVLR Oberpfaffenhofen using spread spectrum techniques. A pseudo noise code signal was transmitted to the ground at 1,53 GHz and retransmitted to the Shuttle at 1,43 GHz. On the ground and in the Shuttle the time intervals between time marks of the arriving signals and of the Cs-clocks there were measured. The on-board measurement result T_S was transmitted to the ground at the same carrier with a low bit rate of 250 bit/s. There the time difference between the on-board clock and the ground clock was determined from the measured values T_G and T_S as shown in Fig. 2.

Some correction terms had to be regarded because of unsymmetries of delay times in transmitters, receivers and in the atmosphere. The necessary information for these corrections was received from calibration and doppler measurements.

Parallel radar range measurements at 5,7 GHz delivered reference values for range determination. The reception of the Shuttle signals at two other sites, SEL Stuttgart and PTB Braunschweig, with well known positions and accurately synchronized clocks allowed the Shuttle position determination by a combined one way/two way method.

Thus the experiment configuration of NAVEX was a scenario of a future ground control segment for controlling time scales and positions of satellites with only three ground stations. One of the main goales of such a control segment is the generation and maintenance of the system time with best possible accuracy at all sites, especially on board of the satellite. The experiment NAVEX had to demonstrate, which accuracy could be achieved with the techniques and methods used.

THE CLOCKS AND THE SYNCHRONIZATION TASK

In Tab. I the clocks are summarized, which participated in the experiment. There was a Cs-clock at each site together with a back up clock. An additional NAVEX reference clock was housed in an air-conditioned cave at DFVLR. PTB supported the experiment with several clocks and with the very stable time scale of its clock ensemble.

The most important parameter of these clocks is their frequency stability, which is characterized commonly by the Allan - or pair - variance $\sigma_y(\tau)$. This frequency stability is a relative measure of the probable random deviation of the clock reading x from an estimated reading $\tilde{x}$ after a time interval τ.

(1) $$\sigma_{\tilde{x}\ \mathrm{clock}}(\tau) = \sigma_y(\tau) \cdot \tau \ ; \qquad \text{e.g.} \quad \sigma_{\tilde{x}}(1\ \mathrm{day}) \approx 10\ \mathrm{ns}$$
$$\text{with } \sigma_y(\tau) = 1 \cdot 10^{-13}, \quad \tau \approx 1\ \mathrm{day}\ .$$

The estimated value $\tilde{x}\ (\tau)$ depends on our knowledge of the clock reading $\tilde{x}\ (t_o)$ at the beginning of the time interval τ and on the knowledge of the mean rate $\tilde{y}$ relative to a reference clock

(2) $$\tilde{x}(\tau) = \tilde{x}(t_o) + \tilde{y} \cdot \tau .$$

This knowledge has to be gained from measurements, which are burdened with uncertainties, too. Thus the estimated value has an additional uncertainty

(3) $$\sigma'_{\tilde{x}}(\tau) = \sqrt{\sigma^2_{\tilde{x}o} + \sigma^2_{\tilde{y}} \cdot \tau^2}$$

and the total uncertainty between two time scales as a function of frequency stability ($\sigma_y(\tau)$,) distance between clock comparisons (τ) and measurement errors ($\sigma_{\tilde{x}o}$, $\sigma_{\tilde{y}}$) can be described as

(4) $$\sigma_{\tilde{x}}(\tau) = \sqrt{\sigma^2_{x\ \mathrm{clock}}(\tau) + \sigma^2_{\tilde{x}}(\tau)}\ .$$

This expression describes the uncertainty of a predicted time scale, τ seconds after the last comparison with a reference clock, as it will be necessary for instance in future navigation satellites.

METHODS OF TIME TRANSFER

The time transfer between the station clocks at the different sites in Germany was performed by four clock transportations and by parallel comparisons using the GPS-commen-view-technique between PTB-Braunschweig and DFVLR-Oberpfaffenhofen [2]. The results agreed within + 5 ns [3]. Frequency transfers between Cape Canaveral and Oberpfaffenhofen via GPS satellites were performed with a precision $\sigma_{\bar{y}} < 1 \cdot 10^{-13}$. On the results and experiences of these measurements a special report will follow. Additional measurements with GPS and with synchronous satellites using PN-code-signals before and after the D1-mission showed also rate uncertainties of $\sigma_{\bar{y}} < 1 \cdot 10^{-13}$.

The main problem in the experiment was the comparison of the Shuttle-borne clocks with the ground clocks at Oberpfaffenhofen by a two-way microwave link. A first report on the results of these measurements and a discussion of special effects was given in [4, 5]. In the following an optimum estimation of clock parameters and of their uncertainties will be given. For this the maximum likelyhood method will be applied. The uncertainty of these estimations will be described by estimated classical variances, which appear to be more adequate to the given problem than the more specialized Allan-variance.

FLYING CLOCK MEASUREMENTS

The main clock onboard was a Cs-FTS 4000 opt 004, in the following designated as Cs_F. The behaviour of this clock is regarded relative to the ground based NAVEX reference clock Cs-hp 5061A opt 004, designated in the following as Cs_R.

Time comparisons between both have been performed at several places in Germany in 1984 during the development- and integration-phase of the NAVEX experiment. The objective of these measurements was to get a good knowledge of the clock behaviour on the ground and to check the status of the clock from time to time. During 1985 the flying clocks were transported to Cape Canaveral and flown in the Shuttle mission D1. Before and immediately after the D1-mission clock comparisons between Cape Canaveral and Oberpfaffenhofen were performed via GPS-satellites, which gave an actual information on the clock rate just before and after the flight. During the Shuttle flight in November 1985 the time signals from the on-board clock Cs_F could be received five to six times per day at the ground control station at Oberpfaffenhofen, giving the opportunity for two-way-clock comparisons. In 1986 the onboard equipment came back to Germany and the onboard clock Cs_F was compared with the reference clock Cs_R in Oberpfaffenhofen directly without any transmission link.

In Table II the environmental conditions during measurements at the different sites are summarized. The influence of different temperatures on the rate y between Cs_F and Cs_R had to be corrected using the equations

(5) $\Delta y_F(\vartheta_F) = -(2{,}43 \pm 0{,}2) \cdot 10^{-14} (\vartheta_F - \vartheta_{FO})\ [^\circ C]$ for Cs_F

ϑ = temperature of the base plate of Cs_F

$\vartheta_{FO} = 37^\circ C$

(6) $\Delta y_R(\vartheta_R) = -(4{,}6 \pm 0{,}2) \cdot 10^{-14} (\vartheta_R - \vartheta_{RO})\ [^\circ C]$ for Cs_R

ϑ_R = temperature of environment of Cs_R

$\vartheta_{RO} = 22^\circ C.$

The constants in these equations were determined by separate measurements for each clock.

A correction of the influence of pressure and relative humidity was not applied because of insufficient calibration data.

At each site measurement series were performed with durations of several days until 3 weeks [4]. By an automatic registration equipment readings were taken every 8(16) minutes. Thus each series consisted of $N \approx 200 \div 2000$ samples

$$x_n = (t_F - t_R)_n \qquad n = 0, 1, 2, \ldots N,$$

taken at the times t_n. After checking that the drift of the series was negligible the regression line with the mean rate $\tilde{y}$ was determined by minimizing the squares of deviations. Thus the estimated value $\tilde{x}_n$ at the time t_n can be described by the equation

(7) $$\tilde{x}_n(t_n) = \tilde{x}_o(t_o) + \tilde{y}\,(t_n - t_o).$$

The deviation between measured values and estimated values is $(x_n - \tilde{x}_n)$ with an estimated variance

(8) $$\tilde{\sigma}_x^2 = \frac{1}{N-1} \Sigma\, (x_n - \tilde{x}_n)^2 .$$

The variance of the estimated value $\tilde{x}_o(t_o)$ is

(9) $$\tilde{\sigma}_{\tilde{x}o}^2 = \frac{\tilde{\sigma}_x^2}{N+1} .$$

Using this result the variance of the estimated mean rate $\tilde{y}$ can be estimated as

(10) $$\tilde{\sigma}_{\tilde{y}}^2 = \frac{\tilde{\sigma}_{xo}^2}{T^2}$$

with

(11) $$T^2 = \frac{1}{N+1} \sum_{n=0}^{N} (t_n - \bar{t})^2$$

and

$$(12) \qquad \bar{t} = \frac{1}{N+1} \sum_{n=0}^{N} t_n .$$

The evaluation method sketched above was applied to each measurement series and yielded the results summarized in Table III.

The overall performance of the time transfer link can be estimated by a comparison of the measurement uncertainties on the ground and during the flight (values of $\tilde{\sigma}_{\tilde{y}}$ and $\tilde{\sigma}_{\tilde{x}}$ in the first two columns). But in both cases these measurment errors were covered by the uncertainties of temperature corrections, and these once more were covered by additional uncertainties, caused by switching the clocks on and off and by other environmental influences.

With the uncertainties of flight measurements in Table III the uncertainties of an on-board clock control can be estimated for daily clock comparisons using equation (4)

$$(13) \qquad \sigma_{\tilde{x}}(\tau) = \sqrt{\sigma_y^2(\tau) \cdot \tau^2 + \tilde{\sigma}_{\tilde{x}}^2 + \tilde{\sigma}_{\tilde{y}}^2 \cdot \tau^2}$$

$$\sigma_x(\tau = 24^h) \approx 12 \text{ ns}.$$

This value does not yet contain the uncertainty of unbalanced equipment delays in the down- and upward link. Measurements of this unsymmetry including antenna delays were performed on ground with an accuracy of $\tilde{\sigma} \approx 18$ ns. The improvement of this calibration measurement by avoiding or eliminating multipath effects will be a separate task in future.

The systematic change of rate during the flight is sufficiently explained by the relativistic time dilatation. The theoretical value given in Table III was determined from NASA orbit data by integrating this effect along the whole Shuttle track. Thus the changes of the Shuttle velocity and height during the mission could be considered as well as the influence of the eart flatness on the gravitational field.

From the results in Table III the relativistic effect can be calculated within the mentioned tolerances.

rate during flight	$-(302{,}64 \pm 0{,}017) \cdot 10^{-12}$
temperature correction	$+(\ 0{,}41 \pm 0{,}061) \cdot 10^{-12}$
mean rate on ground	$+(\ 7{,}54 \pm 0{,}10) \cdot 10^{-12}$
clock switching on board	$\pm\ 0{,}10 \cdot 10^{-12}$
relativistic frequency shift	$-(294{,}69 \pm 0{,}15\ ) \cdot 10^{-12}$

The relative accuracy of this result is 0,5 %. This is slightly less than the accuracy which is determined by the clock's frequency stability.

SUMMARY

In future satellite navigation systems the system time will have to be established at all ground stations and navigation satellites with a maximum possible accuracy.

During the experiment NAVEX the control of ground station time scales was carried out by clock transportations and by a GPS common-view time transfer. Both methods agreed within 5 ns.

The control of the Cs-clock on board of the Shuttle was performed by a microwave two-way link with pseudo noise code correlation technique and spread spectrum data transfer. The results demonstrated that a timekeeping accuracy of $\sigma \approx 12$ ns from a single ground station would be possible with the clocks used and with daily time transfer measurements.

Ground measurements with the flying clock had to be performed at different sites under different environmental conditions. Therefore, a maximum likelyhood estimation of the mean rate and temperature corrections were applied. The estimated uncertainties showed that the measurement uncertainties were small compared with residual temperature errors, and that the latter were smaller than the uncertainties produced by clock switching and other unknown environmental effects.

The frequency difference between ground and onboard operation of the flying clock was determined with an uncertainty of $\sigma_{\bar{y}} = 0{,}15 \cdot 10^{-12}$. Thus the expected relativistic effect could be observed with a relative uncertainty of 0,5 %. Concerning the theoretical value it was assumed that the effect of the uncertainties of used orbit data and theoretical models is smaller than the measurement uncertainty.

ACKNOWLEDGEMENT

The authors are indebted to many colleagues for support and helpful discussions. J. Hammesfahr prepared the ground and mission data of the flying clocks and contributed measurement details. H. Schild prepared all ground clock data. R. Schimmel calculated the theoretical value of the relativistic effect and H. Nau performed time transfers with GPS and determined the temperature coefficient of the Reference Clock.

REFERENCES

[1] S. Starker, H. Nau, J. Hammesfahr
The NAVEX Experiment in the Spacelab D1 Mission Planning and Execution.
ESA-TT-1078 (Nov. 1987), Translation of DFVLR-FB-87-03 Das Experiment NAVEX bei der Spacelab-Mission D1 - Planung und Durchführung -.
DFVLR, NE-HF, Oberpfaffenhofen, West Germany.

[2] H. Schild
Synchronization of ground station atomic clocks using weighted transportable cesium time standards in the space shuttle NAVEX experiment.
Proc. 2nd European Time and Frequency Forum, March 1988.

[3] H. Nau
Sicherung des Zeitbezuges der NAVEX-Hauptuhr und Gangbestimmung durch Zeitvergleiche mit Hilfe des 'Global Positioning System' (GPS).
DFVLR-Mitteilung 87-20, Institut für Hochfrequenztechnik, Oberpfaffenhofen.

[4] J. Hammesfahr
Systematic Performance of the Space Clocks in the NAVEX Experiment.
Proc. 1st European Time and Frequency Forum, March 1987, pp. 71-78.

[5] J. Hammesfahr, H. Nau, S. Starker
Performance of Atomic Clocks flown at the Space Shuttle Experiment NAVEX.
19th PTTI Meeting, Redondo Beach, Dec. 1987.

[6] E. Sappl, A. Zimmermann
Navigationsexperiment NAVEX, Meßergebnisse zum Zeitvergleich.
DFVLR-Mitteilung 87-05.

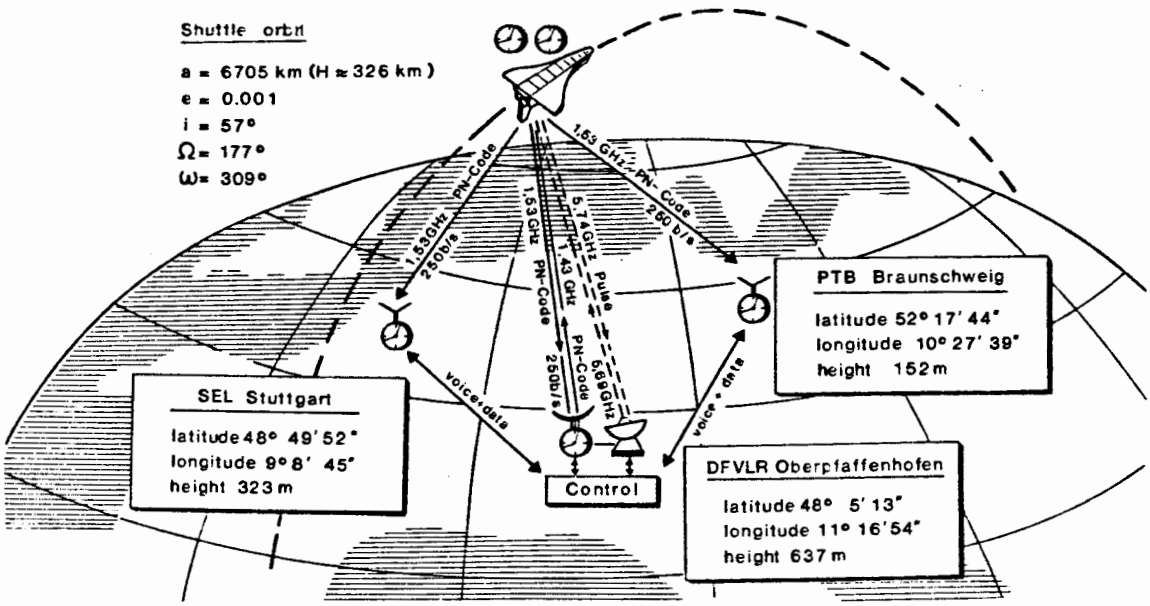


Fig. 1 Experiment Configuration of NAVEX.

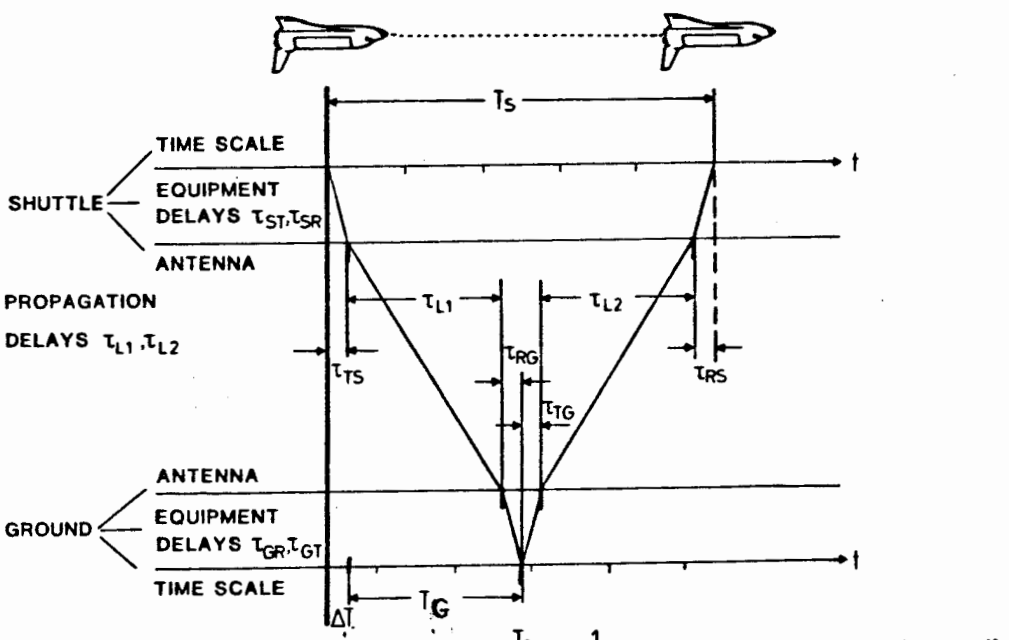


Fig. 2 Principle of the two-way time transfer.

CLOCKS		$\sigma_y(\tau)$	function during NAVEX-experiment	SITES
Flying Clocks	Cs-FTS 4000 opt. 004 ,	$1 \cdot 10^{-13}$ p. day	Main clock on Space Shuttle	• Stuttgart 1984 • Eberdingen 1984 • Bremen 1985 • Cape Canaveral 1985 • Shuttle orbit 1985 • Oberpfaffenhofen 1986
	Rb-FRK-H S.N. 53955	$1 \cdot 10^{-11}$ p. month	Back up clock on Space Shuttle	
Station Clocks	Cs-hp 5061A opt. 004 S.N. 2072	$1 \cdot 10^{-13}$ p. day	NX - reference clock producing UTC (Oberpfaffenhofen)	DFVLR Oberpfaffenhofen Control station: h = 637 m $\phi = 48^{\circ}\ 5'\ 13''$ $\lambda = 11^{\circ}\ 16'\ 54''$
	Cs-hp 5061 A opt. 004 S.N. 1302	$1 \cdot 10^{-13}$ p. day	NX-master clock at control station (DFVLR)	
	Rb - R & S	$3 \cdot 10^{-11}$ p. month	Clock at the radar station (DFVLR)	
	Cs-hp5061 A opt.004	$1 \cdot 10^{-13}$ p. day	Receiving station clock (PTB)	PTB - Braunschweig h = 152 m $\phi = 52^{\circ}\ 17'\ 44''$ $\lambda = 11^{\circ}\ 16'\ 54''$
	Clock-ensemble of UTC (PTB)	$< 1 \cdot 10^{-13}$	Main reference time scale based on a primary standard	
	Cs-hp 5061A S.N. 1283	$3 \cdot 10^{-13}$ p. day	Receiving station clock (SEL)	SEL - Stuttgart h = 323 m $\phi = 48^{\circ}\ 49'\ 52''$ $\lambda = 9^{\circ}\ 8'\ 45''$
	Rb - R & S	$3 \cdot 10^{-11}$ p. month	Station back up	
Transportable Clocks	Cs-hp 5061A S.N. 7562	$3 \cdot 10^{-13}$ p. day	Transp. clock from PTB	DFVLR-Oberpfaffenhofen PTB-Braunschweig SEL-Stuttgart
	Cs-FTS 4010 S.N. 111	$2 \cdot 10^{-13}$ p. week	Transp. clock from PTB	
	Cs-FTS 4010 S.N. 113	$2 \cdot 10^{-13}$ p. week	Transp. clock from DFVLR	

Table I

ATOMIC CLOCKS ENVOLVED IN THE SPACESHUTTLE EXPERIMENT NAVEX

CLOCKS	SITES month/year	TEMP. ($\bar{\delta} \pm \sigma_\delta$) °C environment	TEMP. baseplate of Cs_F	PRESSURE mbar	REL.HUMIDITY %
FLYING CLOCKS	Oberpfaffenhofen 10/83	function tests and temperature influence determination			
	Stuttgart 3/84	19 ± 1,5	33 ± 1	940 ÷ 1000	50 ÷ 80
	Eberdingen 5/84	21 ± 2	37 ± 1	940 ÷ 1000	50 ÷ 80
	Bremen	(integration tests only)		—	—
	Cape Canaveral 7/85	21,5 ± 1	32,5 ± 0,5	980 ÷ 1020	50 ÷ 70
	Shuttle orbit 11/85	—	(50...58)*) ± 1	1000	0
	Cape Canaveral 12/85	21,0 ± 1	31,5 ± 0,5	980 ÷ 1020	50 ÷ 70
	Oberpfaffenhofen 3/86	27 ± 1	46,2 ± 0,5	900 ÷ 980	40 ÷ 60
	Oberpfaffenhofen 4/86	25 ± 1	43,6 ± 0,5	900 ÷ 980	40 ÷ 60
	Oberpfaffenhofen 4/86	25 ± 1	44 ± 0,5	900 ÷ 980	40 ÷ 60
REFERENCE CLOCK	Stuttgart				
	Eberdingen				
	O'hofen 5/84 - 2/86	22 ± 1			
	O'hofen 3/86	27 ± 1			
	O'hofen 4/86	25 ± 1			

*) systematic variations around the mean $\bar{\delta} = 54°$ C were registrated.

Table II ENVIRONMENTAL CONDITIONS OF NAVEX - CLOCKS

Measurement - site - month/year - clocks	Rate measured $\bar{y} = y \pm \bar{\sigma}_{\bar{y}}$	Reading uncertainty $\bar{\sigma}_x$	Corrections a) $\Delta y = \Delta y_R - \Delta y_F$ b) Δy_r = relat. effect	Rate corrected $\bar{y}_c = \bar{y}_c \pm \bar{\sigma}_y$
Stuttgart 3/84 $Cs_F - Cs_{SEL}$ $Cs_{SEL} - Cs_R$ $Cs_F - Cs_R$	$-(7,249\pm0,011)\cdot10^{-12}$ $-(0,30\pm0,01)\cdot10^{-12}$ $-(7,549\pm0,015)\cdot10^{-12}$	3,2 ns	a) $\Delta y = +(0,041\pm0,060)\cdot10^{-12}$	$-(7,51\pm0,054)\cdot10^{-12}$
Eberdingen 5/84 $Cs_F - Cs_R$	$-(7,581\pm0,006)\cdot10^{-12}$	4,6 ns	a) $\Delta y = +(\pm0,046\pm0,074)\cdot10^{-12}$	$-(7,54\pm0,078)\cdot10^{-12}$
D1-Mission 11/85 $Cs_F - Cs_{St}$ $Cs_{St} - Cs_R$ $Cs_F - Cs_R$	$-(302,269\pm0,016)\cdot10^{-12}$ $-(0,373\pm0,005)\cdot10^{-12}$ $-(302,642\pm0,017)\cdot10^{-12}$	5,8 ns	a) $\Delta\bar{y} = +(0,408\pm0,061)\cdot10^{-12}$ b) $\Delta y_r = -294,71\cdot10^{-12}$	$-(7,53\pm0,063)\cdot10^{-12}$
O'hofen 3/86 $Cs_F - Cs_R$	$-(7,466\pm0,003)\cdot10^{-12}$	2,2 ns	a) $\Delta y = -(0,007\pm0,056)\cdot10^{-12}$	$-(7,47\pm0,056)\cdot10^{-12}$
O'hofen 4/86 $Cs_F - Cs_R$	$-(7,741\pm0,003)\cdot10^{-12}$	2,1 ns	a) $\Delta y = +(0,068\pm0,053)\cdot10^{-12}$	$-(7,67\pm0,053)\cdot10^{-12}$
O'hofen 4/86	$-(7,583\pm0,010)\cdot10^{-12}$	1,4 ns	a) $\Delta y = +(0,044\pm0,053)\cdot10^{-12}$	$-(7,54\pm0,054)\cdot10^{-12}$
		Mean rate:		$-(7,54\pm0,10)\cdot10^{-12}$

Table III

RESULTS OF RATE MEASUREMENTS BETWEEN FLYING CLOCK Cs_F AND REFERENCE CLOCK Cs_R

2 ND EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 16 - 17 - 18 MARCH 1988

Session TF-04 Number 4

TRANSFERT DE TEMPS PAR COUPLAGE LORAN-C GPS

M. D'Amico

Techniphone - Le Puy Sainte-Reparade

FRANCE

RESUME

L'utilisation des signaux GPS et LORAN-C permet d'effectuer des transferts de temps dans différents contextes. L'utilisation combinée des deux systèmes offre un service plus performant.

Après avoir donné des résultats d'essais, ce papier présente une analyse comparative des deux systèmes et de leur couplage en matière de transfert de temps.

ABSTRACT

The use of GPS signals or LORAN-C waves allows time keeping in different contexts. Combining both systems improves the availability of time transfers. This paper shows how an hybridization between LORAN-C and GPS leads in an operating mean of synchronisation.

INTRODUCTION

Les techniques de transfert de temps sont destinées à synchroniser des sites distants les uns des autres.

Pour cela les horloges locales de chaque site doivent être comparées entre elles. On utilise différents moyens : transferts physiques d'horloges, visées de satellites naturels ou artificiels, VLBI, réception simultanée d'émetteurs radioélectriques...

Ces émetteurs radioélectriques peuvent être spécialisés dans la diffusion du temps ou bien très souvent des émetteurs de catégories différentes sont utilisés (radiodiffusion, radionavigation, ...).

Il existe des émetteurs terrestres et spatiaux. Nous citerons par exemple les systèmes OMEGA, LORAN-C, TRANSIT et GPS NAVSTAR. L'examen des caractéristiques du LORAN-C et du GPS conduit à noter une complémentarité des deux systèmes en matière de navigation ainsi que dans le domaine de la diffusion du temps.

CARACTERISTIQUES DES DEUX SYSTEMES LORAN-C ET GPS

Le système LORAN-C est composé de nombreuses chaînes d'émetteurs synchronisés essentiellement répartis dans l'hémisphère Nord. La fréquence d'émission est de 100 KHz et la portée des émetteurs de l'ordre de 1500 km.

La précision de localisation dynamique est de quelques 100 m avec toutefois des erreurs systématiques liées aux conditions de propagation de l'onde à la surface de la terre.

Ce système est très largement utilisé depuis de nombreuses années à la fois par les navigateurs et les spécialistes du temps. Les principaux laboratoires participant aux échelles de temps standard (TU, TA, ...) sont équipés de récepteurs LORAN-C.

Le système GPS NAVSTAR est constitué de trois segments.
Le segment spatial constitué de 18 satellites synchronisés sur une échelle de temps GPS conservée par le segment terrestre, assurera la diffusion de messages sur tout le globe à tout instant. Ces messages permettront à l'utilisateur, qui constitue le troisième segment de se positionner en trois dimensions et de synchroniser son horloge locale avec une précision et une exactitude meilleure que le dixième de microsecondes. Ce système est cependant en cours de développement.

SYSTEMES DE RADIONAVIGATION

LORAN-C ET GPS NAVSTAR

	LORAN-C	GPS
BALISES	terrestres	satellites
FREQUENCE D'EMISSION	100kHz	1575MHz (L1)
PORTEE	1000Nm	∞
PRECISION	100m//1000m	25m
COUVERTURE	hémisphère nord	globale
DISPONIBILITE	système opérationnel	en cours de développement

L'analyse comparée des deux systèmes permet de mettre en évidence leurs points forts et leurs points faibles.

Le LORAN-C présente un biais de mesure (erreur lentement évolutive, jour, nuit, saisons). Ce biais est dû à la non homogénéité de la vitesse de propagation de l'onde au cours du trajet émetteurs-utilisateurs. La réception du signal (100 kHz) peut être difficile dans certains contextes (nuit, bruits, ...).

Une très forte onde de ciel cohérente peut venir se combiner à l'onde directe (sol) et conduire à des erreurs de plusieurs dizaines de microsecondes si le récepteur n'effectue pas un traitement adéquat de l'onde reçue. Une ambiguité temporelle (de 100 à 200 ms) doit être levée pour obtenir une information horaire exacte.

Les points faibles du GPS sont essentiellement dûs à l'état actuel du système qui est en cours de développement.
Six satellites seulement sont aujourd'hui utilisables et leurs caractéristiques sujettes à caution. La fréquence d'émission utilisée (bande L : 1575,42 MHz) recquiert une visibilité directe avec les satellites. Il faut donc s'affranchir des masques (feuillages, bâtiments, eau, ...).

On notera cependant que ce système, sous la responsabilité du DOD (US) est destiné à être dégradé (Selective Availability) pour fournir aux utilisateurs du service SPS (code C/A) des performances inférieures à celui du service PPS (code Y). Seul le service SPS sera disponible aux utilisateurs civils.

Les avantages sont une bonne répétabilité des mesure ainsi qu'une disponibilité de fait pour le LORAN-C. Le GPS quant à lui bénéficie d'une exactitude et d'une précision excellentes, d'une couverture qui sera globale et permanente, ses caractéristiques le rendent très immunisé aux bruits présents dans la bande d'émission. Il ne présente aucune ambiguité à priori.

LORAN-C

GPS

INCONVENIENTS

BIAIS

RECEPTION DIFFICILE
(NUIT, TERRE ...)

ONDE DE CIEL

CONSTELLATION INCOMPLETE

RECEPTION MASQUEE

AVANTAGES

BONNE REPETABILITE

SYSTEME OPERATIONNEL

DISPONIBILITE TEMPORELLE

PRECISION ABSOLUE

COUVERTURE

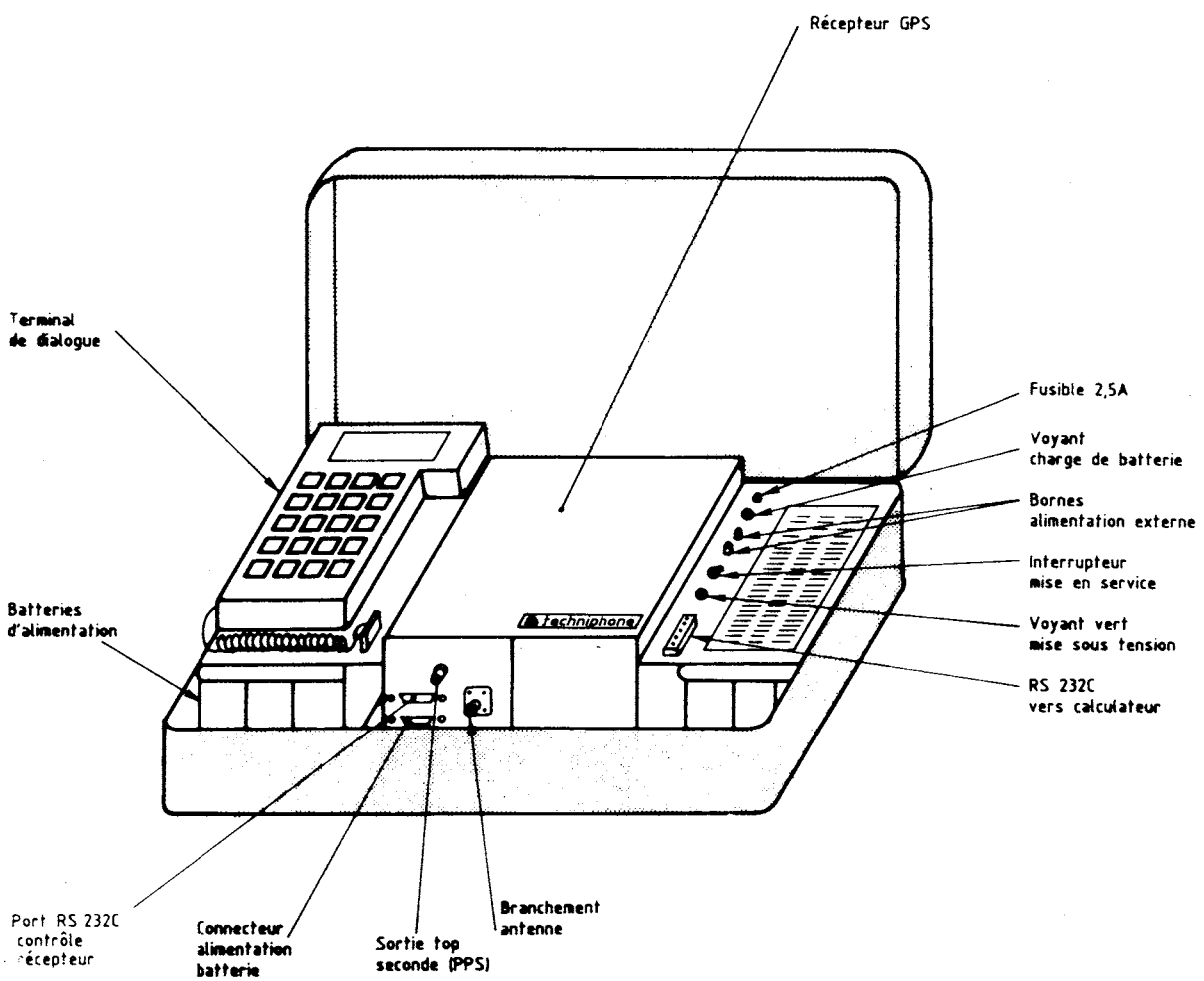
DIFFUSION DU TEMPS UTC

ANTIBROUILLE

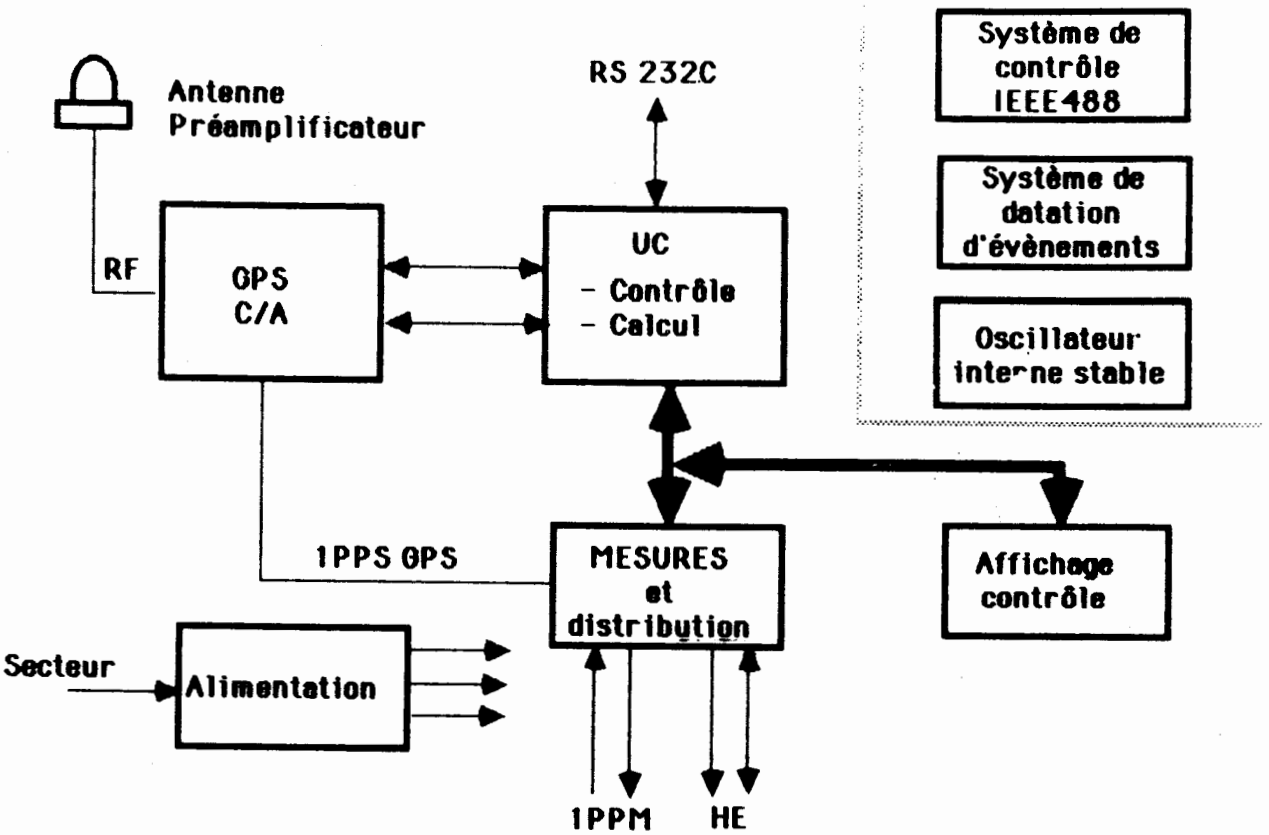
MOYENS OPERATIONNELS DE TRANSFERT DU TEMPS

TECHNIPHONE a développé une gamme de récepteurs LORAN-C pour la Navigation, ainsi qu'un ensemble de génération de fréquences étalons à partir de sources atomiques à jet de césium pour des applications embarquées à bord de mobiles. Une gamme de récepteurs GPS a aussi été développée autour des coeurs de réception GPS de COLLINS (USA). Ces développements sont aussi destinés à la navigation, la localisation et la diffusion du temps à bord de mobiles ou dans tous les cas où l'utilisateur n'a pas le besoin ou la possibilité de mettre en oeuvre des moyens logistiques conséquents pour obtenir des résultats meilleurs que la dizaine de nanosecondes.

Les figures suivantes représentent le récepteur autonome pour Navigation et Temps, alimenté par batterie et le récepteur TE68 orienté vers la diffusion du temps et la métrologie.



GPS TIMING RECEIVER



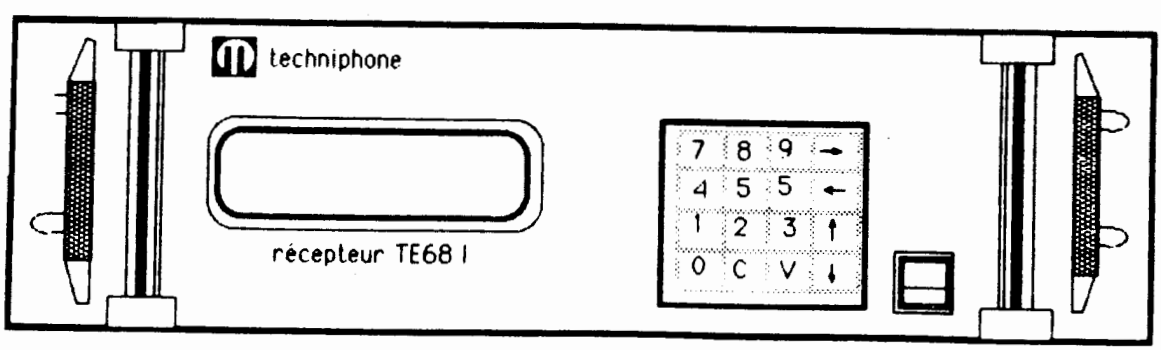
RECEIVER → RELIABLE - ROBUST

RECEIVER → EASY TO USE

TIME GENERATOR

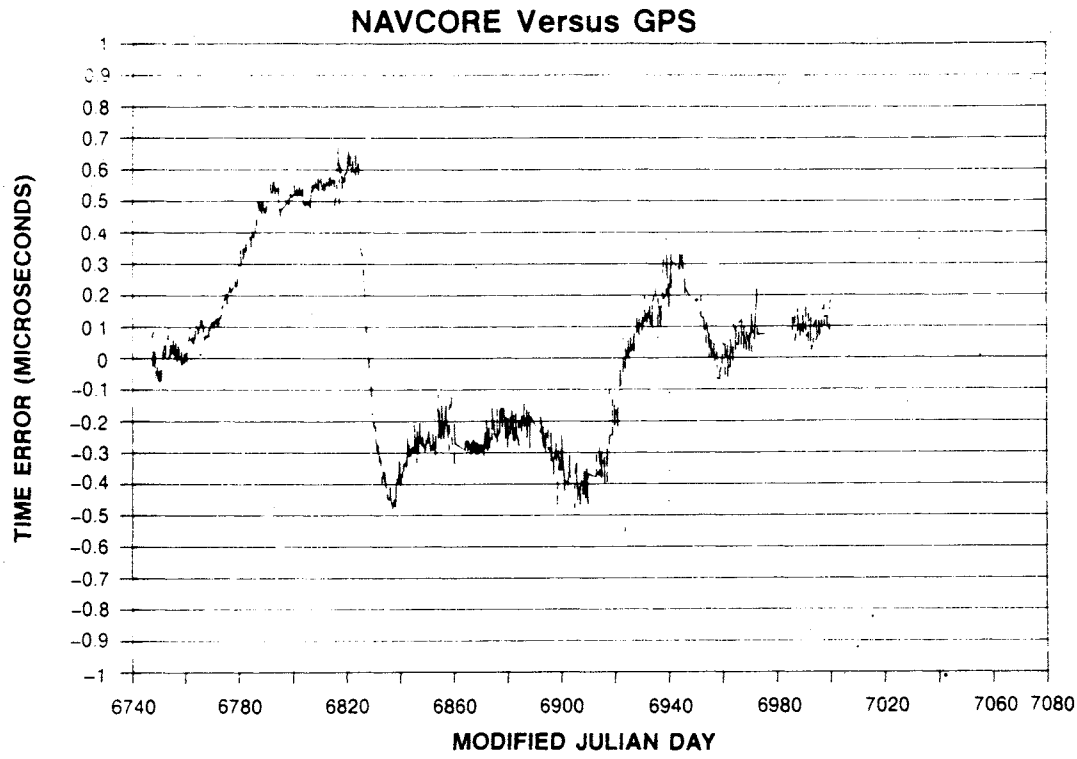
EVENTS TIMING VS UTC

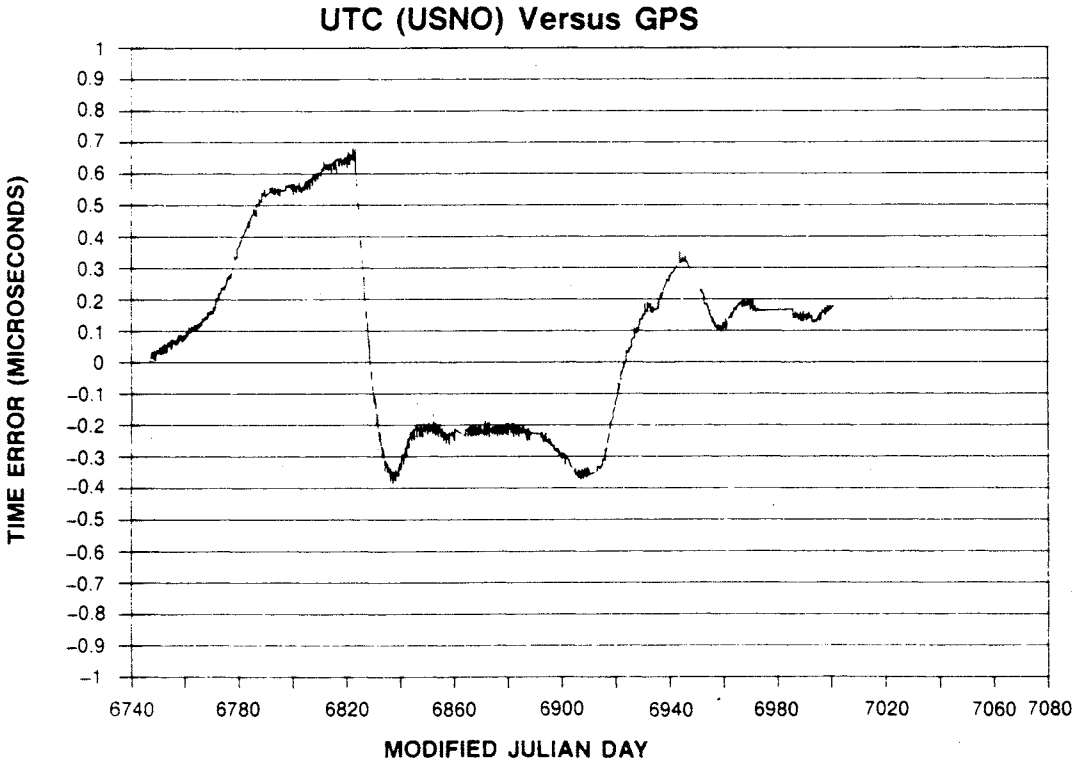
TIME METROLOGY

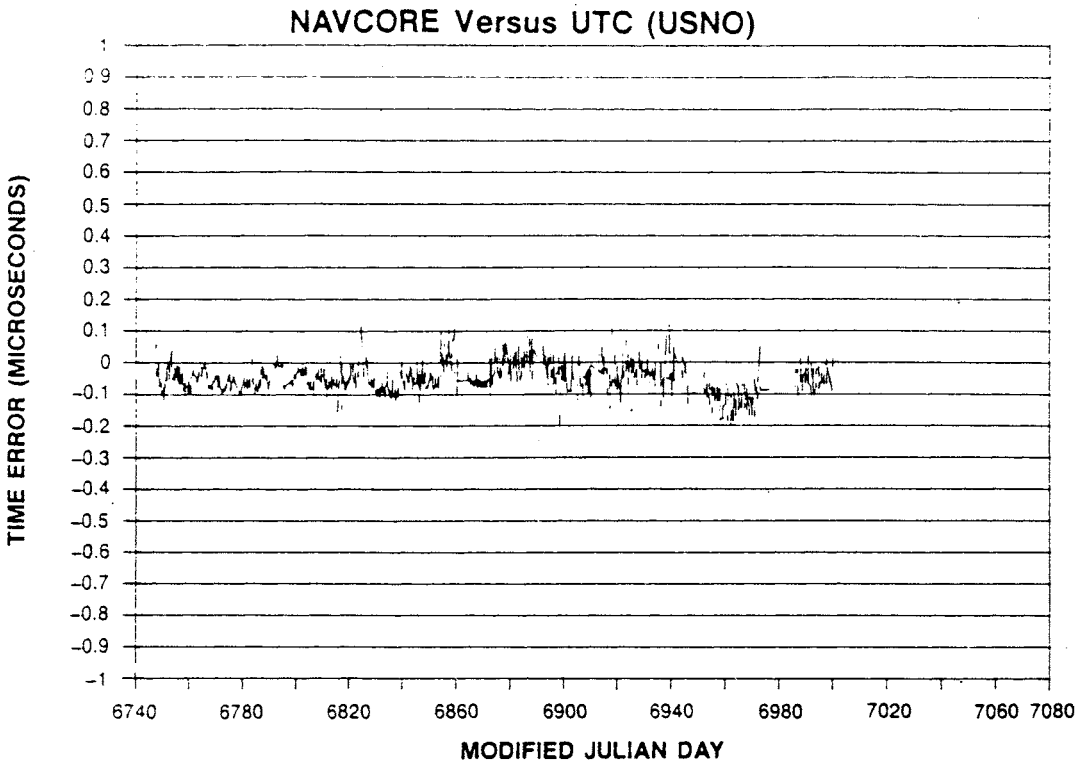


RESULTATS OBTENUS

De nombreux essais ont été effectués avec les récepteurs GPS de toute la gamme. Des résultats issus de mesures faites sur de très longues durées sont représentés sur les courbes suivantes.







On se reportera aux rapports d'essais cités dans les références pour avoir plus de détails.
On peut cependant résumer les essais comme suit :

Des mesures faites par le CNES (Toulouse) en collaboration avec le LPTF (Observatoire de Paris) ont conclu à une exactitude meilleure que 100 ns sur 9 jours dans l'absolu. Une comparaison avec le LPTF a permis de vérifier qu'un majorant de l'erreur en vue commune était de 50 ns.

Des dépouillements plus poussés sur les mesures faites en collaboration avec COLLINS (USA) et le NBS pendant 50 jours ont donné 52,6 ns en moyenne avec un bruit superposé de 28 ns.
Les essais ont continué pendant 8 mois et demi pour donner 42 ns avec 38 ns d'écart type. Ces essais étant effectués sans la connaissance très précise de la position de l'antenne et sans autre calibration que celles effectuées par les récepteurs commercialisés.

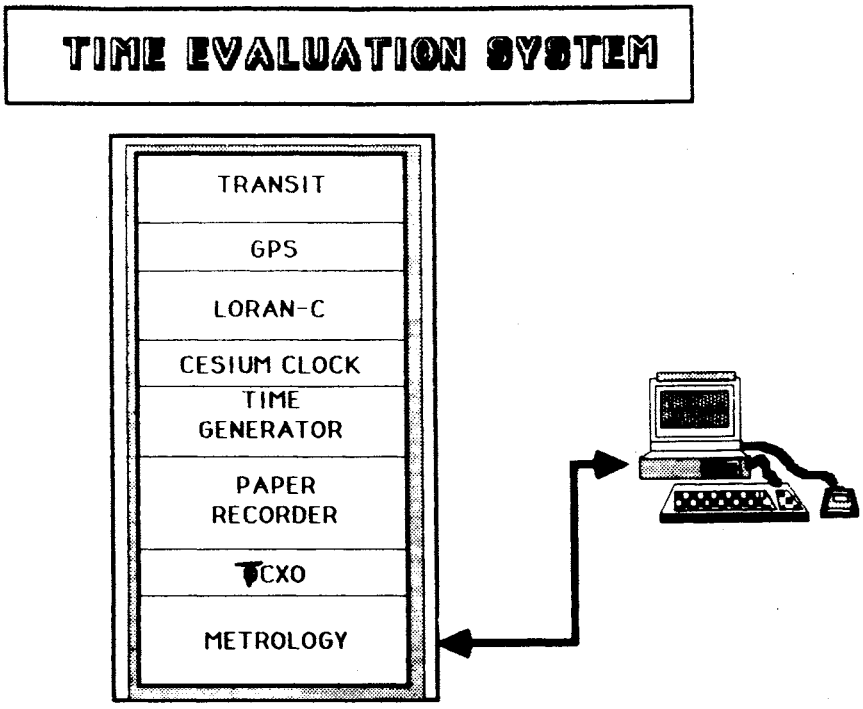
On peut donc en déduire qu'ils sont très bien adaptés dans le cadre d'une exploitation en ambiance difficile ou quand l'utilisateur a besoin de disposer du système en temps réel et avec très peu de logistique.

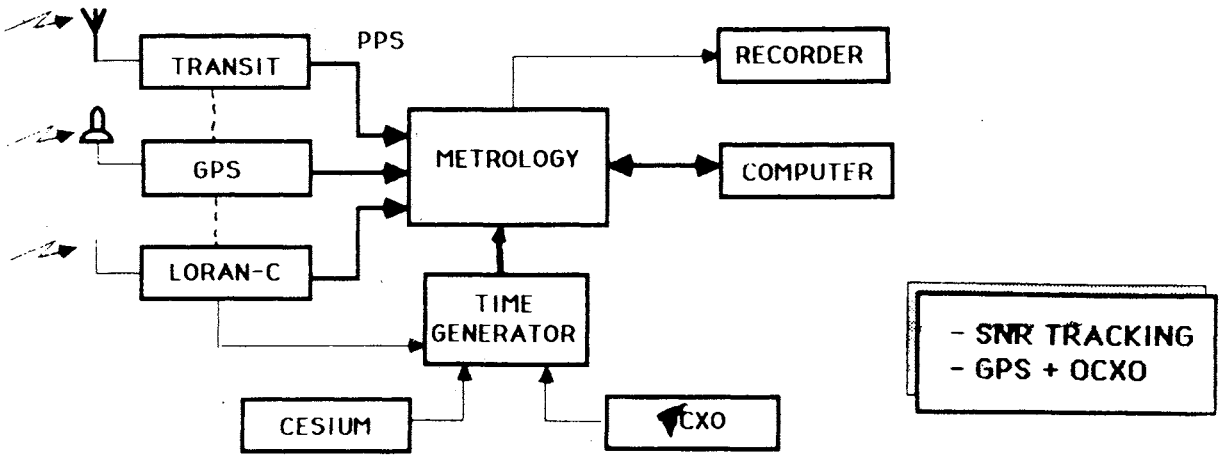
COUPLAGE GPS - LORAN-C

Pour ses besoins propres dans le cadre des développements en RADIONAVIGATION, **TECHNIPHONE** a été amenée à développer une baie TEMPS FREQUENCE. Cette baie architecturée autour d'une référence atomique peut être raccordée aux échelles de temps par l'intermédiaire de 3 canaux radioélectriques (LORAN-C, GPS NAVSTAR ou TRANSIT) ou bien par transport d'horloges.

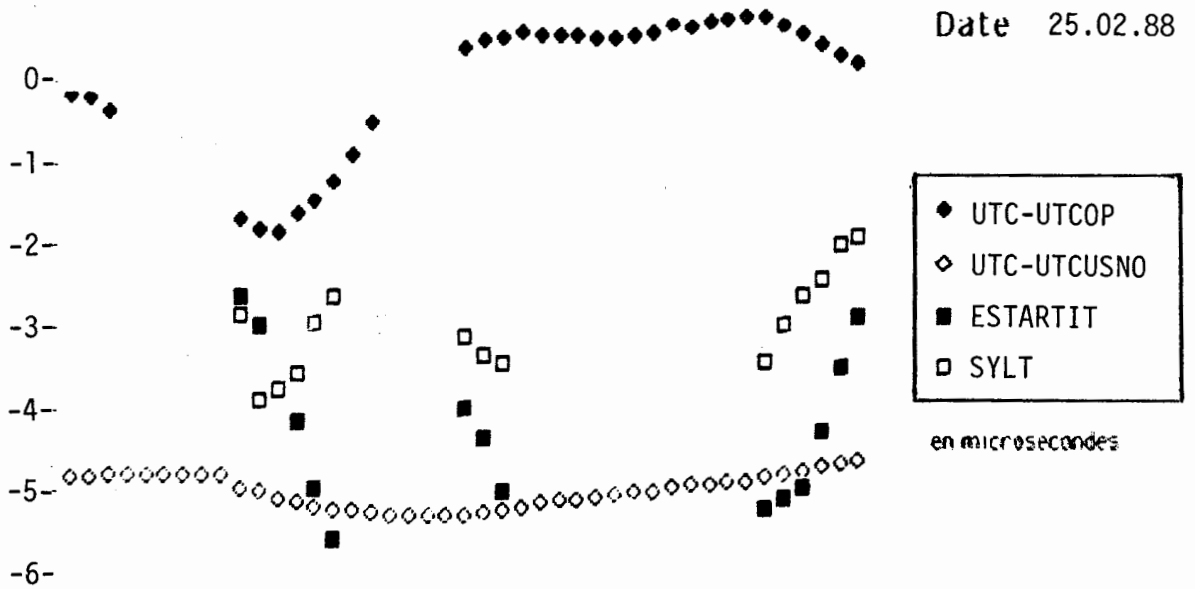
Un calculateur couplé à un système de mesures permet d'effectuer des comparaisons entre les informations élaborées par chaque système.

Des programmes de suivi ont été développés et ont permis d'observer les comportements des systèmes de radionavigation.





La figure suivante montre les courbes des écarts LORAN-C des stations SYLT et ESTARTIT par rapport aux échelles UTC, UTC-OP, UTC-USNO.



Ces observations ainsi que les analyses des différents systèmes mettent en évidence les avantages à disposer d'un couplage GPS + LORAN-C dans le cadre des transferts de temps et synchronisation de systèmes opérationnels.

Pour exploiter au mieux les mesures effectuées par les deux récepteurs GPS et LORAN-C on retient trois modes de fonctionnement.

Mode 1 : Si le GPS est disponible et son fonctionnement nominal, c'est ce système qui offre la meilleure précision et ceci sans ambiguité.

Mode 2 : Le GPS présente des difficultés de réception. La fréquence 1575,42 MHz impose d'être en visibilité directe avec les satellites et empêche l'exploitation du système dans certains contextes (bâtiments, feuillages, fonds marins, ...).

Certaines périodes sont assez pauvres en satellites visibles ; dans ces deux cas le LORAN-C reçoit du GPS la position et l'heure ce qui lui permet de prendre le relais des mesures de temps.

Mode 3 : Le système GPS sera entaché d'erreurs volontaires dégradant les mesures que pourra effectuer l'utilisateur (Selective Availability 385 ns 95%). Les erreurs seront significatives et souvent difficiles à détecter (dérives lentes). Le LORAN-C permet dans une telle phase de détecter et corriger les informations fausses issues du GPS.

CONCLUSION

L'utilisation combinée des deux systèmes GPS et LORAN-C sera indispensable à tout utilisateur désirant une bonne précision de recalage dans des contextes opérationnels.

REFERENCES

1 - Du LORAN-C au GPS M. GRANDVEAUD LPTF
Observatoire de Paris Mars 1986

2 - Hybridation LORAN-C GPS NAVSTAR
Campagne d'essais M. D'AMICO 3ème colloque sur la localisation en mer 1987

3 - Transfert de temps Moyens radioélectriques opérationnels
M. D'AMICO 1er Forum Européen Temps Fréquences 1987

4 - SS - US - 200 System Segment Specifications

5 - STANAG 4294 Issue H

6 - GPS User Equipment PTTI Interface

M. D'AMICO
Responsable du département Navigation Localisation Temps
TECHNIPHONE

TF-05

VLF and LF

1. Envelope Stability of Loran-C Pulses
   E. Rubiola
   Politecnico di Torino, Torino - Italy

2. A simple Loran-C timing receiver for supporting VLBI
   J.E.B. Ponsonby
   University of Manchester, Macclesfield - UK

3. Time dissemination via the LF Transmitter DCF77 using a pseudo-random phase-shift keying of the carrier
   P. Hetzel
   Physikalisch-Technische Bundesanstalt, Braunschweig - FRG

4. Precise determination of UTC in remote regions using OMEGA signals with reference to monitoring performance of a hydrogen maser
   H.J. Gaede
   Hartebeesthoek Radio Astronomy Observatory/NITR, Johannesburg - Republic of South Africa

2 ND EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 16 - 17 - 18 MARCH 1988

Session TF-05 Number 1

Envelope Stability of Loran-C Pulses

Enrico Rubiola
Politecnico di Torino
Dipartimento di Elettronica

Abstract

Some automatic Loran-C receivers base the cycle identification on the comparison between the received pulse envelope and the nominal one. An analysis is performed about the envelope stability of the pulse emitted by three stations of the Mediterranean Sea chain.

Introduction

Hyperbolic Loran-C navigation system has recently found new application areas in land and ground navigation.

In U.S.A., the Federal Aviation Agency will provide coast to coast coverage for air area navigation and non-precision approach; Loran-C system will be kept in operation in the U.S. continental states and along the shoreline well beyond the beginning of the next century. For ground navigation, a number of factories, such as Motorola and II-Morrow and others, are providing vehicle localization and reporting systems based on this navigation system.

As regards Europe, France has installed in 1985 a ρ-ρ navigation network based on Loran-C; the U.K. administration will adopt Loran-C, with a new station, to be located at Bush-Mill in North Ireland. Federal Republic of Germany is considering the possibility to move to an eastern position the Sylt station.

In Italy is under study the interest of providing the support of the system after the 1992 and Norway since four years is considering various options for the future, but it appears that the present North-Sea chain will be kept in operation or probably enlarged.

Consequently the system will remain in operation and new applications, based on inexpensive and fully automated position sensors, will show-up.

For the position sensors, one of the key point to be solved is the cycle selection inside each transmitted pulse. As it is well known, any error in cycle order determination result in a gross position error, whose order is three kilometers or more.

Cycle selection is performed via a microcomputer by comparing the envelope, given by the positive and/or negative cycles, with the theoretical shape.

The aim of this paper is to study the relations between the envelopes observed in experiment and the theoretical ones and to report some preliminary results on the envelope stability with different time scales, from milliseconds to weeks. Data thus considered, are intended to assist the designers of the

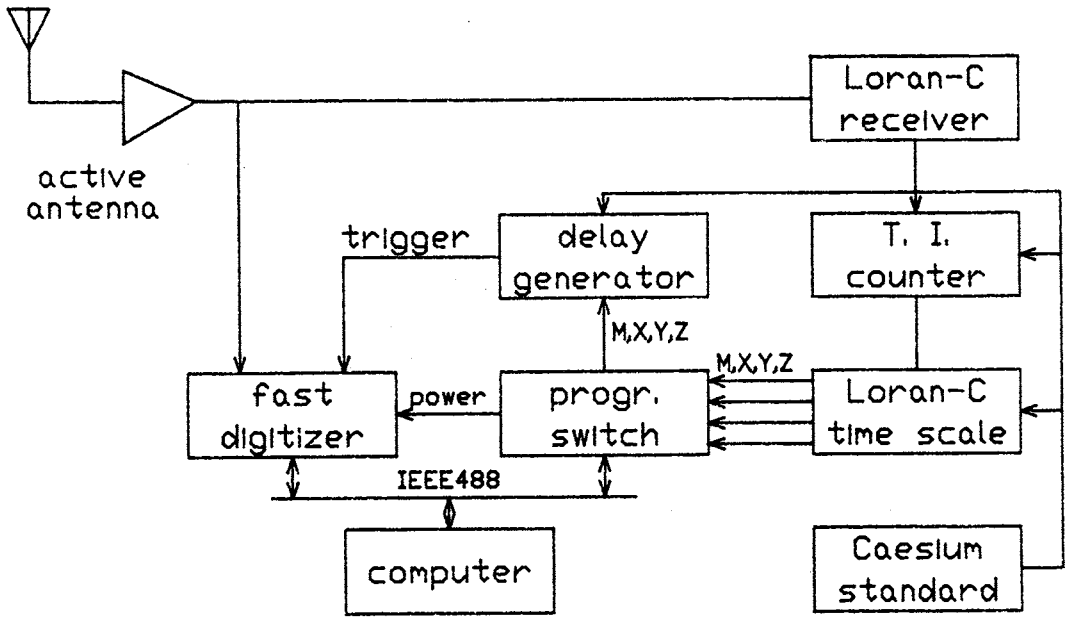


Figure .1: Experimental set-up.

positioning sensors.

The experimental set-up, centered around a fast digitizer, is described in the second section, whereas some experimental results and conclusions are presented respectively in the following third and fourth sections.

Data, gathered in Torino, at Dipartimento di Elettronica of Politecnico, are intended to be preliminary ones, because it is planned to follow some feature, such as the long term envelope-to cycle distortion for about half a year.

No presumption should be inferred that the data here given represent the conditions at emission; they are typical of receiving conditions in an highly noisy environment.

Instrumental Set-Up

The envelope of Loran-C pulses has been measured observing the peak amplitude of half cycles of the carrier.

The system, shown in fig. 1 is based on an high speed digitizer Tektronix 7612-D, that gives 8-bit instantaneous values at a sampling rate of 20 MHz.

The local time scale is compared with Loran-C signals and manually phase locked with small variations of the C-field; so, the long term stability of the sampling point within a Loran-C pulse is guaranteed within 100–200 ns.

The amount of available data is such that only preprocessed values can be hold in a mass storage. This processing is a significant part of the measurement method.

With reference to fig. 2, after each triggering pulse, a record of 2048 bytes is available; if the computer is ready, that occurs with a probability of 40 %, the data are added in an array and averaged after N_a acquisitions. From this averaged array, peak values are detected with a trigonometric best-fit on windows, whose width is 850 ns, centered around their nominal positions and stored as an array of 12 half-cycle peak amplitudes values, corresponding to $2.5,\ 7.5 \ldots 57.5\ \mu s$ from the envelope starting point; in our context, this is one *measure*.

The noise is not a symmetrical phenomenon in a finite field of sampled values if synchronous detection is used; its average value is not zero, despite of the fact that the "true noise" has zero mean. In formula:

$$P\{(s_{(t)} + n_{(t)}) \geq V_{max}\} \neq P\{(s_{(t)} + n_{(t)}) \leq -V_{max}\}$$

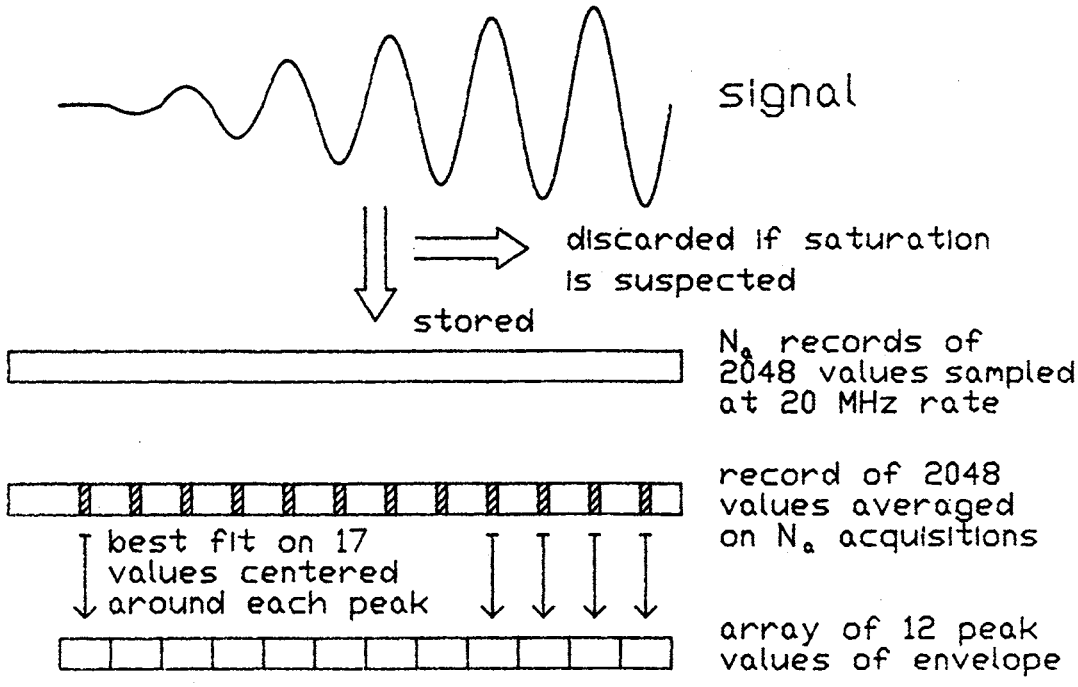


Figure .2: Method used for envelope shape measurements.

The correction needs the exact knowledge of S/N ratio, but this is not a practical approach. If a record has 2 % of extreme values or extreme values appear in the windows used for best-fit purposes, the record is considered affected by pulse noise and discarded; a new record will take its place and measuring rate is just a little slower.

With the above described method, 50 measures of the first pulse have been collected every day, in two sessions 12 and 0 UTC

The value of N_a has been chosen to be 256, improving the S/N ratio by 24 dB. A typical waveform, averaged on 256 records, is shown in fig. 3; at first sight, noise seems to be disappeared.

The short term variations of the peak amplitudes of the same stations have been measured with $N_a = 64$ and $N_a = 256$ only daytime, around 11 UTC

Results

To compare the envelope given by the measurements with the theoretical one, a reference envelope in each case was constructed with a least squares best fit performed on the first four cycles of the pulse. Starting point and shape are known, amplitudes only were fitted.

The bias of the samples was estimated and removed; moreover, since the sampler is fully programmable, the scale factor were chosen in order to avoid saturation phenomena.

Long Term Effects

Piling up data of about one month, the table of fig. 4 and plots of fig. 5 to 7 were obtained.

The table gives the average departure from the reference envelope for the station M (Sellia Marina), X (Lampedusa) and Z (Estartit) of the Mediterranean Sea chain. The distances from the reception point are respectively 1015, 1150 and 490 km; consequently the cycles of the ground wave, free from sky wave interference, are, at least, 5, 4 and 10 respectively.

First column is the order of the cycle, and the two values given for each station, around 12 UTC and 0 UTC, are, for negative and positive half-cycles, the differences between the average values and

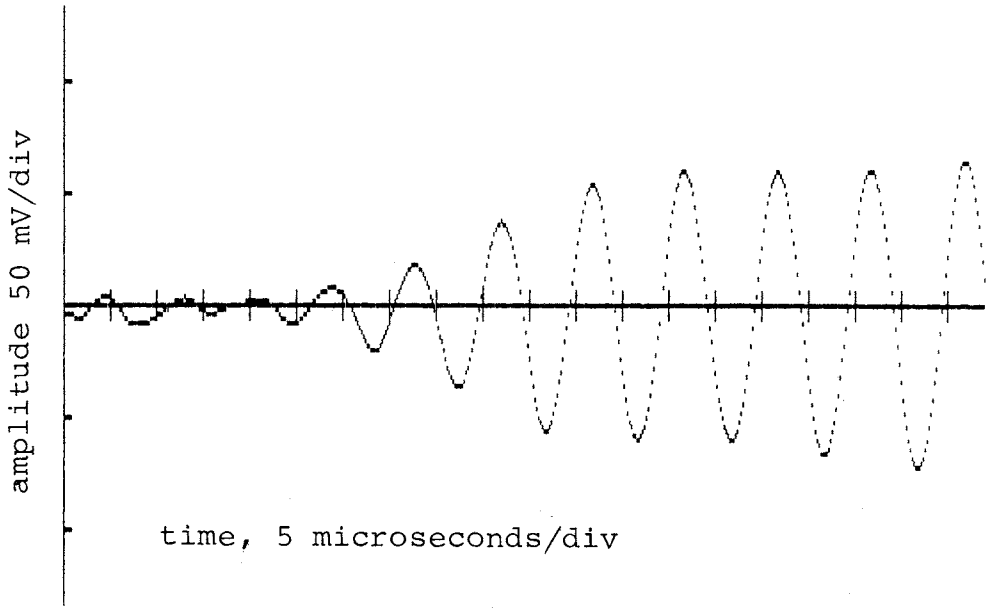


Figure .3: Typical Loran-C pulse received with the system herein described.

CYCLE	ESTARTIT		SELLIA		LAMPEDUSA	
	12 UTC	0 UTC	12 UTC	0 UTC	12 UTC	0 UTC
1 +	3.0	3.8	3.0	3.4	0.9	5.1
1 -	1.6	2.9	5.1	10.5	0.4	-0.5
2 +	-1.9	-2.1	7.3	9.3	-3.6	1.5
2 -	-5.3	-5.8	-6.6	-6.2	1.8	7.9
3 +	-4.9	-2.7	-7.6	-3.4	1.5	2.9
3 -	-1.7	0.2	-5.1	-4.3	-2.5	-8.4
4 +	1.0	1.3	5.8	2.7	-6.0	-7.0
4 -	2.7	2.3	2.9	2.6	7.5	8.0
5 +	8.9	11.9	8.6	9.1	12.4	13.4
5 -	17.8	20.4	4.9	6.8	3.8	7.0

Figure .4: Differences between measured and reference peak amplitudes, given in *percent* of reference amplitude.

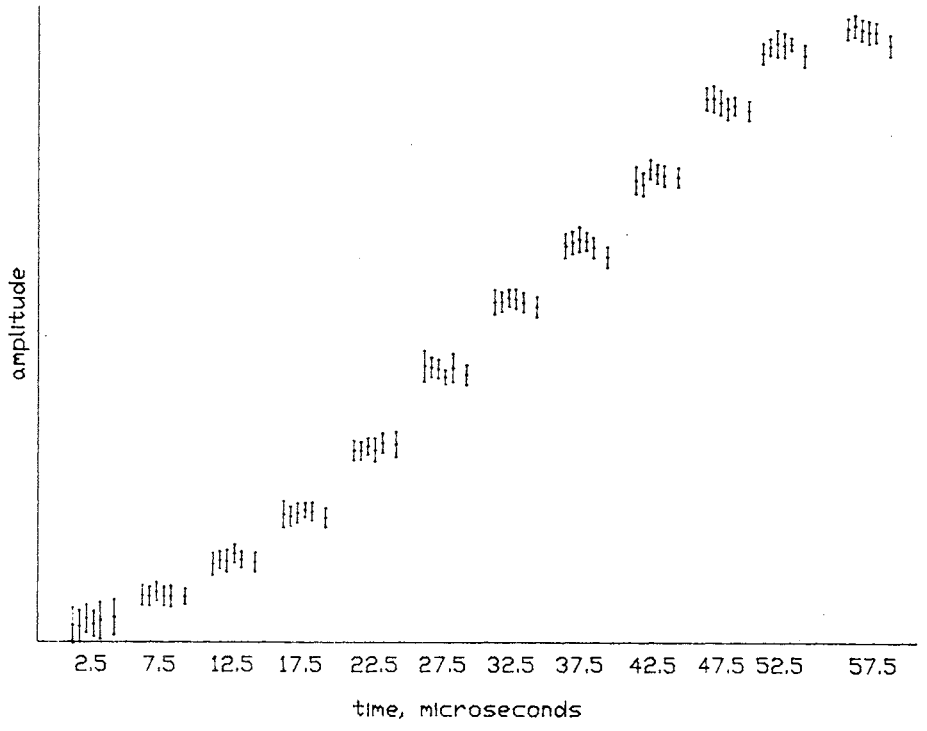


Figure .5: Plot of daily amplitudes of first half-cycles. Estartit, 12 UTC.

reference one; values are given as *percent* of the reference value.

Fig. 5 and 6 represent, for one week, the average value of the individual half-cycle peaks, in an arbitrary linear scale, for station Z Estartit in day and night conditions; each point is a daily value.

Fig. 7 represents, with the same scales, the envelope of master station in daylight conditions; the reference envelope is added for comparison.

By inspection of data in figs. 4 to 7, the following preliminary conclusions can be pointed out:

- a systematic deviation is observed in all the stations for cycles 2, 3 and 5,
- each station seems to show a specific "signature",
- during the night, excluding the first two cycles in which noise contamination is present and cycles in which sky wave contribution could be present, the amplitude is larger by 3–4 %,
- the *one sigma* of the measurements is larger during the night by a factor two.

During a week, typical amplitude variations of the variations of the pulses are of about 2–3 % with reference to an average weekly value. To evaluate this figure, the reader should be aware that it is not easy to separate deterministic events, such as re-tuning of the transmitters or natural or man-made noise.

This remarkable amplitude stability information is not of large interest for designing purposes since far larger amplitude variations are to be fund in a travel.

Short Term Effects

The received envelope for the first five individual pulses, inside a train of eight, was recorded.

Each individual pulse was averaged on 6400 records over a span of 15 minutes.

The results, for station Z, are given in fig. 8, where the amplitude of the same half-cycles in consecutive pulses are plotted.

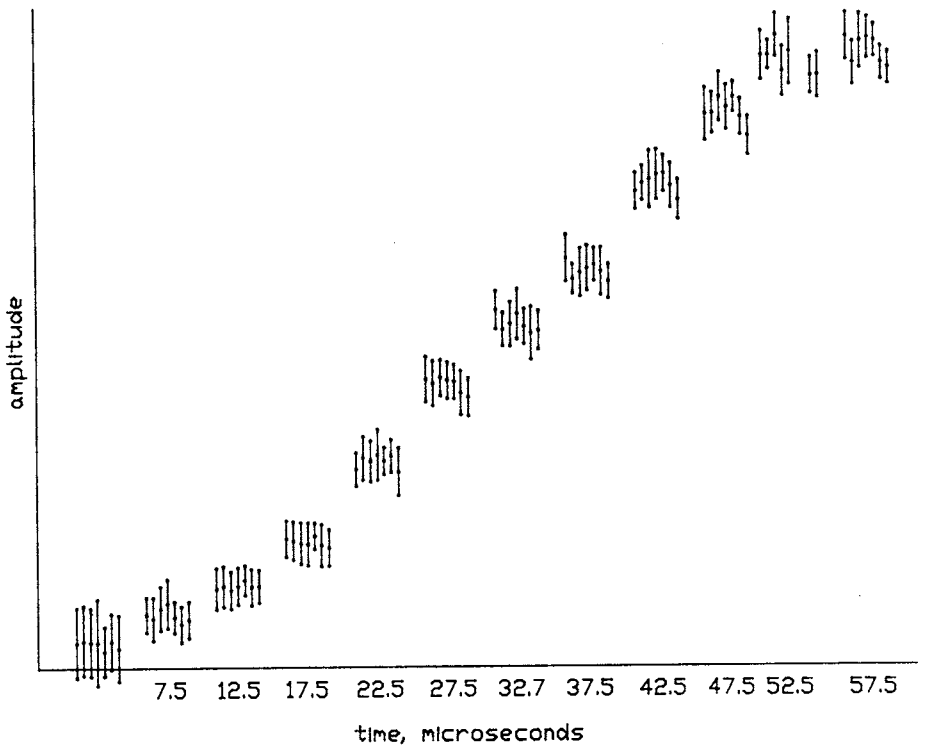


Figure .6: Plot of daily amplitudes of first half-cycles. Estartit, 0 UTC.

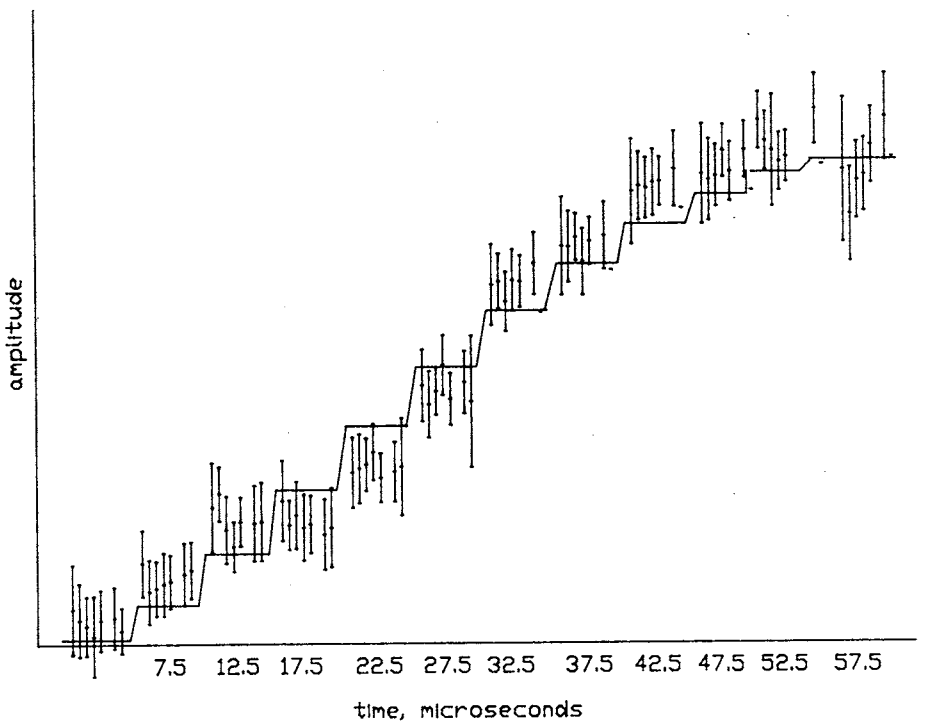


Figure .7: Plot of daily amplitudes of first half-cycles. Sellia, 12 UTC. The reference envelope is added for comparison.

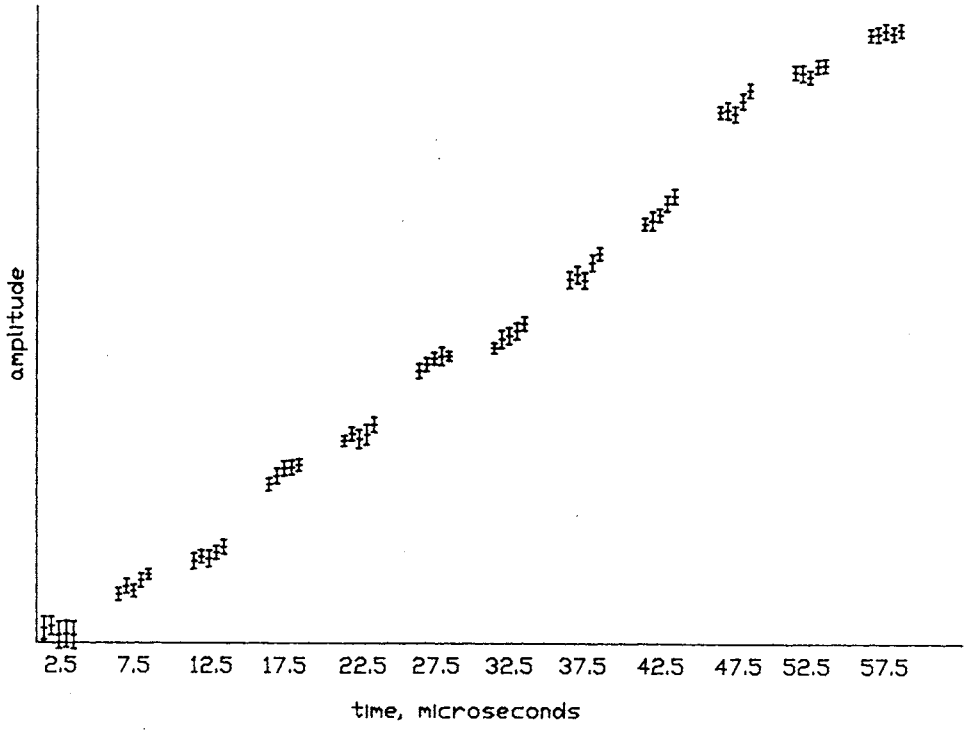


Figure .8: Plot of variations of various half-cycle amplitudes within the train of a single station (Estartit, 11 UTC). Only first five pulses are plotted.

Near the envelope maximum, a sort of flattening is observed, possibly due to saturation problems in final stage.

For other half-cycles, a sort of amplitude increasing is observed. At cycle 4, the variation between the first and fifth pulses are of about 8 %. This effect seems to be systematic.

In other instances, a variation of about 5 % in one hour was observed.

Final Remarks

It is intended to further investigate in these phenomena. At any rate the maximum discrepancies in the envelope shape insofar observed are less than 2 dB. This latter figure is to be compared to the theoretical differences between the third and fourth cycles (receivers usually lock at the beginning of the fourth cycle) that is 3.7 dB.

2 ND EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 16 - 17 - 18 MARCH 1988

Session TF-05 Number 2

A simple LORAN-C timing receiver for supporting VLBI

John E.B. Ponsonby

University of Manchester
Nuffield Radio Astronomy Laboratories
Jodrell Bank

A single station phase tracking LORAN-C receiver has been developed which gives 1 p.p.s. to within 1 μS of the integral second on the LORAN-C time scale. These pulses are generated by division of the 1 MHz drive from the local frequency standard but track the LORAN-C signal in 1 μS steps. The drift of the LORAN signal relative to the local standard produces an analogue triangular waveform of 1 μS range for display for instance on a chart recorder. These outputs are sufficient for supporting Mk II and Mk III VLBI (Very Long Baseline Interferometry).

An external 1 p.p.s. accurate to better than the LORAN GRI (Group Repetition Interval) (0.0797 sec) and knowledge of the propagation delay to the site and of TOC's (Time of coincidence) are required to set it up. The receiver has been developed from a three station tracking receiver designed and built by the author for navigating his yacht.

The RF bandpass is 3rd-order Butterworth with -3 dB points at 90 and 110 KHz defined by a filter in the aerial preamplifier. Two manually tuned notch filters are provided in the RX itself. The salient features of the TRACKING and AGC loops, the timing mechanization and of the manual ACQUISITION circuits are described.

Four receivers have been built. Two are for use in monitoring rubidium standards at Jodrell Bank and its outstations, one is in use in the Dept of Physics, University of Manchester, and one is being installed at another radio observatory to facilitate participation in the work of the EVN (European VLBI Network).

1. Introduction.

At Jodrell Bank we are engaged in Very Long Base Line Interferometry (VLBI) which is a technique for mapping radio sources with angular resolution of order of milliarc seconds. To do it each independent participating radio observatory needs to control the frequency of its receivers to very high accuracy, and the signals need to be recorded against a time scale within a few microsec of UTC to allow them to be played back in synchronism to measure the cross correlation between them. For this reason we need atomic frequency standards and access to UTC at all observatories. At present we at Jodrell use HP rubidium standards and obtain UTC via the Sylt LORAN-C transmitter. My main work is to develop hydrogen masers to supercede the rubidiums. For many years we have had an AUSTRON 2000-C receiver tracking Sylt.

Having become familiar with LORAN-C in this connection I was moved ten years ago to design and build for myself a very simple low power three station tracking time difference receiver for navigating my small sailing yacht. This receiver has served me well in the Irish sea and amongst the Western Isles off Scotland.

A few years ago at a time when our AUSTRON was going through a long period of recalcitrant and intermittent malfunction, the question was raised of how we could get hold of a backup for it and also get several more receivers for use at our MERLIN out-stations and elsewhere. I pointed out that I had to hand the design for my navigational receiver and I undertook to modify it to turn it into a single-station-tracking Timing receiver. It is this receiver that I now describe, and it is shown in Plate 1.

Because the timing receiver has evolved from the navigational version it has several features that one would not have chosen if it have been designed as a timing receiver from the start. The two most prominent such features are that is uses a vertical whip aerial whose omni-azimuthal reception pattern is essential for navigation but is not optimal for single station reception for timing, and that it has automatic gain control (AGC). For navigation each of the three station channels has its own AGC loop, since the three signals can differ enormously in strength, and the receiver proper is time multiplexed round the three loops spending 10ms on each. Fixed gain however would be quite satisfactory for timing and might even be preferable.

2. The receiver.

A block diagram of the timing receiver is shown in Fig 1. The CMOS digital circuits run at 1 MHz which comes from an internal 1MHz VCXO which is phase-locked to an external 1MHz coming from a local external frequency standard such as a hydrogen maser or a rubidium standard. The characteristics of this PLL are such that the receiver can maintain lock on the incoming LORAN signal without an external source of 1MHz, so it is possible to transfer it from one atomic standard to another without

loosing lock and having to go through the rather tedious signal aquisition procedure. The Austron 2000-C doesn't have an internal oscillator, so it cannot have its drive transferred to another source without stopping it completely and that has proved to be extremely annoying.

The 1MHz drives a synchronous decade divider chain which cycles at the Group Repetition Interval (GRI), which is 79700µs for the Norwegian Sea chain, and the states of this counter define a Digital Cyclic Time-Base (DCTB). All the sampling and gating and phase-decoding waveforms required to match the incoming LORAN-C pulses are derived from this counter. A second SECONDS counter divides the 1MHz down to 1pps to provide approximate UTC seconds pulses. Both counters share the first stage of division, and whilst nominally a decade, it can be set from time to time to divide by either 9 or 11 for one cycle, and when this happens the DCTB is advanced or retarded by a 1µs step relative to the incoming LORAN-C signal. Since the SECONDS divider chain shares this first decade its 1pps is likewise advanced or retarded by 1µs at the same time.

The receiver (RX) itself consists of a remote preamplifier where a 1m whip aerial is directly coupled to a tuned ciruit (TC) of Q=5 to provide initial selectivity before the first active device which is a low noise FET. After the FET further TC's shape the overall response to be maximally flat with -3dB points at 90 and 110 KHz. The main gain block is preceded by two manually tunable notch-filters. These are essential because of DECCA chain and other strong signals in the vicinity of the band. The RX has no further significant selectivity, it is all in the preamp.

The phase-decoding of the pulses is done at high signal level at the output of the RX, immediatly before the coherent detecting sampling circuits. An incoherent linear detector feeds a TEN CHANNEL SYNCHRONOUS INTEGRATOR (TCSI) which in turn drives an array of 10 LEDs which provide the indication for initial aquisition of the LORAN pulses. The signal sampling and tracking scheme is shown in the timing diagram of Fig 2. Every incoming LORAN-C pulse is sampled in 3 places, with 1µs sampling windows. The first is at the INDEX point which is on a negative peak of the RF waveform. The second is at the AGC point, 10µs later and also on a negative peak, and the third is at the TRACKING point which is nominally 2.5µs later again on the 3rd positive going zero-crossing of the RF pulse. The TRACKING point sampling pulse adjusts itself to sit at the zero-crossing by means of an analog feedback loop via a LINEAR VOLTAGE CONTROLLED VARIABLE ONE-SHOT (LVCVOS) which has a range of + 0.5µs about its nominal 2.5µs position. If the local external frequency standard runs fast, then the LORAN pulses drift on the DCTB until the LVCVOS hits its LATE 3µs end stop. When this occurs the first decade counter is set to divide by 11 for one cycle, thereby retarding both the DCTB and the 1pps by 1.s and so keeping them in step with the LORAN signals. At the same time the TRACKING point analog integrator is "clawed back" to set the LVCVOS to 2.0µs so that the TRACKING point sampling window continues to sit on the positive going zero-crossing. The voltage going into the LVCVOS traces out a triangular wave each cycle of which corresponds to 1µs of drift of the local standard relative to the LORAN time scale. If the

standard runs slow then the LVCVOS hits its EARLY end stop and 1μs steps are deleted from the DCTB. The result is that the 1pps pulses follow the LORAN-C time scale in steps of 1μs, but the drift of the local frequency standard can be monitored on the analog output with a resolution dependent on the S/N ratio, but certainly to 100ns. This step-wise approximation to the UTC second is adequate in support of MKII and MKIII VLBI.

The signal amplitude at the AGC sampling point is held constant via the AGC loop regardless of the S/N ratio. The idea is that by standardizing the amplitude at the AGC negative peak , the amplitude at the previous INDEX negative peak should then be an indication of onto which RF cycle the receiver is locked, since the build up of the pulse at the RX output is ideally linear with time, and therefore the ratio of amplitudes at successive peaks varies along the pulse. This scheme has not proved as reliable an indicator of cycle identification as I should like, and particularly on the navigation receiver correct cycle identification has been a continuing worry.

The sampling and integrator ciruits are shown in Fig 3. Three CMOS analog switches are used to sample the RF each for 1μs. These charge up "sample and slowly forget" circuits which stretch the samples in time long enough for the subsequent op-amps to have time to respond. The INDEX integrator is a simple smoothing ciruit to indicate the signal level at the INDEX sampling point. The AGC integrator is a true integrator which has a bias input which the mean signal level at the AGC sampling point adjusts itself to equal. The TRACK integrator drives the LVCVOS, which in turn drives a 1μs one-shot which provides the TRACKING point sampling pulse. The sense of the loop is such that the TRACKING point is driven towards the +ve going zero-crossing. Additional inputs to this integrator provide the "claw-back" currents. Claw-back takes ~ 400μs, it being limited by the maximum output current capability of the op-amps. Another current input is provided to SLEW the DCTB during initial signal aquisition.

3. Signal aquisition.

Initial aquisition is effected using the TCSI shown in Fig 4. This consists in effect of a ten-way rotating switch driving 10 HOLD capacitors and each of these, through an amplifier, drives one of the 10 LEDs. In the FAST SLEW mode when the DCTB can be slewed at up to +1ms/sec, the ten positions correspond to 10 contiguous 1ms windows. When, in this mode, the 8 left-hand-most LEDs are bright it indicates that the 8 LORAN pulses of a group are within their correct 1ms time slots. On SLOW SLEW the slew rate is ~+12μs/sec and the windows though still 1ms apart are narrowed to 100μs. When the DCTB is adjusted so that the 8 LH-most LEDs are again bright the DCTB is correctly aligned within a few 10's of μs. Final aquisition is achieved by advancing the DCTB to beyond the beginning of the RF pulses, so that the pulses are moved out of the 100μs windows and the LEDs are extinguished, and then slowly retarding the DCTB till first the LEDs come on again and then by counting the RF cycles up to 3 as they are traced out on the front panel meter. This aligns the

receiver to within half an RF cycle and the TRACK loop pulls it into precise lock. This process enables the receiver to be set up on signals of such poor S/N ratio that they cannot be seen on an oscilloscope.

The TCSI has a reference channel which samples the RX noise level between the LORAN pulses and it is the amplified difference between the signal levels in the individual windows from this reference level which drives the array of 10 LEDs. It is this arrangement together with proper choice of time constants which enables the LEDs to display signals well below the receiver noise level.

4.Setting the 1pps.

Having aquired the LORAN signal and locked the receiver onto the correct +ve going zero-crossing the 1pps SECONDS counter may be set up. An external 1pps signal accurate to better than one GRI (0.0797s in our case), is needed to initialize the SECONDS counter. This is done just as it is with the AUSTRON 2000C, at a TIME OF COINCIDENCE (TOC) or so called NULL SECOND, which occur every 797 sec of time, by setting the SECONDS counter to the known propagation delay of the RX site at the instant of arrival of the correct LORAN pulse. Thereafter "spill over" times of the SECONDS counter occur within 1µs of the UTC second. This initialization is analogous to meshing two gear wheels, one with 979 teeth and one with 10,000 teeth, and with both rotating at 10,000 teeth per sec. At the TOC one particular tooth on one wheel has to be be meshed with one particular slot in the other. Once correctly meshed the 1pps output remain correctly timed relative to the LORAN time scale even if the local frequency standard drifts with time.

5.Conclusion.

Four of these receivers have been built. One is installed at Jodrell Bank to act as back up to the AUSTRON 2000-C. One is to be used in the Physics Dept. of Manchester University for monitoring the master oscillator in a molecular spectroscopy laboratory. One is going to another VLBI Radio Astronomy observatory in Europe and one is to be used at Jodrell's remote MERLIN sites accompanying a portable HP rubidium standard.

Finally I should say that I eventually found the obscure fault in the AUSTRON 2000-C which is now working reliably again.

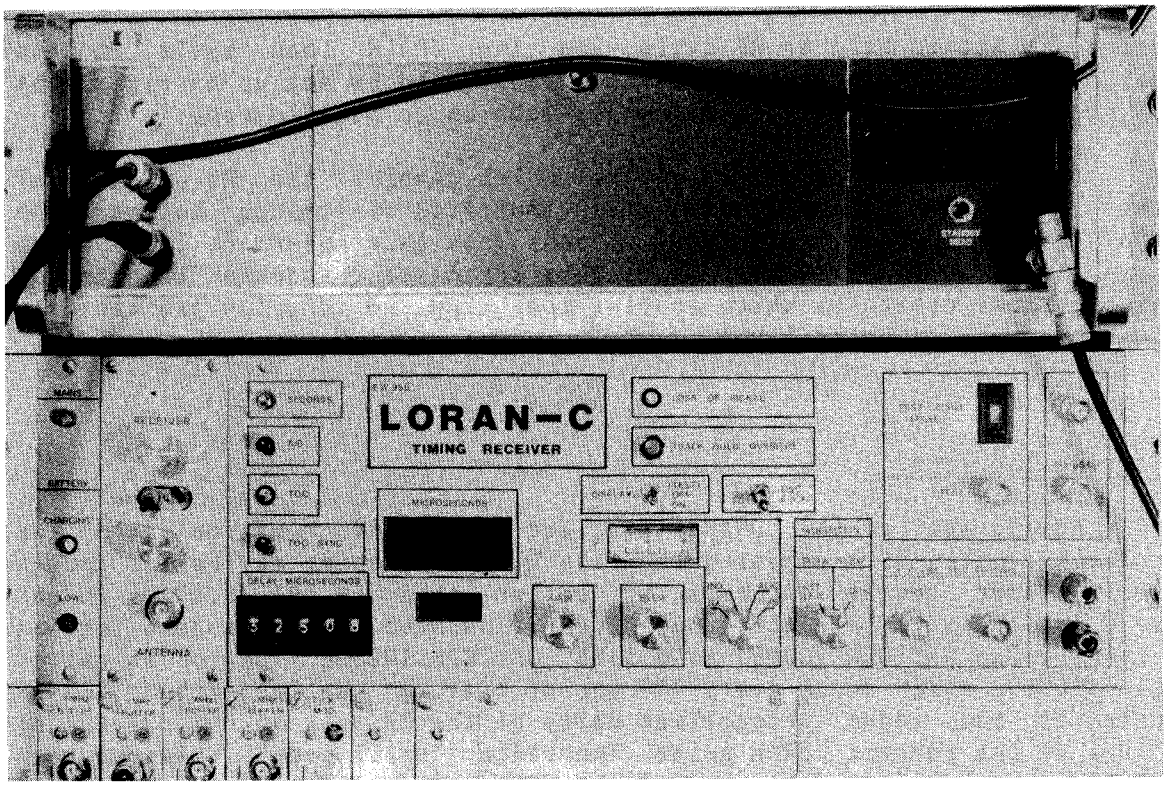


PLATE 1

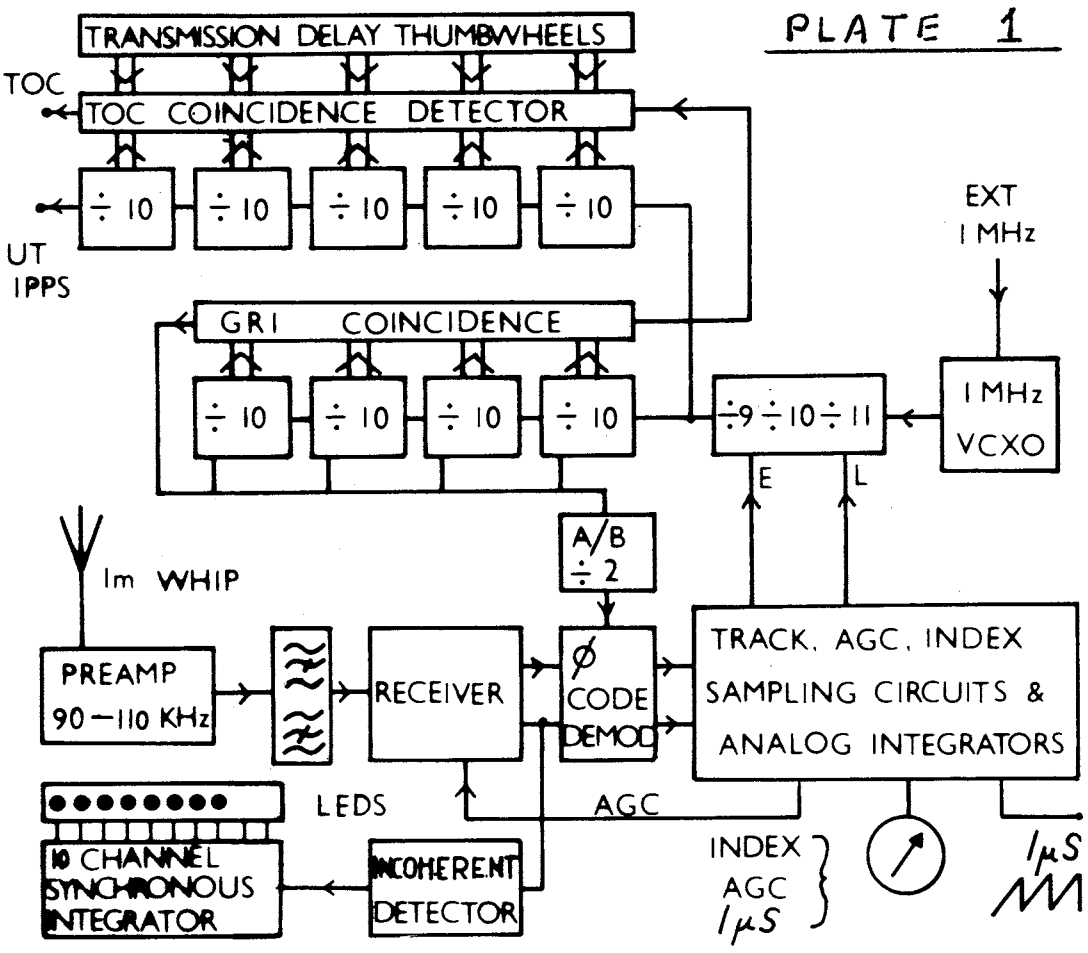


FIG. 1 BLOCK DIAGRAM OF LORAN—C TIMING RECEIVER

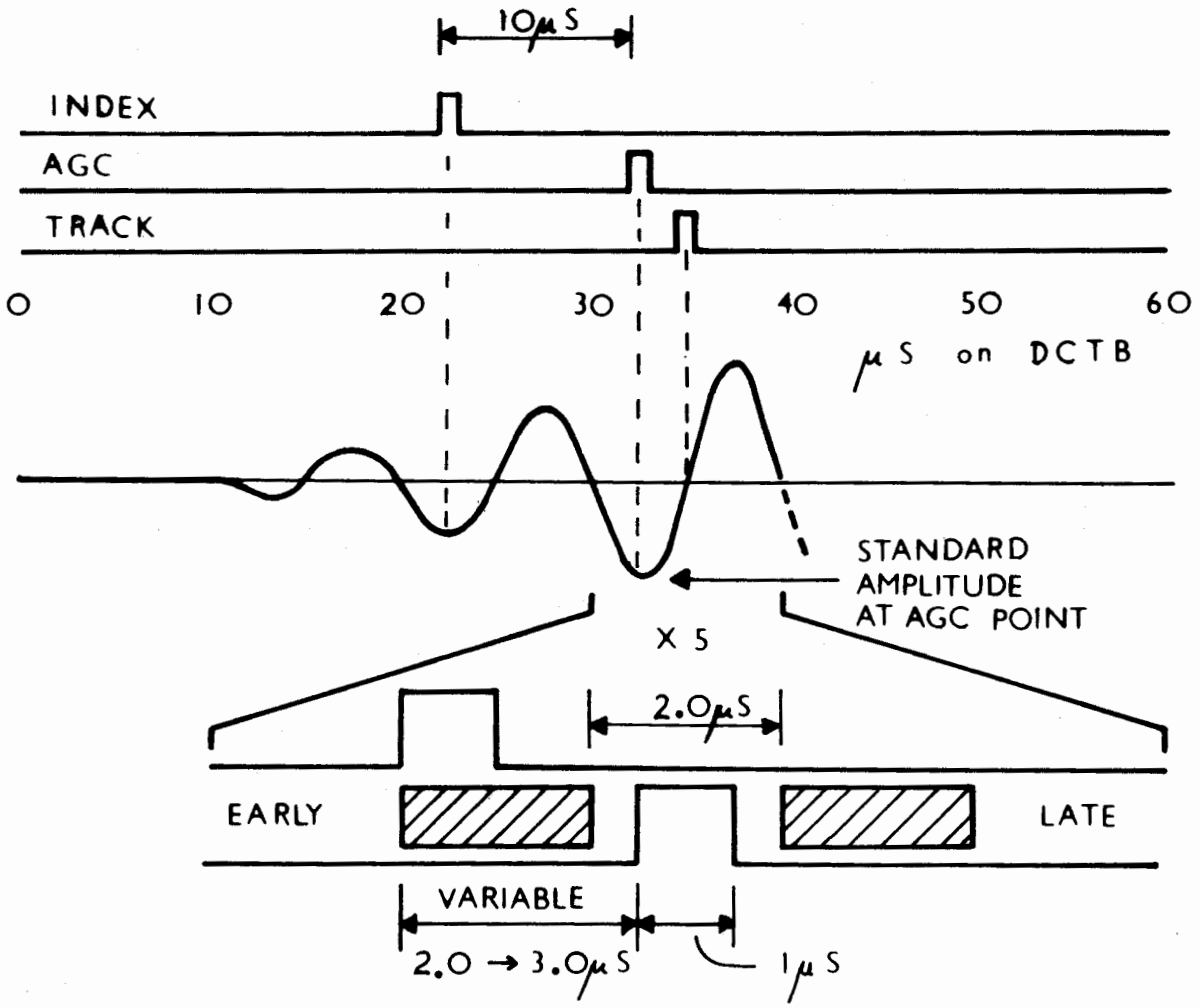


FIG 2 INDEX AGC AND TRACKING POINT SAMPLING PULSES

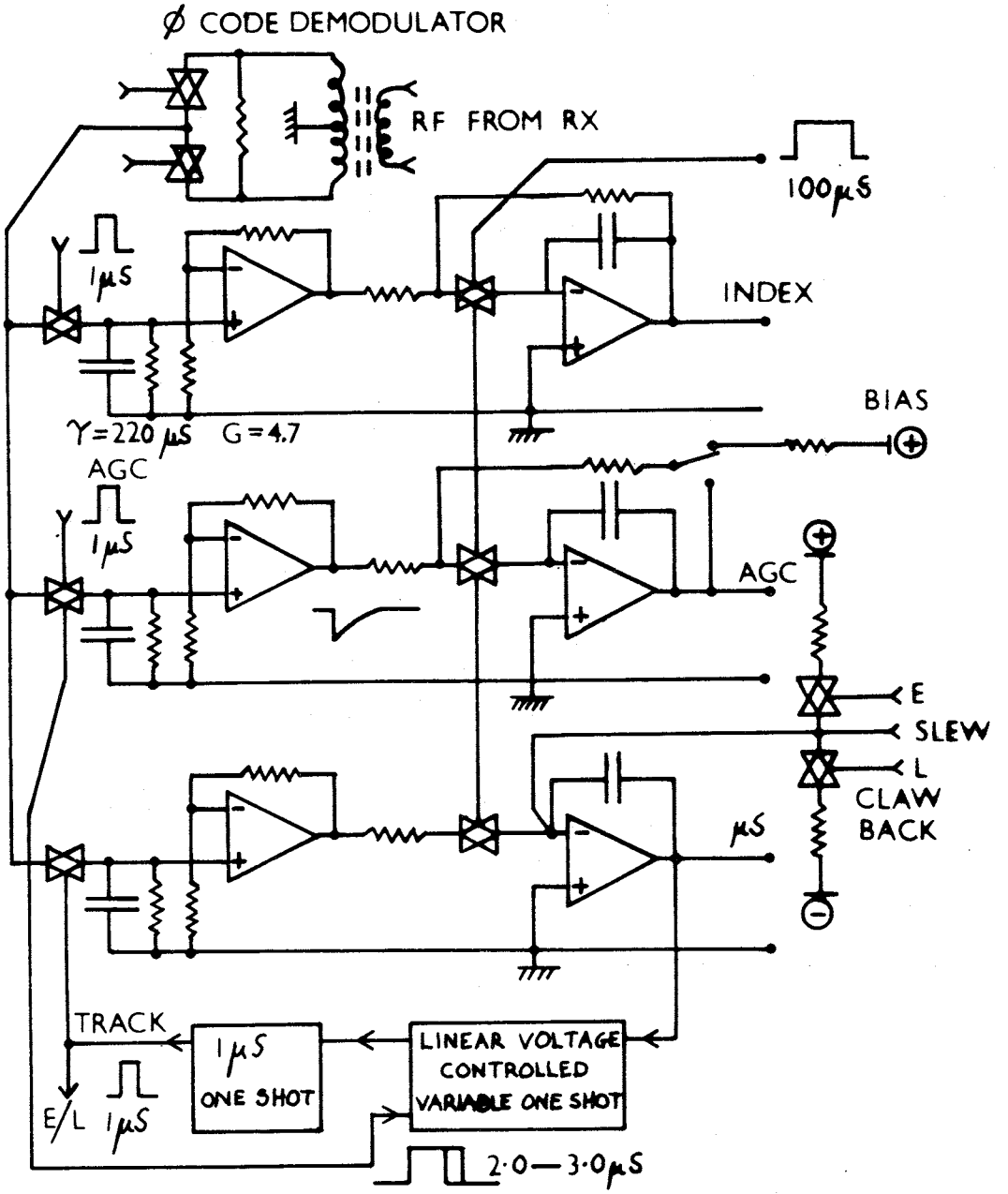


FIG. 3 TRACK AGC INDEX SAMPLING CIRCUITS AND ANALOG INTEGRATORS

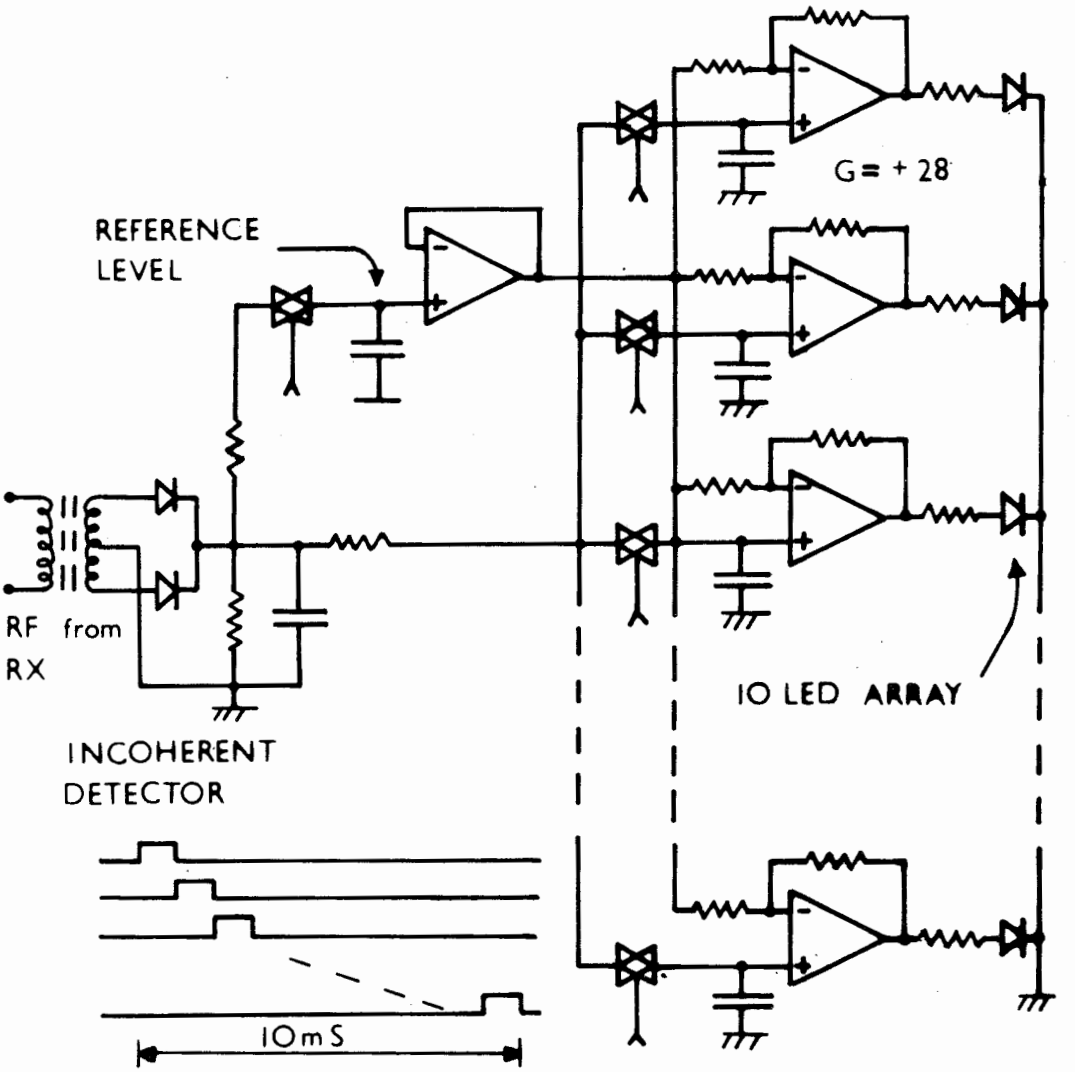


FIG 4 10 CHANNEL SYNCHRONOUS INTEGRATOR FOR COARSE AND FINE SIGNAL AQUISITION

2 ND EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 16 - 17 - 18 MARCH 1988

Session TF-05 Number 3

Time dissemination via the LF transmitter DCF77 using a pseudo-random phase-shift keying of the carrier

P. Hetzel
Physikalisch-Technische Bundesanstalt, Braunschweig, FRG

Summary

In addition to the amplitude modulation by second markers the carrier of DCF77 is phase modulated using a pseudo-random phase-shift keying (PRPSK). The PRPSK is accomplished according to a pseudo-random binary sequence of maximal length with 2^9 states. Therefore the mean value of the carrier phase is not influenced. At the receiver side the arrival time of the PR cycles can be determined by cross correlation. Time comparisons with reference to UTC(PTB) demonstrate the capability of the cross correlation method combined with PRPSK also in the LF band. This technique makes better use of the frequency spectrum available and results in a more precise LF time distribution. The arrival times of the pseudo-random cycles received 273 km from the transmitter fluctuate during day time (except in winter) within a standard deviation of less than half a period (< 6.5 µs) of the carrier.

1. Introduction

Due to a great number of harmful interferences, atmospheric disturbances and man-made noise, amplitude-modulated time signals in the LF band can often be recorded only with narrow bandwidth. Narrow-band reception has, however, the consequence that the accuracy with which the arrival times of the amplitude-modulated time signals are ascertained decreases with the degree of filtering. When the arrival times of the amplitude-modulated time signals received are to be determined with as low an uncertainty as possible, it is therefore necessary to use a broad band. Broad-band reception results, however, in a high liability to interferences. This can in part be compensated for by suitable demodulation techniques, averaging methods and other methods of noise suppression. All conventional reception and averaging techniques for amplitude-modulated time signals have, however, the disadvantage that any phase-synchronous harmful interference which falls within the effective bandwidth results in spurious modulation which superposes the time signals and cannot be separated from them. A technique with which this disadvantage is avoided because it allows various signal components in the same frequency range to be separated and which also excels by a high immunity from disturbance is the cross correlation /1/. The advantages which are known from the theory of the correlation technique can, however, be made full use of only when pseudo-random noise (PRN) signals are used /2/. In contrast to the conventional second markers, the duration of correlation is shorter for PRN signals, which means a higher resolution in time. For this reason, in the following, a correlation method adapted to the transmitter DCF77, combined with PRN signals is described which optimally utilizes the bandwidth available.

2. Cross correlator configuration for time transmission

Fig. 1 shows the principle of a correlator configuration for time transfer. Essential elements of this configuration are the delay network and the multplier with subsequent integrator.

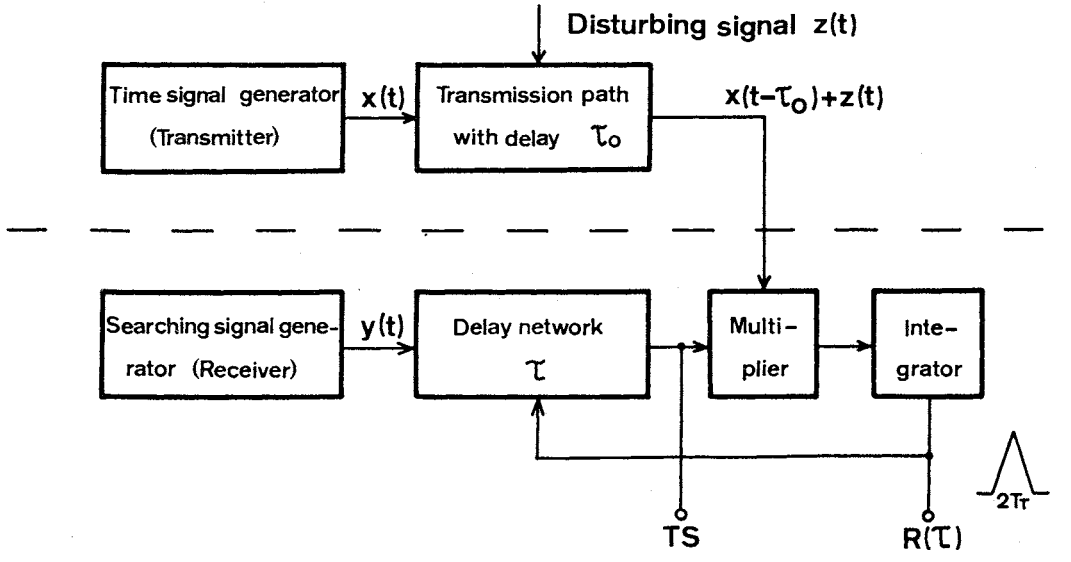


Fig. 1: Simplified cross correlator block diagram: Arrival time determination of the received time signal (TS) by means of the cross correlation function $R(\tau)$.

Mathematically, the measuring operation which is carried out by the cross correlator can be described by the following equation:

$$R(\tau) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{+T} \left[x(t - \tau_0) + z(t)\right] \cdot y\,(t - \tau)\,dt \qquad (1)$$

When the averaging is expressed by a bar instead of the time integral, eq. (1) may be written as:

$$R(\tau) = \overline{x(t - \tau_0) \cdot y(t - \tau)} + \overline{z(t) \cdot y(t - \tau)} \qquad (2)$$

Equations (1) and (2) describe the principle of the cross correlator: On the transmission path, the time signal x(t) emitted by the time signal transmitter is superposed by a harmful interference z(t). Furtheron, it undergoes a delay τ_0 before it arrives at the receiver end at an input of the multiplier. To the second input of the multiplier, via a delay network, the search signal y(t) is supplied which is generated in the receiver and has the same properties as x(t).

In the subsequent averaging of the product signal over a sufficiently long integration time, the second term in eq. (2) disappears when y(t) and z(t) are incoherent for all values of τ and are statistically independent. At its output the correlator thus provides a correlation function $R(\tau)$ which is dependent only on x(t) and y(t) and free from superposed interferences. When the time displacement τ due to the delay network is equal to the delay τ_0 on the transmission path, $R(\tau)$ becomes a maximum and the search signal y(t) tracked by the delay line represents the time information which is emitted by the transmitter and largely free from interferences.

Pseudo-random noise signals have proved to be particularly suitable for the correlation technique. Such signals have a triangular correlation function with a correlation time of the duration of 2 chips. As will be shown in the following, this correlation function allows to easily obtain a control signal with correct sign in order to adjust the delay network with the aid of a control circuit to the maximum of $R(\tau)$.

3. AM-compatible, pseudo-random phase-shift keying of the DCF77 carrier

For DCF77, the modulation with a phase noise is made as pseudo-random phase-shift keying of the carrier according to a binary random sequence. In order to affect the use of DCF77 as a standard frequency transmitter to as small an extent as possible, a small phase swing $\Delta\varphi = 10°$ is used. At a period duration of 12.9 µs, this corresponds to changes in the phase time of $\pm$ 0.4 µs.

For the production of the pseudo-random sequence, a nine-stage shift register is used which is feedbacked according to Fig. 2 and clocked with a clock frequency $f_T = 645.8\overline{3}$ Hz. This clock frequency is as close as possible to the upper frequency limit at which the transmitting antenna is still in a position to follow the phase-shift keyings. It is a subharmonic of the carrier frequency and can easily be derived from it by frequency division (77500:120).

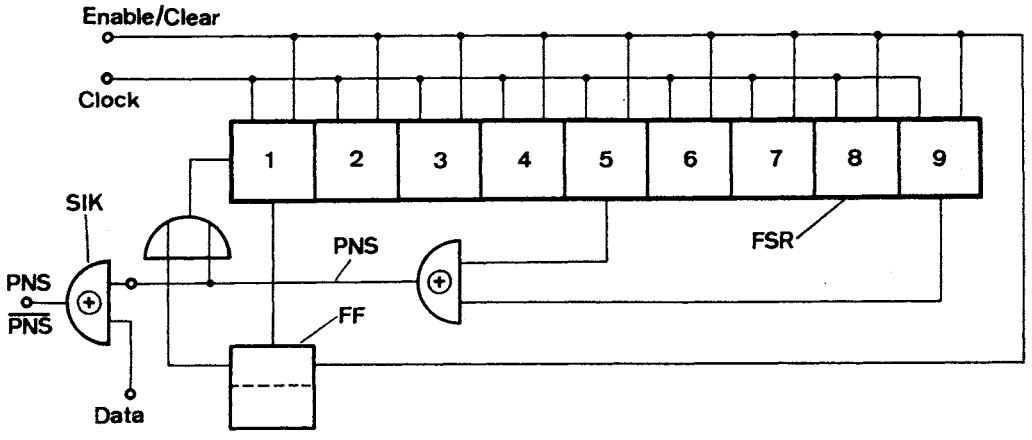


Fig. 2: Nine-stage feedback shift register FSR for generation of the pseudo noise sequence PNS. Flipflop FF forces the FSR out of the "all zero state". Sequence inversion keying (SIK) for data modulation.

With the specified number of stages n = 9 and the duration for 1 chip of $T_T = 1/f_T \approx 1.55$ ms the following values result for the length N and the duration T_K of one noise cycle: $N = 2^9 = 512$, $T_K = N \cdot T_T \approx 793$ ms.

Because of the signal structure of DCF77, the pseudo-random generator is not operated continuously. An enable/clear pulse serves to start it 0.2 s after the beginning of each second marker and to stop it after completion of one complete noise cycle, about 7 ms before the beginning of the next second marker. This method ensures that the falling edge defined as beginning of the second remains undisturbed and the noise cycles fall within the range of the 100 % amplitude also in the case of extended 0.2 s second markers. As in each noise cycle the states zero and one occur equally frequently (256 times each), the mean value φ_m of the carrier phase subjected to pseudo-random phase-shift keying between $\varphi_m + \Delta\varphi$ and $\varphi_m - \Delta\varphi$ also remains unchanged.

The flipflop FF ensures that the pseudo-random generator can start from the state zero. At the beginning of the cycle, FF is set to one so that at the first clock pulse after the release, a one is read into the first register stage. As soon as this one has traversed the first stage, FF is set back to the state zero.

The feedbacked shift register is followed by an exclusive OR gate with which binary data can be added by modulo 2 adding to the pseudo-random sequence generated /3/. In the case of the data state zero, the pseudo-random sequence PNS traverses this data adder SIK without a change, whereas in the data state one, it takes the inverted form $\overline{\text{PNS}}$. With each noise cycle, one bit is transmitted, the binary information transmitted with SIK (sequence inversion keying) being the same as the conventional one which is emitted by pulse duration modulation of the AM-second markers /4/. In the case of SIK, only the minute marker identification is different: Instead of

omitting the 59th second marker, 10 inverted $\overline{PNS}$ sequences are transmitted in the seconds 0 to 9.

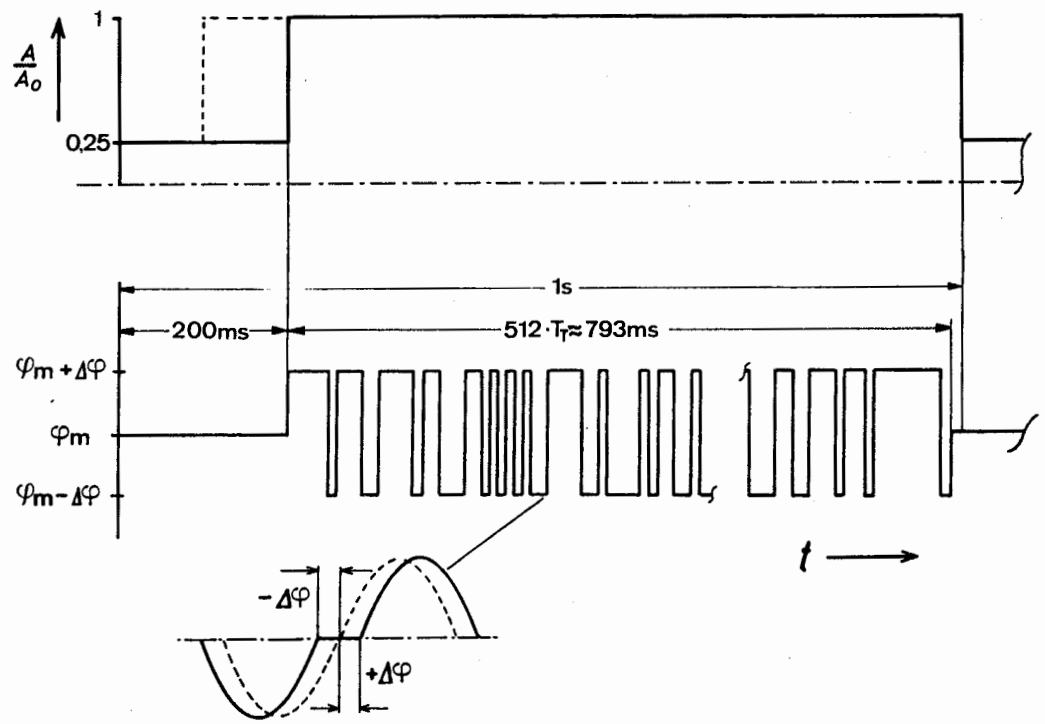


Fig. 3: Amplitude and phase of the DCF77 carrier during one second.

To illustrate the relations described, in Fig. 3 the amplitude response and the phase response are plotted over the duration of a second for the data state zero. To the state "low" at the output of the data adder the phase $\varphi_m + \Delta\varphi$, and to the state "high" the phase $\varphi_m - \Delta\varphi$, were assigned. In the steering signal the phase shift keyings and the amplitude changes at the beginning and at the end of the time markers take place in positive zero crossings of the carrier at the defined times. This ensures that the carrier, the amplitude-modulated time signals and the noise cycles are always phase-locked. At the receiving end, these sudden changes are not, however, recognizable, as phase and amplitude build up only with great time constants.

4. Frequency spectrum of the DCF77 signal

In the frequency domain the spectral density envelope of the DCF77 signal follows a (sinx)/x distribution with minima at multiples of the clock frequency f_T. The envelope is filled with spectral lines whose frequency distance is f_T/N. The amplitudes of the line components are dependent on $\sin\Delta\varphi$ while the carrier amplitude is a function of $\cos\Delta\varphi$ /5/. Figure 4 shows the measured frequency spectrum of the DCF77 PN code. Due to the resolution bandwidth of the spectrum analyser the spectral lines degenerate to a continuous spectral density envelope.

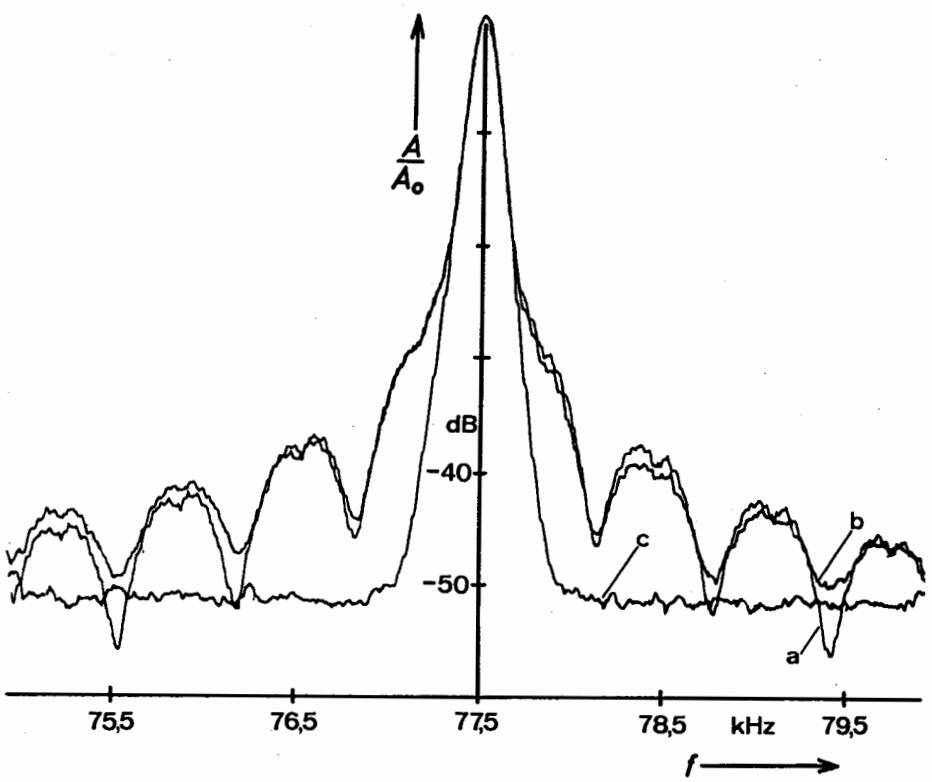


Fig. 4: Spectrum of the DCF77 signal measured with 100 Hz resolution bandwidth at the transmitter station:
a Spectrum of the generated steering signal
b Spectrum of the radiated signal
c Carrier without modulation

5. DCF77-phase noise reception, tracking and data readout

Fig. 5 shows the basic circuit of the receiving arrangement for the evaluation of the carrier noise developed at the PTB. The voltage-controlled quartz oscillator 1 provides the output frequency 6.2 MHz which allows the carrier frequency 77.5 kHz and the decadic frequencies 10 Hz and 1 Hz to be obtained by integer frequency division. After frequency division by 80 in the divider 2, the oscillator signal converted to 77.5 kHz is supplied to the phase detector 3 where it is compared with the DCF77 signal received. The received signal is supplied to the detector 3 via the band pass 5 ($\Delta f = 4\ f_T$) and the amplifier 4. If there is a phase difference between the converted oscillator signal and the received carrier signal, the phase detector 3 supplies a control voltage of corresponding size and sign which readjusts the quartz oscillator 1 through a PI control system element 6 until there is phase coincidence at the phase detector 3. This method allows a voltage to be obtained at the phase detector which on average becomes approximately zero. In the adjusted state it is superposed only by the rapid fluctuations produced by the pseudo-random phase-shift keying. The phase detector output signal thus represents the noise sequence received which is required as reference signal for the correlation control loop in the lower part of the circuit.

The phase detector 8 mixes the oscillator signal shifted by 90° with the signal received. The mixed product is a voltage which is proportional to the envelope of the carrier. The signal shaper 9 following the phase detector, which consists of a comparator, converts the envelope signal into the pulse sequence Z which according to the coding of DCF77 consists of second pulses 0.1 s and 0.2 s long. These circuit elements practically represent a synchronous demodulator for the amplitude-modulated time signals.

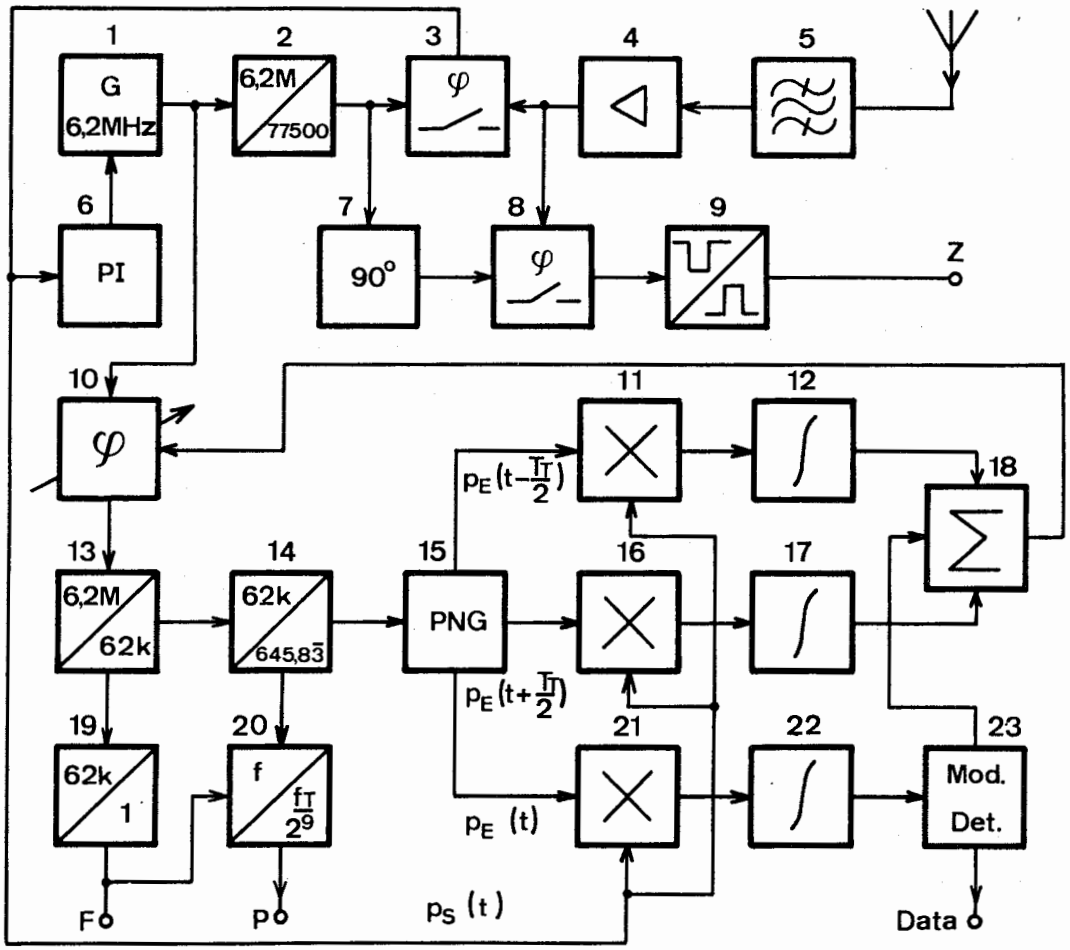


Fig. 5: Block diagram of a delay-lock tracking receiver with data readout.

The lower part of the block diagram represents a delay-lock loop with data readout. The pseudo-noise generator 15 produces three random sequences displaced in time. The two sequences $p_E(t - T_T/2)$ and $p_E\ (t + T_T/2)$ displaced by T_T are multiplied in the multipliers 11, 16 by the sequence $p_S(t)$ coming from the transmitter. For averaging, the product signals are subsequently supplied to the integrators 12, 17. The subtraction 18 of the integrator output voltages finally provides a correction signal to drive the phase shifter 10. If the loop is closed, the sequence $p_E(t)$ is brought to coincidence with the received sequence $p_S(t)$ and the multiplication in 21 and integrating in 22 of both sequences yield a maximum (corre-

lation) or a minimum (anticorrelation) depending on whether a data zero or a data one is being transmitted. At the end of each noise cycle, the modulation detector detects whether correlation or anticorrelation is concerned and outputs the corresponding binary data. Furthermore, in the summation stage 18, it changes the incoming integrator output polarity so that regardless of whether a data zero or a data one is being transmitted, the control voltage will always be of the proper polarity for tracking. The clock frequency for the PN generator 15 is derived from the divider 13 (divided by 100) and the divider 14 (divided by 8·12 = 96).

Divider 19 and binary counter 20 provide the 1 Hz pulses F and P (cf. Fig. 6). These pulses enable and clear the PN generator, and before the beginning of each cycle, they erase the integrators. Furthermore, they cause the data readout at the end of each cycle.

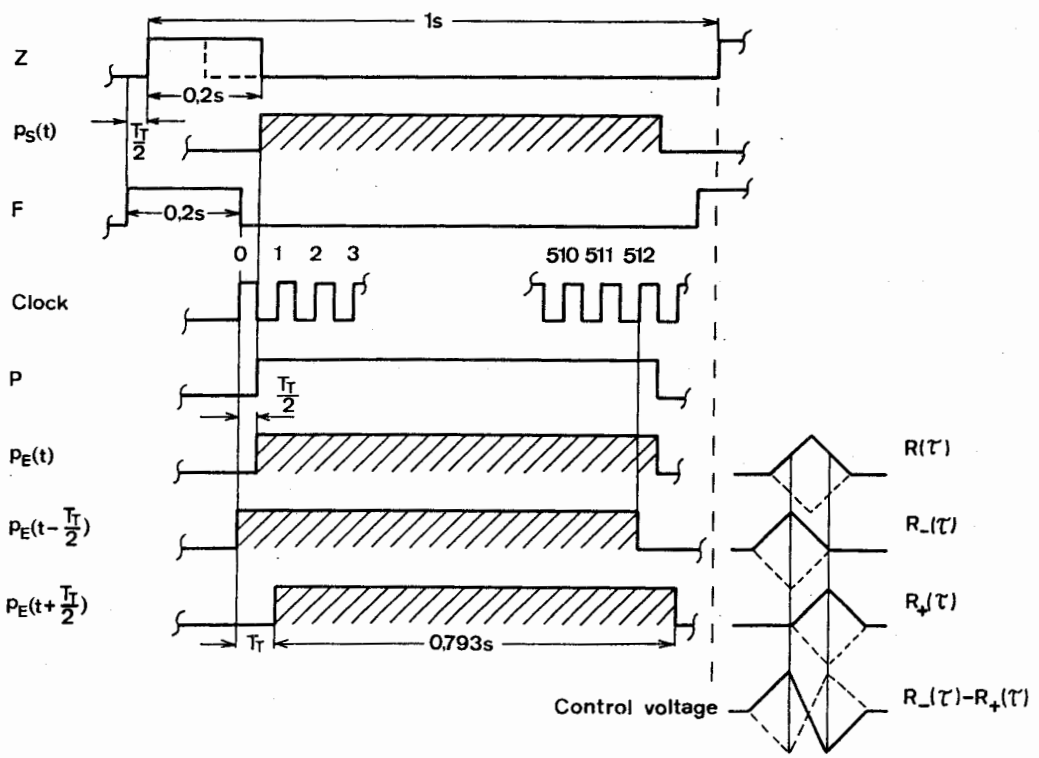


Fig. 6: Timing diagram of the individual pulses Z, F, P and the pseudo random noise sequences generated and received in the delay-lock receiver of Fig. 5.

6. Measurement results

Investigations and measurements carried out at the PTB using the PRPSK method have confirmed that both the binary data can be reliably recognized and the arrival times of the noise signals can be determined with a small uncertainty. The superiority of the pseudo-random noise signals chiefly appeared on days with strong atmospheric disturbances due to lightning discharges. Fig. 7 gives some measurement values. They show the time differences UTC(PTB) - UTC(DCF77) + K measured daily at 11.00 h and 23.00 h over about 200 days in Braunschweig (situated at a distance of 273 km from the transmitter). UTC(DCF77) designates the rising edge of the pulse F defined in Fig. 6 and K means a delay constant. It can be seen that due to the influence of the sky wave, the fluctuations occuring in the night are greater than those occurring in the day-time. It can further be seen that the scatter increases during the day when from about mid-October, the sky-wave share increases. The following table gives some typical standard deviations for various day-times and seasons at a distance of about 300 km from the transmitter:

	Typical standard deviation		
	of an individual value from the daily mean value (10000 s values)	of an individual value from the nightly mean value (1000 min values)	of a daily mean value from the monthly mean value
summer	2 µs	14 µs	4 µs
winter	9 µs	22 µs	10 µs

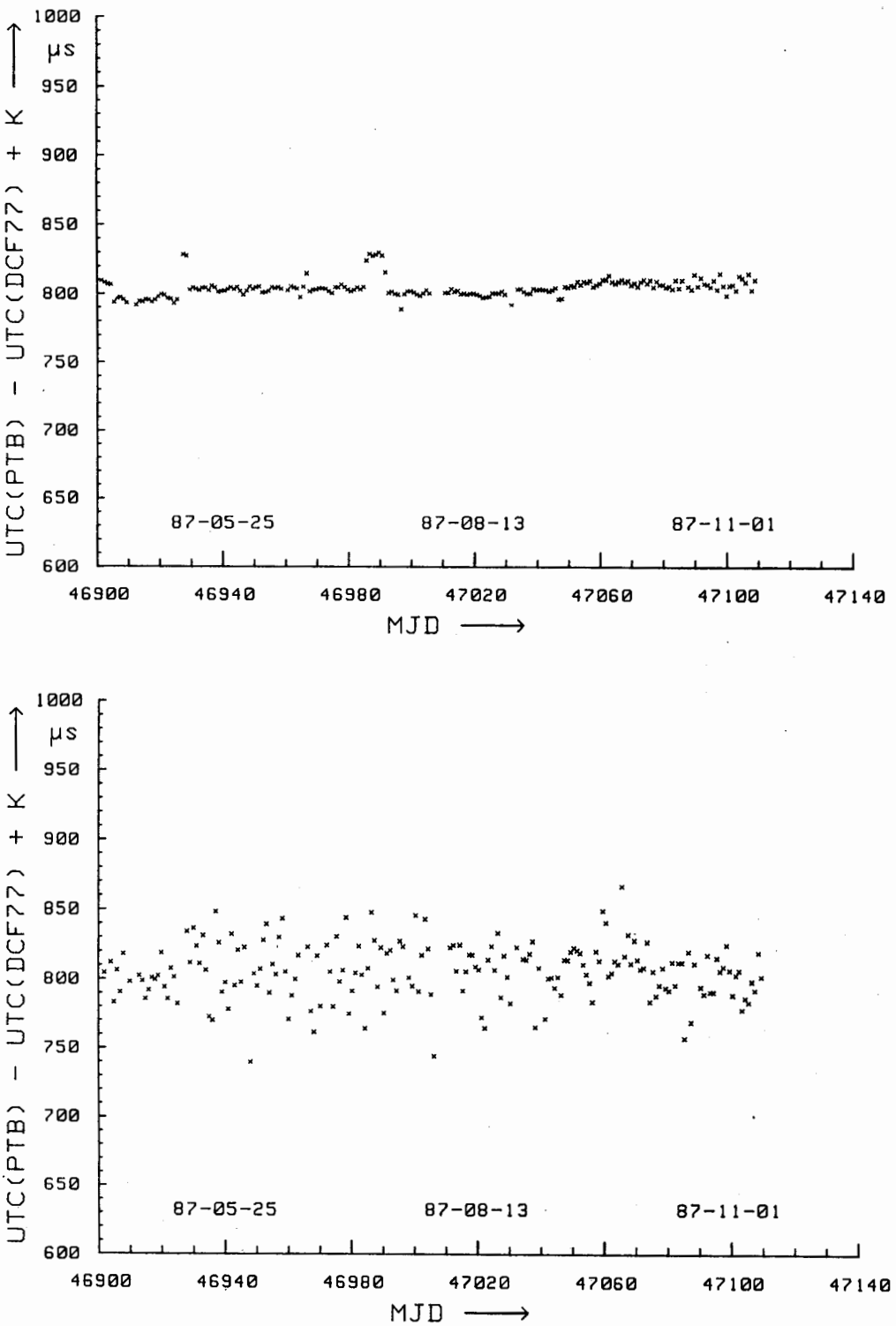


Fig. 7: Time differences UTC(PTB) - UTC(DCF77) + K measured in Braunschweig, at a distance of 273 km from the transmitter.
Upper part: single measurement values at 11 h UTC
Lower part: single measurement values at 23 h UTC

7. Conclusions

The cross correlation technique combined with pseudo-random phase-shift keying of the carrier can successfully be used also in the LF band. This technique makes better use of the frequency spectrum available and results in a more precise LF time distribution and increased reliability. Under usual receiving conditions, the arrival times of the transmitted pseudo-random noise cycles can be obtained with a smaller uncertainty than is possible with the common AM second markers.

References

/1/ Lange, F.H.: Methoden der Meßstochastik
Akademie-Verlag Berlin 1978

/2/ Leonhard, W.: Statistische Analyse linearer Regelsysteme.
Teubner Studienbücher
Teubner Verlag Stuttgart 1973, S. 144 - 158

/3/ Ward, R.B.: Digital Communications on a Pseudonnoise Tracking Link Using Sequence Inversion Modulation.
Spread Spectrum Techniques, Edited by R.C. Dixon
IEEE Press 1976, S. 163 - 172

/4/ Hetzel, P.: Mitteilungen über den Zeitsignal Sender DCF77 und sein aktuelles Sendeprogramm.
Funkuhrtechnik (Hrsg.: W. Hilberg), Oldenbourg Verlag München, Wien, 1988, S. 29 - 45

/5/ Hetzel, P.: Zeitübertragung auf Langwelle durch amplitudenmodulierte Zeitsignale und pseudozufällige Umtastung der Trägerphase.
Dissertation, Universität Stuttgart, 1987

Session TF-05 Number 4

Precise determination of UTC in remote regions using OMEGA signals with reference to monitoring performance of a hydrogen maser

H.J. GAEDE
Hartebeesthoek Radio Astronomy Observatory/NITR
P O Box 3718
JOHANNESBURG
2000
Republic of South Africa

SUMMARY :

Inexpensive but efficient electronic techniques were used to develop a time transfer method based on OMEGA VLF transmissions.

The transmissions are used both as a timing reference and as a transfer instrument, therefore the cooperation of other parties is normally not needed and the method may readily be used in remote locations world wide.

In dedicated operations clocks may be held to within a few microseconds of UTC without much difficulty. Conformity to about 0.5 microseconds can easily be achieved in a regional network of clocks by applying the method differentially.

1. Introduction:

The time transfer method to be described here uses the OMEGA navigation system and was developed at HartRAO (Hartebeesthoek Radio Astronomy Observatory). HartRAO takes part in VLBI operations (Very Long Baseline Interferometry) which require initial synchronization of the clocks at all participating observatories. Knowledge of time to within four microseconds is normally needed, which implies that reference must be made regularly to some superior international reference of UTC.

A few years ago we considered common methods for time dissemination, all of which were found to be lacking in our particular situation. For example :

a) Ground wave signals from Loran C stations are out of range
b) Time transfers by TV could not be used unreservedly as potential partners also had no access to precise time, eg UTC (USNO).
c) International time transfers by travelling clocks were found to be too cumbersome and too expensive for our needs.
d) Time transfers via satellites were considered and only the GPS system seemed to offer prospects for a long term solution. But at that time GPS receivers were both unaffordable and virtually unobtainable.

Against this background, in 1982, a fresh start was made to provide HartRAO with a precise time reference derived from modern format OMEGA transmissions.

We became aware of some practical work with OMEGA timing receivers (1,2,3) done prior to 1977. It seems that interest in that work faded when the first timing satellites appeared. Transmissions on a special frequency (12.85 kHz) on which these receivers depended came to an end when the present OMEGA format was adopted.

Although GPS timing receivers are more readily available now (1988), they are complex and still expensive. There are also some doubts concerning the continuing availability of the GPS services. Therefore we believe that our work may be of interest to the time-keeping community particularly in the southern hemisphere.

2. A new look at OMEGA :

OMEGA is a well established navigation system operating world wide in the VLF portion of the radio spectrum. It is capable of providing navigational services with only eight strategically located transmitters.

Strict controls within the OMEGA network ensure coordination and synchronisation of transmissions. Each transmitter uses five frequencies of which four are shared among all transmitters, while multiplexing ensures that mutual interference is avoided.

An examination of the shared frequencies shows the following relationships :

36 x 283.333 ...	=	10.200	kHz
39 x 283.333 ...	=	11.050	kHz
40 x 283.333 ...	=	11.333	kHz
48 x 283.333 ...	=	13.600	kHz

With a view to the remaining frequencies which are each assigned to one station ("unique" frequencies) this can be taken further :

36 x 17 x 16.666 ...	=	10200	Hz
39 x 17 x 16.666 ...	=	11050	Hz
40 x 17 x 16.666 ...	=	11333.33	Hz
48 x 17 x 16.666 ...	=	13600	Hz

and

n x 6 x 16.666 ...	=	nx100	Hz

These simple relationships show that all OMEGA carriers are harmonics of the fundamental frequency 16.666 Hz. More specifically we may assume that OMEGA signals are five lines of the spectrum ("comb") that is generated by infinitely narrow pulses at 60 millisecond intervals (PRF = 16.666 Hz).

The timing of such pulses at certain points, say at the bases of OMEGA transmitting antennas, may then be used to define the timing of the OMEGA system.

The OMEGA system may thus be seen, not only as a multiplexed CW system but also as a system based on pulses. In the literature we found nothing to support explicitly such a view of the OMEGA system. "Epoch" of OMEGA is occasionally mentioned (4,5) but there is a conspicious absence of operational specifications and the relationship between "OMEGA EPOCH" and "UTC" is nowhere defined.

Nevertheless we now postulate that 60 ms time signals (related to UTC) are contained in broadcasts from OMEGA transmitters.

From our point of view, we need not concern ourselves any further with signals within and around OMEGA stations. We are concerned only with signals available at receiving sites, particularly the electric field of the waves travelling in the dominant wave guide mode. We now assume, rather arbitrarily, that signals emerge at the transmitter antennas with their positive zero crossings coincident with 60ms pulses, which in turn are coincident with UTC.

Such a hypothetical view of the OMEGA system has proved to be a most useful concept for the development of a precise timing receiver with an ambiguity of 60 milliseconds. Ambiguity of timing can be reduced to 30 seconds by evaluating also the RF envelope of OMEGA transmissions. This feature is often of no interest, however, and therefore this option has been omitted from the following description of an OMEGA based time transfer instrument.

3. Description of an "OMEGA Time Transfer Experimental Receiver" (OTTER):

The aim of the OTTER is to extract the timing signal embedded in OMEGA signals. Fig 1 is the block diagram of a receiver, which is essentially a tunable narrow band RF amplifier with a sample-and-hold detector at its output. The detector samples at the 60 ms rate and the timing of the sampling may be shifted for homing in on positive zero crossings. The phase fidelity of the receiver is accurately controlled with the aid of calibration signals which are generated in conjunction with the sampling pulses. The OTTER also contains averaging circuits which may be used as required when the signal to noise ratio becomes very poor.

The antenna unit consists of a whip, an amplifier and a pedestal. It must be placed into open terrain to ensure that the input signals are derived from an undistorted RF-field. Signals from the antenna unit are fed via a cable to the OTTER which is usually operated next to the clock under test.

The acquisition procedure is based on the relationship among OMEGA frequencies discussed in the previous section and may be based on the following stratagem.

Starting with the frequency of 11.333 kHz the sampling point is homed in on positive zero crossings. Continuous adjustments of up to 1/11333.3 = 88.235 microseconds may be needed for this step.

Then the 11.050 kHz signal is tuned in and again zero crossings are sought. In order to preserve the coincidence with zero crossings at the first frequency, the search must consist of a series of steps of 88.235 microseconds (one full cycle at 11.333 kHz) and up to 39 steps may be needed. At the end of that search the time pulse marks the instant of simultaneous zero crossings of two frequencies.

The unique frequency is then tuned in and once more zero crossings are obtained. In order not to loose the point of zero crossings with the first two frequencies the steps now have to be as large as 3.5294 milliseconds and up to 17 steps may be needed. (40 x 88.235 = 39 x 90.498 = 3529.4 ; 17 x 3529.4 = 60 000)

When the acquisition has been completed the sampling pulses mark points of zero crossings of three OMEGA carriers from one station. The remaining frequencies (10.2 kHz and 13.6 kHz) are essentially redundant but can be used to good advantage in this stratagem in an intermediate role. Their inclusion virtually eliminates the probability of blundering.

The acquisitions are somewhat more complicated in actual fact because of the dispersion which OMEGA signals undergo during propagation. The effect is significant but its compensation is easily achieved by advancing or retarding from the sampling point as required. Fortunately dispersions are also remarkably stable and can be predicted well enough without difficulty.

Practical measurements with the OTTER require time interval measurements between the clock under test and 60 ms pulses from the OTTER. Measurements for the various frequencies are made independently but ought to be completed before the effects of clock drift or propagation changes can become significant. Data from these measurements is then typed into the computer (or transferred via an IEEE bus) and the computer makes the acquisition. The error of the clock is then calculated and displayed along with other auxiliary information.

For time determinations the geographical coordinates of the OTTER have to be entered into the computer. We do not generally recommend the normal navigational use of the OMEGA system for finding the OTTER coordinates. This would have entailed the use of two additional OMEGA stations and, in most cases, some unnecessary dilution of timing accuracy. Rather we recommend a method which refers to some landmark of known coordinates. The procedure then calls for entering the distance and the azimuth of the antenna unit as seen from the landmark.

Calculation of distances and delays between OMEGA stations and OTTER positions are performed on the basis of a very simple model of the earth : a sphere with a circumference of 40 000 km and a phase velocity of 300 000 km/s were assumed. Such a model agrees well with actual midday measurements at 11.333 kHz.

The program for calculating the error of the clock under test makes allowance for the fixed delay inside the OTTER and, if required, also compensates for seasonal changes of the propagation velocity.

Finally a value C, which is an overall correction, has still to be added in order to transfer to UTC (USNO).

Values of C ought to be determined experimentally. They correct the deliberate simplifications contained in the model as well as known and unknown phenomena affecting the result of time transfers with the OTTER.

4. OTTER Development :

In January 1983 we obtained results for the first time from OMEGA Reunion and OMEGA Liberia which differed almost exactly by the theoretical difference of propagation times from these two transmitters. This seemed to prove that we had acquired the same epoch criterion in both signals and that our method was working as intended. This kind of proof, as it is independent of knowledge of UTC, became the standard procedure for verifying OTTER time transfers.

Furthermore, results clearly vindicated not only our methods but also our criterion of epoch : UTC preceeded the 60 ms pulses from the OTTER by almost exactly the expected propagation delays between transmitters and the location of HartRAO.

On September 07, 1983 we linked the whole system to UTC with the aid of a cesium clock flown into South Africa; the values for C we obtained then have been vindicated again and again by successful VLBI operations (6).

We were then able to satisfy the specific needs of radio astronomers at HartRAO and began to look for formulations and specifications by which the performance of our method could be presented with a view to more general applications.

We concluded that it would be useful to conduct a field test away from the laboratory environment of HartRAO and that a tentative survey of the spread of C over some large area should be attempted.

In the event we spent twelve days travelling 7 300 km throughout South Africa and made single day observations at six selected positions comparing time on a cesium clock with time obtained from OTTER measurements. Fig. 2 and Fig. 3 show lines of equal C (C contours) on a map also indicating the positions at which these observations were made. Coherence of C contours is clearly demonstrated; had there been significant errors the disruptions of smoothness would have been most obvious.

Measurements were made normally at times when the propagation path is fully illuminated by the sun. At such times the propagation delay goes through a broad minimum and is also more stable than at other times of the day.

For our work, we found the OMPUB 224(XXX-C)X series of OMEGA propagation correction tables quite useful. These tables, available from the U S Defence Mapping Agency, provide seasonal and diurnal profiles (24x24 points) of timing variations for about 4000 defined regions of the world. They also identify areas and times of unreliable reception due to unwanted propagation modes. We used these tables mainly for estimating seasonal effects and for finding the hours best suited for time transfer measurements.

Finally, from an operators point of view two kinds of operating modes have emerged : "time determination" concentrates on initial acquisitions and occasional checking of time scales, whereas "time keeping" concentrates on the refinement of accuracy of well established clocks. The latter employs statistical methods and the measurements are confined to a few averaged measurements of one frequency (11.333 KHz) at certain times of the day.

5. OTTER measurements with a hydrogen maser :

Since 1985 the HartRAO clock has been derived from a hydrogen maser and it is now much easier to distinguish between deficiencies of the OTTER system and deficiencies of the HartRAO clock.

In the course of measurements we soon discovered that the ageing parameters of the hydrogen maser became measurable. On the basis of experience we now find it convenient to make calculated frequency adjustments at 90 day intervals, thus keeping the mean maser frequency in agreement with the clock rate of UTC (USNO). As a result we obtain a distinct pattern in the clock error curve consisting of parabolic segments which maintain an acceptable timing error.

Fig. 4 is an example of recent measurements. With the adjustment on DOY 320/1987 we aimed for a minimum frequency error for JAN/FEB 88, anticipating a campaign of VLBI experiments which were scheduled for that time of the year.

A maintenance shutdown of the Reunion transmitter is seen on DOY 243/87. When the transmitter came back into service one month later, there was clearly a need for some adjustments to return it to the "correct" phase value. From DOY 286/87 to DOY 5/88 the traces of both Reunion and Liberia are typical for normal operation of the OTTER system, some small divergence of slopes is probably due to seasonal effects. Around DOY 6/88 the Reunion signals were advanced for reasons not obvious at the time of writing.

Fig. 4 also shows a second order polynomial fit through the first section as calculated from the Liberia transmission. Data from Reunion, which is usually less noisy and less prone to seasonal effects, was not used in this instance because of the shutdown for maintenance during September 1987. Although data was relatively poor in this example, the equation of the fit describes the ageing of the hydrogen maser well enough for most practical time keeping applications.

The HartRAO clock was also compared with the time scale of the National Physical Research Laboratory (NMSUT) in Pretoria, which is now monitored via GPS . The comparison was made by TV based time transfers using a common view TV transmitter. The ageing of the HartRAO clock as seen by GPS was the same as that seen by the OTTER system. A small offset (less than 2 microseconds) was probably due to TV link errors, seasonal effects and adherence to non-identical time scales (NBS, NMSUT and USNO).

6. System performance :

As the OTTER is not a commercially developed system it would be both improper and premature to lay claim to some definite specifications, but the following comments are in order :

Navigational use of OMEGA implies continuous operation in moving craft and the use of at least three transmitters. Uncertainties of a few miles are usually quoted for navigating, and there are seldom incentives to improve upon such performance. Corresponding uncertainties of timing would mostly be unacceptable, but performance of a well suited OMEGA station, as perceived by a stationary time keeper, is indeed **v e r y** much better than might be suggested by specifications for navigational applications.

At HartRAO, within preferred operating conditions already described, we normally have long runs of good time keeping data in which RMS deviations from the running average are of the order 0.5 microseconds or less. One must assume that the propagation conditions, the OMEGA stations and the clock under test are well behaved at such times. Such performance seems to be a baseline from which more detailed specifications for various applications can be derived.

It goes without saying that the C values (with due regard to seasonal effects) ought to be determined with the greatest possible care : remaining uncertainties of C are translated directly into the statements describing the offset of the clock under test with respect to UTC.

The effects of the uncertainty of C can be excluded in many applications, however. Simple differential measurements may be used when compliance with UTC is not needed but precise synchronisation within some network of clocks has to be achieved and maintained. Measurements over small distances have shown that residual errors are as small as the uncertainty of one count occurring in the time interval counters (100 nano seconds).

With increasing distances between clock sites, the timing errors begin to depend only on the differential errors contained in the contour maps. Without using contour maps, spreads of less than one microsecond ought to be achievable almost at any time anywhere for time transfers between surveyed points not more than 50 Km apart.

OMEGA transmitters are normally phase-stepped every four hours to maintain synchronism with UTC. Such steps are imperceptible to OTTER users, but larger steps of the order of one microsecond are seen in exceptional cases and may be cause for concern to time keepers. By monitoring a second OMEGA station or other independent checks, it is usually possible, however, to distinguish such steps from "jumps" in the clock under test. The occurrence of such steps seriously limits the knowledge of time attainable by OTTER users in remote regions.

Overall adherence of OMEGA to UTC is very well enforced, however : "at any time any OMEGA station is likely to be synchronised to UTC as represented at the USNO to within one or two microseconds"(4). Monitoring reports (8) and our experience with VLBI observations suggest that this is being achieved.

We have observed some inconsistencies over the years in the difference between signals from OMEGA Reunion and OMEGA Liberia which cannot be explained as seasonal effects. The changes are of the order of one microsecond (from one year to the same season in the next year) and it is an open question whether they were due to long term propagation effects or transmitter timing.

Performance of the OTTER system is sometimes impaired by phenomena well known from navigational OMEGA applications (sudden ionospheric disturbances (SID), severe static, transmitter breakdowns, etc). The regular OTTER user will learn to recognise these phenomena and would allow a postponement of measurements for a day or two whenever they occur. The shutdown for maintenance, sometimes for weeks, of a preferred OMEGA station is to the regular user probably the most troublesome of all interruptions.

Finally there should be a hint about prices : A commercial OTTER-like product would cost little in relation to comparable time transfer systems and may interest time keepers who cannot afford expensive systems which are on overkill for their particular application.

7. Acknowledgements :

The author wishes to express his thanks and appreciation to the management and staff of the Hartebeesthoek Radio Astronomy Observatory who gave their able support to this work in numerous ways. Special thanks is due to Messrs Grinaker Electronics Agencies (Pty) Ltd. who arranged the loan of a cesium clock from Oscilloquartz SA of Switzerland, for the field tests.

8. References :

"Literature supporting OMEGA is extraordinary voluminous and will be found scattered in a number of areas including the geophysical, propagational, electronic and statistical literature as well as within the specifically navigational literature." (4)

We would like to draw attention to items (4, 7, 9) for background information, other items (10, 11, 12, 13, 14) address matters more closely related to the subject of this paper.

1. Wilson, J.J., Britt, J.E. and Cassellman, C.J. "Omega Timing Receiver Design and Results of Preliminary Tests Conducted October-November 1972", NELC Technical Note 2340, 4 April 1973.

2. Wilson, J.J., Britt, J.E. "Omega Timing Receiver (Mod 2, Superheterodyne) Technical Description", NELC A 106, 19 March 1975.

3. Chi, A.R. and Wardrip, S.C. "Clock Synchronization Experiments using OMEGA Transmissions", Proceedings of Fifth Annual NASA and Department of Defense Precise Time and Time Interval (PTTI) Planning Meeting December 4-6, 1973, Goddard Space Flight Center, Greenbelt, MD.

4. Swanson, E.R., "Omega" Proceedings of the IEEE Vol. 71, No. 10, October 1983.

5. Swanson, E.R. and Kugel, C.P. "Omega Synchronization and Control", NELC/TR 1757, 19 March 1971, Naval Electronics Center, San Diego, California. 92152.

6. Nothnagel, A. "Report on Geodetic Mark III Campaign, January/February 1986", HartRAO/NITR/CSIR, P O Box 3718, Johannesburg 2000, South Africa.

7. U S Department of Transportation, United States Coast Guard "Omega Global Radionavigation, A Guide for Users", COM DT INST P16566.2 November 1983.

8. U S Naval Observatory, Washington, D.C. "Time Service Announcement Series 4, (Daily Phase Values and Time Differences)" (Weekly Publication).

9. Internatioal Omega Association "Bibliography of Omega Publications", I.O.A., P O Box 2324, Arlington, Virginia 22202, U S A.

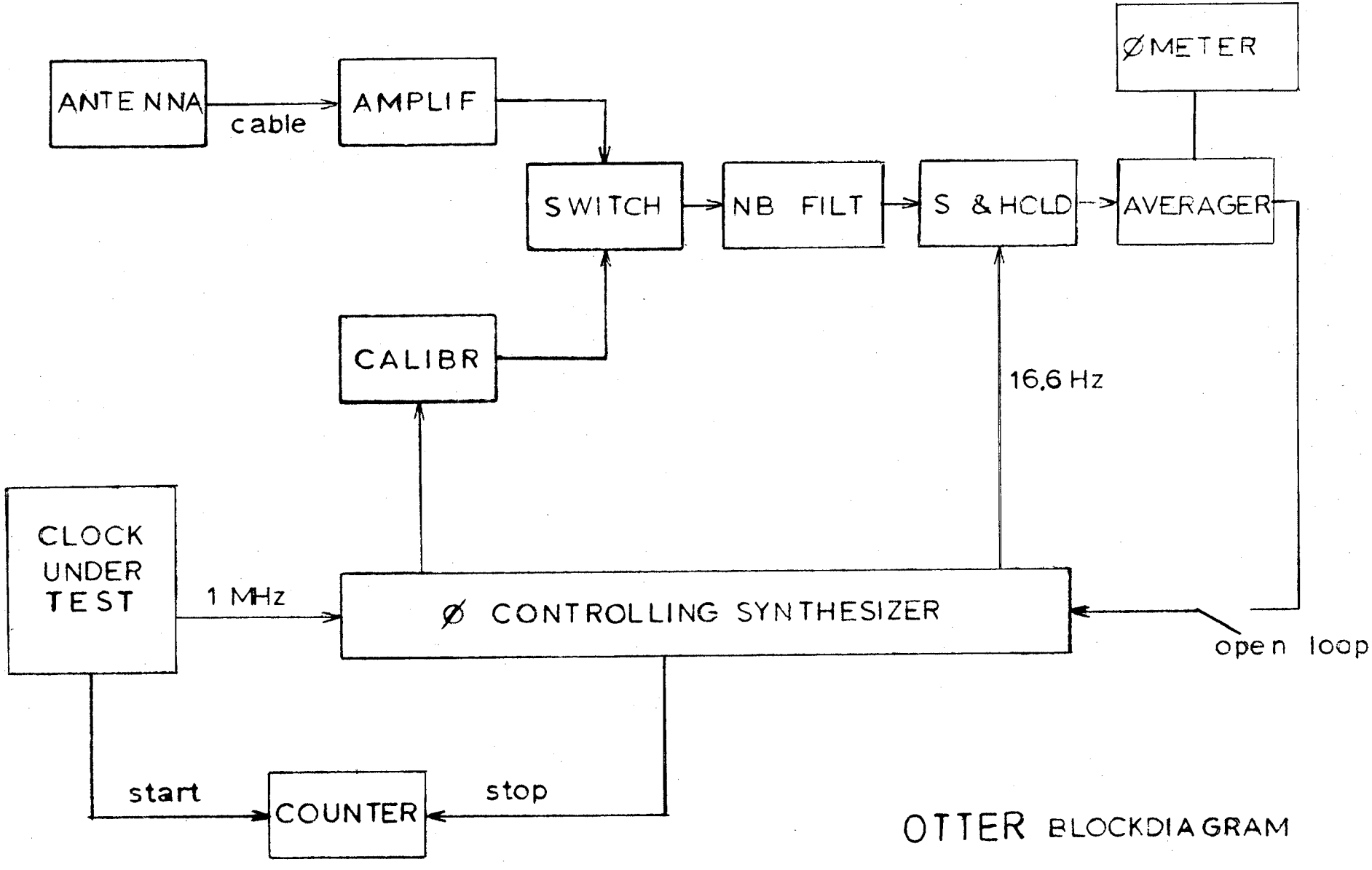
10. Lavanceau, J.D. "Omega Signals and Universal Time" presented at Omega Symposium Tokyo, Japan, 12 June 1975.

11. Klepczynski, W.J. "Modern Navigation Systems and their Relation to Time Keeping", Proceedings of the IEEE, Vol. 71, No. 10, October 1983.

12. Swanson, E.R. "Interrelation of Navigation and Timing" JIETE, Vol. 27, No. 11, Nov 1981.

13. Schane, R.N. "Linear, Dynamic Estimation and Control of OMEGA Radio Navigation System Synchronization", IEEE Transactions on Automatic Control, Vol. AC-22, No. 2, April, 1977.

14. Scientiae "Precise Time Transfer", Publicity Division of the CSIR, P O Box 395, Pretoria 0001, South Africa, Vol. 25, No. 3, September 1984.



OTTER BLOCKDIAGRAM

Fig 1

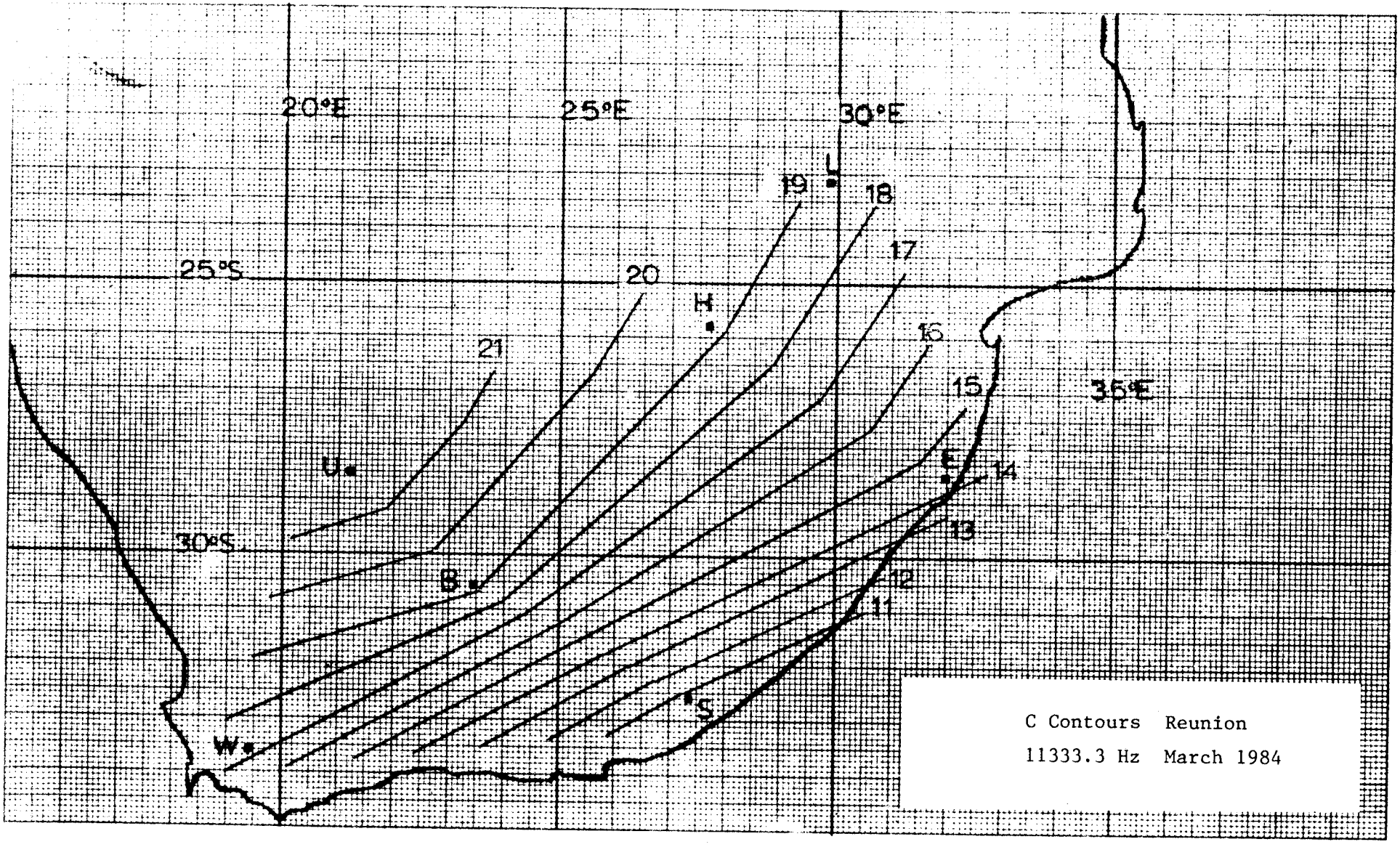

20°E
25°E
30°E
35°E
25°S
30°S
19
18
17
20
21
16
15
14
13
12
11
U
H
E
B
W
S
C Contours Reunion
11333.3 Hz March 1984

Fig 2

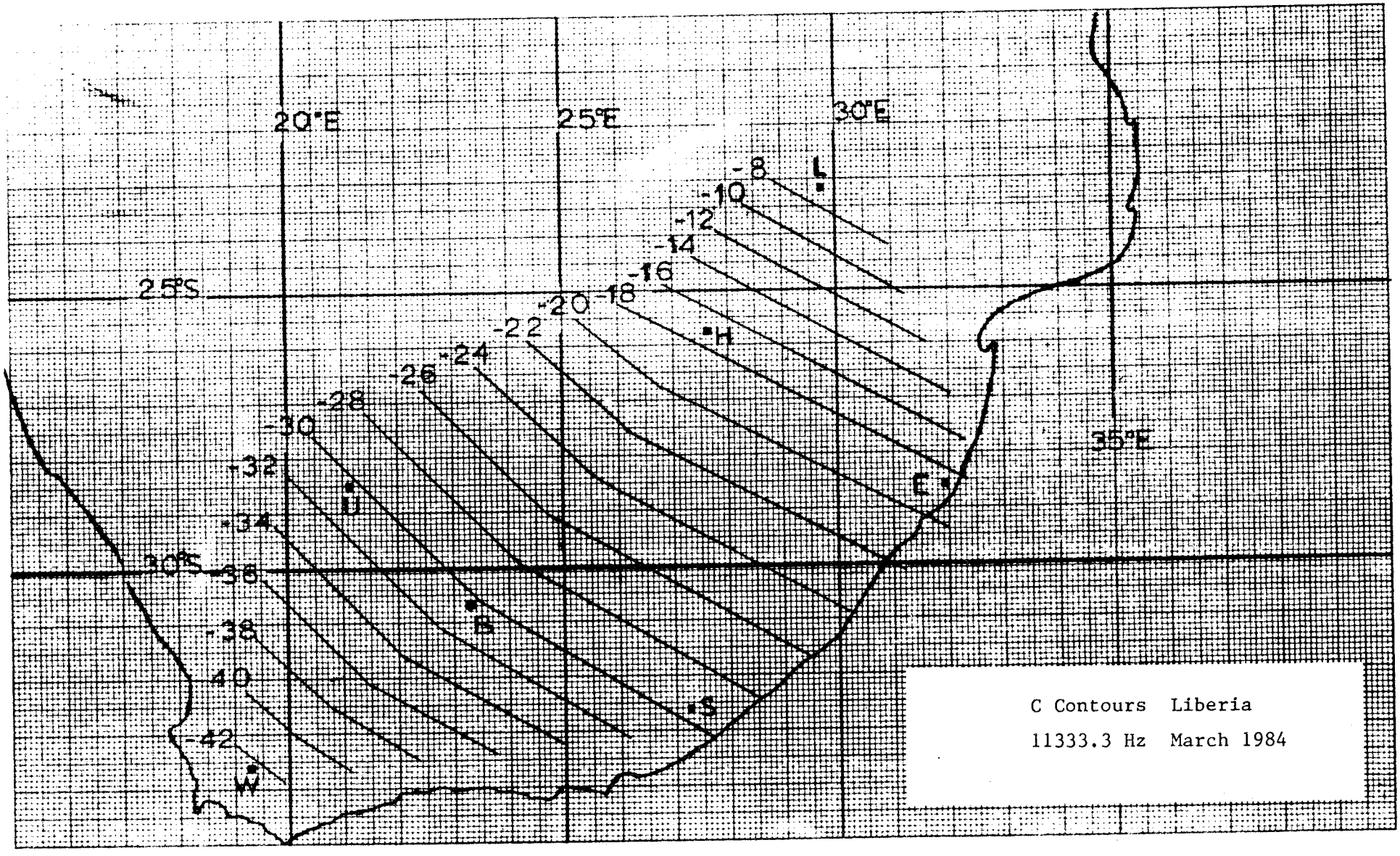
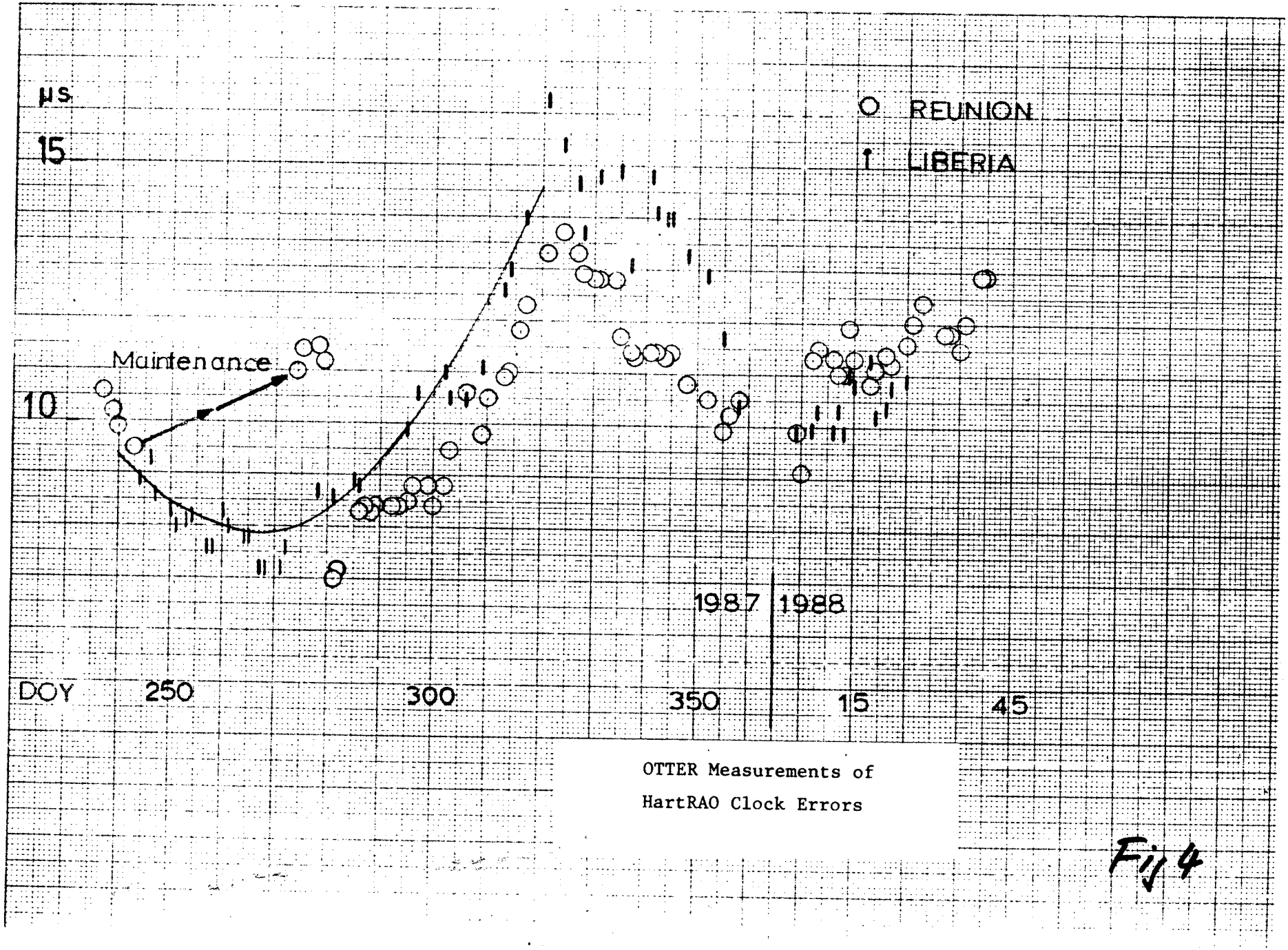

20°E
25°E
30°E
35°E
25°S
30°S
-8
-10
-12
-14
-16
-18
-20
-22
-24
-26
-28
-30
-32
-34
-36
-38
-40
-42
L
H
E
J
B
S
W
C Contours Liberia
11333.3 Hz March 1984

Fig 3



OTTER Measurements of
HartRAO Clock Errors

Fig 4

Atomic Frequency Standards

		Chairman
AFS-01	Various atomic frequency standards	H. de Boer
AFS-02	Cesium Beam Frequency Standards	C. Audoin
AFS-03	Optically Pumped CS Beam Frequency Standards	G. Busca
AFS-04	Atomic Frequency Standards	S. Leschiutta

AFS-01

Various atomic frequency standards

1. The Recirculating Cryogenic Hydrogen Maser*
   M. Hürlimann, W.N. Hardy
   University of British Columbia, Vancouver - Canada

2. Studies of Dielectrically Loaded Cavities for Small Size Hydrogen Masers
   N. Yahyabey, P. Lesage, C. Audoin
   Laboratoire de l'Horloge Atomique, Orsay - France

3. Doppler Effect in Mercury Ion Clock
   C. Meis, M. Jardino, M. Desaintfuscien
   Laboratoire de l'Horloge Atomique, Orsay - France

4. Miniaturization Versus Performances in Rubidium Frequency Standards
   G. Busca, L. Pan
   Oscilloquartz SA, Neuchâtel - Suisse

5. Performance of Disciplined Rubidium Oscillators*
   G. Huebner
   Ball Efratom Elektronik GmbH, München - FRG

6. Frequency shift of an atomic or molecular resonance, bound to the interaction time between an atom and the electromagnetic field
   J. Balbi
   Laboratoire d'Hélioénergétique, Université de Corte - France
   D. Gillet
   Laboratoire de Piézo-électricité et d'Electronique, La Bouloie, Besançon - France
   G. Lomaglio, J.G. Théobald
   Laboratoire de Spectoscopie Hertzienne et d'Electronique, La Bouloie, Besançon - France

2 ND EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 16 - 17 - 18 MARCH 1988

Session AFS 01 Number 1

The Recirculating Cryogenic Hydrogen Maser

Martin Hürlimann and Walter N. Hardy

Department of Physics, University of British Columbia, Vancouver, B.C., V6T 2A6, Canada

A new type of hydrogen maser, operated at low temperatures, has been developed and successfully tested at the University of British Columbia. It has the potential to improve the frequency stability of the best conventional hydrogen masers by 2 to 3 orders of magnitude. In this maser, which operates at .5 K and uses superfluid liquid helium coated walls throughout, H atoms circulate back and forth between a microwave pumped state selector and the maser cavity. We describe the design, theoretical limitations and the performance of our prototype maser which already achieved a frequency stability of better than 10^{-13} for a measuring time of 1 second.

2 ND EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 16 - 17 - 18 MARCH 1988

Session AFS-01 Number 2

STUDIES OF DIELECTRICALLY LOADED CAVITIES FOR SMALL SIZE HYDROGEN MASERS.

N. YAHYABEY, P. LESAGE and C. AUDOIN

Laboratoire de l'Horloge Atomique
Equipe de Recherche du CNRS associée à l'Université Paris-Sud
Bât. 221 - Université Paris-Sud
91405 ORSAY - France

SUMMARY

A dielectrically loaded cavity, made of a hollow alumina cylinder, used in order to reduce the hydrogen maser size is considered. We report the results of a theoretical study of the effect of the cavity geometry on the frequency stability capability of the maser. The active mode as well as the passive mode of operation are considered and the achievable performances are compared.

1. INTRODUCTION

In previously published analytical studies of the effect of internal thermal noise on the frequency stability of the hydrogen maser, the conditions for achieving the optimum frequency stability have been derived[1,2].

In the work reported here, we apply these conditions to specify the influence of the microwave cavity geometry on the frequency stability, in the case where the cavity is made of a hollow alumina cylinder. Such a dielectrically loaded cavity is used in reduced size hydrogen masers [3,6].

Various parameters of interest, like the filling factor and the cavity quality factor are computed for several cavity diameters and lengths, in order to show the effect of the cavity geometry on the level of the white frequency noise. Charts showing the expected power spectral density of the maser white frequency noise versus the radius and length of the cavity are presented for each possible mode of operation, passive or active.

2. BACKGROUND AND DEFINITIONS

In the present work, we focus on the ultimately achievable medium and long term frequency stability of small size hydrogen masers, which is related to the thermal noise in the microwave cavity. The noise considered is white frequency noise, whose power spectral density of fractional frequency fluctuations will be denoted as S_y.

We shall recall, at first, the expressions of S_y when the maser conditions of operation are optimized with the aim of achieving the best possible frequency stability. It will be assumed that the effect of the microwave cavity thermal noise is the dominant one. This means that the microwave receiver added noise is negligible in the case of the passive operation and that the effect of the noise added by the amplifier in the feedback loop is also negligible in the case of the active operation. These assumptions can be fulfilled in practice.

2.1. Passively operated hydrogen maser.

It can be shown that, assuming the fast modulation method[3,5], the value of the power spectral density of the frequency fluctuations of the frequency locked quartz crystal oscillator depends on the phase modulation index of the interrogation signal. However, it exists an optimum value of this index for which S_y shows a minimum [1,2], the latter being approximately the same for sine wave frequency modulation, square wave frequency modulation and square wave phase modulation [8]. Furthermore, the value of S_y depends on the hydrogen beam flux, but as a consequence of the spin exchange line broadening effect, there is a value of the beam flux for which S_y shows a minimum.

Finally, the minimum achievable value of S_y for a passively operated hydrogen maser can be set equal to S_{yp} given by :

$$S_{yp} = 850 \frac{k \theta_c}{\mu_o \mu_B^2 \omega_o^3} \left(\sigma \bar{v}_r \frac{T_b}{T_t} \frac{I_{tot}}{I} \right)^2 x \frac{1}{\eta' Q_c V_b} \quad , \qquad (1)$$

where k is Boltzmann constant, θ_c is the absolute temperature of the cavity, μ_o is the magnetic permeability of vacuum, μ_B is Bohr magnetron, ω_o is the atomic resonance angular frequency, σ is the spin exchange cross-section, $\bar{v}_r$ is the average relative hydrogen velocity, T_b is the bulb storage time constant, I_{tot} is the total flux entering the bulb, I is the flux of hydrogen atoms entering the bulb in the F = 1, m_F = 0 state, T_t is the geometrical mean between longitudinal and transverse relaxation times for I = 0, Q_c is the quality factor of the loaded cavity, V_b is the volume of the storage bulb and η' denotes the filling factor of the cavity. It is defined by :

$$\eta' = \frac{\langle H_z \rangle_b^2}{\langle H^2 \rangle_c} \frac{V_b}{V_c} \quad . \qquad (2)$$

In equation (2), H_z and H are respectively the axial component and the modulus of the magnetic field inside the cavity, V_c is the volume

of the cavity. Subscripts b and c refer to the bulb and cavity volumes, respectively.

Equation (1) shows that the influence of the cavity geometry on the frequency stability arises from the quantities η', Q_c and V_b. For optimal performances in the passive mode of operation, the product $\eta' Q_c V_b$ must take values as large as possible.

2.2. Actively operated hydrogen maser.

Losses in the dielectric part of the cavity lead to a value of the cavity quality factor which does not enable, usually, oscillation to occur. However the active mode of operation can be obtained by enhancing the quality factor by means of a feedback loop [9].

The power spectral density of fractional frequency fluctuations of the maser oscillation still depends on the beam flux. Again, it exists a value of this flux for which S_y has a minimum value. In the case of an actively operated maser with a Q-enhanced cavity, this minimum value is S_{ya} given by :

$$S_{ya} = \frac{16\, k\, \theta_c}{\hbar\, \omega_o^3} \; \frac{\sigma \bar{v}_r}{Q_c V_b} \; \frac{T_b}{T_t} \; \frac{I_{tot}}{I} \; \frac{Q_e}{q_e^2 - 6q_e + 1} \quad , \tag{3}$$

where Q_c is the quality factor of the loaded cavity, Q_e is the quality factor of the Q-enhanced cavity, $\hbar$ is Planck constant divided by 2π and q_e is the spin exchange maser parameter which is presently defined by :

$$q_e = \frac{\sigma \bar{v}_r \hbar}{2\mu_o \mu_B^2} \; \frac{T_b}{T_t} \; \frac{I_{tot}}{I} \; \frac{1}{\eta' Q_e} \quad . \tag{4}$$

Equations (3) and (4) show that the frequency stability of a small size active hydrogen maser depends on the cavity geometry through the parameters η', Q_c, V_b and Q_e. The minimum value of S_{ya} which is achievable is obtained when the quantity $Q_e / Q_c V_b (q_e^2 - 6\, q_e + 1)$ is minimized.

3. STRUCTURE AND PROPERTIES OF THE MICROWAVE CAVITY

We consider the TE_{011} mode of a small size cavity made of a hollow aluminia cylinder. The cavity shows an axial symmetry. A metallic boundary around the dielectric material confines the electromagnetic fields in the empty central region and in the coaxial dielectric material. In the following we shall assume that the cavity wall is made of copper and that the relative permittivity of the dielectric material is equal to 9.53.

Let R_i and R_e the internal and the external radius of the dielectric hollow cylinder and L the length of the cavity. The values of the parameters R_i, R_e and L are bound by the condition that the cavity resonance frequency is equal to the atomic resonance frequency.

The relation between R_i, R_e, L and the cavity resonant frequency is obtained from boundary conditions applied to the electric and magnetic fields in the two cavity regions [10, 11, 12]. We have :

$$\frac{1}{\alpha_o R_i} \frac{J_1(\alpha_o R_i)}{J_o(\alpha_o R_i)} - \frac{1}{\alpha_1 R_i} \frac{G_1(\alpha_1 R_i)}{G_o(\alpha_1 R_i)} = 0 \quad , \tag{5}$$

$$G_1(\alpha_1 R_e) = 0 \quad , \tag{6}$$

where J_o and J_1 are first kind Bessel functions. G_o and G_1 are defined by :

$$G_o(x) = J_o(x) + \gamma N_o(x) \quad , \tag{7}$$

$$G_1(x) = J_1(x) + \gamma N_1(x) \quad , \tag{8}$$

where N_o and N_1 are second kind Bessel functions.
The parameter γ is given by :

$$\gamma = \frac{\alpha_1 R_i J_1(\alpha_o R_i) J_o(\alpha_1 R_i) - \alpha_o R_i J_1(\alpha_1 R_i) J_o(\alpha_o R_i)}{\alpha_o R_i N_1(\alpha_1 R_i) J_o(\alpha_o R_i) - \alpha_1 R_i J_1(\alpha_o R_i) N_o(\alpha_1 R_i)} \quad . \tag{9}$$

The coefficients α_o and α_1 are defined by :

$$\alpha_o = \sqrt{\omega_o^2 \varepsilon_o \mu_o - \left(\frac{\pi}{L}\right)^2} \quad , \tag{10}$$

$$\alpha_1 = \sqrt{\omega_o^2 \varepsilon_o \varepsilon_r \mu_o - \left(\frac{\pi}{L}\right)^2} \quad , \tag{11}$$

where ε_o represents the permittivity of vacuum and ε_r the relative permittivity of the dielectric material.

For a given set of R_i, we have computed the values of R_e as a function of L. Figure 1 depicts the variation of R_e versus R_i for L = 11 cm and 23 cm. It shows that the value of the external radius of the cavity depends strongly on that of the internal radius.

Figure 2 shows the variation of the filling factor η' versus R_e for different values of the bulb volume. It can be seen that the value of η' shows a maximum for R_i closely equal to 5 cm.

Session Number

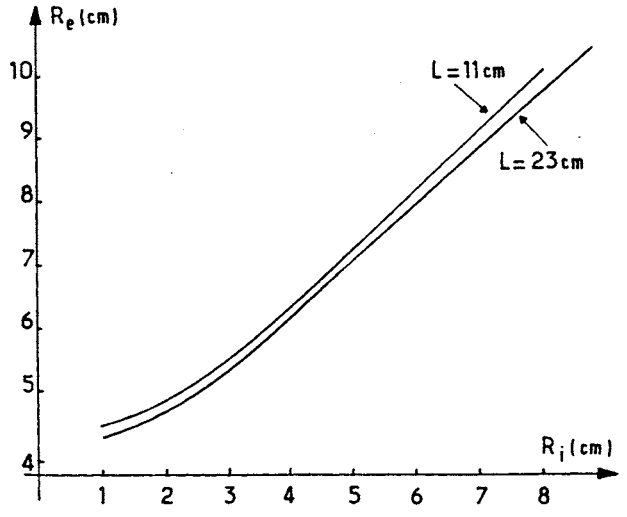


Fig. 1. Variations versus the internal radius R_i of the external radius R_e for cavities made of a hollow cylinder of alumina, having length L and resonating at 1.42 GHz.

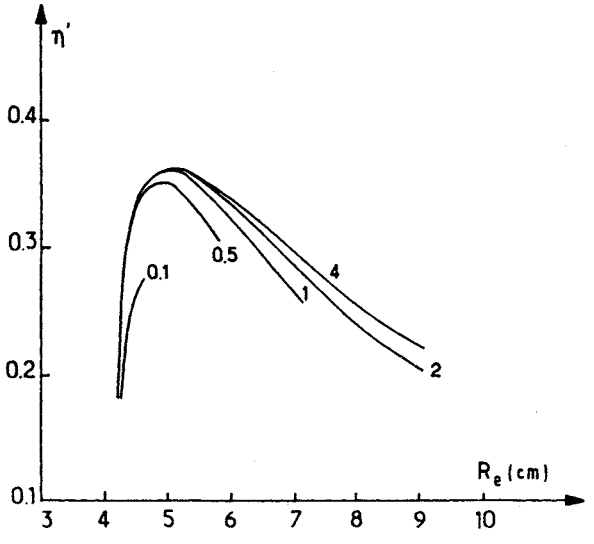


Fig. 2. Variations versus the external radius R_e of the filling factor η' for several storage bulb volumes. Bulb volumes are given in cubic decimeters.

We introduce the unloaded quality factor, Q_o, defined by :

$$\frac{1}{Q_o} = \frac{1}{Q_P} + \frac{1}{Q_D} \quad , \tag{12}$$

where Q_P and Q_D are the quality factors associated with the ohmic losses in the cavity boundary and with the dielectric losses, respectively. They are given by :

$$Q_P = \sqrt{2\,\omega_o\,\mu_o\,\sigma} \cdot \frac{\int_{V_c} H^2 dV}{\int_S H_t^{\,2}\, dS} \quad , \tag{13}$$

$$Q_D = \frac{1}{tg\,\delta} \cdot \frac{\int_{V_c} E^2\, dV}{\int_{V'} E^2 dV} \quad . \tag{14}$$

In equations (13) and (14), σ represents the conductivity of the metallic wall, S is the surface of the metallic boundary, H_t is the component of the magnetic field tangent to the cavity walls, $tg\delta$ defines the losses inside the dielectrical material, E is the electric field and V' the volume of the dielectrical material.

Figure 3 shows the variations of Q_D, Q_P and Q_o versus $tg\,\delta$ for different values of the internal radius R_i. In this figure, we have assumed that the bulb volume is equal to $5 \times 10^{-4}\ m^{-3}$.

Figure 4 shows the variation of Q_o versus R_i, assuming $tg\,\delta = 1 \times 10^{-4}$. It is observed that Q_o is an increasing function of R_i and that it does not depend much on the length of the cavity.

The value of the loaded quality factor, which appears in equations (1) and (3), depends on the value of Q_o and on the coupling constants of the cavity to its external circuits. Nevertheless, for simplicity, we shall give to Q_c the value of the unloaded quality factor. This approximation does not change significantly the conclusions concerning the effect of the cavity geometry on the frequency stability. Similarly, and taking into account the assumptions made in the derivation of equations (1) and (3), it does not invalidate the result of the comparison of the frequency stability capability of passively or actively operated small size hydrogen masers.

Session Number

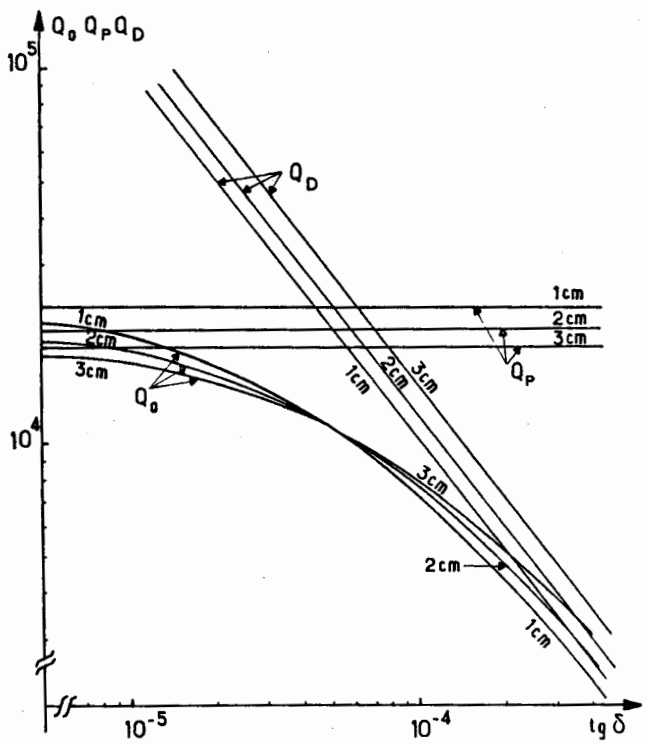


Fig. 3. Variations versus dielectric losses of the cavity quality factors Q_P, Q_D and Q_o, for several internal radius R_i. It is assumed that the bulb volume is equal to 5 x 10^{-4} m^3, and that the cavities resonate at 1.42 GHz.

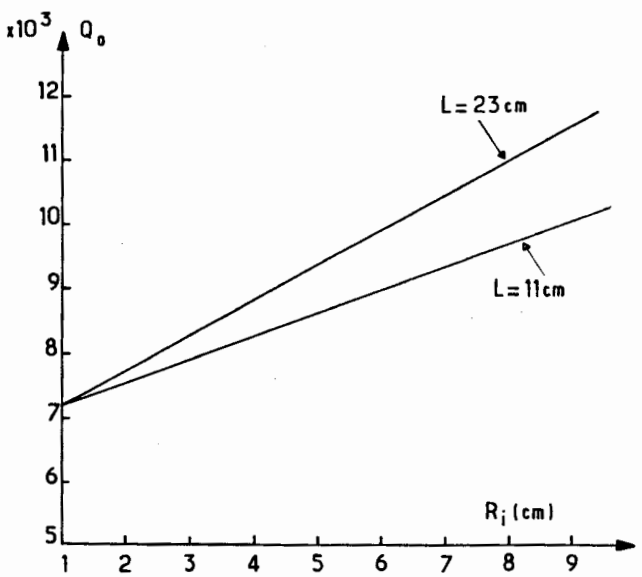


Fig.4. Variations versus the internal radius R_i of the unloaded cavity quality factor Q_o for two different values of the length L.

Session Number

4. INFLUENCE OF THE CAVITY GEOMETRY IN THE CASE OF THE PASSIVE MODE OF OPERATION.

Considering equation (1), we have derived the variations of the product $\eta' Q_o V_b$ for different values of R_i, R_e and L. The method used is the following. For a given set of values of R_i and L, we have computed the value of R_e which gives a cavity resonant frequency equal to the atomic resonant frequency. From the knowledge of (R_i, R_e, L), we have computed the product $\eta' Q_o V_b$ for a cavity limited by copper walls and loaded with alumina having a permittivity equal to 9.53 and losses such as $tg\delta = 10^{-4}$. The obtained results are reported in figure 5. The solid lines show variations of the product $Q_o V_b \eta'$ versus R_e for different values of the cavity volume. The dashed lines represent the loci of points for which the ratio $L/2R_e$ is a constant. Figure 5 clearly shows that it exists optimum values of R_e for which the product $Q_o V_b \eta'$ presents a maximum. Extrema are obtained for values of $L/2R_e$ roughly equal to 1.

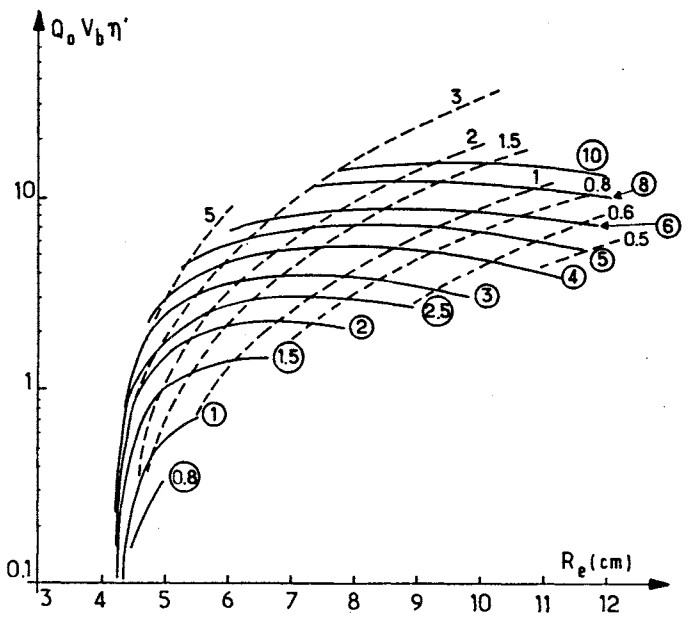


Fig. 5. Variations of the product $Q_o V_b \eta'$ with external radius R_e. Circled values are cavity volumes corresponding to the solid lines. They are given in cubic decimeters. The dashed lines are the loci of the points for which the ratio $L/2R_e$ has the specified values.

Session Number

From figure 5, we have deduced the variations of S_{yp} versus R_e for different cavity volumes and different values of $L/2R_e$. For that purpose we have considered the following maser physical parameters : θ_c = 313 K, $\sigma = 23.3 \times 10^{-20}$ m^2, $\bar{v}_r = 3.63 \times 10^3$ ms^{-1}, $T_b/T_t = 1.3$, $I_{tot}/I = 2$ and $\omega_o = 2\pi \times 1.42 \times 10^9$ rad. s^{-1}. The obtained results are shown in figure 6 for different cavity volumes (solid lines) and different values of $L/2R_e$ (dashed lines). The mixed line represents the limit of validity of the solutions and the thin line the locus of the minimum values of S_{yp} as a function of the cavity volume.

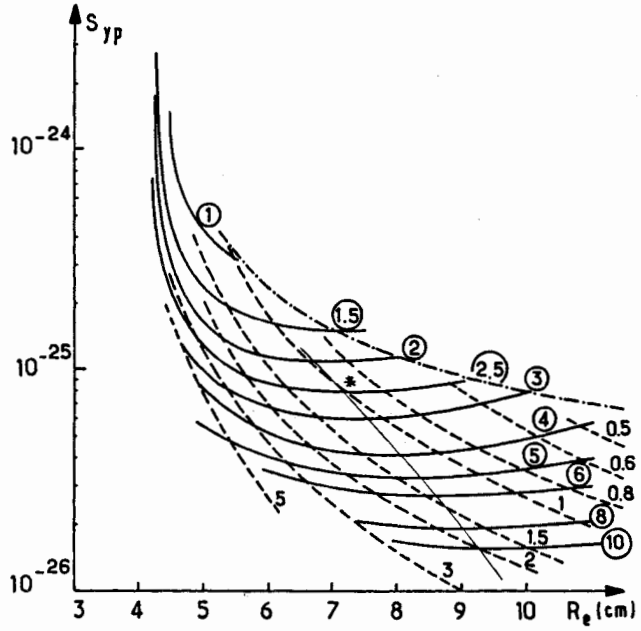


Fig. 6. Variations versus the external radius R_e of the white frequency noise level for passively operated masers. Circled values are cavity volumes corresponding to the solid lines. They are given in cubic decimeters. The dashed lines refer to geometries for which the ratio $L/2R_e$ has the specified values. The asterisk refers to the particular geometry considered in the text.

5. INFLUENCE OF THE CAVITY GEOMETRY IN THE CASE OF THE ACTIVE MODE OF OPERATION.

In this Section, we shall assume that the material properties are the same as previously and that the physical parameters have the values given in the preceding Section.

In the present case, we have to consider an additional parameter. This is the ratio Q_e/Q_c which characterizes the degree of Q-enhancement achieved by means of the feedback loop.

For several cavity geometries, we have computed the optimum value, $Q_{e,opt}$, of Q_e, i.e. the value of Q_e which minimizes the quantity

$Q_e/(q_e^2 - 6q_e + 1)$. Figure 7 shows the variations of $Q_{e,opt}$ versus R_e, for different values of the bulb volume. Fig. 8 shows the related variations of $Q_{e,opt}/Q_o$. This quantity is comprised between 8 and 14 at least, depending on the cavity geometry. Such values can be easily achieved.

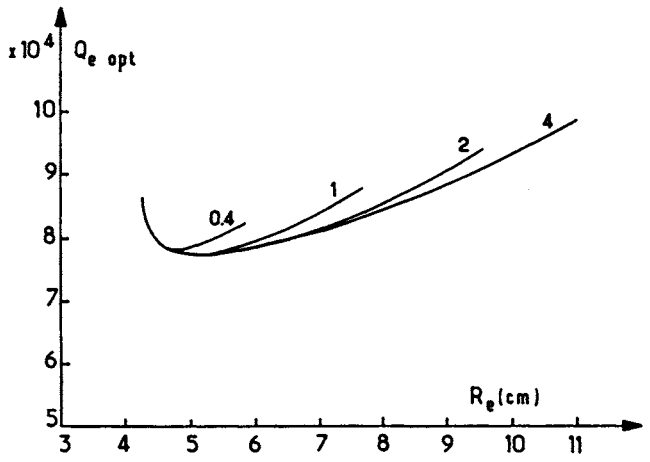


Fig. 7. Variations versus the external radius R_e of the enhanced cavity quality factor Q_e for which S_{ya} shows a minimum value. Solid lines refer to different bulb volumes which are given in cubic decimeters.

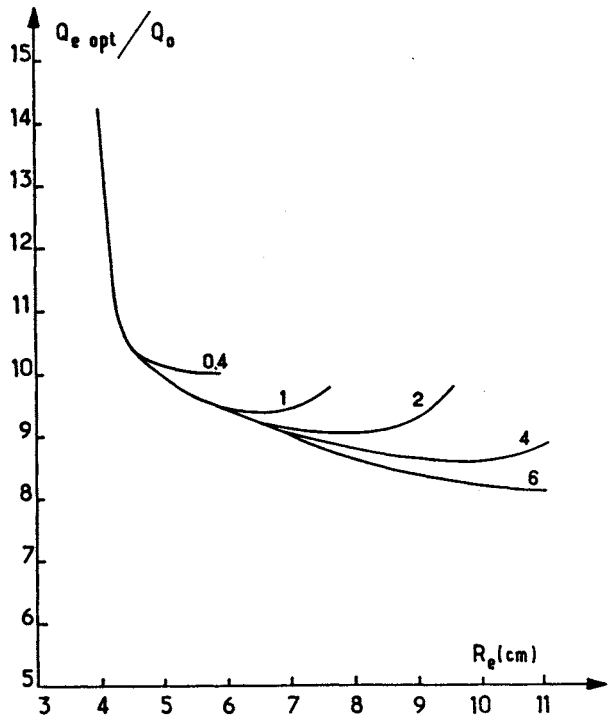


Fig. 8. Variations versus the external radius R_e of the ratio $Q_{e,opt}/Q_o$ for actively operated masers. Solid lines refer to different bulb volumes, which are given in cubic decimeters.

Session Number

Figure 9 illustrates the dependence of the parameter q_e, defined by equation 4 on the cavity geometry. One sees that q_e shows a minimum for R_e close to 5 cm.

In figure 10, the variation of S_{ya}, given by equation (3), are summarized. The solid lines represent cavity geometries for which the cavity volume is a constant, while the dashed lines refer to constant values of the ratio $L/2R_e$. The mixed line shows the limit of validity of the solutions.

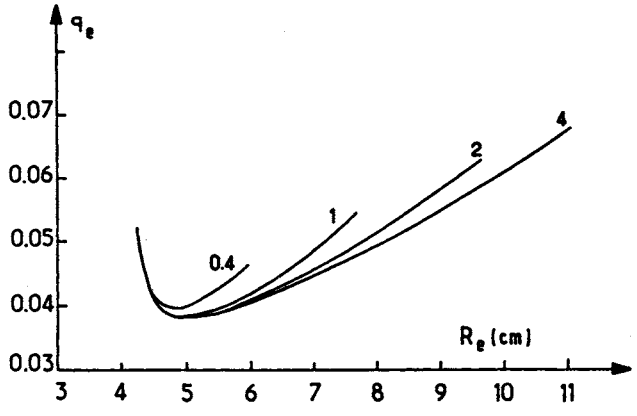


Fig. 9. Variations versus the external radius R_e of the parameter q_e for different bulb volumes, given in cubic decimeters.

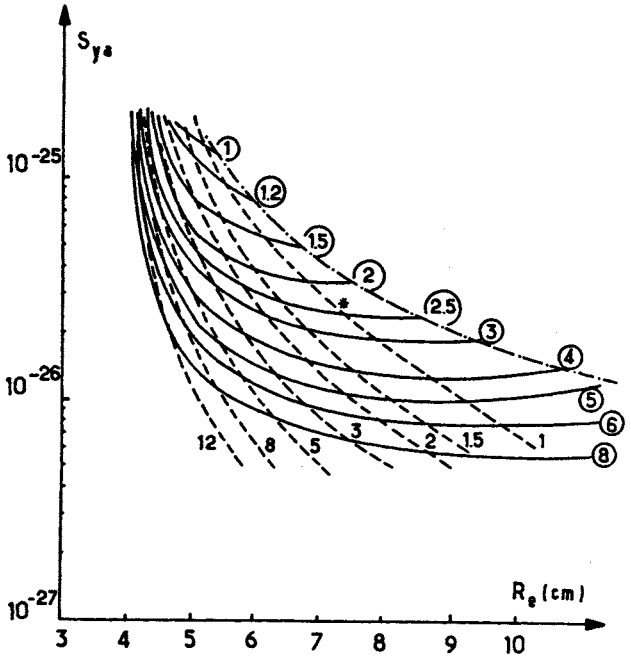


Fig. 10. Variations versus the external radius R_e of the white frequency noise level for actively operated masers. Circled values are cavity volumes corresponding to the solid lines. They are given in cubic decimeters. The dashed lines refer to geometries for which the ratio $L/2R_e$ has the specified values. The asterisk refers to the particular geometry considered in the text.

6. COMPARISON OF THE FREQUENCY STABILITY CAPABILITY OF ACTIVE AND PASSIVE MODES OF OPERATION.

The frequency stability capability of actively and passively operated small size hydrogen masers can be discussed from figures 6 and 10.

S_{ya} and S_{yp} are decreasing functions of the bulb volume. Optimal geometries are obtained for $L/2R_e$ close to unity, for both active and passive modes. For a given bulb volume, variations of S_{ya} and S_{yp} versus R_e show a minimum which is more pronounced for the passive mode.

For example, let us consider a cavity with [4] $R_e = 7.3 \times 10^{-2}$ m and $V_c = 2.3 \times 10^{-3}$ m^3. We see, in figures 6 and 10 that, for the given cavity volume, the value of the cavity radius is very close to its optimum value, for both the active and the passive modes of operation. Furthermore, we have : $S_{ya} \simeq 2.8 \times 10^{-26}$ Hz^{-1} and $S_{yp} \simeq 9 \times 10^{-26}$ Hz^{-1}.

In the time domain, the frequency stability can be characterized by means of the Allan standard deviation $\sigma_y(\tau)$. For white frequency noise, we have :

$$\sigma_y(\tau) = (S_y/2\tau)^{1/2} \qquad (15)$$

We therefore obtain :

$\sigma_y(\tau) \simeq 1.2 \times 10^{-13}\ \tau^{-1/2}$ for the active mode and
$\sigma_y(\tau) \simeq 2.1 \times 10^{-13}\ \tau^{-1/2}$ for the passive mode.

It must be pointed out that these computed frequency stabilities depend strongly on the cavity quality factor and, therefore, on the dielectric losses. The results given show that, for a given cavity geometry, the frequency stability capabilities are not very much different according as the passive or the active mode of operation is considered. However, the latter is able to give a slightly better frequency stability.

7. CONCLUSION

We have considered the cavity thermal noise limited frequency stability of small size hydrogen masers whose cavity is made of an alumina hollow cylinder. We have given the power spectral density of the best achievable frequency stability for the passive and the active modes of operation. We have shown that the frequency stability capabilities does not differ very much for both modes of operation with, however a slight advantage to the active mode.

Session Number

REFERENCES

1. P. Lesage, C. Audoin and M. Têtu. Proc. of the 33rd Annual Symposium on Frequency Control, (1979), p. 515.
2. C. Audoin, J. Viennet and P. Lesage. Journal de Physique 12 (1981), p. C8-159
3. D.A. Howe, F.L. Walls, H.E. Bell and H. Hellwig. Proc. of the 33rd Annual Symposium on Frequency Control (1979), p. 554
4. F.L. Walls. IEEE Trans. on Instr. and Meas. IM-36, (1987), p. 596.
5. E.M. Mattison, E.L. Blomberg, G.U. Nystrom and R.F.C. Vessot. Proc. of the 33rd Annual Symposium on Frequency Control, (1979), p. 549
6. J. White, A. Frank and V. Folen. Proc. of the 12th Annual Precise Time and Time Interval (PTTI). Applications and Planning Meeting (1980), p. 495.
7. G. Busca. Proc. of the 33rd Annual Symposium on Frequency Control, (1979), p. 563.
8. J. Vanier and C. Audoin in "The Quantum Physics of Atomic Frequency Standards", Adam Hilger, Bristol (1989). To be published.
9. H.T.M. Wang. Proc. of the 34th Annual Symposium on Frequency Control (1980), p. 364.
10. G. Boudouris, "Cavités Electromagnétiques", Dunod, Paris (1971).
11. T. Sphicopoulos, L.G. Bernier and F. Gardiol, IEEE Proc. 131 Pt. H, (1984), p. 94.
12. F.S. Rusin, B.A. Gaigerov and S.F. Ponomarenko. Izmeritel'naya Tekhnika 5 (1985), p. 15.
13. J. Robert (1967) unpublished report.
14. P. Louche (1986) unpublished report.

2 ND EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 16 - 17 - 18 MARCH 1988

Session AFS-01 Number 3

DOPPLER EFFECT IN MERCURY ION CLOCK

C. MEIS, M. JARDINO and M. DESAINTFUSCIEN

Laboratoire de l'Horloge Atomique
Equipe de Recherche du CNRS associée à l'Université Paris-Sud
Bât. 221 - Université Paris-Sud

91405 ORSAY - France

ABSTRACT

The mercury ion atomic frequency standard, uses as the clock transition the hyperfine splitting at 40,5 GHz of the ground level $^2S_{1/2}$ of singly ionized 199 mercury isotopes, stored in an rf trap. The main source of systematic error of such a standard is the second order Doppler shift, a relativistic effect which produces a relative frequency shift equal to the ratio of the mean kinetic energy to the rest mass energy of an ion. In this paper, an analytical expression of this shift is given involving physical parameters which can be easily measured namely the macromotion secular frequency and the total number of ions. This expression is valid in the case of a spherically symmetrical ion cloud and has been established considering a thermalized model of the stored ions taking space charge effects into account.

Experimentally, we measured the secular frequency from the first order Doppler spectrum of the hyperfine transition and the total ion number from image currents due to the periodic motion of the stored ions.

1. INTRODUCTION

The use of rf traps [1] in high resolution microwave hyperfine spectroscopy led, among others, to the device of $^{199}Hg^+$ stored ions atomic frequency standard [2,3,4]. The ions are stored by an inhomogeneous electric field created by the voltage $V = V_{DC} + V_{AC} \cos \Omega t$ applied on the trap electrodes. They experience a small amplitude oscillation (micromotion) at the frequency of the confining field $\Omega/2\pi$, superimposed to large thermalized orbits governed by an effective potential [5]. Thus, the microwave hyperfine absorption spectrum of the ions is phase modulated at the secular frequency, (1^{st} order Doppler effect), which is shifted from its theoretical value because of space charge effect [6]. The wavelength of the hyperfine transition (λ = 7,4 mm, ν_0 = 40,5 GHz) is of the same order of magnitude than the mean amplitude of the secular motion (Lamb-Dicke limit). Consequently, the observed microwave spectrum [7-10] consists of a narrow central carrier, unshifted and unbroadened by first order Doppler effect

and periodic side-bands, spaced by the secular frequency and broadened by the ion velocity and trajectory distributions [7,8]. The overall motion of trapped ions causes the most significant absolute offset of the central carrier, due to the second order Doppler effect. We describe a method to measure it in order to improve the accuracy of this trapped ion frequency standard.

2. THEORETICAL BACKGROUND

In the case of a cylindrical rf trap, the instantaneous electric potential due only to the trap electrodes writes in terms of the components of the trapping voltage in cylindrical coordinates :

$$\psi(\rho,z,t) = (V_{DC}+V_{AC}\cos\Omega t)\,\frac{2r_o}{z_o}\sum_{p=o}^{\infty}\frac{(-1)^p}{m_p}\cos(m_p\frac{z}{r_o})\,\frac{I_o(m_p\frac{\rho}{r_o})}{I_o(m_p)} \qquad (1)$$

with $m_p = (p+\frac{1}{2})\pi\frac{r_o}{z_o}$; r_o, z_o the radius and half height of the trap and I_o the modified Bessel functions of first kind of zero order.

$$I_o(\alpha) = \sum_{k=o}^{\infty}(-1)^k\,(i\alpha)^{2k}/2^{2k}\,k!\,\Gamma(k+1) \qquad (2)$$

The effective potential [5] $V_{eff}(\rho,z)$, is a fairly complicated function of Bessel series :

$$V_{eff}(\rho,z) = \frac{1}{V_{DC}+V_{AC}\cos\Omega t}\left[V_{DC}\,\psi(\rho,z,t)+\frac{q}{4m\Omega^2}(-\vec{\nabla}V_{AC}\psi(\rho,z,t))^2\right] \qquad (3)$$

where m is the mass of the ion.

We have demonstrated by numerical calculation [11] that $V_{eff}(\rho,z)$ cannot be perfectly spherically symmetrical. But a good approximation can be obtained in the central volume of the trap (within a sphere of a radius of about 6 mm which is typically the volume of a hot ion cloud of a few million ions) when

$$V_{DC} = 0.74\sqrt{2}\,q\,V_{AC}^2/m\Omega^2\,(r_o^2+2z_o^2) \qquad (4)$$

Consequently, in this space of the trap the effective potential can be expressed in spherical coordinates

$$V_{eff}(r) = \frac{1}{2}\,m\,\omega_s^2 r^2 \quad \text{with} \quad r^2 = \rho^2+z^2 \qquad (5)$$

where $\omega_s/2\pi$ is the ion macromotion secular frequency [11] :

$$\omega_s = 1.11\,(2q\,V_{AC}/m\Omega\,(r_o^2+2z_o^2)) \qquad (6)$$

Thus, a single ion in the potential (5) undergoes a three-dimensional harmonic motion at the secular frequency $\omega_s/2\pi$, upon which is superimposed a small amplitude micromotion at frequency $\Omega/2\pi$. When a large number of ions is present in the trap, equation (5) is modified by space charge and becomes :

$$V_{eff}(r) = \frac{1}{2} m\omega_s^2 r^2 + q\ \phi_{SC}(r) \qquad (7)$$

where we suppose space charge potential is a radial function for a spherical ion cloud. Assuming Gaussian charge distribution, we obtain [12] :

$$V_{eff}(r) = \frac{1}{2} m\omega_s^2 r^2 + \frac{Nq^2}{4\pi\varepsilon_o} \frac{1}{r} \text{erf}\ (\frac{r}{R}) \qquad (8)$$

when N is the total number of ions, R is a parameter such as : $\ln(\frac{n(o)}{n(R)}) = 1$ and n(r) the charge density distribution.

Equation (8) shows that space charge results in a diminution of the effective potential and of the secular frequency which takes the value ω_m. If the ions are thermalized with a temperature T, it can be shown that for a given total number of ions N, we have :

$$T = N^{2/3} (\frac{m\omega_m^2}{2\pi K_B}) \left[\frac{q^2|\mu|}{m\varepsilon_o(\omega_s^2 - \omega_m^2)}\right]^{2/3} \qquad (9)$$

where $\mu = \sum_{n=1}^{\infty} (-1)^n\ 2\Gamma(n + \frac{1}{2}) / n!\ (2n + 1)\sqrt{\pi} \simeq -0.23$

ε_o is the vacuum permittivity and K_B the Boltzmann's constant. This result is valid only when $\omega_m \gtrsim \omega_s/2$.

The second order Doppler shift due to both micromotion and macromotion can be expressed in terms of the secular motion temperature :

$$\frac{\Delta f}{f} \sim -\frac{K_B T}{mc^2} (\frac{3}{2} + \frac{\omega_s^2}{\omega_m^2}) \qquad (10)$$

Numerical simulations [7] are in good agreement with these results.

3. EXPERIMENTAL SET-UP AND RESULTS

$^{199}Hg^+$ ions are created by electron impact with neutral 199 mercury isotopes inside a cylindrical rf trap (Fig. 1). The radius of the trap is r_o = 19 mm and its half height $z_o = r_o/\sqrt{2}$ = 13,4 mm. The whole system is kept in ultra-high vacuum with a residual pressure of 10^{-9} Torr which increases up to 10^{-8} Torr when neutral mercury is introduced in the trap by a variable leak. The voltage components applied on the trap's electrodes are chosen so as to satisfy (4) V_{DC} = 31 V ; $V_{AC} \simeq$ 341 V and $\Omega/2\pi$ = 257 kHz.

The theoretical secular frequency (6) is $\omega_s/2\pi = 50$ kHz. We optically pump the $^{199}Hg^+$ ions to the $|^2S_{1/2}\ F = 0>$ hyperfine level with a 194.2 nm radiation from a $^{202}Hg^+$ lamp (Fig. 2). We sweep the hyperfine transition by a microwave signal introduced in the trap by a horn and observe the variation of the intensity of the ion fluorescence at 194.2 nm. No buffer gas was used in this experiment. The linewidth of the hyperfine resonance at 40.5 GHz is of a few Hz at sufficiently low microwave power. Side-bands are observed (Fig. 3), at maximum microwave power. In these conditions, the saturation factor for the central line is about 650. The measured secular frequency determined from the lateral lines is $\omega_m/2\pi = 36$ kHz.

The total number of stored ions is measured by observing image currents induced in the trap electrodes by the motion of the ion cloud when a wobulated signal near ω_m is applied between the two endcaps of the trap [13]. The detected signal is shown in Fig. 4. The number of ions is proportional to the surface of this signal according the relation :

$$N = (Y_{max} \cdot \Delta\omega)\ \left(\frac{4mz_o^2}{\Gamma^2 q^2}\right)\ \left(\frac{1 + R_f G_o}{b\ R_f}\right) \qquad (11)$$

Y_{max}, $\Delta\omega$, Γ, R_f, G_o and b are defined in [13]. $\Gamma \sim 1$ when the ratio of the potential well depth to the ion mean kinetic energy is superior to 5, which is the case in our experiment.

In our experimental conditions, we have $z_o = 13.4$ mm ; b = 0.93 ; $\Gamma = 1$; $R_f = 2M\Omega$; $G_o = 4.9\ 10^{-7}\ \Omega^{-1}$; $Y_{max} = 0.25$; $\Delta\omega = 2\pi \cdot 120$ Hz. We obtain $N = (1.9 \pm 0.1)\ 10^6$ ions.

Injecting the experimental values of ω_m and N in (9) and (10), we obtain the ion cloud temperature and the second order Doppler shift.

$$T = 3584 \pm 170°\ K \text{ and } \frac{\Delta f}{f} = -(5.7 \pm 0.2)\ 10^{-12}$$

The contribution of the second order Doppler effect to the budget error is consequently of the order of

$$2\ 10^{-13}$$

It could be reduced by improving the precision of the measurement of R_f and G_o.

Session Number

REFERENCES

(1) W. Paul, O. Osberghaus and E. Fischer, Forschunger. Wirtsch. Verkehrsministerium, Nord-Rhein-Westphalen, n° 415 (1958)

(2) M. Jardino, M. Desaintfuscien, R. Barillet, J. Viennet, P. Petit, C. Audoin, Appl. Phys. 24, 107 (1981)

(3) L.S. Cutler, R.P. Giffard, M.D. Mc Guire, Proc. 37th Annual Frequency Control Symposium, p. 32 (1983)

(4) L.S. Cutler, R.P. Giffard, M.D. Mc Guire, Appl. Phys. B 36, 137-142 (1985)

(5) A.V. Gaponov, M.A. Miller, JETP, October 15 (1957)

(6) F. Vedel, J. André, M. Vedel, J. de Physique 42, 1611 (1981)

(7) M. Jardino, F. Plumelle, M. Desaintfuscien, J.L. Duchêne, Proc. 38th Annual Frequency Control Symposium, IEEE Press. 84 CH 2062-8, p. 431 (1984)

(8) L.S. Cutler, C.A. Flory, R.P. Giffard and M.D. Mc Guire, Appl. Phys. B 39, 251-259 (1986)

(9) H.S. Lakkaraju and H.A. Schuessler, J. Appl. Phys. 53, n° 6, 3967 (1982)

(10) H.G. Dehmelt, in "Advances in Atomic and Molecular Physics" 3, 53 (1967)

(11) H. Lagadec, C. Meis, M. Jardino, submitted for publication to Int. Journ. of Mass Spectroscopy and Ion Phys.

(12) C. Meis, M. Desaintfuscien, M. Jardino, Appl. Phys. B 45, 59-64 (1988)

(13) M.N. Gaboriaud, M. Desaintfuscien, F.G. Major, Int. J. of Mass Spectroscopy and Ion Phys. 41, 109-123 (1981)

Session Number

FIGURE CAPTIONS

Fig. 1. Experimental set-up.

Fig. 2. Comparison of the first energy levels of $^{199}Hg^{+}$ and $^{202}Hg^{+}$.

Fig. 3. Doppler side-bands of the microwave hyperfine transitions of $^{199}Hg^{+}$.

Fig. 4. Radiofrequency resonance absorption curve of the ions oscillating at 36 kHz.

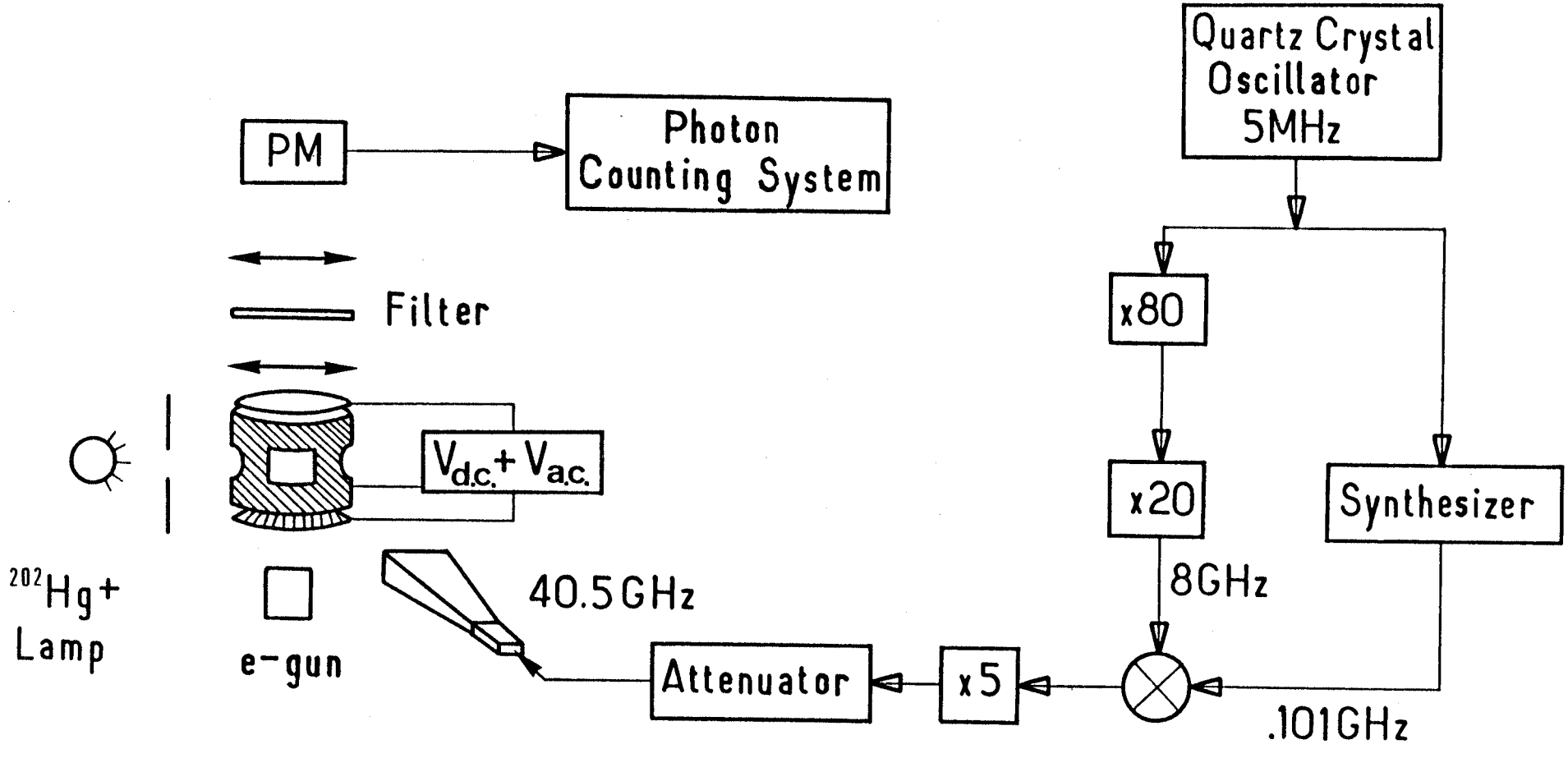
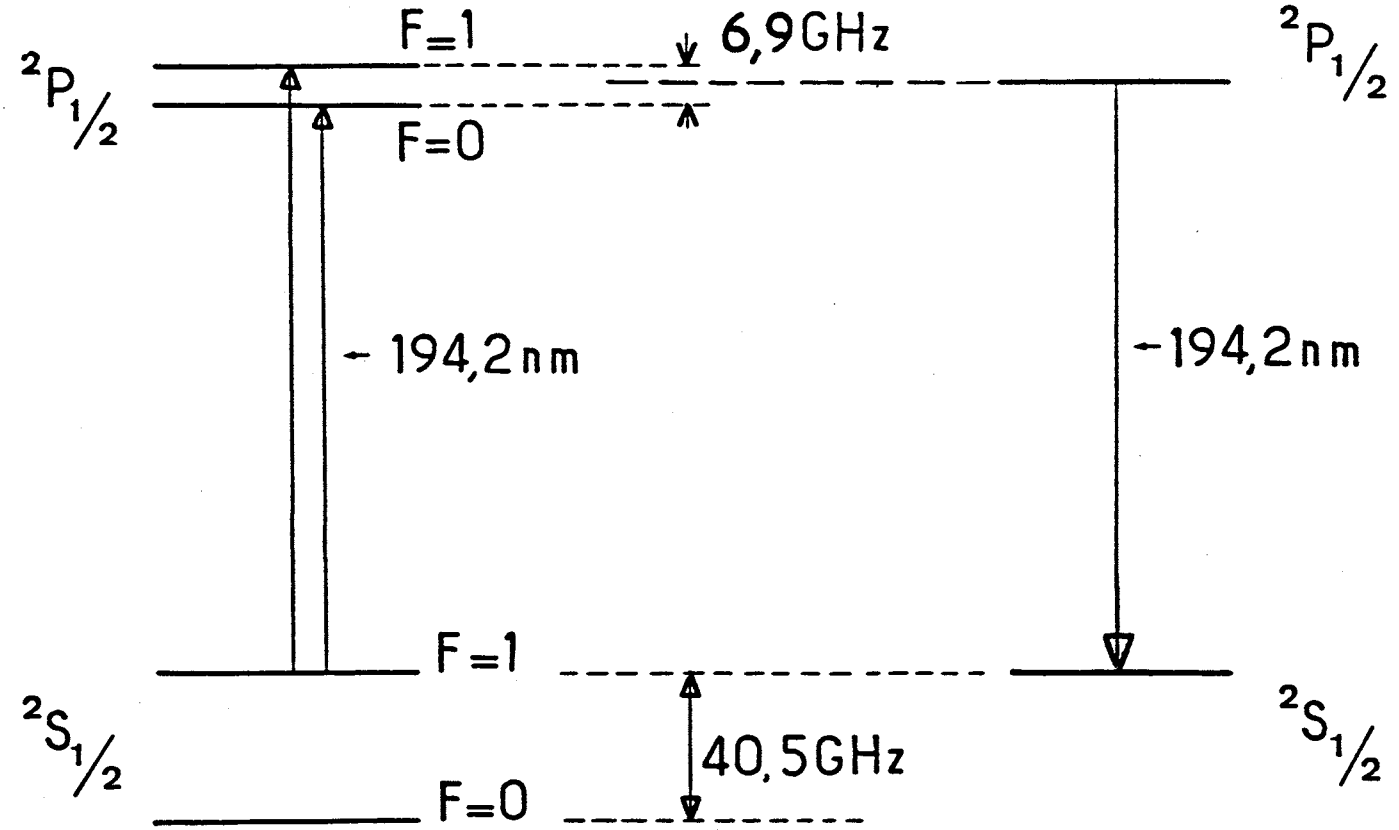

Quartz Crystal Oscillator 5MHz
PM
Photon Counting System
Filter
x80
x20
Synthesizer
$V_{d.c.} + V_{a.c.}$
$^{202}Hg^{+}$ Lamp
e-gun
40.5GHz
8GHz
Attenuator
x5
.101GHz

Fig. 1



$^{199}Hg^+$

$I=\frac{1}{2}$

$^{202}Hg^+$

$I=0$

Fig. 2

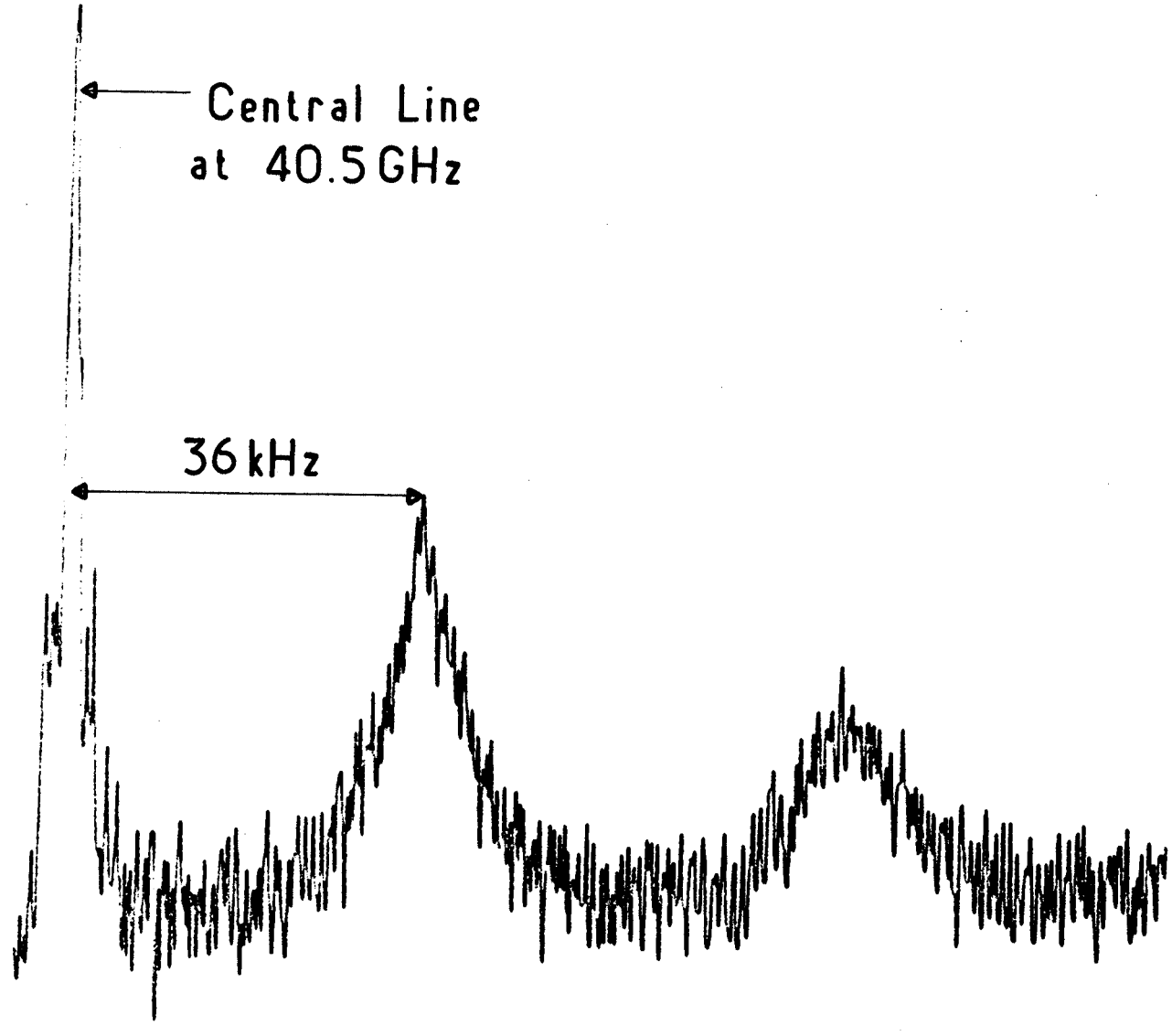


Fig.3

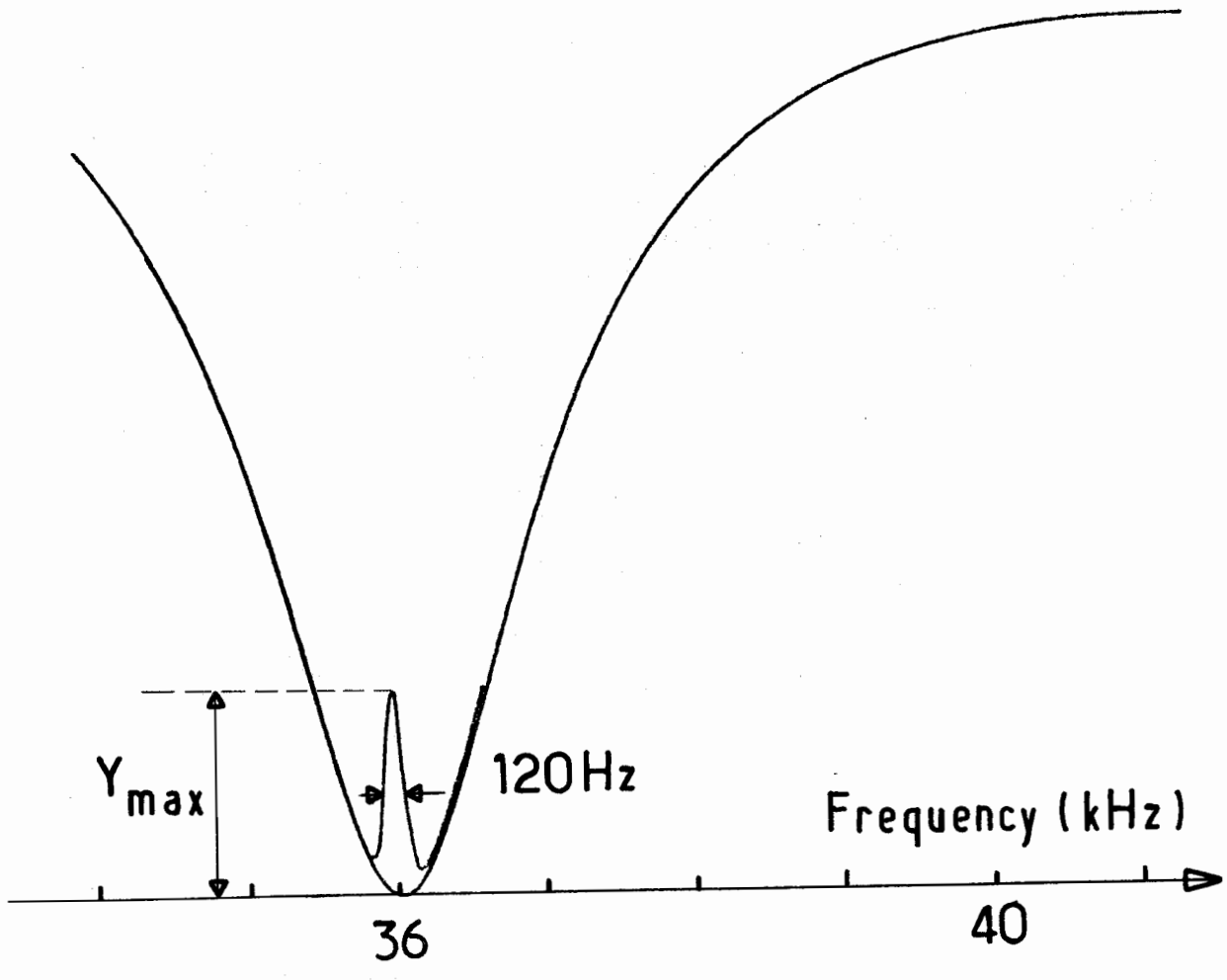


Fig. 4

2 ND EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 16 - 17 - 18 MARCH 1988

Session AFS-01 Number 4

MINIATURIZATION VERSUS PERFORMANCES IN RUBIDIUM FREQUENCY STANDARDS

G. Busca, L. Pan *

OSCILLOQUARTZ SA, 2002 Neuchâtel, Switzerland

Abstract

The relationship between size and performances in passive Rubidium standards is studied experimentally mostly in connection with the short term stability. The analysis involves several physical packages of different concept and design, based upon the buffer gas or wall coating, and having resonance cell volumes ranging from 800 cm^3 to 0.4 cm^3. The theoretical stability over 1 second, evaluated from the experimental discriminator slope, is shown to range from 1×10^{-14} for the biggest to 6×10^{-12} for the smallest cell.

Introduction

Miniaturization is the trend in the development of passive Rubidium standard today. The existing, commercially available, Rubidium standards can be divided in two categories : the "laboratory" type having an overall volume of several liters and the "miniaturized" type having typically a volume of 0.5 liters. The "laboratory" type standards are characterised by an Allan Variance given by $\sigma(\tau) = 5 \times 10^{-12} \tau^{-1/2}$ and a frequency drift less than 1×10^{-11} / month. The "miniaturized" standards show today a performance degradation of almost one order of magnitude compared to the "laboratory" type in discrepancy with the results obtained in the development phase yelding performances even superior the laboratory standards (1). The work reported here is intended to clarify this situation and shows that high performance miniaturized rubidium standards are actually faisible.

* permanent address : WUHAN INSTITUTE OF PHYSICS - ACADEMIA SINICA - P.R. CHINA

1) Theoretical considerations

The population difference Δ between the 2 levels involved in the Rubidium clock transition, can be expressed as a function of non-dimensional parameters Γ', β', Ω' - Ω_o' given by :

$$\Gamma' = \frac{\Gamma}{\gamma_1} \quad , \quad \beta'^2 = \frac{\beta^2}{\gamma_1 (\gamma_2 + \frac{\Gamma}{2})} \quad , \quad \Omega' - \Omega'_0 = \frac{\omega - \omega_0}{\gamma_2 + \frac{\Gamma}{2}} \tag{1}$$

Where : Γ is the pumping rate; $\beta = \mu_0 \frac{Hz}{\hbar}$ is the RABI frequency; H_z is the amplitude of the microwave interrogating magnetic field, γ_1 and γ_2 are respectively the longitudinal and the transverse relaxation rates; γ_2 + Γ/2 is the light broadened half line-width (in rad/sec), ω_o is the Rubidium angular frequency and ω the interrogating frequency (2).

The expression $\Delta = \Delta (\Gamma', \beta', \Omega' - \Omega'_0)$

for ω = ω_o is reduced to :

$$\Delta = \frac{\Gamma' \left(\frac{\Gamma' + 1}{5\Gamma'^2 + 13\Gamma' + 8} \right)}{1 + 4\beta'^2 \frac{2\Gamma'^2 + 9\Gamma' + 8}{5\Gamma'^2 + 13\Gamma' + 8}} \tag{2}$$

the full linewidth being :

$$\Delta \nu_L = \frac{\gamma_2 + \Gamma/2}{\pi} \tag{3}$$

The discriminator slope (D) of the resonance cell, used as frequency discriminator in the frequency standard, can be approximated by the eq. :

$$D = N \frac{2\Delta}{\Delta \nu_L} \tag{4}$$

Where N is the total number of atoms in the cell. The quantity N x Δ is also denoted as the signal S, representing the deep observed in the photocurrent, when the interrogating microwave frequency is varied. It is clear that resonance cells of different volumes, operating at different temperatures, based on buffer gas or wall coating, manifeste a big variation of γ_1, γ_2 and N values. However if enough pumping rate Γ and interrogation power β^2 is available, the reduced non dimensional parameters Γ' and β'^2 can be maintained constant through the different physical situations insuring a constant value for Δ. In this case the discriminator slope is determined by the variable $N/\Delta\nu_L$. Other constraints existing in real standards (as operation at zero light shift, maximum discriminator slope and minimum temperature coefficient) prevent in practice the realization of such a constant population difference Δ through different physical situations.

In addition the actual frequency stability of a standard is theoretically limited by the shot noise of the DC photocurrent I, through the equation :

$$\sigma_y(\tau) = \frac{1}{\nu_0} \frac{\sqrt{2eI}}{\sqrt{2}\,D} \tau^{-1/2} = \frac{n}{\sqrt{2}\,D\nu_0} \tau^{-1/2} \qquad (5)$$

where υ_o is the Rubidium frequency; e the electron charge, n the photodetector shot noise in Amp./$\sqrt{\text{Hz}}$ and D the discriminator slope in Amp./Hz.

The total photocurrent I is also a parameter which is difficult to predict depending critically upon the geometry of the unit and pumping efficiency.

2) Experimental results

Table 1 summarizes the experimental results and is self-explanatory. The theoretical short term stability is derived from the experimental discriminator slope and the theoretical shot-noise associated with the experimental DC photocurrent. The expected short term stability is computed from the experimental discriminator slope and the experimental

value of the photocurrent noise measured at the output of the current to the voltage converter. The experimental short term stability is the actual stability measured for the operating standard and, in our experiments, is mostly determinded by a detection noise much higher than the theoretical shot noise. This point need to be studied in more details, in particular by selecting better photodetectors, in order to obtain an experimental performance closer to the theoretical value.

In column 1 data of the Rubidium Masers operating at Laval University, Quebec, Canada and Wuhan Institute of Physics, Wuhan, P.R. of China, have been elaborated in order to compute the theoretical short term stability achievable with such a standard, when used as conventional passive standards. The calculation is based on an output power of 2×10^{-10} Watts, a coupling factor of 0.1, a photodetector radiometric sensitivity of 0.5 Amp/Watt, a DC photocurrent one order of magnitude greater than the photocurrent measured with the one inch diameter photocell used in the TE_{011} cavity system and with the assumption that each microwave photon emitted by the maser is detected as an optical photon in the passive operation.

- Column 2 refers to a TE_{011} cavity and integrated filter approach. (3)
- Column 3 refers to the same apparatus involved with column 2 data but with the addition of a separate isotopic filter.
- Column 4 refers again to the same apparatus involved with the data of column 2 by replacing the buffer gas with wall caoting.
- Column 5 refers to the approach based on the TE_{111} cavity mode and integrated filter, which is the mostly used today.
- Column 6 refers to the "utility"Rubidium described later, employing a 0.4 cm^3 cell, the smallest cell used up to now (to the author's knowledge)

3) Comments on the experimental results

Comparing the data shown in column 1 and 3, related to physical packages based on similar concepts even if very different in size, we make the following remarks :

- The ratio of the total number N of atoms as deduced from the volumes ratio at constant temperature is N3/N1 = 1/100. The ratio of the signal $S_3/S_1 = N3\Delta3/N1\Delta1 = \Delta_3/100\Delta_1$ is found to be $S_3/S_1 = 1/30$ which implies $\Delta_3/\Delta_1 = 3$. This last relation shows that an higher population difference exists in correspondance with the smaller cell, probably due to higher optical pumping rate achievable within the smaller cell. From the previous data together with the experimental ratio of the linewidth $\frac{\Delta\nu3}{\Delta\nu1} = 5$, we can compute :

$D_3 = S3/\Delta\nu3 = 1/150\ D_1$ in agreement with the discriminator slope reported.

- Comparing the data in columns 2,3,4 and 5 for which an approximately constant number of atoms N is realized, it appears that, independently of the approach and cavity mode used, a theoretical stability of $\cong 2 \times 10^{-12}$ at 1 sec is obtainable. A special case is represented by the data shown in column 3, which shows, to the knowledge of the authors, the highest discriminator slope experimentally measured up to now. It corresponds to a buffer gas mixture composition of 30 mbars of N_2 and 30 mbars of CH_4.

- Comparing the data of column 4 and column 6, the ratio of the total number of Rb^{87} atoms corresponding to different volumes, temperatures and mixtures is evaluated as : $N_6/N_4 = 1/20$. With an experimental linewidth ratio $\frac{\Delta\nu6}{\Delta\nu4} = 5$ we compute a discriminator slope ratio : $\frac{D6}{D4} = \frac{1}{100}\ \frac{\Delta6}{\Delta4}$ which is, at first sight, much lower than the experimental ratio $\frac{D6}{D4} = 1/4$.

Also for this case we think that part of the disagreement can be explained by supposing that $\Delta 6/\Delta 4$ is of the order of 10 due to the better pumping efficiency of smaller cell.

4) The Utility Rubidium

The concept of the Utility Rubidium is based upon the use of a very small cell having a volume of less than 1 cm^3 in order to investigate the feasibility of a drastic miniaturization. For such a small cell a microwave cavity cannot be used conveniently and a simple loop is preferred. A Dicke narrowing can be obtained directly by the confinement operated by the cell itself so that no buffer gas is needed for this purpose. A coating is then applied to the cell walls and a separate isotopic filter is imperative. The Fig. 1 shows the overall physical package which includes 2 levels of magnetic shields. Fig. 2 shows the frequency shift as a function of the light intensity emitted by the lamp (using neutral density filters) for two different isotopic filters and as a function of the cell temperature, the cell temperature being equal to the filter temperature. From Fig. 2 we deduce a wall shift of - 500 Hz for a "new" polyethylene coating and an upper limit for the wall shift temperature coefficient of $10^{-10}/C^{o}$. Very important Light-shift appears, probably due to the exceedingly high photon flux available inside the cell. A zero Light-shift operation appears faisible in principle with both isotopic filters, the preferred situation being the one corresponding also to a maximum signal. The use of coating represent a simplification from the point of view of the distributed Light-shift and line homogeneity. However the filter temperature coefficient is present (4) and is dominating the temperature behaviour of this physical package.

So no practical advantage can be found with the use of wall coating for reducing the standard temperature coefficient. On the contrary higher long term drift have to be expected due to wall coating pollution by the Rubidium metal (3). The modifications of the coating characteristics are already evident in the data shown in Fig. 2, for the higher cavity temperatures used. In fact, the values of the wall shift shows an hysteresis effect for temperature cycling above $80^{o}C$.

Conclusion

Size reduction is accompanied by stability performances reduction governed by the quantity $\frac{N\Delta}{\Delta\nu}$ and the photodetector shot noise. The parameter N is totally under control and is a function of the design. The population difference Δ actually increases for smaller cells due to higher pumping rate available in small volumes. On the contrary the linewidth is uniformly deteriorated with miniaturization. The data reported here show that, also with a 0.4 cm^3 cell, is possible to obtain in principle a frequency stability of the order of 6×10^{-12} for 1 sec averaging time close to the best commercially available performance today.

Acknowledge

This work has been sponsored by the Swiss :

Commission pour l'Encouragement de la Recherche Scientifique (CERS)

References

(1) T.J. LYNCH and W.J. RILEY. Proceeding of the fifteenth Annual Precise Time and Time Interval (P.T.T.I) Applications and Planning Meeting. Page 269, (1983).

(2) J. VANIER, Basic theory of Lasers and Masers, GORDON and BREACH SCIENCE PUBLISHERS N.Y., (1971)

(3) G. BUSCA, L. PAN. 1st EUROPEAN TIME and FREQUENCY FORUM, 331, (1987)

(4) J. VANIER et al. Third Symposium on Frequency Standards and Metrology. J. de Physique. C8-139 (1981).

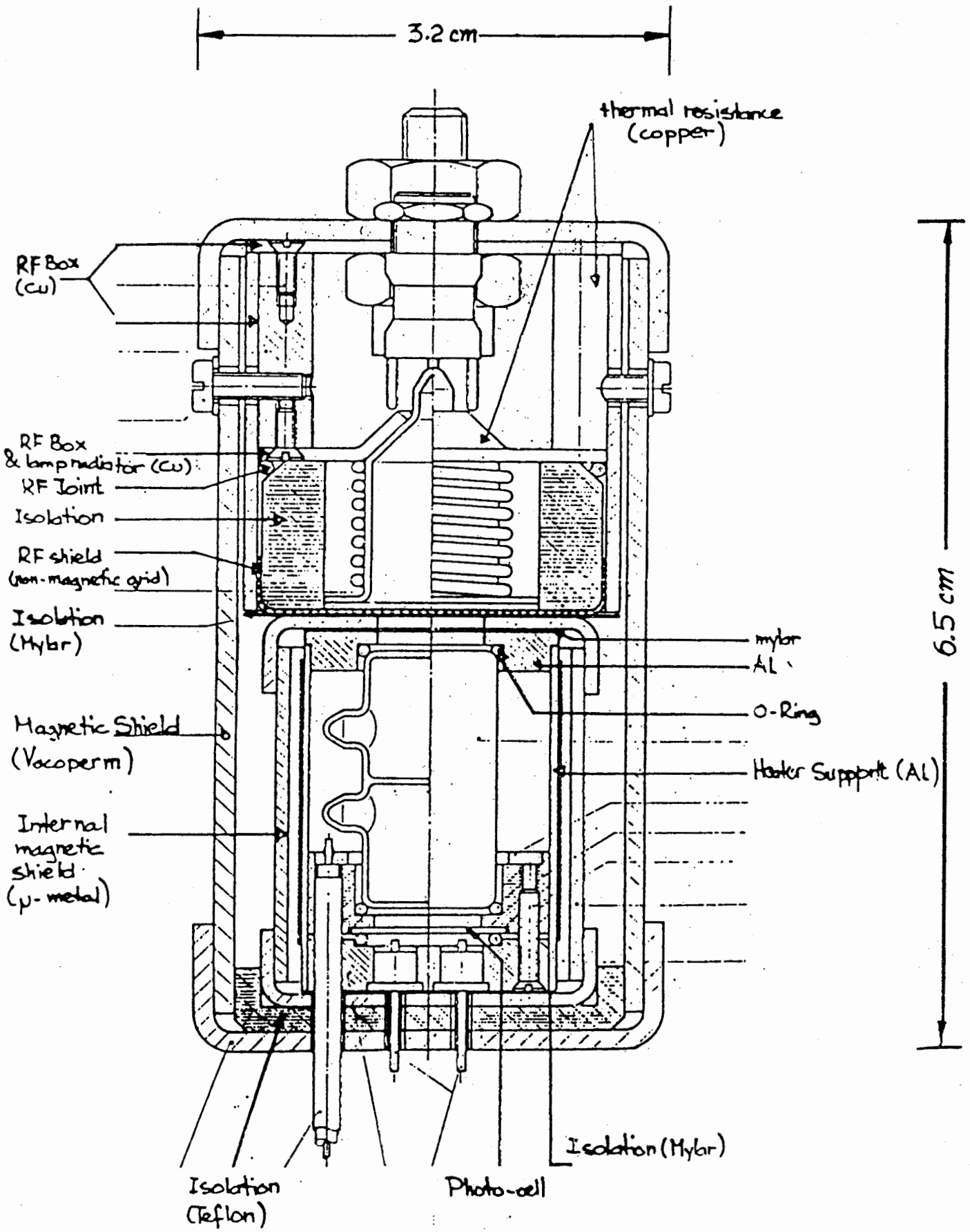
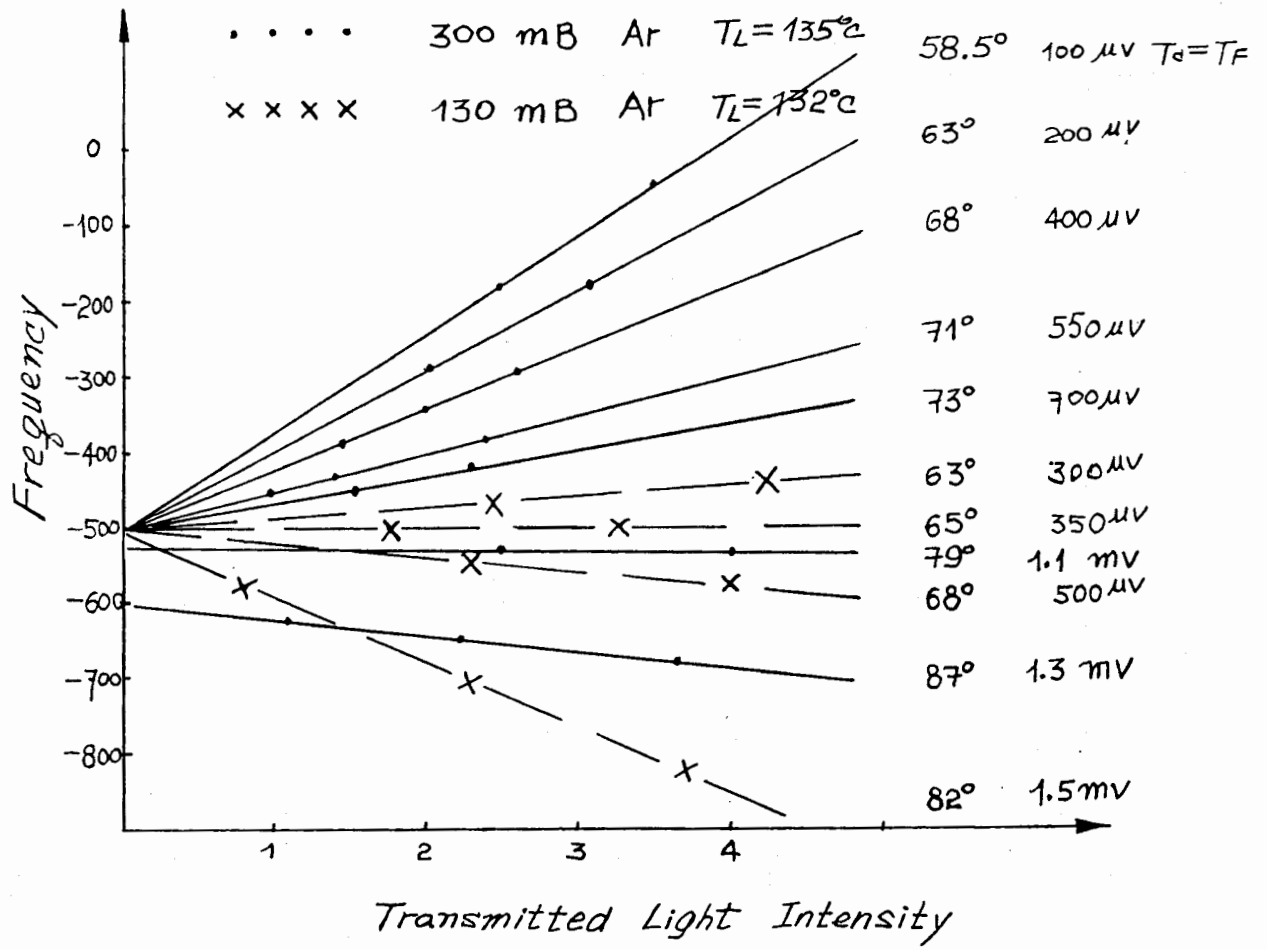


Fig. 1 UTILITY RUBIDIUM PHYSICAL PACKAGE



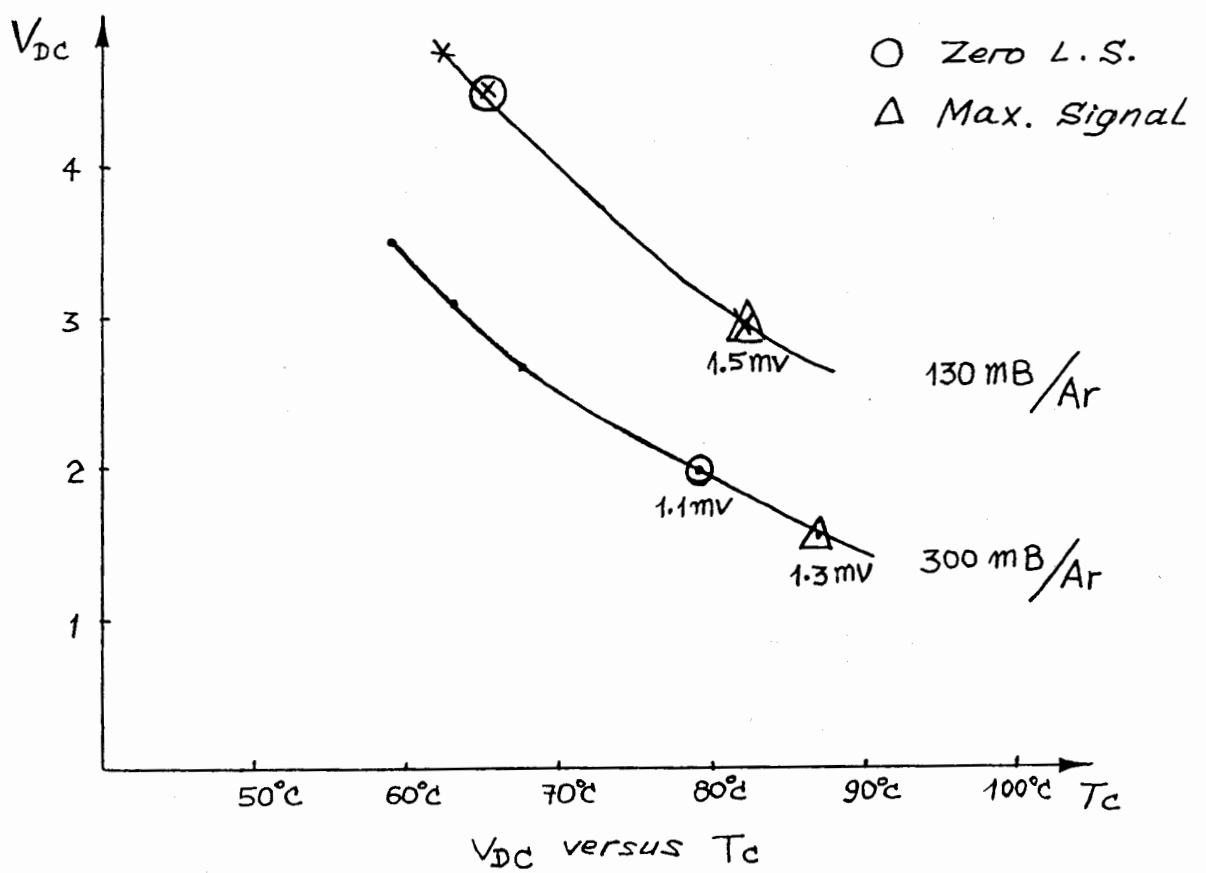


Fig. 2 Frequency shift as a function of the light intensity and as a function of the cell temperature

Performance v/s Size

	1	2	3	4	5	6
		Passive Rb Frequency Standard				
	Rb^{87} Maser	TE_{011} Rb NAT Int. Cell	TE_{011} Rb 87 Buff. Gas	TE_{011} Rb 87 Coating	TE_{111} Rb NAT Int. Cell	Utility Rb NAT
Volume of the cell (cm^3)	800	16	8	8	16	0.4
Temp. of the cell (°C)	60	80	60	60	80	87
Operating Linewidth (Hz)	100	500	517	800	700	4500
Exp. Discriminator Slope (PA/Hz)	2×10^5 *	400	1400	350	400	100
Exp. Detection Noise (PA/$\sqrt{Hz}$)			100	100		100
Theoretical Detection Noise (PA/$\sqrt{Hz}$)	16.8 *	4.73	5.6	5.7	7.55	5.6
Theoretical short-term Stability (τ = 1 sec)	1×10^{-14}	1.2×10^{-12}	4×10^{-13}	2×10^{-12}	1.96×10^{-12}	6×10^{-12}
Expected short-term Stability (τ = 1 sec)			7.4×10^{-12}	3×10^{-11}		1×10^{-10}
Exp. short-term Stability (τ = 1 sec)		5×10^{-12}	8×10^{-12}	3×10^{-11}		2×10^{-10}

* Calculated from the Maser parameters - for operation as passive standard.

2 ND EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 16 - 17 - 18 MARCH 1988

Session AFS 01 Number 5

Performance of Disciplined Rubidium Oscillators

Gerhard Hübner

Ball Efratom Elektronik GmbH, Munich West-Germany

The paper describes approaches to lock a precision local rubidium oscillator to a remote reference to improve accuracy and long-term stability. Various sources such as standard time and frequency transmitters, the Omega navigation system and the Global Positioning System (GPS) are discussed. Performance data of rubidium oscillators disciplined by LF and VLF standard frequency transmissions will be given.

Session AFS 01 Number 6

FREQUENCY SHIFT OF AN ATOMIC (OR MOLECULAR) RESONANCE BOUND TO THE INTERACTION TIME BETWEEN THE ATOM AND THE ELECTROMAGNETIC FIELD

J.H. BALBI*, D. GILLET**, Mme G. LOMAGLIO***, J.G. THEOBALD***

* Laboratoire d'Hélioénergétique - Université de Corte - 20250 CORTE

** Laboratoire de Piézo-Electricité et d'Electronique - ENSMM - La Bouloie 25030 BESANCON

*** Laboratoire de Spectroscopie Hertzienne et d'Electronique Faculté des Sciences - Route de Gray - La Bouloie - 25030 BESANCON

The width of the transitions used in atomic or molecular frequency standards is related to the interaction time between the atom and the electromagnetic field. But a frequency shift may also occurs, which is not bound to the irradiation level but to the interaction time.

Using the fictitions spin 1/2 and the Larmor theorem, which applies when atomic or molecular levels are pairs, we find that the transition probability induced by a field $\omega_1 = \gamma H_1 \sin(\Omega t + \varphi)$ applied during a time t ($\Omega t = n\pi$) is :

$$p = \frac{\Omega_1^2}{(\delta^2 - \Omega^2)^2} \left(\Omega^2 \cos^2\varphi + \delta^2 \sin^2\varphi\right) F^2\left(\frac{\delta t}{2}\right)$$

F = sin if n is even
F = cos if n is odd

$$\Omega_1 = \gamma H_1$$

The formula applies :

a) When the irradiation field has a sinusoïdal form: the resonance occurs when

$\Omega/\delta = 1 - 3\,!/n^2\pi^2$ if $\varphi = 0$; $\Omega/\delta = 1 - 3.3\,!/(n^2\pi^2 - 3\,!)$ if $\varphi = \frac{\pi}{2}$.

δ is then the steady magnetic field, that means the natural frequency.

b) When the irradiation field is submitted to a sinusoïdal modulation with frequency $\Omega/2\pi$ there are two resonances arround the natural frequency with

$$/\delta/ = \Omega \pm \delta_m \qquad \delta_m/\Omega = \pm 3\,!/n^2\pi^2$$

The sign is - if $\varphi = 0$ and + if $\varphi = \pi/2$

The experimental verification is performed on a Rabi-type formaldehyd molecular beam spectrometer on 11,5 and 4,3 lines with low values of n. In the case of a cesium beam, whe can obtain relative frequency shifts of 10^{-13} to 10^{-16}.

The theoretical method can also be applied to high irradiation levels and we give shortly results obtained on the Λ doublets of nitric oxyd with two-pulses irradiations.

AFS-02

Cesium Beam Frequency Standards

1. Long-term performance of the PTB's primary clock CS2
   A. Bauch, H. de Boer, B. Fischer, T. Heindorff, R. Schröder
   PTB, Braunschweig, FRG

2. The Effect of Feeding Junction Asymetries on End to End Phase Difference in Ramsey Cavities
   A. De Marchi
   Universita di Ancona, Ancona - Italy

3. Characterization of a Cesium Beam Detector
   H.S. Schweda, P. Thomann, G. Busca
   Oscilloquartz S.A., Neuchâtel - Switzerland

4. Full Digital-Processing in a New Commercial Cesium Standard
   J. Rabian
   Sté Sercel, Carquefou - France
   P. Rochat
   Oscilloquartz SA, Neuchâtel - Switzerland

5. Improving the Long-Term Stability of a Time Scale Generated by Commercial Cesium Clocks
   F. Cordara, V. Pettiti
   Istituto Elettrotecnico Nazionale Galileo Ferraris, Torino - Italy
   P. Tavella
   Associazione Nazionale Galileo Ferraris Grant, Torino - Italy

2 ND EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 16 - 17 - 18 MARCH 1988

Session AFS-02 Number 1

Long-term performance of the PTB's primary clock CS2

A. Bauch, H. de Boer, B. Fischer, T. Heindorff, R. Schröder
Physikalisch-Technische Bundesanstalt, Braunschweig, FRG

Summary

PTB's second primary clock, CS2, is compared with the first one, CS1, over a period of 700 days. The relative average frequency difference was found to be $25 \cdot 10^{-15}$ in agreement with the clocks' uncertainty figures of $15 \cdot 10^{-15}$ and $30 \cdot 10^{-15}$. A reduction of the C-field, and the servo system contribution to the total 1σ uncertainty of CS2 is reported. Cavity phase difference measurements carried out several times support the corresponding uncertainty estimate. The long-term frequency instability of CS2 versus CS1 for periods of 70 d is $10 \cdot 10^{-15}$. A floor of flicker noise modulation is not yet in evidence.

1. Introduction

Primary clocks with total relative uncertainties of a few parts in 10^{14} are gaining increasing importance in daily applications in the time and frequency community and no longer spend their lives more or less out of the public eye. This satisfactory development is the result of the continual progress in research work in all metrological institutes and the installation of the Global Positioning System, GPS. Primary clocks now claim uncertainty limits comparable to their long-term stability figures, while GPS offers direct access to these primary clocks without the need of expensive and laborious clock transportations. Nevertheless, industrially manufactured atomic clocks with their high flexibility still demonstrate their tremendous benefits such as the driving oscillator in a standard time and frequency transmitter or the on-board clock in a GPS space vehicle. In contrast primary clocks are stationary and require a group of skilled scientists and technicians for their operation.

If the number of primary clocks running without interruption over several years is growing, interesting questions might be answered. Misunderstandings arising from the non-uniform evaluation of the uncertainty of primary clocks can be eliminated. Close connections between various primary clocks will strengthen confidence in these clocks and eventually lead to an unambiguous consideration of the error budget.

In the operation of a primary clock, the uncertainty figure is usually determined as an estimate at the end of a limited time interval and future investigations will eventually improve it. Furthermore, frequency stability measurements with reference to a commercial cesium atomic clock normally show a flicker floor of a few parts in 10^{14}, usually associated with the latter one. But what happens if two primary clocks with almost equal uncertainty figures are compared? What about the flicker floor? If two primary clocks are evaluated independently,

their output frequencies should agree within the limits of their combined uncertainty figures. This is not proof of, but a necessary condition for a correct uncertainty figure estimate. Does a non-stationary noise process like the flicker noise, now prevent this verification, because in this case the average difference frequency value and its random uncertainty (precision) increase with the number of measurements? /1/, /2/. In the following we report the measuring results of continual frequency comparisons between the two primary clocks of the PTB: CS1 with a relative 1σ uncertainty of $30 \cdot 10^{-15}$ /3/ and CS2 with $15 \cdot 10^{-15}$ /4/. In addition, we discuss the uncertainty contributions of the C-field, the servo system and the cavity phase difference.

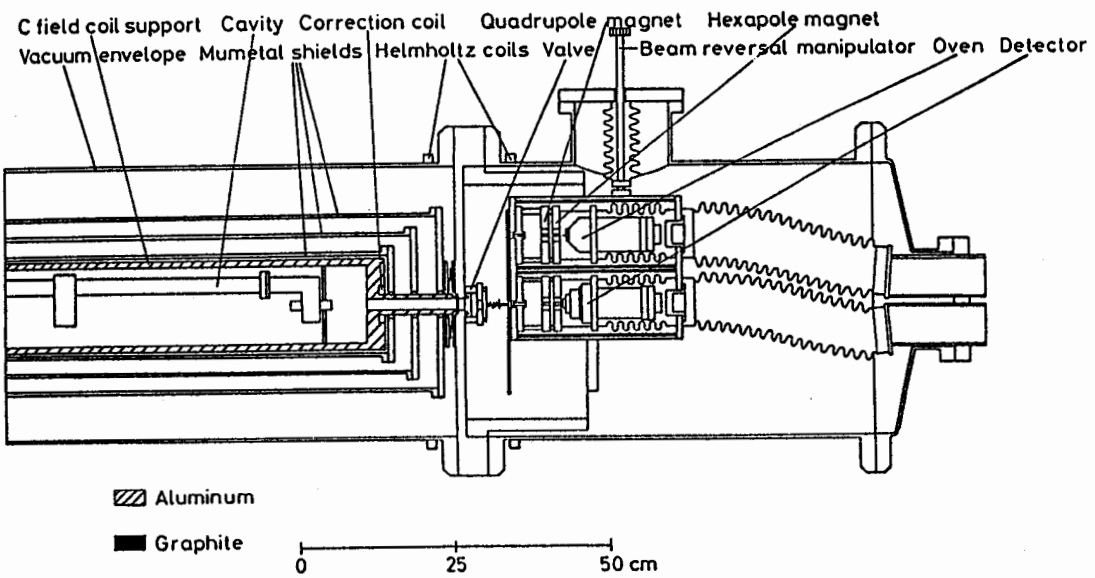


Fig. 1: Schematic drawing of one half of the PTB's second primary clock, CS2 (horizontal section, cavity rotated by 90°).

2. Error budget of CS2

Fig. 1 shows the design of PTB's second primary clock, CS2, while Table I depicts its systematic frequency corrections together with their uncertainty contributions. Compared with

Table I: Systematic CS2 frequency corrections and uncertainties in parts in 10^{15}

Type of Frequency Bias	Relative Frequency Bias	Relative Uncertainty
1. C-field	- 326254	5
2. Second-order Doppler effect	49	1
3. Cavity phase difference	246	10
4. Microwave power	-	3
5. Cavity pulling	-	<0.1
6. Spectral impurities	-	4
7. Rabi pulling	-	<0.1
8. Servo system	-	<2

an earlier paper /4/, further experiments yielded a change of the servo system uncertainty contribution from $< 10 \cdot 10^{-15}$ to $< 2 \cdot 10^{-15}$, the C-field contribution from $15 \cdot 10^{-15}$ to $5 \cdot 10^{-15}$, the spectral impurities contribution from $6 \cdot 10^{-15}$ to $4 \cdot 10^{-15}$ and the second-order Doppler effect contribution from $0.2 \cdot 10^{-15}$ to $1 \cdot 10^{-15}$. Thorough and extensive investigations of the Rabi pulling influence on CS2 and CSX (PTB's experimental atomic beam resonance apparatus) supported the corresponding contribution in the error budget as reported elsewhere /5/. In the following, we discuss in detail only the remarkable improvements in uncertainty contributions.

2.1. Servo system

Two kinds of experiments were carried out to eliminate possible failures of the servo system. In the first experiments which took place during 1986 - 1988, the connection between the cesium detector and the servo electronics was cut and the output voltage U_{Int} of the integrator controlling the VCO was observed over about one hour. Fig. 2 shows that U_{Int} has no systematic component but fluctuates randomly in a 17 µV interval. This is equivalent to possible frequency deviations of CS2, δF_{Int}, not larger than $\pm\ 1 \cdot 10^{-15}$.

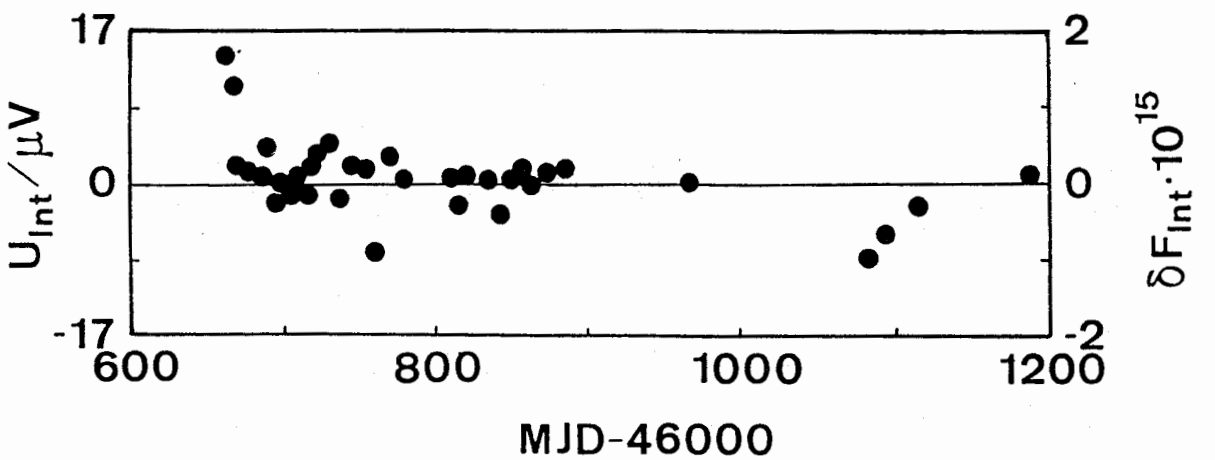


Fig. 2: Integrator output voltage U_{Int} with interrupted Cs ion current signal during the time 1986 - 1988. 8.5 µV integrator output voltage corresponds to a $1 \cdot 10^{-15}$ relative frequency offset of CS2.
MJD: Modified Julian Date.

Secondly, the servo system was operating with the detector current as an input, but the cesium atomic beam was shadowed out. The U_{Int} variations observed in a time interval of about 24 h turned out to result in corresponding frequency deviations of CS2 not exceeding $\pm\ 2 \cdot 10^{-15}$. Therefore the total servo system contribution in table I was estimated at $< 2 \cdot 10^{-15}$.

2.2. C-field

In the case of CS2, the Zeeman frequency is determined routinely with an uncertainty of 0.1 Hz. This amounts to an uncertainty contribution of $1.1 \cdot 10^{-15}$. After demagnetization, C-field reversal or C-field power supply exchange, the C-field strength, i.e. the Zeeman frequency was readjusted: Deviations δf_z of the Zeeman frequency from this value adjusted once were recorded in the following period till the

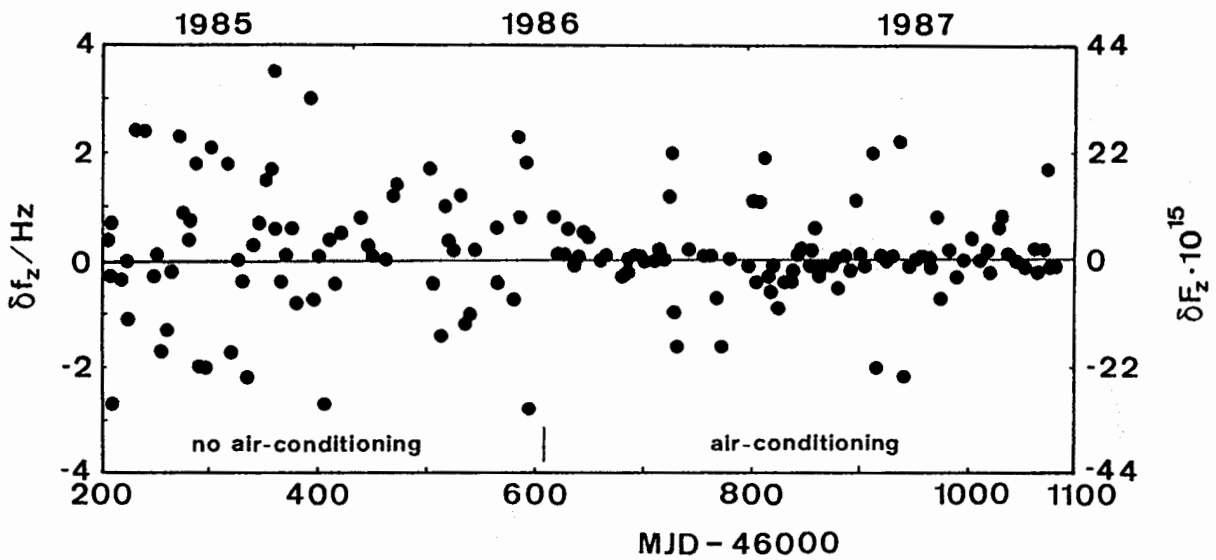


Fig. 3: Zeeman frequency changes of CS2, δf_z, in Hz and expressed as relative transition frequency corrections, δF_z. MJD: Modified Julian Date.

next break and so on. Fig. 3 gives an impression of the instability of the C-field of CS2. Obviously, the environmental conditions strongly influence this instability via temperature fluctuations working on the length of the C-field solenoid support. It becomes clear from Fig. 3 that air-conditioning improved the standard deviation of δf_z. The CS2 frequency is corrected for these Zeeman frequency changes twice a week. Summing up the contributions mentioned and that component which is caused by the temperature-dependent current flowing in the C-field coil (it yields $1 \cdot 10^{-15}$), one gets $5 \cdot 10^{-15}$, the value listed in table I.

2.3. Cavity phase difference

CS2 is a double beam device. The beam direction is reversed and the cavity phase difference φ is determined by measuring the frequency difference of CS2 between both beam directions, $2\delta F_\varphi$. Just before and after a beam reversal, a 10-day average of the frequency difference between CS2 and a reference clock is calculated. The reference clocks are CS1 and the least unstable industrial available cesium atomic clock of the PTB, model HP 5061 A, option 007 (super tube).

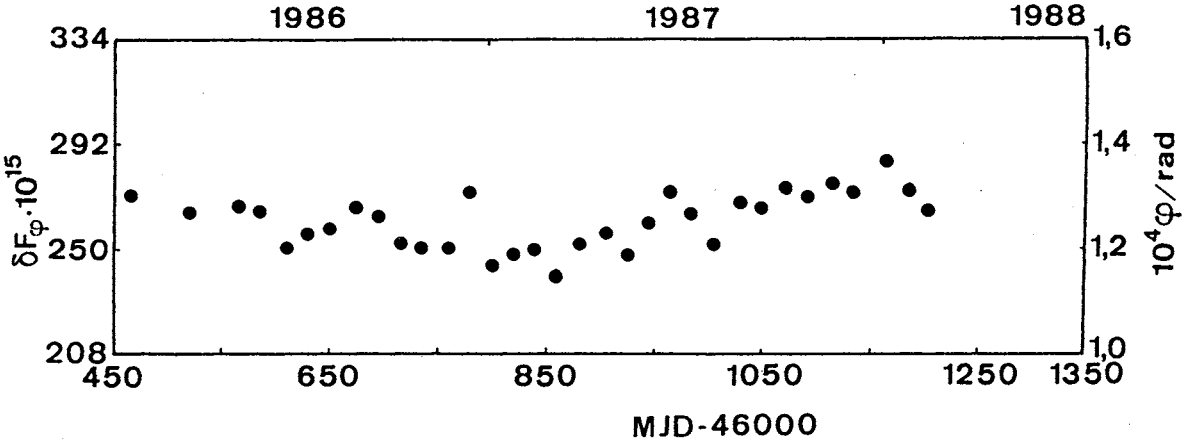


Fig. 4: The cavity phase difference φ and the corresponding frequency correction δF_φ of CS2 between 1986 and 1988. CS1 is used as reference clock (error bars of δF_φ: $\pm$ $11 \cdot 10^{-15}$). MJD: Modified Julian Date.

Fig. 4 demonstrates that the cavity phase difference φ of CS2 or the corresponding frequency correction δF_φ is very stable. φ is calculated from δF_φ by the equation $\varphi = \pi \cdot Q_l \cdot \delta F_\varphi$. Q_l is the line Q. If it is assumed for the moment that the maximum fluctuations of $\pm$ $20 \cdot 10^{-15}$ are only caused by mechanical imperfections of the adjustment of oven and detector positions, then it might be concluded from the phase gradient of $9.4 \cdot 10^{-5}$ rad/mm or $1.9 \cdot 10^{-13}$/mm /4/ measured in CS2, that oven and detector positions are uncertain

by approximately 0.1 mm. In reality, this is not more than a measure of reproducibility and at some future date a systematic error undiscovered up to now, may be revealed. Furthermore, these observed fluctuations include cavity phase difference variations and noise produced by beam reversals of the CS1 reference clock. In CS2, the beam is reversed every three weeks, in CS1 after a period of six weeks.

The experimental results of Fig. 4 confirm the estimated uncertainty contribution of the cavity phase difference of CS2 in table I.

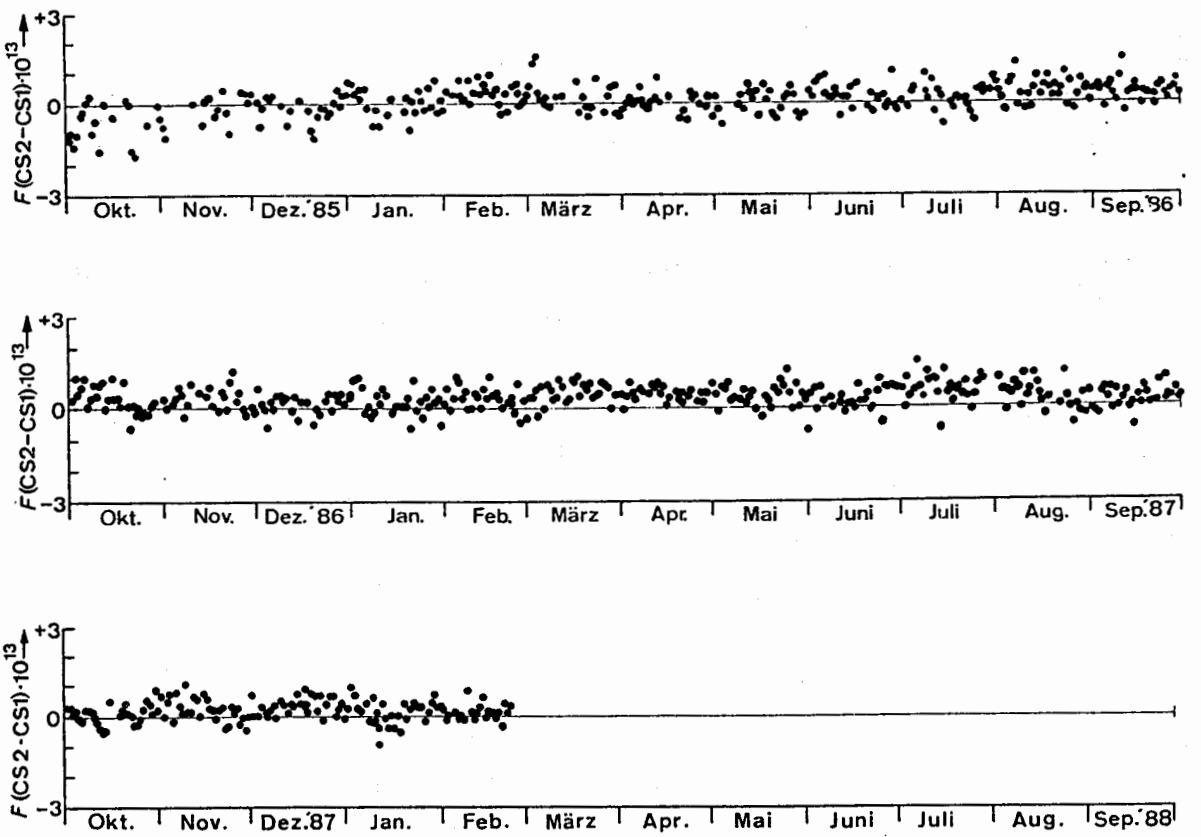


Fig. 5: 1d-average values of the relative frequency difference between PTB's two primary clocks, CS2 and CS1, F(CS2-CS1). MJD: Modified Julian Date.

3. Long-term time and frequency comparisons between two primary clocks

Fig. 5 gives a survey of the time and frequency measuring results of CS2 versus CS1 since 1985. The 1d-averages of the frequency difference between CS1 and CS2 were measured by the well-known start-stop method. It is remarkable that the relative frequency difference values F(CS2-CS1) indeed generally oscillate in the limits of the root of the sum of the squared total 1σ uncertainty values, $34 \cdot 10^{-15}$, or of the sum of these figures of CS2 and CS1, $45 \cdot 10^{-15}$, respectively. The over-all average amounts to $(25 \pm 2) \cdot 10^{-15}$. Its precision is correctly calculated if all the averaged values are independent of each other (white frequency noise modulation). A plot of the 2-sample standard deviation as a function of the sample time may decide the question of the underlying noise process and its power spectrum.

Fig. 6 shows in a log-log description the 2-sample standard deviation $\sigma_A(2,\tau)$ calculated on the basis of the frequency values of Fig. 5 as a function of the sample time τ. The signal-to-noise ratio and the linewidth yield a theoretical frequency instability for CS2 over 1s, $\sigma(2,1s) = 3 \cdot 10^{-12}$. Therefore the main part of $\sigma_A(2,1s) = 10600 \cdot 10^{-15}$ in Fig. 6 seems to be produced in CS1, as can be derived from a three-corner-hat analysis (a comparison between CS1, CS2, and a HP super tube). Between 2000 s $< \tau <$ 5 d, the frequency instability shows the expected behaviour of a white frequency noise process due to the shot noise of the cesium atomic beams. The average frequency values are not correlated and the 2-sample standard deviation has the same value as the usually calculated standard deviation /1/. For $\tau >$ 5 d the measuring points no longer follow the $\tau^{-0.5}$ slope. It is evident that the actual slope may be approximated by an exponent $0 \geq \mu \geq -0.5$. $\mu = 0$ is valid for flicker noise. The increasing error bars for the largest values of τ have made

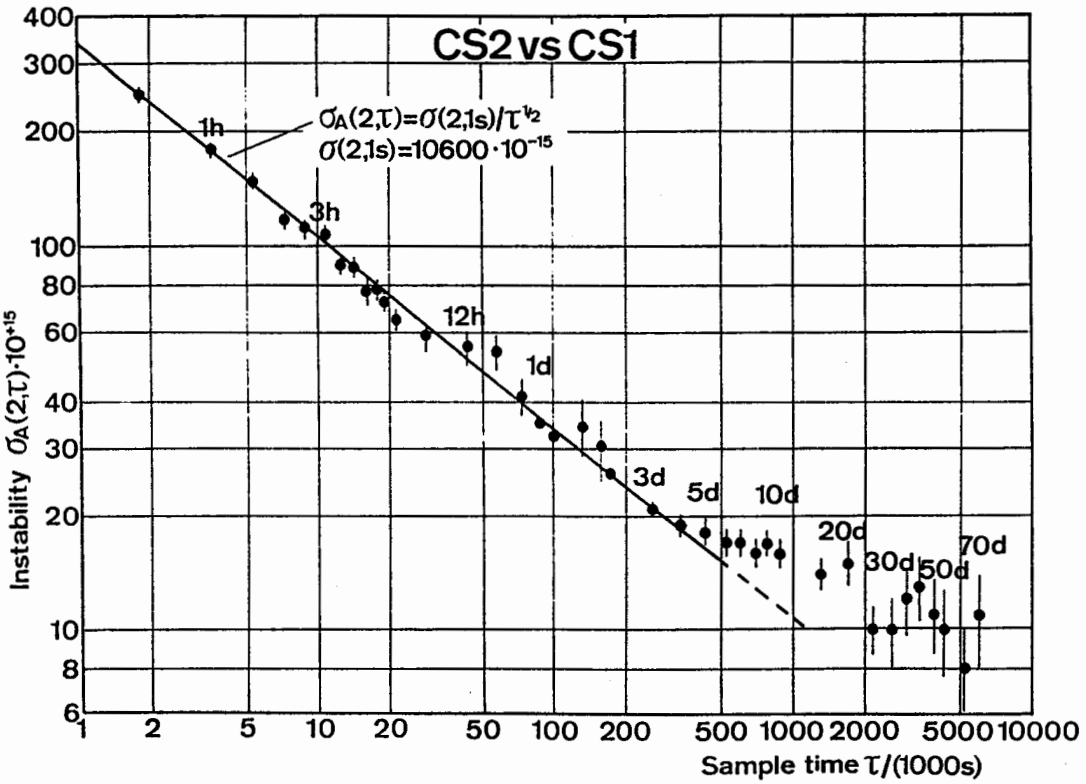


Fig. 6: 2-sample-standard deviation $\sigma_A(2,\tau)$ of CS2 versus Csl as function of the sample time τ.

it impossible up to now to answer the question whether there is a flicker floor in this diagram or not. Nevertheless, the usual estimate of the standard deviation of average frequency values in Fig. 5 is no longer a reliable procedure for $\tau > 5$ d and must be modified. The assumption of a stationary process might also be called into question /1/, /2/.

4. Conclusions

The operation and comparison of both PTB's primary clocks, CS1 and CS2, over a period of 700 days demonstrate their superb reliability and the very good agreement in the duration of their seconds. But the residual small systematic difference of $25 \cdot 10^{-15}$ remains the aim of our future investigations.

References

/1/ Allan, D.W.: Statistics of atomic frequency standards. Proc. IEEE 54 (1966), p. 221-230

/2/ Barnes, J.A.: Atomic time keeping and the statistics of precision signal generators. Proc. IEEE 54 (1966), p. 207-220

/3/ Dorenwendt, K.: Development and performance of the PTB's CS1 primary clock. Proc. IEEE 74 (1986), p. 137-140

/4/ Bauch, A.; Dorenwendt, K.; Fischer, B.; Heindorff, T.; Müller, E.K.; Schröder, R.: CS2: The PTB's new primary clock.
IEEE Trans.Instr.Meas. IM-36 (1987), p. 613-616

/5/ Bauch, A.; Dorenwendt, K.; Heindorff, T.: The PTB's atomic frequency standards CS2 and CSX: Frequency shifts by pulling due to neighbouring transitions. Metrologia 24 (1987), p. 199-203

2 ND EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 16 - 17 - 18 MARCH 1988

Session AFS-02 Number 2

THE EFFECT OF FEEDING JUNCTION ASYMMETRIES ON END TO END PHASE DIFFERENCE IN RAMSEY CAVITIES.

Andrea De Marchi

Dipartimemto di Elettronica e Automatica
Universita' di Ancona, Italy.

Abstract

An approximate formula has been traditionally used to express end-to-end phase difference in Ramsey cavities, as a function of their length and arm asymmetry [1]. Little consideration has been given to the feeding structure, which indeed contributes just as much to the overall asymmetry of the cavity.

In this paper a complete formal treatment of the whole problem is given with the help of the Scattering Matrix formalism.

As a consequence the untold approximations and assumptions which must be accepted for the traditional formula to hold are clarified, as well as the role of cavity tuning and junction asymmetry in determining the actual end-to-end phase difference.

Compensation of arm length asymmetry by the introduction of a suitable junction asymmetry appears feasible in principle.

1 Introduction

The end-to-end cavity phase shift has been considered a limitation toward improvement of accuracy as well as long term stability in commercial Cesium beam frequency standards [2]. It has been shown recently that other effects are the cause of many of the problems previously attributed to phase shift [3-5]. These studies have suggested that Accuracies of the order of 10^{-13} or better may be achievable in well aligned commercial standards.

However, in order to further study the limit to which this Accuracy can be pushed in standards where the cavity is symmetrized and the tube is sealed by the manufacturer, and beam reversal experiments cannot therefore be performed, a few additional considerations are in order, which will be spelled out in this paper.

The main question we are trying to answer is: 1) How well can the symmetry of the cavity be evaluated from the outside of the sealed tube? To this we will add the other question: 2) By any chance would it be possible to introduce from the outside of the sealed tube a minor adjustment to the symmetry of the cavity?

The latter question leads to assessing what is the effect of the asymmetries of the T-junction on the overall asymmetry of the cavity, which is the subject of this work as indicated in the title.

In the process of establishing the mathematical groundwork for this task, a full transmission line analysis of a complete Ramsey cavity will be given, with the help of the Scattering Matrix formalism. It will be shown that a Desired mode (mode D) is strongly coupled to the outside world, and that an Undesired mode (mode U) is weakly coupled, only through the asymmetries of the whole structure, when they exist. The latter has a high Q (basically the intrinsic Q of the unloaded cavity) and therefore it will be optimally excited at very low coupling, so that the existing asymmetries may be enough to transfer significant power into it.

This latter mode is responsible for the existence of a phase difference between the two ends of the cavity.

In order to simplify calculations it will be assumed that the transmission lines on the two sides are uniform, so that the only sources of asymmetry for the whole structure are their electrical length and average propagation constant γ ($\gamma=\alpha+j\beta$), the end short circuits, and the central T-junction feed.

2 Model and Analysis

The model of the Ramsey cavity and the definitions which will be used are shown in Fig.1, where the symbols are self explanatory.

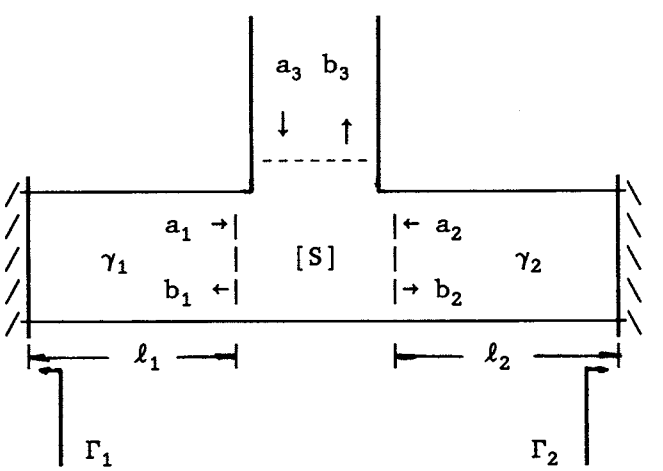


Fig.1 Transmission line model of the Ramsey cavity.

It is assumed that the cavity is made with standard X-band Copper waveguide, and that the T-junction is an E-plane or H-plane "symmetric" Tee. In this case the the junction will be accurately described by the Scattering Matrix S+E, where:

$$S = \begin{pmatrix} R & M & T \\ M & R & \mp T \\ T & \mp T & K \end{pmatrix} \quad ; \quad E = \begin{pmatrix} \epsilon_R & 0 & \epsilon_T \\ 0 & -\epsilon_R \pm \epsilon_T & \\ \epsilon_T \pm \epsilon_T & 0 & \end{pmatrix} \tag{1}$$

The upper sign is relative to the E-plane Tee case, and the lower sign is relative to the H-plane Tee. This result is obtained in [6] for an E-plane Tee with small displacements from symmetry. A relation between ϵ_R and ϵ_T is also obtained there in the case of small losses. Here however (1) will be considered a first order expansion of an exponential, and Matrix elements of the slightly asymmetric Tee will be written in the form $Te^{\epsilon_T/T}$ because this will make formulae easier to handle due to notation uniformity.

In fact, as we have done in other occasions, we will express also the reflection from the short circuits in a form which is homogeneous with the propagator $e^{-\gamma x}$ along x, that is as $\Gamma_s=-e^{-2\theta}$. In this way, by going through the same algebra as done in [6,7], the waves entering the two arms of the cavity from an E-plane junction are found to be

$$\begin{cases} b_1 = b\, e^{\gamma d}\, e^{\gamma_\delta \ell}\, e^{\epsilon'_T/T}\, e^{\theta_\delta}\, \dfrac{\cosh(\xi+\psi)}{\cosh \xi} \\ b_2 = -b e^{-\gamma d}\, e^{-\gamma_\delta \ell}\, e^{-\epsilon'_T/T}\, e^{-\theta_\delta}\, \dfrac{\cosh(\xi-\psi)}{\cosh \xi} \end{cases} \tag{2}$$

where $(b_1+b_2)/2=b$, and with $R+M=e^{-j\delta}$, δ being then real; $\ell_1-\ell_2=2d$ and $\ell_1+\ell_2=2\ell=L$; $\gamma_1-\gamma_2=2\gamma_\delta=2(\alpha_\delta+j\beta_\delta)$ and $\gamma_1+\gamma_2=2\gamma$; $\theta_1-\theta_2=2\theta_\delta$ and $\theta_1+\theta_2=2\theta$. Here and in the rest of this paper the single and double apex obviously indicate real and imaginary part of the quantity they are attached to. For an H-plane junction the cosh become sinh.

As for ξ and ψ the two following definitions have been used in (2):

$$\xi = \gamma\ell+\theta+j\delta/2 \tag{3}$$

$$\psi = -\gamma d-\gamma_\delta \ell-\theta_\delta+j\epsilon''_T/T \tag{4}$$

and it should be pointed out that ψ vanishes at symmetry.

Since the magnetic field $H_i(\ell_i)$ at the extremity of the arm i is given by

$$H_i(\ell_i) = b_i\, e^{-\gamma \ell_i}\, (1-\Gamma_{si}) \ , \tag{5}$$

with

$$1-\Gamma_{si} = e^{-\theta_i} \cosh \theta_i \ , \tag{6}$$

then the exponential terms in (2) relative to the asymmetry of arms and shorts disappear from the expression of H, and the ratio of magnetic fields at the two ends becomes:

$$\frac{H_1(\ell_1)}{H_2(\ell_2)} = e^{2\epsilon_T'/T}\,\frac{\cosh\theta_1}{\cosh\theta_2}\,\frac{\cosh(\xi+\psi)}{\cosh(\xi-\psi)}\,, \tag{7}$$

for the E-plane Tee case, with the two cosh in the last fraction substituted by sinh for the H-plane Tee.

The first term of (7) is just an amplitude term, and we are not interested in it as we are looking for end-to-end phase difference, in fact it causes only a second order effect. The second term is also very nearly real, as its phase is at most $4r\delta r$, r being the normalized surface impedance of the short circuit. Since the latter is smaller than 10^{-4} for Copper, even if $\delta r/r$ were as big as 20%, the phase of this term would still be smaller than .1μrad, and its module very close to one. Therefore the whole expression is equal to the third term. The third term in turn can be rewritten by taking into account that $|\psi| \ll 1$, and as a consequence

$$\cosh(\xi\pm\psi)/\cosh\xi \simeq 1\pm\psi\,\mathrm{tgh}\xi \quad ; \quad \sinh(\xi\pm\psi)/\sinh\xi \simeq 1\pm\psi\,\mathrm{ctgh}\xi \tag{8}$$

for the two cases of E and H-plane Tee respectively.

The term $\mathrm{tgh}\xi$ (or $\mathrm{ctgh}\xi$) represents the undesired resonances of the structure (U mode), and it has no effect when $\psi=0$, in which case then the end-to-end phase difference will be zero. The U mode resonance term can be written in the form $B\,e^{-j\sigma}$, and it is:

$$B^2 = \frac{\cosh 2\xi' - \cos 2\xi''}{\cosh 2\xi' + \cos 2\xi''} \tag{9}$$

$$\mathrm{tg}\sigma = -\sin 2\xi''/\sinh 2\xi' \tag{10}$$

for the E-plane Tee case. For the H-plane Tee case, B is the reciprocal of (10) and σ is changed of sign. The behaviour of B and σ as a function of frequency is shown qualitatively in Fig.2.

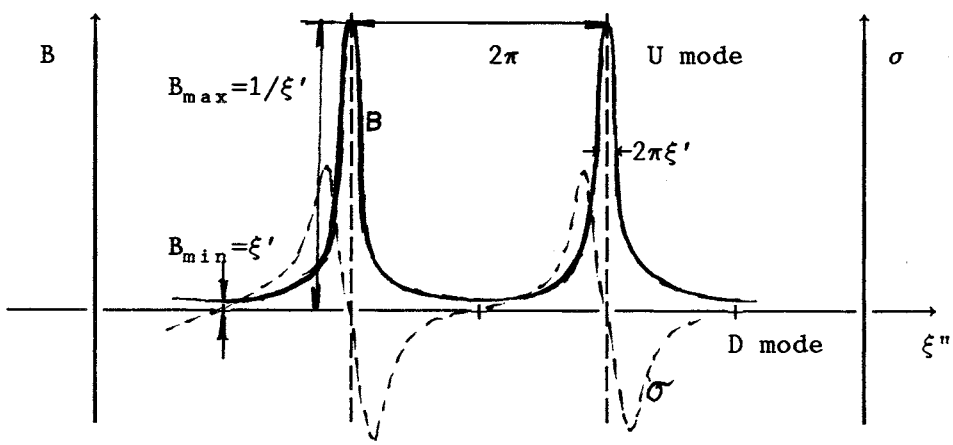


Fig.2 Qualitative plot of modulus and phase of $\mathrm{tgh}\xi$.

The resonances of the U mode appear at $2\xi''=2\beta\ell+2\theta''+\delta=(2n+1)\pi$ for an E-plane feed Tee (odd number of half wavelengths), and at $2\xi''=2n\pi$ for an H-plane feed Tee (even number of half wavelengths). In both cases the finesse is $1/\xi'$, as can be seen in Fig.2.

At resonance for the U mode it is $B=1/\xi'$ and $\sigma=0$; at its antiresonance, which is the resonance of the D mode, it is also $\sigma=0$, but $B=\xi'$.

Therefore close to the D mode, where the cavity is going to be operated, it is $1\pm\psi Be^{-j\sigma}=1\pm\xi'[\psi'+\sigma\psi''+j(\psi''-\sigma\psi')]$, with phase $\pm\xi'(\psi''-\sigma\psi')/[1\pm\xi'(\psi'+\sigma\psi'')]$. The phase difference between the two ends of the cavity can now be expressed for the optimum cavity tuning (half way between contiguous U modes). It is:

$$\phi_1-\phi_2 \simeq 2\xi'\psi'' = -2(\alpha\ell+\theta')(\beta d+\beta_\delta\ell+\theta''_\delta-\epsilon''_T/T) \qquad (11)$$

The latter term of the equation coming from (3) and (4). It is clear that if the asymmetry of the cavity is such as to compensate the asymmetries of arms and end shorts, in the sense of making the last parenthesis vanish, then the end to end phase difference is zero. This condition can be monitored from the outside of the cavity because then the sharp U modes don't show up in reflection.

This is interesting because it means that the end-to-end phase difference is determined by the symmetry of the whole structure, and not just by the arms asymmetry as previously understood. It should be pointed out here that by taking only the first term of each parenthesis in (11) one obtains the usual formula that is always quoted for the end-to-end phase difference.

It is also clear that (7) depends on frequency trough γ. Since (11) was obtained for operation at antiresonance for the U modes, if the tuning of the cavity were to deviate from this point, the actual value of the phase in operation difference could vary from what found here.

3 Correction of end-to-end phase difference

Since the symmetry of the junction can be affected by modifications introduced in the "symmetric" port waveguide, then (11) shows that it is possible to modify end-to-end phase shift from the outside of the vacuum sealed tube. Now the question is how much can be done in this direction.

We have to point out here that being able to affect the symmetry from the outside also means risking to spoil with a less than perfect junction the symmetry of a well trimmed cavity. For example an asymmetry of 10^{-3} can be introduced by angular deviations from perpendicularity of about 5 mrad in the E-plane of an E-plane Tee junction [8]. This is not very big and manifacturing imperfections may easily introduce it. Since $\alpha\simeq 2\times10^{-2}$ for a standard Copper X band waveguide, and in commercial tubes with a line Q of the order of 3×10^7 is $2\ell\simeq0.25$m, it can be seen from (11) that such an asymmetry corresponds to a frequency correction of about 10^{-13} (10^{-5} rad).

In practice in order to ensure an accuracy of 10^{-13} one might want to correct a phase shift a few times that much. It looks feasible to introduce intentionally a Tee junction asymmetry of the order of magnitude needed to compensate end-to-end phase shifts of several 10^{-13}.

This can be done for example by inserting in an asymmetric way a waveguide tuning post in the "symmetric" arm of the junction. The distance of this post from the Tee must be small enough for the superior modes not to be yet "completely" attenuated, but great enough for the effect not to be overwhelming. What this means is that the distance should be a certain number of penetration lengths of the first asymmetric superior mode. In principle this number can be calculated.

For an E-plane Tee for example one could insert a slug as shown in Fig.3, and introduce an asymmetry by the excitation of the TE_{20} mode, which is the first superior mode, with a cutoff frequency of about 26 GHz, about 2.5 times higher than the operating frequency. The penetration depth of this mode into the standard X band waveguide is then about 2 mm. A discontinuity at about 15 mm from the junction is about 8 penetration depths away from the point where an effect is felt on the asymmetry of the Scattering Matrix. The attenuation of the effect at this distance is then about a factor of 10^{-3}, or 60 dB. This may seem like a very high attenuation, however the remaining asymmetry needs only to be of the order of a few times that much to be able to take care of most of the arm asymmetries that one might want to compensate. Therefore it is not out of the question to think about such a possibility.

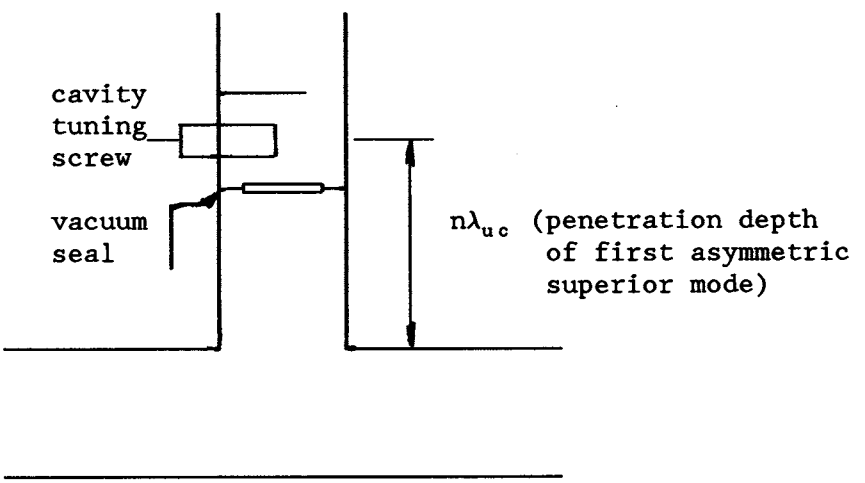


Fig.3 Example of a comfiguration capable of introducing an asymmetry in an E-plane waveguide junction. The actual arrangement in many commercial tubes is not too far from what shown here.

The arrangement of the cavity tuning slug in HP commercial tubes is not far from what schematically shown in Fig.3, and in fact an indication that such a compensation may be possible for those tubes can be obtained from frequency measurements at Optimum Power with different positions of the tuning slug that we are discussing. In fact a preliminary indication that this may be the case has been obtained with the fact that the frequency changes. This should not happen at Optimum Power if cavity pulling were the only effect to change. It is the author's feeling that a check of this hypothesis made by reversing the sign of the asymmetry without touching the tuning slug should give a positive answer. Unfortunately this check could not be performed as yet.

5 Conclusions

A full transmission line analysis of a complete Ramsey cavity was given, with the help of the Scattering Matrix formalism. It was shown that a Desired mode (mode D) is strongly coupled to the outside world, and that an Undesired mode (mode U) is weakly coupled, only through the asymmetries of the whole structure, when they exist. It was also shown that the U mode is responsible for the existence of an end-to-end phase difference if the two arms are uniform, so that the existence of an end-to-end phase difference can be inferred from the outside of the cavity by the observation of the sharp U mode.

Compensation of the latter can be tried from the outside of the cavity by introducing an artificial asymmetry in the feeding junction to minimize the observed U mode.

Limits to the effectiveness of this technique may be set by the existence in the cavity of asymmetric reflections, and an experimental analysis was suggested which should be able to provide information on the magnitude of the problem.

References

[1] - R.F.Lacey: "Phase shift in microwave Ramsey structures" - Proc. 22nd Freq.Contr.Symp., p.545 (1968).

[2] - D.W.Allan, H.W.Hellwig, S.Jarvis Jr., D.A.Howe, R.M.Garvey: " Some causes and cures of frequency instabilities (drift & noise) in Cesium beam frequency standards" - Proc. 31st Freq.Contr.Symp., p.555 (1977).

[3] - A.De Marchi, G.P.Bava: "On cavity phase shift in commercial Cesium beam frequency standards" - Metrologia 20, p.33 (1984).

[4] - A.De Marchi, G.D.Rovera, A.Premoli: "Pulling by neighbouring transitions and its effects on the performances of Cesium beam frequency standards" - Metrologia 20, p.37 (1984).

[5] - A.De Marchi: "Rabi pulling and long term stability in Cesium beam frequncy standards" - IEEE Trans.on Ultrasonics, Ferroel. and Freq. Contr., UFFC-34, p.598 (1987).

[6] - A.De Marchi: "A novel cavity design for minimization of distributed phase shift in atomic beam frequency standards" - Proc. 40th Freq.Contr.Symp., p.441 (1986).

[7] - A.De Marchi, J.Shirley, R.E.Drullinger: "A new cavity configuration for primary Cesium beam frequency standards" - IEEE Trans.on Instr. and Meas., (1988).

[8] - L.Lewin: Private Communication, (1987).

2 ND EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 16 - 17 - 18 MARCH 1988

Session AFS-02 Number 3

CHARACTERIZATION OF A CESIUM BEAM DETECTOR

H.S.Schweda, P.Thomann, and G.Busca

Oscilloquartz SA, Neuchâtel, Switzerland

ABSTRACT

The development, characterization and optimization of a standard Cs beam detector consisting of a hot wire detector, mass filter and electron multiplier is presented. Frequency response, gain, noise behavior and long-term stability of the system are investigated. The influence of air exposure, Cs ion and electron bombardment on the performance of the electron multiplier is studied.

INTRODUCTION

In the final part of a compact Cs-tube after microwave interrogation sits the cesium beam detector consisting of a hot wire ionizer, mass filter and electron multiplier. Before the atomic beam flux is seen as an electrical signal at the input of the phase-sensitive detector it has undergone a twofold transformation, from atoms to ions and from ions to electrons, and it has been mass selected and amplified. The various physical processes involved introduce different characteristics in efficiency, time response, and noise behavior which have to be studied carefully to optimize the performance of the detector. Above all the stability of the different parameters is of utmost importance, since commercial Cs-tubes are designed for continuous operation over many years.

In the course of a redesign of a Cs-tube we had the chance to have a fresh look at the different components of the detector and the materials used which stayed virtually unchanged since the early days of Cs-tube development.

HOT WIRE IONIZER

The detection of neutral atoms by ionizing them on a suitable heated surface is known for many centuries. The probability α that a positive ion rather than a neutral particle will be vaporized is given by the Saha-equation (1)

$$\alpha = \frac{N_+}{N_a} = \frac{g_+}{g_a} \exp\left(\frac{\phi - I}{kT}\right)$$

where g_+/g_a is the ratio of the statistical weights of the ground states of the ion and the atom, respectively (= 1/2 for Cs), I the ionization potential of the atom, ϕ the work function of the surface and k the Boltzmann constant. It is evident that for $\phi > I$ the emission of ions is highly favored. The ionization coefficient β is the ratio of ions N_+ leaving the surface and the number of incident atoms N

$$\beta = \frac{N_+}{N} = \frac{\alpha}{\alpha + 1}$$

Due to the low ionization potential of Cs (I = 3.88 eV) many metals can be used as efficient ionizers. However, in order to avoid adsorption of Cs an elevated temperature at a low vapor pressure has to be maintained. Possible Cs ionizer materials are listed in the following table.

Metal	Work function [eV]	Vapor pressure [torr]	Electrical resistivity [$\mu\Omega$cm]	Thermal conductivity [cal/cmKs]
Ir	5,0	2.10^{-9}	5,3	0,14
Mo	4,5	1.10^{-9}	5,2	0,34
Ni	5,1	6.10^{-3}	6,9	0,22
Os	4,8	$<10^{-11}$	9,5	-
Pt	5,6	1.10^{-6}	10,6	0,17
Re	5,7	$<10^{-11}$	19,3	0,17
Rh	5,0	2.10^{-6}	4,5	0,21
Ta	4,25	$<10^{-11}$	12,5 51*	0,13 0,16*
W	4,6	$<10^{-11}$	5,6	0,4
		[at 1500°c]		* at 1200K

Tantalum and rhenium combine two advantages : the higher electric resistivity allows to dissipate the heating power at a relatively low current and the low thermal conductivity assures a more homogeneous temperature along the ionizer filament. Other alkalis are sometimes present in the

Cs beam and almost always in the filament itself. Their work functions are ; Li : 5.4eV, Na : 5.12eV, K : 4.32eV, Rb : 4.12eV. In order to be selective against other alkalis the Saha equation shows that one should choose a not too high work function. Here tantalum is much more selective than rhenium. Fig. 1 and Fig. 2 show the ionization coefficients of the two metals for Cs, K and Na.

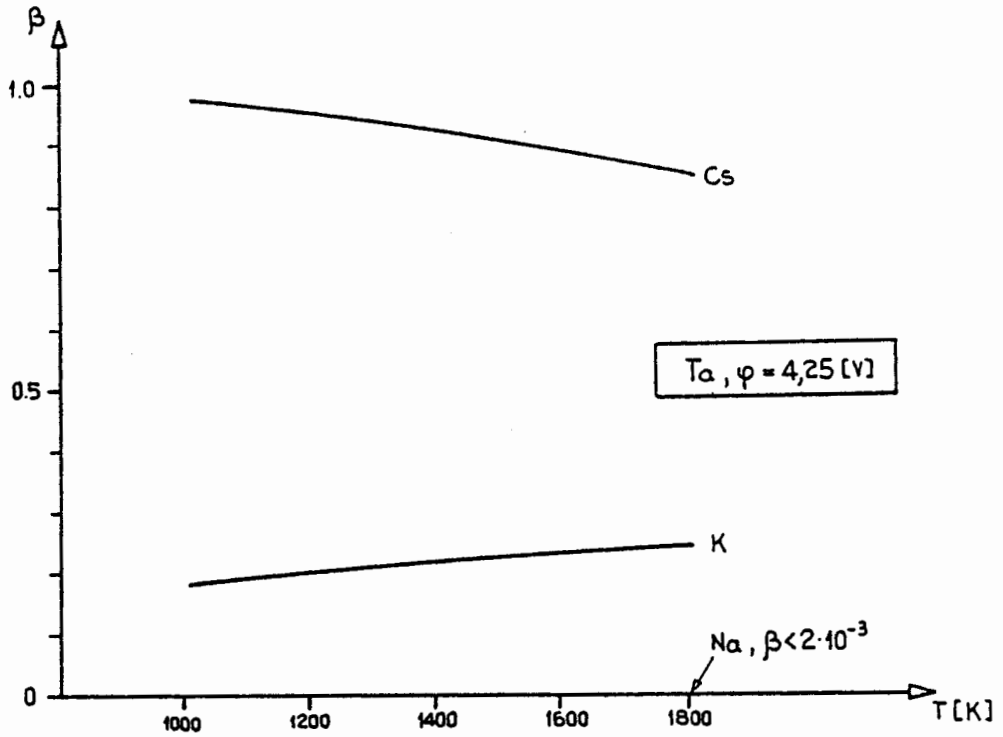


FIGURE 1

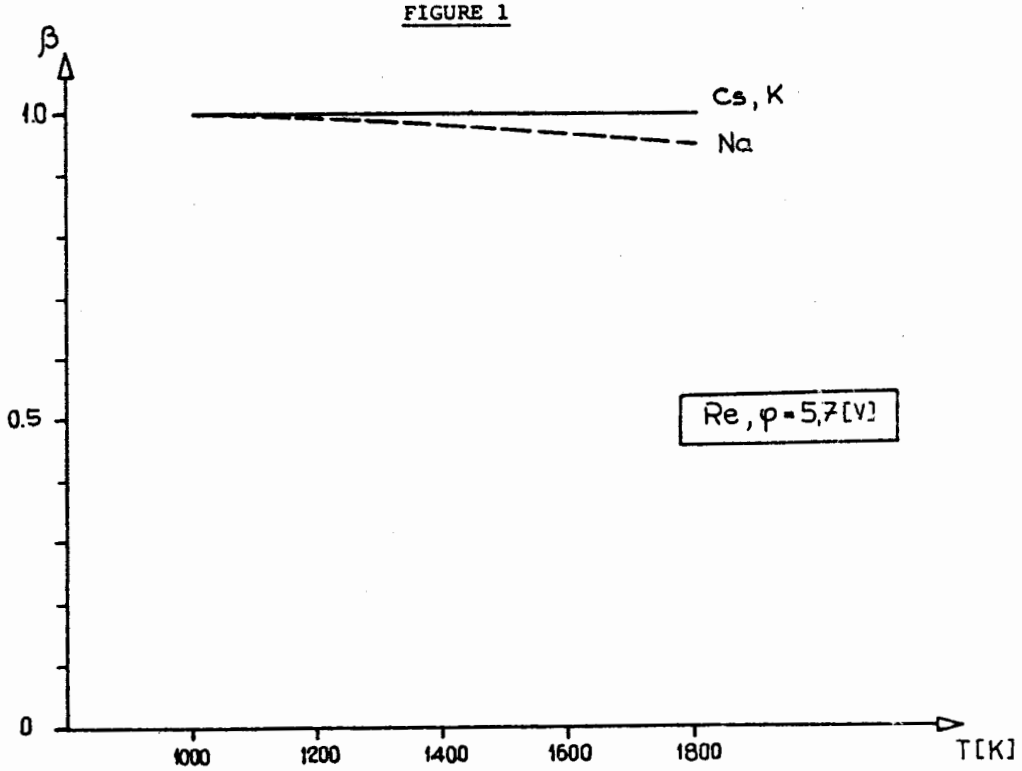


FIGURE 2

Another important property is the time response of the filament. It is well known that surface ionization is not an instantaneous phenomenon and that the dwell time of the ion on the filament can be quite large. The temperature dependence is given by

$$\tau = \tau_{\infty} \exp (Q_+ / kT)$$

with Q_+ being the ionic heat of desorption. The response of a tantalum filament (12mm x 1mm x 25μm) to a Cs beam modulated at 264Hz as function of the heater current (Fig.3) shows that the current has to be greater than 2.2A (T = 1200°C) for the modulation to be fully transmitted. A more detailed treatment shows that the filament behaves similar to a simple low-pass filter with cut-off frequency $\nu_{co} = 1/2 \pi \tau$ (2)

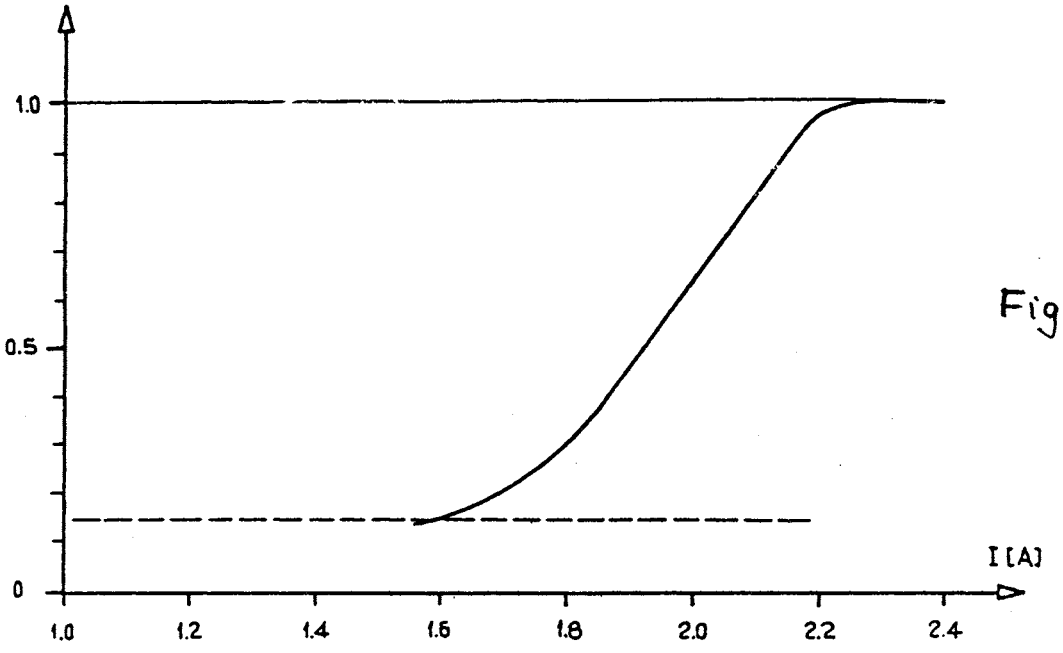


Fig. 3

ELECTRON MULTIPLIER

The effective Cs-beam flux at the exit of the atomic beam magnetic resonance set-up in the Cs-tube is about 5 x 10^6 atoms/s under typical operation conditions. The hot-wire detector converts this into a signal current of about 7.5 x 10^{-13}A. Phase modulation of the microwave field at 137 Hz causes the Cs-beam intensity to be modulated at this frequency. The noise current of the FET amplifier at the entrance of the measuring circuit is as low as 1 x 10^{-16}A/ $\sqrt{\text{Hz}}$ with a band width of 500 Hz. The

thermal noise of the input resistor dominates the noise level. The thermal noise is reduced by increasing the resistance, while the frequency response decreases. With an input resistance of $10^8 \Omega$ the thermal noise formula

$$i_N^2 = 4kT \, \Delta f / R$$

gives

$$i_N = 2.14 \times 10^{-14} A / \sqrt{Hz}$$

For a bandwidth of 300 Hz the noise level is $i_N = 3.7 \times 10^{-13} A$ which is in the order of the signal current. In order to achieve a signal-to-noise ratio of 1000 or better the signal has to be boosted by at least this factor with no significant additional noise.
This amplification step is usually provided by an electron multiplier (EM) which is inclosed in the Cs-tube. The characteristics of this EM should fulfil the following conditions :

1) Gain of 10^3-10^5 at 2-2.5 kV
2) High gain stability under continuous operation for many years
3) Detection efficiency for Cs^+ ions $\geqq 1$
4) Low noise and low background current
5) Band width $\geqq$ 300 Hz
6) Bakability in vacuum up to 400^oC
7) Small dimensions.

In compact Cs-tubes, one uses for this purpose discrete dynode EMs with 6-9 stages of the so-called box-and-grid type. The secondary electron emitting (SEE) material is Ag-Mg(MgO). These units fulfil more or less the conditions given above, with the gain stability being the weakest point. Good units have an initial gain of about 10^5 and show a relative gain degradation of 50% per year of operation. Readjustment of the gain by using a higher overall voltage is only possible in a limited range and leads in general to a faster decay. In frequent cases the gain degradation of the EM is the limiting factor in the life of a Cs-tube.

Our original idea to replace the discrete dynode EM by a continuous dynode channeltron had to be given up for several reasons. By studying the literature and the data of commercially available products it turned out that the long-term gain stability of the most advanced units was still inferior to the discrete dynode EM. Moreover many channeltrons do not withstand the temperature of the bake-out cycle. An own development did not seem to be advisable.

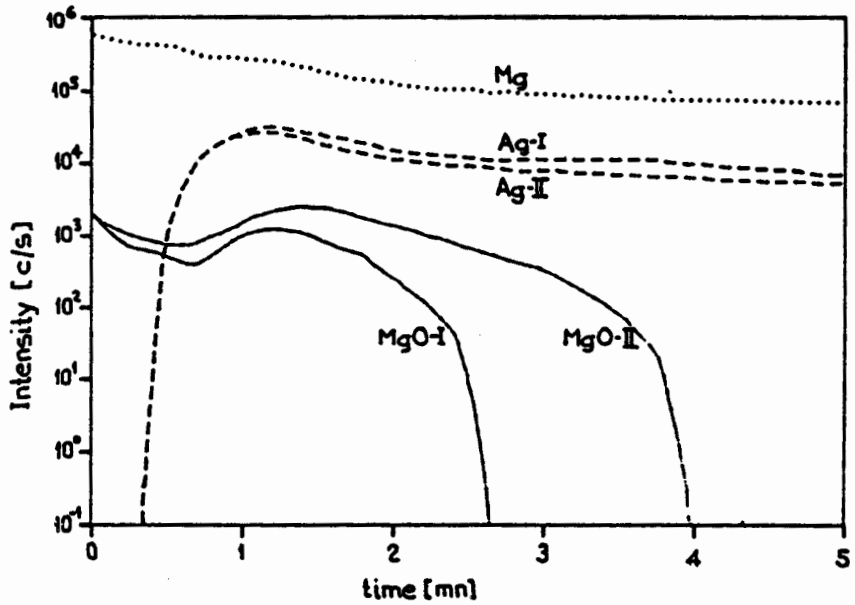
Metal oxides, i.e. MgO, BeO, Al_2O_3, are the usual choice as dynode material for open electron multipliers. These materials show not only a good SEE yield δ at relatively low energies but also a reasonable ion-to-electron conversion efficiency γ . With them the EM can be used as both converter and amplifier with the first dynode acting as conversion electrode. Unlike alkali compound dynodes which can only be used in sealed photomultipliers, these materials can be exposed to air and baked out.

Among the different oxides MgO has the highest SEE and Cs^+-ion conversion efficiency, although BeO and Al_2O_3 are slightly more stable under air exposure. Traditionally MgO dynodes are produced by selective surface oxidation of an Ag-Mg alloy at elevated temperatures in a low pressure atmosphere of oxidant gases. Only some percent of Mg are contained in the alloy. Careful adjustment of pressure, temperature and time is necessary to obtain good results. A number of alchimistic recipes are given in the literature. The gist of the process is to selectively oxidize Mg only at the surface of the alloy.

We have tried to rationalize this formation process. Following the ideas of Hirashima (3) on parabolic oxidation, we present a simple model of surface oxidation of Ag-Mg in the appendix.

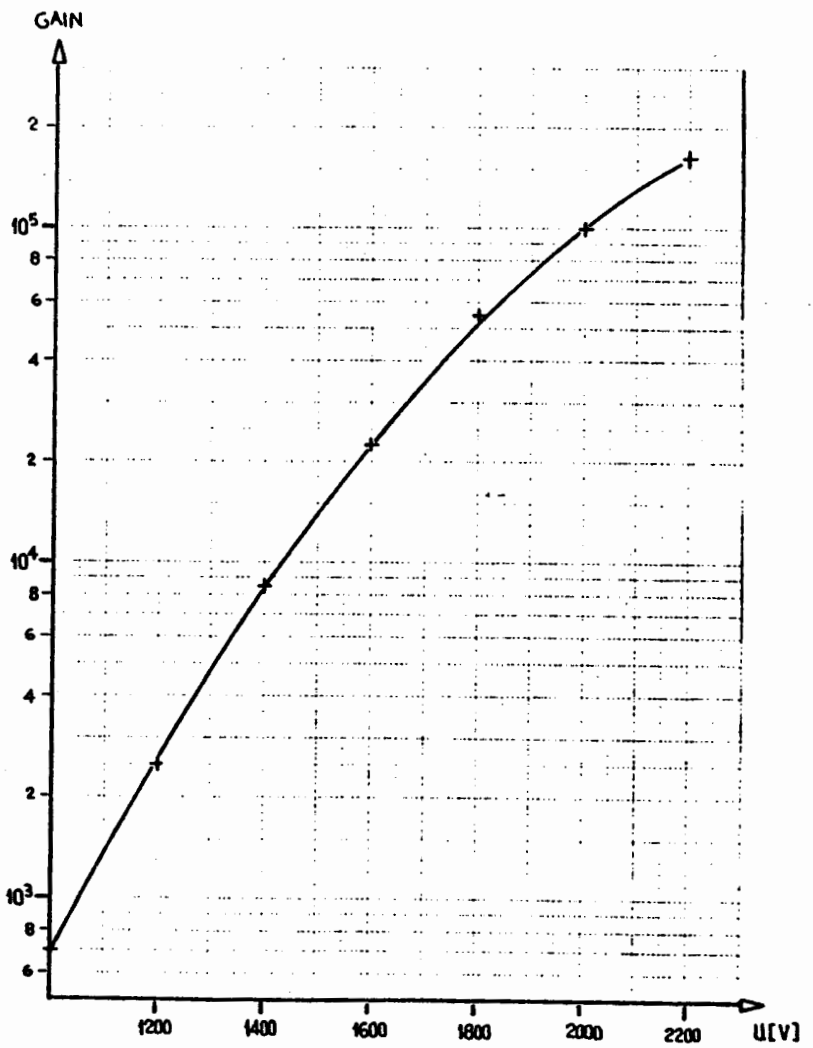
We have investigated dynode surfaces produced in the way described by secondary ion mass spectrometry (SIMS) and electron microprobe techniques. The results indicate a surface layer of almost stoichiometric composition on top of a region with higher Mg content. Fig. 4 compares two SIMS depth profiles of dynodes produced with one and two-step oxidations respectively. The abscissa is the sputtering time, an uncalibrated measure of the depth. The higher MgO content in deeper regions of the two-step oxidized sample is clearly visible. The complete MgO depth is several hundred ångströms.

Dynodes produced with this technique showed variable maximal yields $\delta_m = 5 - 6.5$ at 550 eV. Electron multipliers with 8 dynodes had a gain of $G = \gamma \delta^7$ up to 10^5 at 2 kV overall voltage. The Cs^+-to-electron conversion efficiency γ was close to unity. However, the long-term stability under operation was still unsatisfactory. Figure 5 shows the initial EM gain as function of the overall voltage.



I- one step activation II- two step activation

Fig. 4



Initial gain of 8-stage EMT vs. voltage

Fig. 5

Sputtered dynodes

Dynodes of superior quality have been obtained by sputtering off a high purity magnesium oxide target on a electropolished stainless steel sheet. With this technique thickness, composition and structure of the films can be controlled by varying process parameters, i.e. deposition rate, target temperature, pressure etc. Highly stoichiometric MgO films of

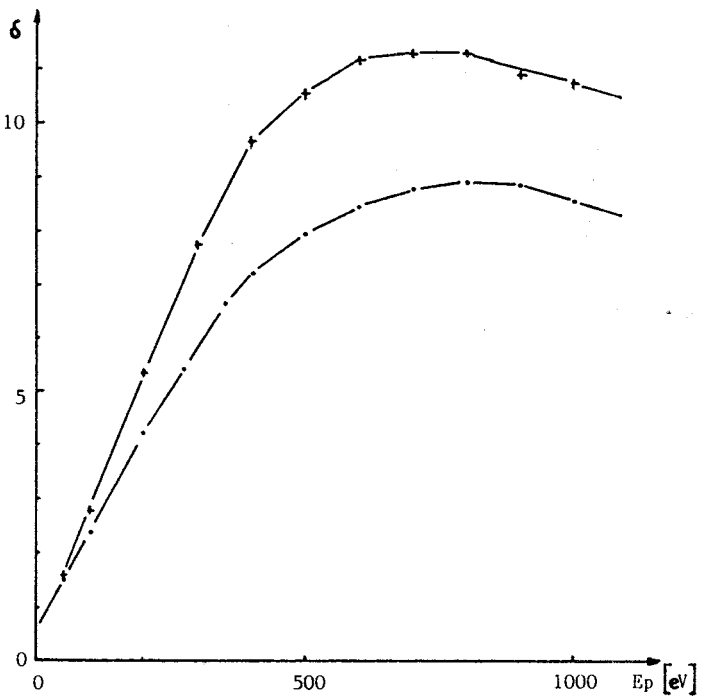


Fig. 6

SEE yield of two sputtered MgO dynodes as function of primary energy

polycrystalline structure have been obtained this way. Fig. 6 shows secondary electron emission yields δ of two samples as function of the energy of the primaries. The long-term behavior under operation is depicted in Fig. 7. . The abscissa shows the time of bombardement and the

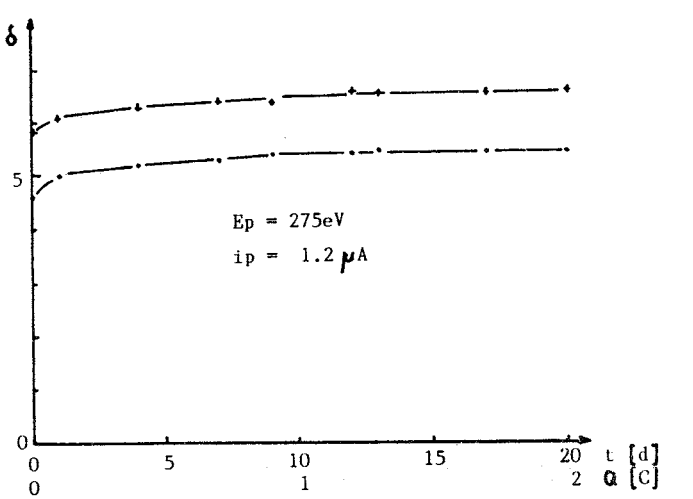


Fig. 7

SEE yield as function of electron bombardment time

complete charge laid down at the dynode. The degradation of SEE under e^--bombardement is due to incongruent dissociation of the oxide and proportional to the accumulated charge (4). The complete charge uptake of the last dynode of the EM under typical operation conditions in the Cs-tube is 2 coulombs in 6 years. Complete EMs with such dynodes had electronic gains of about 4×10^5 at 2kV and showed no gain degradation under comparable charge uptake. The Cs^+-to-electron conversion as function of the energy of the primaries is displayed in Fig. 8.

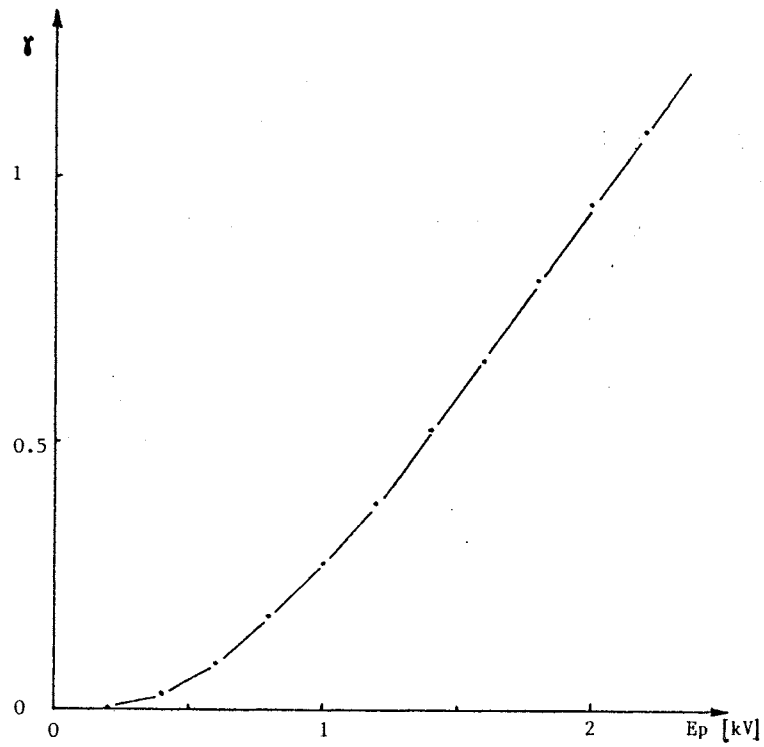


Fig. 8

Cs^+- to-electron conversion efficiency as function of ionic energy

MgO is among the best insulators known. With no conduction through the film the surface would charge up positively under Cs^+ ion bombardement and $\gamma > 0$ or electron bombardement and $\delta > 1$. In layers with thicknesses > 2000 Å field-enhanced or sustained secondary electron emission (Malter effect) is observed which make such films useless as real EM dynode materials. In thinner films of some hundreds ångströms thickness often some sort of conduction mechanism occurs, i.e. breakdown along grain boundaries, tunneling between residual metal precipitates, hole hopping

etc. Nevertheless significant surface charging has been observed for thermally grown MgO layers of this thickness which led to a delayed SEE component which persists as long as 10^{-2}s after interruption of the primary beam (5). Although such films can be used as dynode materials they are limited to slow time response applications.

Sputtered MgO dynodes could be produced so that surface charging was negligible. No influence on the time response of the electron multiplier could be observed in the modulation frequency range used in the Cs-tube.

After the film production, the dynode has to be exposed to air at least once, when the EM is assembled. Although air exposure does not destroy a MgO dynode, the oxide is rather hygroscopic and adsorption and surface reaction with water deteriorates the SEE yield. Fig. 9 displays the yield decay of a dynode after prolonged exposure to room air (25°C, approx. 50% rel. hum.) and subsequent measurements. Unfortunately, the Cs^+-to-electron conversion efficiency seems to be affected more strongly by this type of degradation, although no systematic measurements have been made so far. The uptake of water in MgO films can be in form of hydroxyl groups or adsorbed water which has been determined by infrared spectroscopy (6).

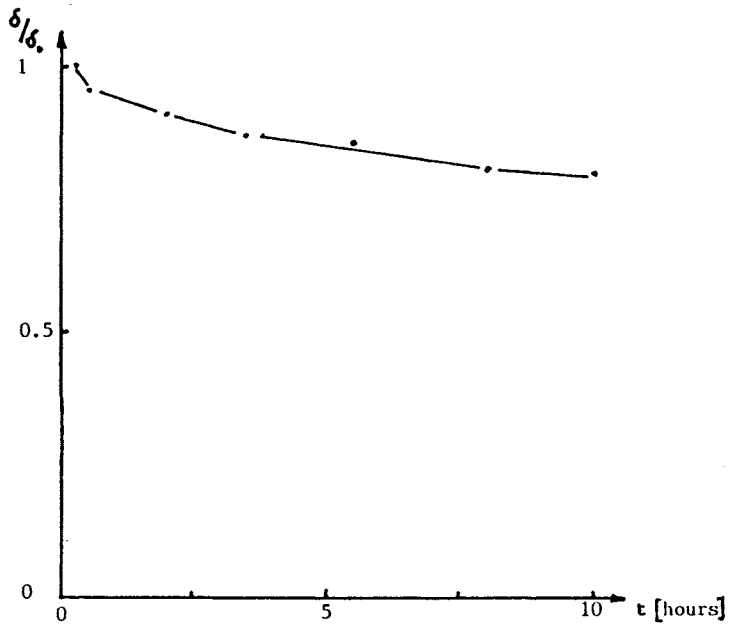


Fig. 9

Decay of SEE yield under air exposure

Noise considerations

The average Cs^+ ion current I_{Cs} coming from the hot-wire ionizer is assumed to be shot-noise limited. The rms noise current at the input of the EM is then given by

$$(\Delta I_{Cs})_{rms}=(2eI_{Cs}\Delta f)^{1/2}$$

The signal-to-noise ratio for a 1-Hz noise bandwidth is

$$(SNR)_i = \frac{\text{signal power}}{\text{noise power}} = \frac{I_{Cs}^2}{2eI_{Cs}} = \frac{I_{Cs}}{2e}$$

The processes of Cs^+-to-electron conversion and secondary electron emission degrade the SNR by some factor F , called the noise figure. At the the output of the EM we have therefore

$$(SNR)_o = \frac{I_{Cs}}{2eF} = \frac{I_o}{2eFG} = \frac{(I_o)^2}{(\Delta I_o)^2_{rms}}$$

where $I_o = GI_{Cs}$ and G the overall gain of the EM.
With a Cs^+-to-electron conversion efficiency γ and the same SEE yield δ for each of the succeeding n-1 dynodes the overall gain of an EM is given by

$$G = \gamma\delta^{n-1}$$

If one assumes that both Cs^+-conversion and SEE statistics have a Poisson probability distribution, the noise figure is given by

$$F = 1 + \frac{1}{\gamma}\frac{\delta}{\delta-1}$$

Since γ is always much smaller than δ it is evident that the noise figure is dominated by the value of γ . For example $\gamma = 1$ and $\delta = 5$ the noise figure is F = 2.25. On the other hand if $\gamma = \delta = 5$, which would be the case for electron input the noise figure is only F = 1.25
By measuring the input and the output currents I_{Cs} and I_o as well as the rms output current $(\Delta I_o)_{rms}$ one can always determine the gain G and the noise figure F. Using the theoretical expressions for G and F one can calculate δ and γ . Successive measurements before and after air exposure or certain operation times provide information on the

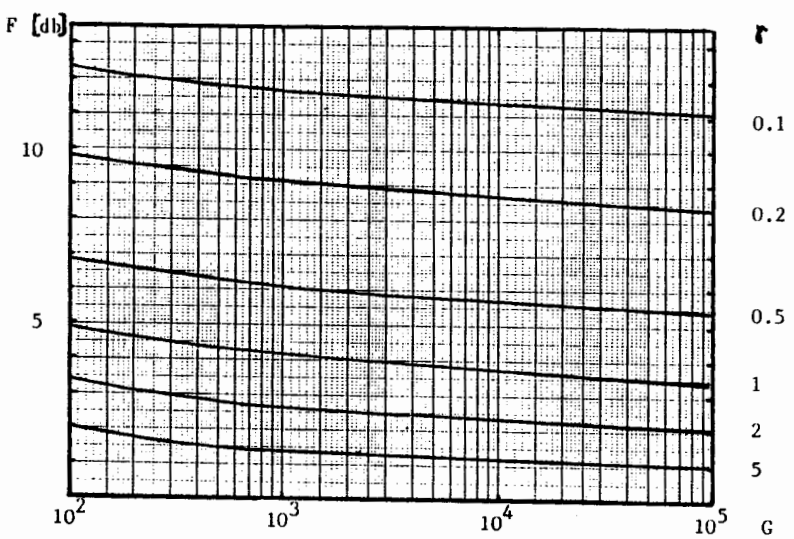


Fig. 10

Noise figure F of an electron multiplier with Poisson statistics.

$F=10\ \log(1+\frac{1}{\gamma}\frac{\delta}{\delta-1})$

$G=\gamma\delta^7$

evolution of γ and δ . Fig.10 displays the relationship of G and F with γ as a parameter for an EM with eight dynodes.

For quick non-invasive characterization of the noise performance of a Cs-tube it has been suggested (7) only to measure I_o and $(\Delta I_o)_{rms}$ which determine the product FG. A loss of tube current can be caused by a reduction of γ or δ which have a different impact on FG.

The electron multipliers assembled with sputtered MgO dynodes showed electronic gain statistics close to Poisson. Long-term studies of the noise behavior in the Cs tube are under way.

APPENDIX

Selective surface oxidation of Ag-Mg

When the surface of the alloy is subjected to an oxidizing gas, i.e. O_2, CO_2, H_2O, ..., the magnesium component will be oxidized according to

$$2\,Mg + O_2 \longrightarrow 2MgO$$
$$Mg + CO_2 \longrightarrow MgO + CO$$
$$Mg + H_2O \longrightarrow MgO + H_2$$

Silver is not oxidized below 1000°C. The goal is the production of a dense stoichiometric surface layer of MgO.

When Ag-Mg is baked in O_2, in general the O_2 diffuses into the alloy much faster than the Mg can diffuse to the surface. The O_2 on its way into the bulk Ag-Mg material reacts with any Mg atom that it contacts. The MgO formed is immobile and "frozen in place" . This prevents the formation of a surface MgO film. A careful adjustment of pressure and temperature causes the reaction to start at the metal/gas interface so that a protective MgO layer is formed at the surface. If this is the case, the oxidant gas is greatly inhibited to penetrate the MgO film and the reaction is further on governed by the diffusion of Mg through MgO. In the following, the film builds up at the outer MgO/gas interface. This is further faciliated by using CO_2 or H_2O as oxidants. These molecules are larger than O_2 and diffuse not so easily through MgO and Ag. On the other hand, if the gas pressure is too low or the temperature too high, more Mg atoms diffuse to the surface than can be oxidized. The rest will evaporate. The rate-determining factor is then the reaction of the oxidant molecules with magnesium.

Although the exact reaction cross sections of the above reactions are unknown, experiments indicate that the rate of arrival of oxidant mole-

cules at the surface should be at least 100x higher than the diffusion flux of Mg to the surface in order to prevent evaporation. Under such conditions, every Mg atom reaching the surface is readily oxidized. The rate of arrival of gas molecules R depends on the temperature T pressure P and the molecular weight M of the oxidant.

$$R \left[\frac{\text{molecules}}{\text{cm}^2\ \text{sec}}\right] = \frac{1}{4} n v_m = 3.513 \times 10^{22} \frac{P\,[\text{Torr}]}{\sqrt{MT[K]}}$$

In the first approximation, the Mg-concentration gradient across the MgO film during the reaction is constant. For the diffusion flux of Mg through MgO one can then write

$$\frac{dN}{dt} = - D \frac{n_{Mg}}{x}$$

N : number of Mg-atoms that reach the unit surface

n_{Mg} : concentration of Mg-atoms in the Ag-Mg alloy

x : thickness of MgO layer

$D = D_o \exp\ (-\frac{Q}{RT})$: diffusion constant

For the diffusion of Mg through MgO Hirashima (3) found

$$D = 1.87 \times 10^{-8} \exp\ (\frac{20.7\ \text{kcal/mole}}{RT\ [K]})$$

at $375^oC < T < 550^oC$

The layer thickness can be expressed as

$$x = \frac{N}{n_{MgO}} \qquad n_{MgO} = 5.4 \times 10^{22}\ \text{cm}^{-3}$$

(concentration of MgO-molecules in MgO film)

giving

$$\frac{dx}{dt} = D \frac{n_{Mg}}{n_{MgO}} \frac{1}{x} \qquad n_{Mg} = 4 \times 10^{21}\ \text{cm}^{-3}$$

(for Ag-Mg (1.7%))

Integration leads to a relation between layer thickness and oxidation time.

$$x^2 = 2 D \frac{n_{Mg}}{n_{MgO}} t + C \qquad t < t_0$$

This parabolic relation is characteristic for oxidation processes through a protective oxide layer. The constant C represents the layer which has been formed before the onset of protection at t_0. Fig. 11 displays results of this simplified calculation for an alloy containing 1.7% Mg. An average value $C = 3.3 \times 10^{-12}$ cm^2 has been used.

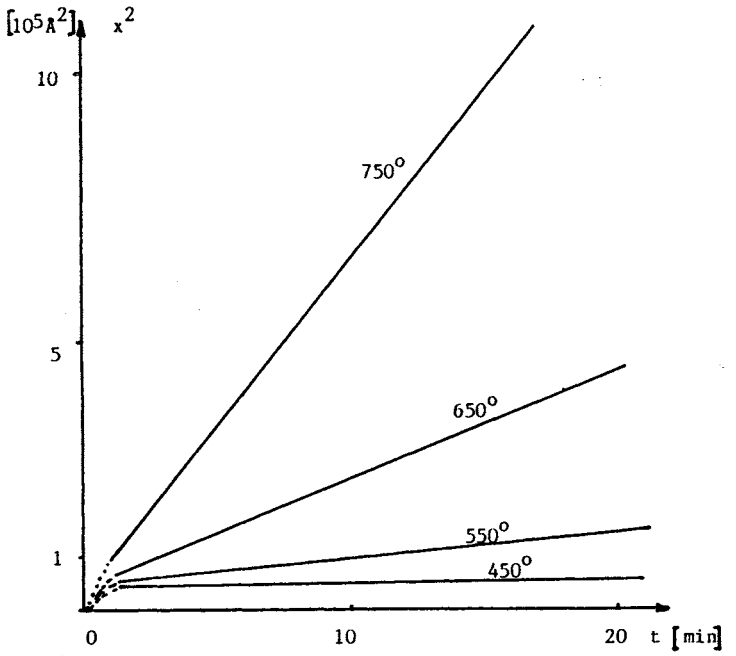


Fig. 11
MgO layer thickness as function of oxidation time and temperature

Usually not all magnesium in the Ag-Mg alloy is oxidized in the activation process. An additional heating in vacuum, i.e. exhaust baking of the EM or of the complete Cs tube, causes Mg diffusion to the surface. This leads to Mg enrichment in the MgO film and a subsequent decrease in δ. Evaporating Mg can also contaminate the ceramic parts of the EM and cause insulation problem under high voltage operation. To avoid this problem, Rappaport (8) suggested a two-step activation process. After surface oxidation at low pressure, a second step with O_2 at atmospheric pressure is added. At this high pressure O_2 diffuses into the bulk material and oxidises the residual Mg in place thus preventing future Mg diffusion.

REFERENCES

1) M.N. Saha, Phil. Mag. 46, 534 (1923)

2) B. Jaduszliwer, IEEE Trans. UFFC-34, 690 (1987)

3) M. Hirashima, J.Appl. Phys. 25, 288 (1954)

4) J. Dresner, J. Appl. Phys. 47, 1038 (1976)

5) J. Dresner and B. Goldstein, J. Appl. Phys. 47, 4760 (1977)

6) M.O. Aboelfotoh and O. Sahni, IEEE Trans. ED-28, 645 (1981)

7) J.P. Hurrel, W.A. Johnson, S.K. Karuza and F.J. Voit IEEE Trans. UFFC-34, 602 (1987)

2 ND EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 16 - 17 - 18 MARCH 1988

Session AFS-02 Number 4

FULL DIGITAL-PROCESSING IN A NEW COMMERCIAL CESIUM STANDARD

Jean RABIAN — *Sté SERCEL*
B.P. 64 - 44471 CARQUEFOU CEDEX - FRANCE

Pascal ROCHAT — *Sté OSCILLOQUARTZ S.A.*
Brévards 16.2002 - NEUCHATEL
SUISSE

1 - FOREWORD

Early in 1983, OSCILLOQUARTZ S.A., Neuchatel, Switzerland, and SERCEL, Nantes, France, decided a joint venture to design and build the first European commercial cesium standard.

This unit is called EUDICS (EUropean DIgital Cesium Standard).

OSCILLOQUARTZ S.A., a company of well-known experience in the Time & Frequency field, is in charge of designing and building a new cesium tube along with its power supply and a 10 MHz - to - 9.2 GHz frequency multiplier. It also has to provide the master crystal oscillator for the unit.

SERCEL, interested in the Time & Frequency field too, has lately come up with GPS satellite receivers. Its know-how results from 25 years in the design of high-accuracy radiopositioning systems and range-finding equipment. Time & Frequency techniques are directly involved in all those systems.

In this joint venture, SERCEL is in charge of the electronic circuitry and signal processing as well as of the software involved in that new standard. The electronic circuitry makes use of modern techniques in digital signal processing and frequency synthesis. This paper deals above all with frequency synthesis.

2 - BASIC PRINCIPLES

For an electronic engineer a cesium tube is equivalent to a four-pole circuit in which the frequency of the input signal varies around the atomic response, and that gives rise to a direct microcurrent, with variable intensity, at the output. It has the same function as a filter

with very narrow passband, associated with an amplitude detector. For a cesium tube, the transfer function of the four-pole equivalent circuit is known as Ramsey pattern. The frequency corresponding to centre resonance is called ν_0. Using a synthesizer with microprocessor-programmable frequency and fast response time, we can periodically scan various characteristic points in the Ramsey pattern.

1st principle (fig. 1) :

If we alternately program two frequencies symmetrically located on either side of ν_0 (centre peak), we get two current intensities at the tube's output and the difference between those two values stands for the frequency deviation of the crystal oscillator associated with the tube. Those signals, converted into digital data, make up the basic feedback information for controlling the frequency of that oscillator.

2nd principle (fig. 2) :

If the time constant of the feedback control loop is long enough with respect to the measurement cycle, we can safely "steal" one of the measurement cycles, periodically, in order to program other frequencies that will allow us to scan other characteristic points in the Ramsey pattern.

Thus, benefiting from the high flexibility of this special frequency synthesis technique combined with the various advantages of a microprocessor-controlled system, this new standard can permanently and in real time perform the following interesting functions (fig. 3) :

- Checking the centring of the Ramsey spectrum line, by symmetrically testing the Rabi.
- Measuring the maximum level of the Ramsey peak.
- Measuring the mean background level.
- Checking the Ramsey-to-Zeeman frequency deviation.

All those operations are performed by digitally processing the tube's output signals. Such permanent scrutiny will allow accurate correction of those parameters affected by the tube's ageing or outside environment. Corrections are brought about by 8 digital control loops, of which 6 are operating simultaneously without mutual interference.

3 - GENERAL THEORY OF OPERATION

Oscillator's main control loop :

During the startup sequence, only the Rabi centring measurement is carried out. For that purpose, two frequencies (ν_0 – 5 kHz and ν_0 + 5 kHz) are programmed. The oscillator's control loop is driven by that measurement until the frequency error is less than ± 500 Hz with respect to ν_0. Then the control loop goes to the Ramsey line tracking sequence. This is done by programming two frequencies (ν_0 – 193 Hz, ν_0 + 193 Hz) in order to measure the centring of the Ramsey peak. This measurement cycle is repeated 27 times a second, and each of the states lasts a half-cycle, that is 18 ms.

Signal integration begins 2 ms after switching the frequency, so as to avoid the effect of transient states. It proceeds over 16 ms in the analog form until the signal is digitized, then in the digital form until the reach of the chosen time constant.

The oscillator's control loop may operate according to two modes (fig. 4) :

- 1st order loop with a filtering constant adjustable from 0.1 to 10 s.
- 2nd order loop with a filtering constant adjustable from 0.5 to 100 s. This operating mode allows removal of the phase error that may arise when the oscillator undergoes a thermal shock or gravity change, especially with high filtering-constants.

Performance checks :

Operation of the standard is checked using the "cycle stealing" technique, which is performed during one cycle every nine cycles, that is, during 36 ms every 322 ms. During this test cycle, eleven different measurements are made at characteristic points in the Ramsey pattern.

Possibility of implementing a loop for controlling the RAMSEY-ZEEMAN frequency deviation. :

This loop operates in the same way as the main control loop :

- Pre-centring on the Rabi of the upper Zeeman spectrum line.
- Tracking of the Zeeman fine spectrum line every 645 ms. The 30-s time constant, first order loop corrects the intensity of the C field-generating current.

This user-selectable function is of a remarkable efficiency against most external disturbances, which are chiefly :

- Earth gravity field
- Temperature effects
- Tube magnetization
- C-field current instability.

Checking tube transfer function :

This check, which is performed every 3,225 ms, provides the following information :

- Ramsey peak level measured with respect to background. This level is regulated through an automatic correction of the input gain applied to all the tube output signals, which affords stabilization of both loop gains and alarm levels despite transfer loss due to tube ageing. This digital Automatic Gain Control also allows a very fast starting of the source : once the OCXO warm-up time has been allowed for, the frequency stability is 2×10^{-12} after 3 min at ambient temperature, the recovery delay is less than 1 min after a one-hour power outage.

- Evaluation of the tube S/N ratio, standard deviation calculated over 30 s.

- Zeeman peak level.

-Symmetry of the Ramsey peak foot within ± 386 Hz.

The other loops deal with :

- Offset voltage at preamplifier input.
- 12.6 - MHz signal level needed to maintain modulation depth at J1 max.
- Microwave power needed for a maximum signal at tube output.

Alarm levels :

In addition to the above functions, the microsystem measures fifteen service voltages and the temperature inside the case. It also ensures continuous clock monitoring by calculating twenty alarm levels among which five are safety levels intended to turn off the tube power supply in case of major fault.

F (t°) correction pattern (Fig. 5) :

The temperature data can be used to perfect the F (t°) frequency correction from a pattern kept in memory.

Fine adjustment of the frequency :

To allow a fine adjustment of the frequency while maintaining the advantages of the C-field control loop, an additional system, acting directly at the synthesizer stage, has been developed to adjust the frequency without affecting the C-field. Its resolution is 1×10^{-13} within a range of $\pm\ 1 \times 10^{-10}$. This function also ensures phase setting, that is time setting. The resolution is 1 ns within a range of $\pm$ 10µs. It may be operated under the control of the RS232 I/O line.

4 - INPUT /OUTPUT MESSAGES

System remote control can be implemented through an RS232C serial line, using one of the following two modes :

- Console mode with system / operator dialogue in English or French language.

- Remote control mode using structured messages and reply signals. This mode makes it possible to establish long-distance links via modem and to simultaneously monitor several standards connected to a microcomputer through a single line.

5 - MANUFACTURE AND MAINTENANCE (FIG. 6, 7)

The functions described above reduce the drifts due to both tube ageing and environment to a considerable extent. They also ensure a permanent monitoring of the standard. Being of an evident interest in R & D laboratories, they afford an even greater interest for applications on board mobiles.

For this reason, an important industrial effort was consented to introduce new manufacturing techniques and improved maintenance procedures : continuous and numerous checkings, quick access and easy dismounting of the subassemblies. The Cesium Standard is fitted in a solid and compact cabinet in aluminium casting. A sealed cabinet option is also proposed. The cabinet dimensions are : 188 x 117 x 363 mm. The ready-to-operate cabinet weighs 22 lbs (10 Kg) The "floating" DC power supply can range from 18 V to 64 V DC for any embarked applications or telecommunications.

Maintenance is facilitated thanks to the microsystem-aided diagnostic. Each subassembly can be replaced in the field in a few minutes For this purpose, the tube is delivered with a pluggable module in charge of setting parameters automatically.

6 - CONCLUSION

For the first time, modern signal processing methods are fully used in a Cesium Standard. EUDICS is furthermore the first completely European-designed frequency standard intended for industrial manufacturing.

Our experimenting with the numerous capabilities of the Cesium Standard is still too recent for us to appraise all the advantages it can afford the scientists and manufacturers. Yet, we are convinced that the possibilities of that Standard are unsuspectedly important and we count upon the users' cooperation and opinion to specify them.

7 - ACKNOWLEDGEMENTS

We would like to thank Mr. GRAF and Mr. BUSCA from OSCILLOQUARTZ and Mr. NARD from SERCEL for their contribution during the design of EUDICS.

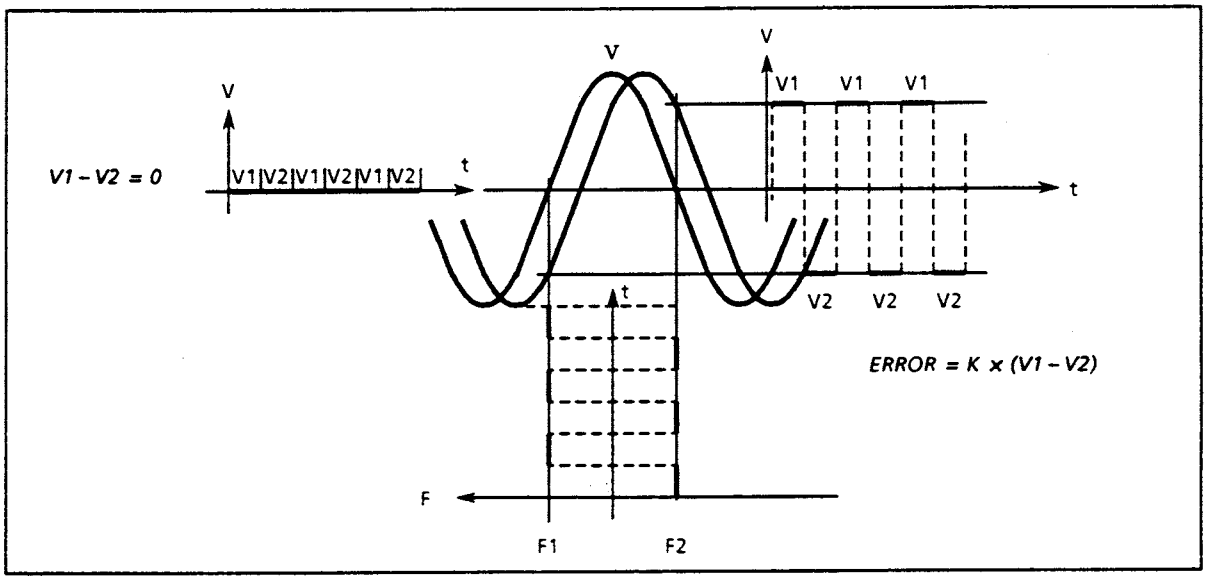


FIGURE 1
OUTPUT SIGNAL VERSUS FREQUENCY STEPS

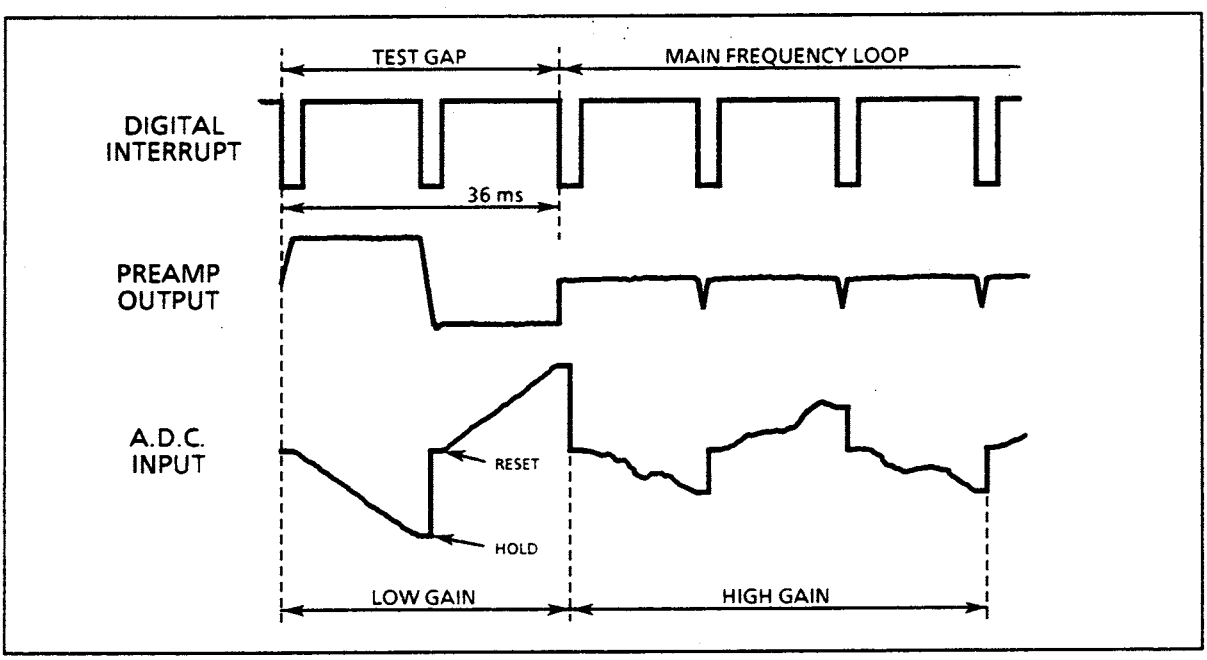


FIGURE 2
ANALOG PROCESSING SEQUENCE

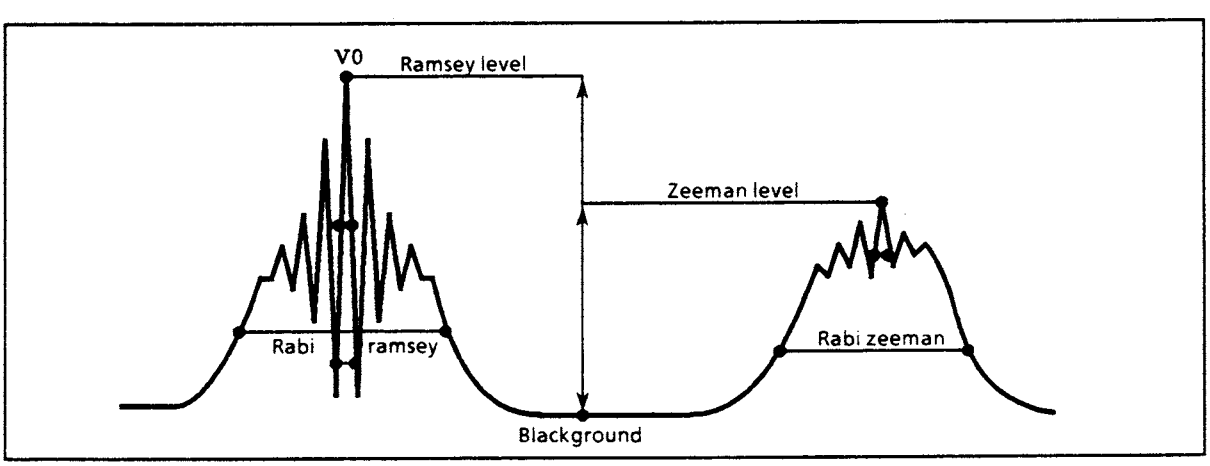


FIGURE 3
13 FREQUENCY STEPS

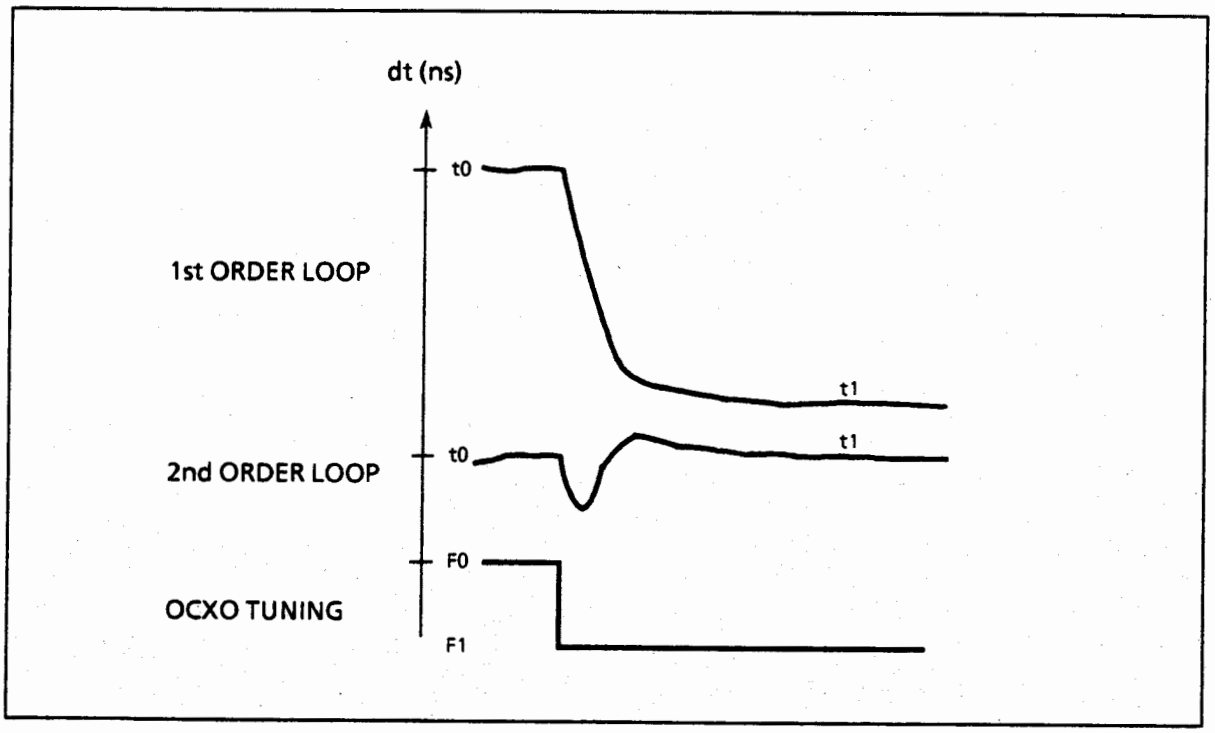


FIGURE 4
TIME RESPONSE FOR A FREQUENCY STEP ON O.C.X.O.

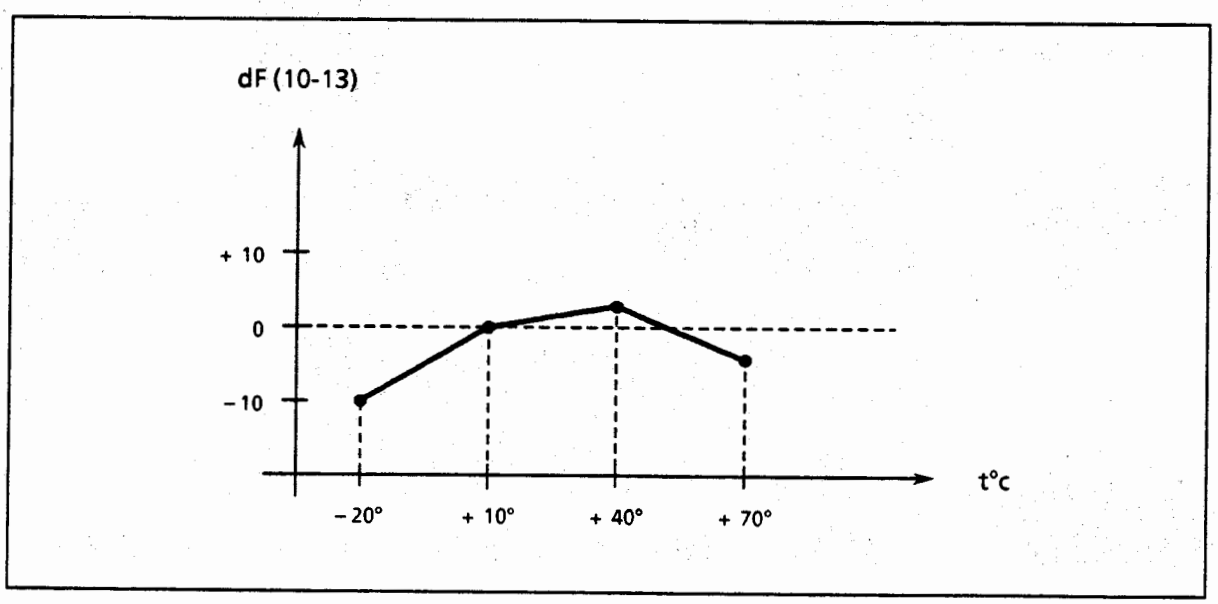
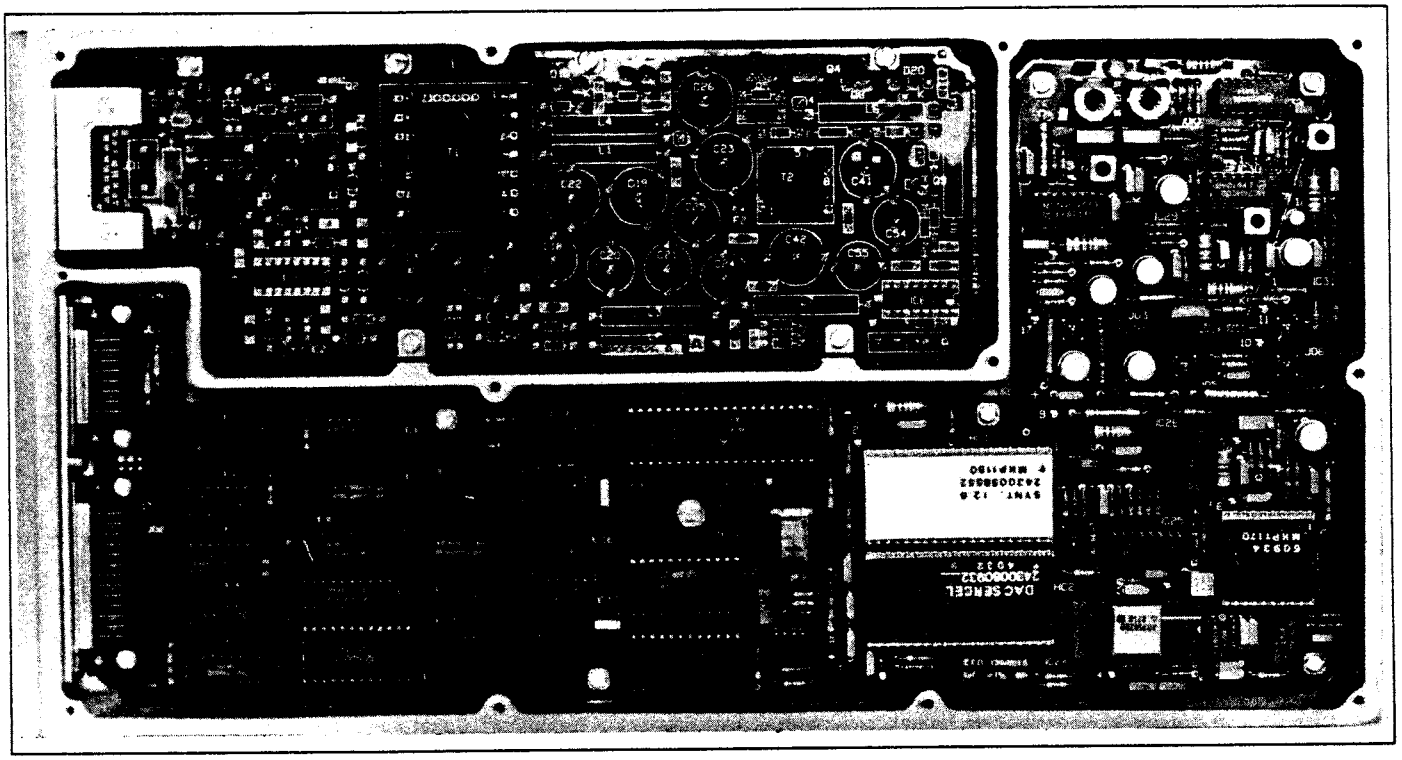
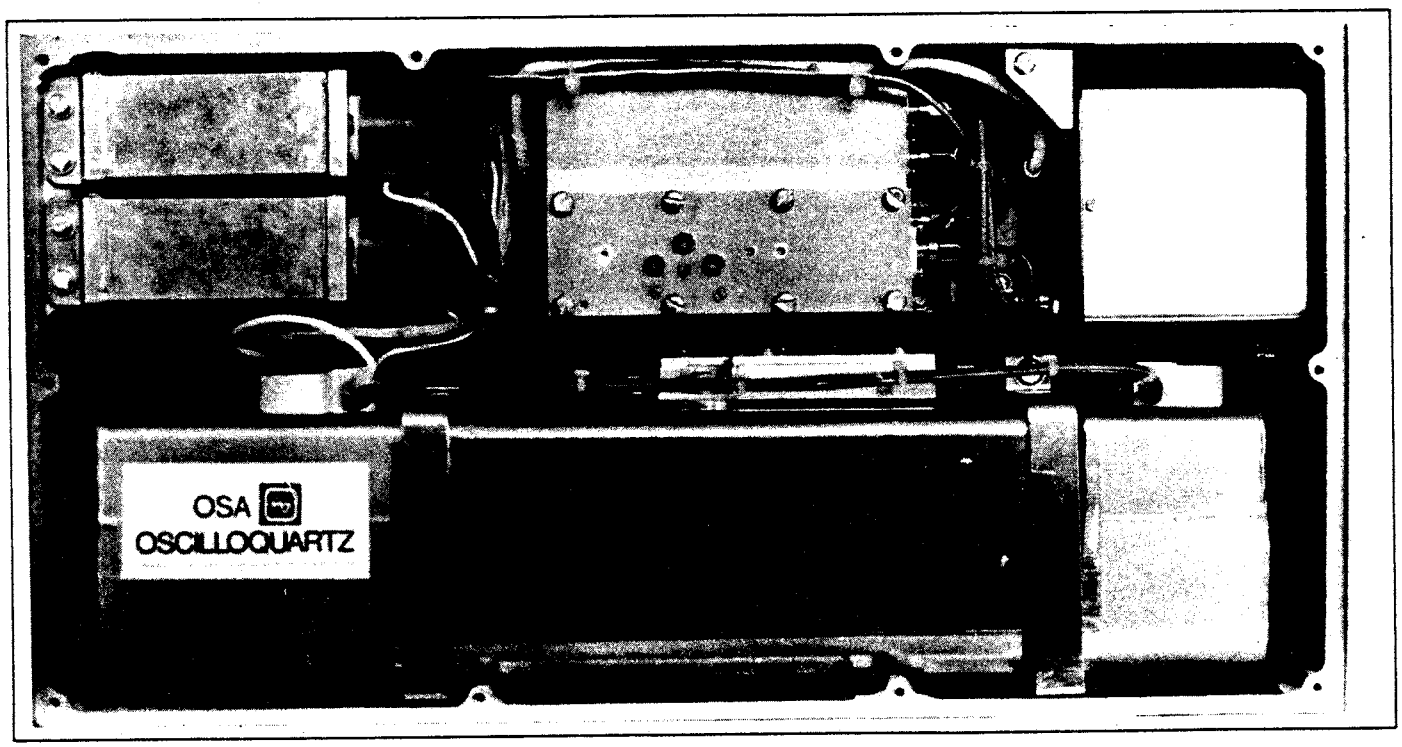


FIGURE 5
EXAMPLE OF F (t°c) CORRECTION MODEL




OSA
OSCILLOQUARTZ

2 ND EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 16 - 17 - 18 MARCH 1988

Session AFS-02 Number 5

IMPROVING THE LONG-TERM STABILITY OF A TIME SCALE GENERATED BY COMMERCIAL CESIUM CLOCKS

F. Cordara, V. Pettiti, P. Tavella(*)

Istituto Elettrotecnico Nazionale Galileo Ferraris
C.so Massimo d'Azeglio, 42 - 10125 Torino, Italy
(*) Associazione Nazionale Galileo Ferraris Grant

Abstract

The seasonal frequency variations of the cesium clocks maintained at the Istituto Elettrotecnico Nazionale of Torino (Italy) have been analyzed over a period of two years by means of statistical tools. Some simulations show how the long-term stability of a time scale generated by commercial cesium clocks can be improved compensating the deterministic frequency variations either using a phase microstepper or a statistical process.

1 - Introduction

The influence of environmental conditions on commercial cesium clock behaviour has been analyzed in the literature where the sensitivity mainly to temperature and humidity variations was pointed out [1,2,3,4].

Nevertheless, the results obtained through careful investigations on a large number of clocks [5], have demonstrated that the sensitivity to the variations introduced in the environmental conditions is different for each clock and therefore each device must be investigated separately to estimate its long-term behaviour

This evaluation is essential in maintaining the local UTC time scale of a laboratory, generated either by a single cesium clock with some kind of rate correction or by a clock ensemble through a suitable algorithm, in close agreement with UTC.

2 - Long-term behaviour of the UTC(IEN) time scale

The time scale of IEN, relying at present on five cesium clocks, has been generated since 1973 by a single clock whose rate is periodically adjusted by means of a phase microstepper to maintain UTC(IEN) in close agreement with UTC.

In Fig. 1 is depicted the equipment set-up for the IEN time scale realization [6]; it can be noticed that a GPS receiver is used to relate the IEN clocks to the international atomic time scale.

In Fig. 2 have been reported three year values of UTC-UTC(IEN) supplied by BIPM starting from 1985. Looking at the shape of the graph, it can be seen that a seasonal effect is present and that the rate compensation performed by the phase microstepper has not eliminated this effect.

This consideration showed the importance of a careful characterization of the behaviour of each IEN clock to find out whether the seasonal effects were detectable in order to optimize the estimate of the rate of the clocks for the improvement of the long-term stability of the time scale.

A first attempt in analyzing the long-term behaviour of some IEN cesium clocks, showed the presence of deterministic effects [7]; however it turned out necessary to improve the investigation tools by adopting suitable mathematical models and statistical treatments. The results, recently obtained on this subject and reported in a previous paper [8], have suggested further studies to establish a better use of the phase microstepper corrections.

3 - Analysis of the clocks data

As reported in [8], the behaviour versus UTC of three clocks, Cs5, Cs6, Cs9 (see Table 1) was observed over a period of two years (mid 1985 - mid 1987), and the deterministic fluctuations of their rates were determined with two different approaches that gave results in close agreement.

Table 1

Cesium clock	Model	Serial number	Observation period (days)	Observation period from - to (MJD)
Cs5	HP5061A	609	660	46229 - 46889
Cs6	HP5061A/004	893	730	46229 - 46959
Cs9	OSA3200	84	830	46229 - 47059

Ten-day averaged time differences between UTC and each IEN clock and, in some cases, the mean normalized frequency departures were used for the computations.

Supposing to approximate the normalized frequency departure of a cesium clock versus a reference frequency with the function:

$$h(x) = B + C\,x + D\sin\omega(x + E)$$

expressed in units of 10^{-13}, where x and E are expressed in days and ω is equal to $2\pi/365$ (rad/d), the following results were obtained.

Table 2

Clock	Parameters B $[10^{-13}]$	C $[10^{-13}/yr]$	D $[10^{-13}]$	E [d]
Cs5	8.0	-0.2	1.2	0
Cs6	2.5	-0.4	1.1	0
Cs9	15.2	0.6	-2.4	40

The coefficients B,C,D,E represent respectively:

B : the initial mean frequency departure;

C : the frequency drift:

D : the amplitude of the sinusoidal term;

E : the initial phase (divided by ω) at x = 0.

Fig. 3 reports for each clock the normalized frequency departure versus UTC with a constant value subtracted and the frequency drift removed. The dashed lines represent the results obtained fitting the experimental data by a least squares approximation. The lower side shows the temperature and the absolute humidity variations of the IEN clocks room with the fitting curves obtained as before.

The next step consisted in subtracting from the original clock data the deterministic effects evaluated as before and analysing the residuals by means of the Allan variance to determine the remaining noise.

Fig. 4 reports the results obtained calculating the Allan variance for τ = 10,20,40,80,160 days, using the original clock data. In Fig. 5 the same computations, repeated with the deterministic noises removed, are shown. The predominant noise of the residuals of each clock appears to be white frequency.

4 - Simulation of rate corrections to compensate the deterministic effects

Using the same set of data of Cs6, the IEN master clock in the period considered, an optimization of the use of a phase microstepper, either hardware or software, was attempted.

The first year (from mid 1985 to mid 1986) was examined, the parameters of the fitting curve were evaluated and then a continuous phase correction for the next six months was applied accounting for the behaviour of the past period.

After six months, the entire period of real data was observed again, the parameters B, C, D were evaluated and a new continuous phase correction for the next six months was applied.

The time differences between UTC and the IEN corrected clock, obtained extending the above process up to the end of 1987, are shown in Fig. 6. The solid line represents the real data corrected for the mean rate, and the dashed line the rate that can be obtained by the compensation method. For the first period examined (MJD 46229 to MJD 46599), no correction was applied.

The simulation was repeated considering periods of three months for the evaluations and corrections.

It is clear from the results reported in Fig. 7(a) that the second attempt, with an evaluation interval of three months, is more satisfactory than the first one and that this process reduces of one third the total time error of the clock considered.

Obviously this process cannot completely compensate the seasonal phenomenon because the amount of the effects observed can change every year, depending on several factors that cannot be predicted exactly (environmental conditions, aging effects on the clocks, etc.).

A final attempt to find a better prediction of the clock rate correction was made analyzing the mean frequency departure of the clock and the amplitude of the seasonal term only over the last twelve months, while for the frequency drift evaluation, the data of the whole period were used.

This procedure was repeated every three months and the correction found was used for the following period.

As can be seen in Fig. 7(b), the behaviour of the time scale obtained by this last simulation shows a small improvement that is probably due to a quicker response of the fitting method adopted to the variations of the influence factors.

5 - Conclusions

The presence of seasonal effects on commercial cesium clocks has been analyzed and the relative amount has been reported.

Some simulations have been performed to correct the deterministic variations of the rate of a cesium clock, showing that it is possible to maintain a time scale in agreement with a time reference within 2 microseconds or better and to improve its long-term stability.

References

[1] S. Iijima, K. Fujiwara, H. Kobayashi, T. Kato: Effect of environmental conditions on the rate of a cesium clock - Annals of the Tokio Astronomical Observatory, Second Series, vol. XVIII, n. 1, p. 50-67.

[2] G. Becker: Zeitskalenprobleme: Jahreszeitliche Gangschwankungen von Atomuhren, PTB Mitteilungen 92 2/82, p. 105-113.

[3] K. Dorenwendt: Das verhalten Kommerzielle cäsium-atomuhren und die Zeitskalen der PTB - PTB Mitteilungen 94, p. 35-43.

[4] A. De Marchi: Understanding environmental sensitivity and ageing of cesium beam frequency standards - Proceedings 1st European Frequency and Time Forum, Besançon (France), March 1987.

[5] J. Kusters: New 5061A Specifications: Temperature Performance - Hewlett-Packard - The Standard, October 1983.

[6] F. Cordara, V. Pettiti: Time keeping and time and frequency dissemination at the Istituto Elettrotecnico Nazionale (IEN) - 18th Precise Time and Time Interval Meeting, Washington D.C., Dec. 1986.

[7] F. Cordara, V. Pettiti, G. Tignanelli: Environmental effects on cesium clocks kept at IEN - Torino and ISPT - Roma Time and Frequency laboratories - Proceedings 1st European Frequency and Time Forum, Besançon (France), March 1987.

[8] E. Bava, F. Cordara, V. Pettiti, P. Tavella: Analysis of the seasonal effects on cesium clocks to improve the long term stability of a time scale - 19th Precise Time and Time Interval Meeting, Los Angeles, Dec. 1987.

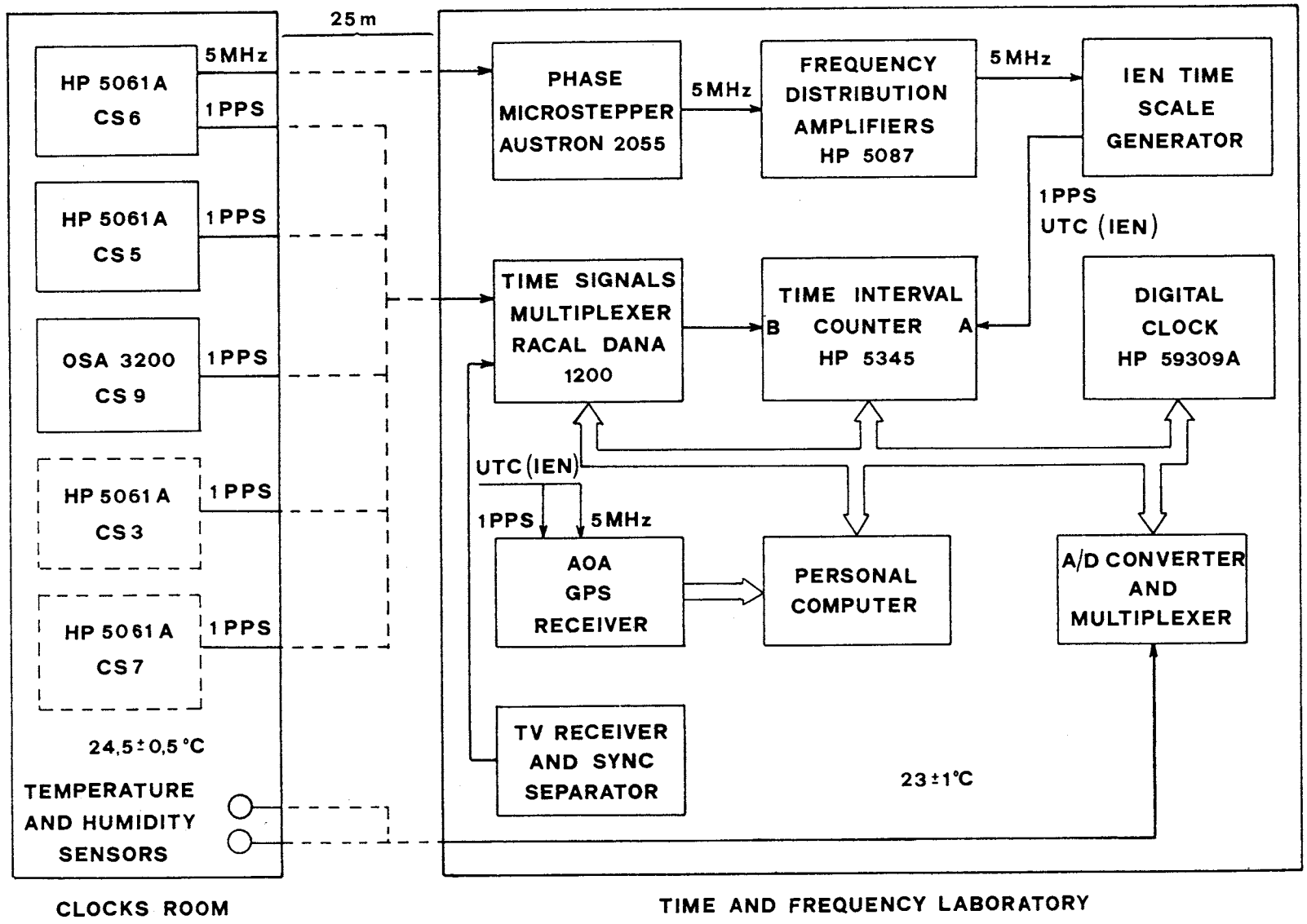


Fig. 1 - Equipment used to realize UTC(IEN) and to perform clocks intercomparisons.

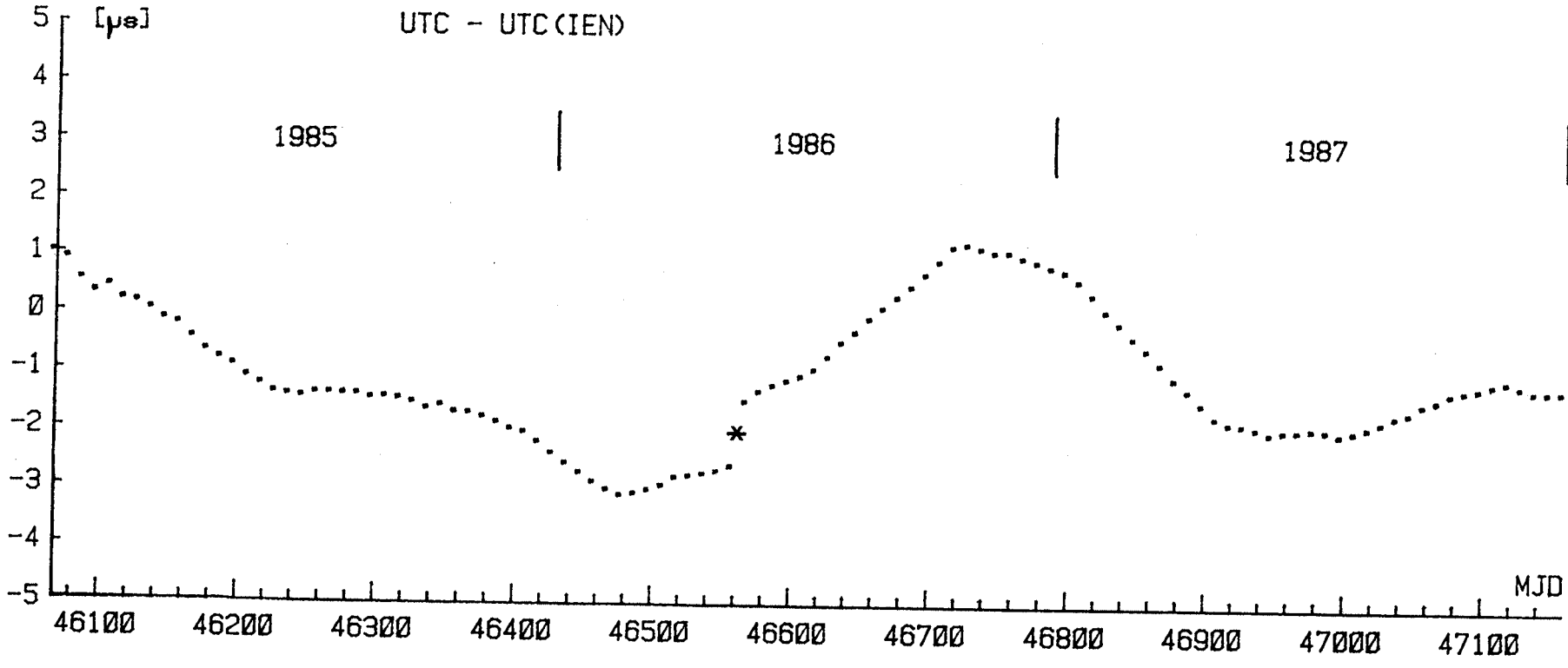


Fig. 2 - UTC(IEN) vs.UTC from BIH and Circular D. (* time scale correction)

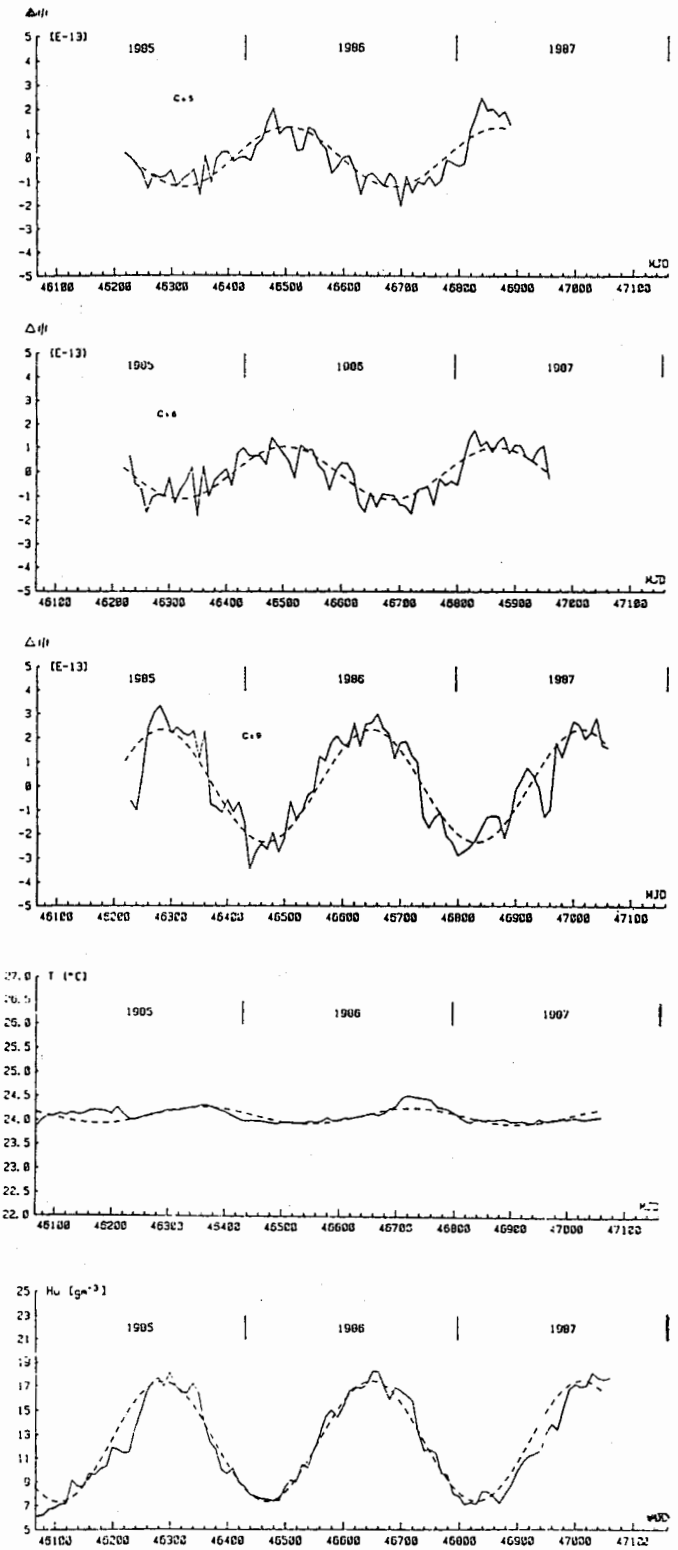


Fig. 3 - Normalized frequency departures of IEN cesium clocks vs. UTC, temperature (T) and absolute humidity (Hu) of the clocks room.

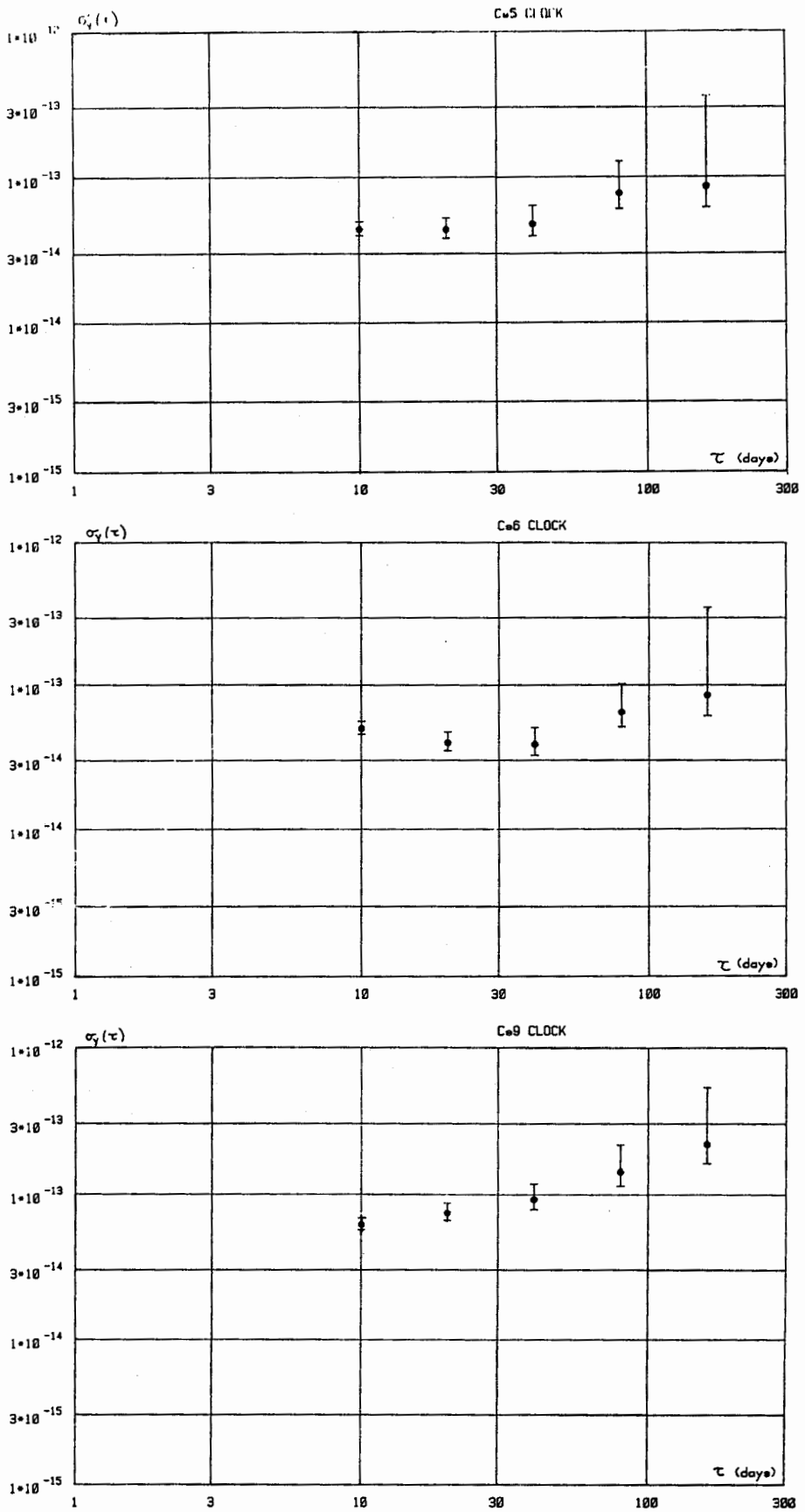


Fig. 4 - Allan variance of IEN cesium clocks vs. UTC.

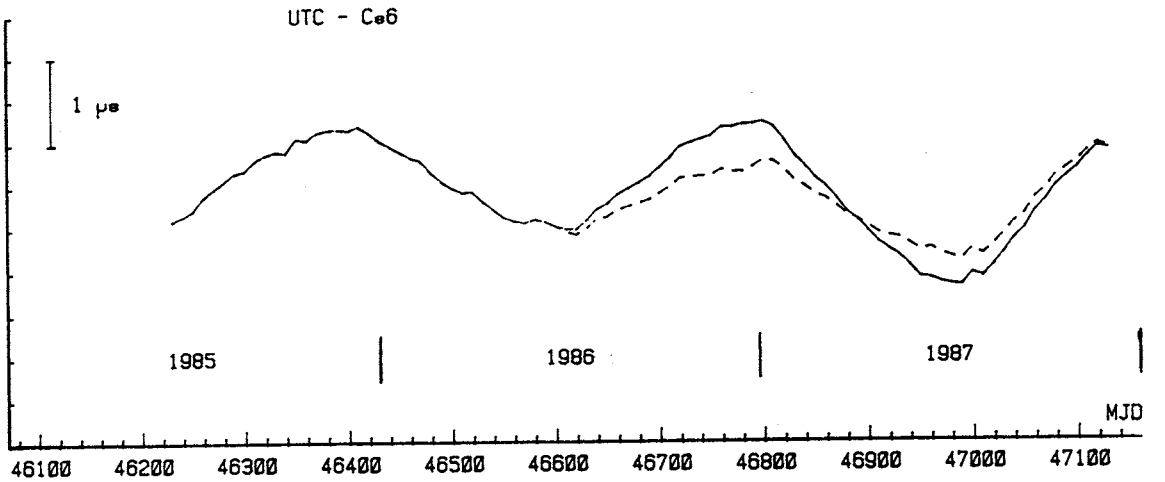
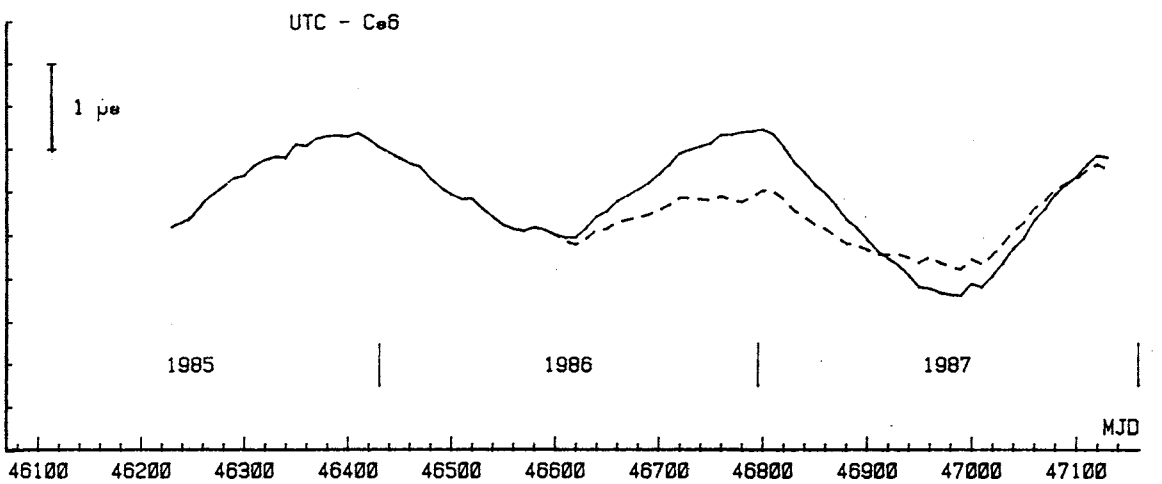


Fig. 6 - IEN Cs6 vs. UTC: original data (solid line) and simulated data (dashed line) with evaluation period of 6 months.

a)



b)

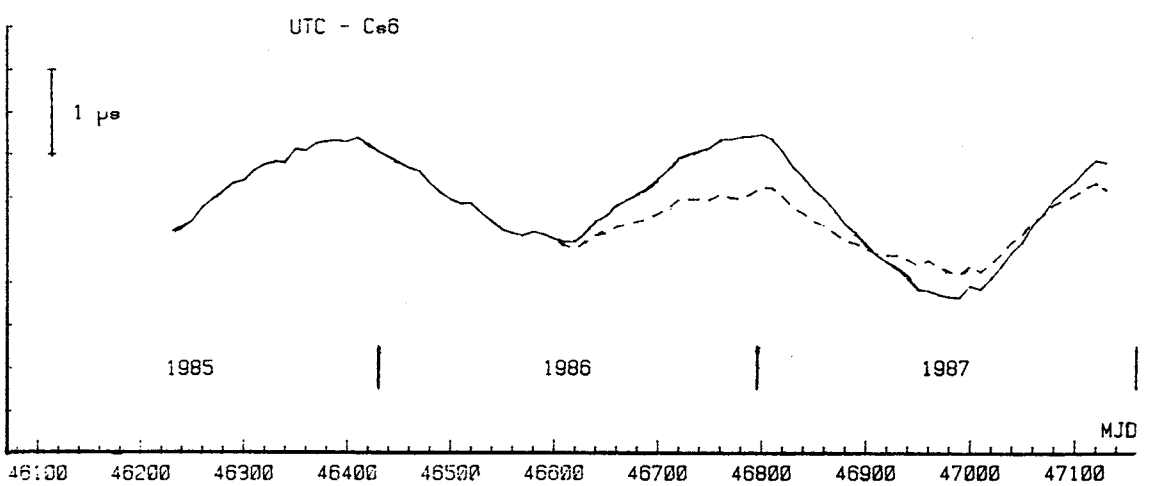


Fig. 7 - IEN Cs6 vs. UTC: original data (solid line) and simulated data (dashed line) with evaluation period of 3 months.

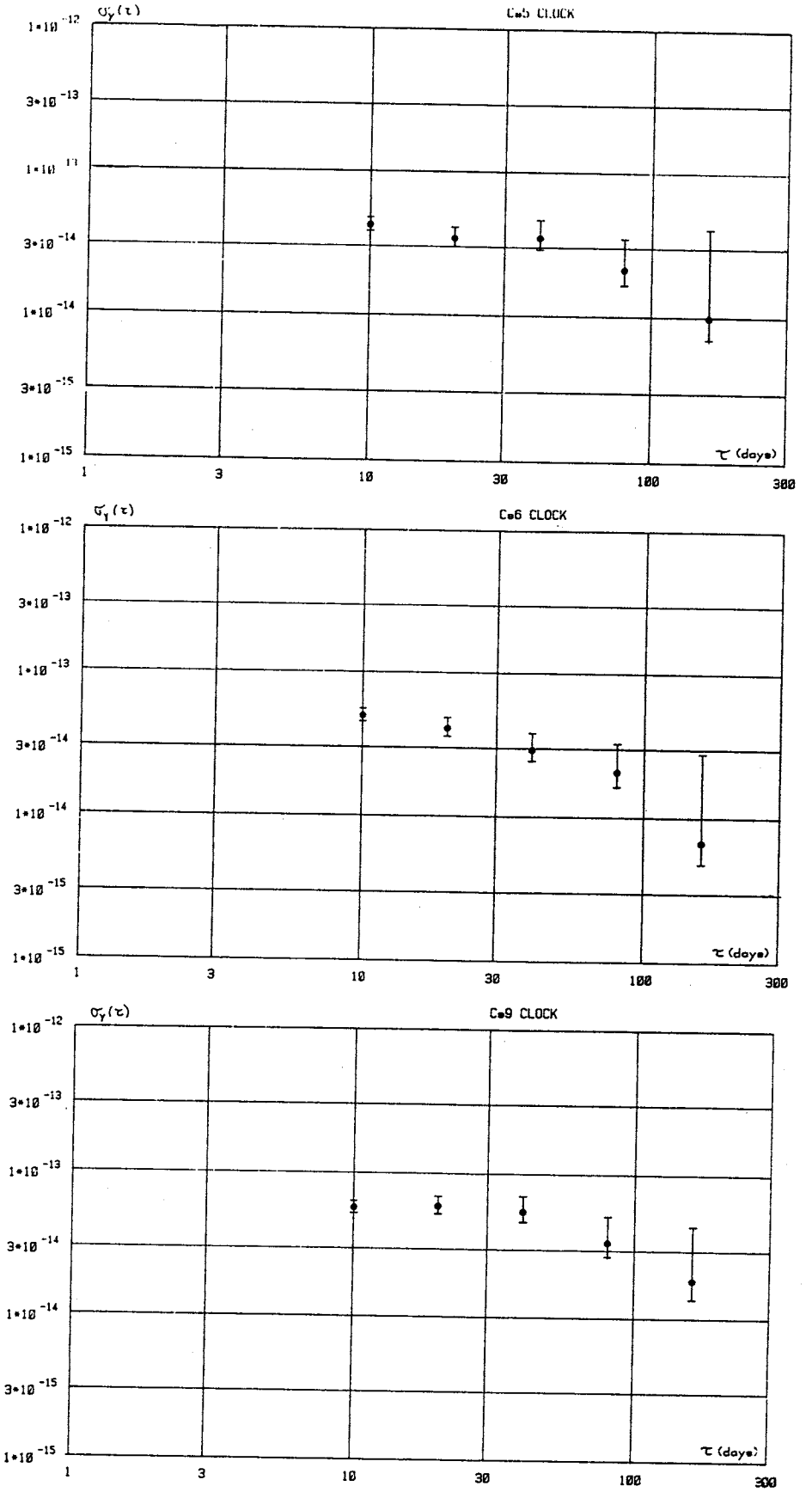


Fig. 5 - Allan variance of the residuals of IEN cesium clocks vs.UTC.

AFS-03

Optically Pumped CS Beam Frequency Standards

1. Frequency Stability of an Optically Pumped Cesium Beam Frequency Standard : Preliminary Results
   V. Candelier, V. Giordano, A. Hamel, G. Theobald, P. Cerez, C. Audoin
   Laboratoire de l'Horloge Atomique, Orsay - France

2. Light Pressure effects on an Optically Pumped Cesium Beam Frequency Standard
   A. Clairon, E. de Clercq, B. Dahmani, A.H. Gérard
   Laboratoire Primaire du Temps et des Frequences, Paris - France

3. Commercial Size Optically Pumped Cesium Beam Resonator
   P. Thomann, H. Schweda, G. Busca
   Oscilloquartz S.A., Neuchâtel

4. Experiments on optically pumped cesium beam frequency standard*
   J. Umezu, B. Komiyama, H. Saitho, Y. Ohta
   Radio Research Laboratory Koganei-Shi, Tokyo - Japan

5. A FM-stabilized semiconductor laser for optical pumping of Cesium atoms and length measurements
   H.R. Telle
   Physikalisch-Technische Bundesanstalt, Braunschweig - FRG

6. The Effect of Laser Line-Norrowing on the Performance of Optically Pumped Cesium Atomic Beam Frequency Standards
   S. Ohshima, Y. Koga, Y. Nakadan
   National Research Laboratory of Metrology, Ibaraki-ken - Japan
   L. Hollberg, R. Drullinger
   National Bureau of Standards, Boulder - USA

Session AFS-03 Number 1

FREQUENCY STABILITY OF AN OPTICALLY PUMPED CESIUM BEAM FREQUENCY STANDARD : PRELIMINARY RESULTS.

V. CANDELIER[1], V. GIORDANO, A. HAMEL, G. THEOBALD,
P. CEREZ and C. AUDOIN

Laboratoire de l'Horloge Atomique
Equipe de Recherche du CNRS associée à l'Université Paris-Sud
Bât. 221 - Université Paris-Sud

91405 ORSAY - France

SUMMARY

The study of the frequency stability capability of an experimental optically pumped cesium beam frequency standard has been pursued. An electronic system ensuring the frequency locking of a 10 MHz quartz crystal oscillator to the microwave resonance at 9.2 GHz has been realized. The generation of the atomic transition interrogation signal includes innovative features. They are aimed at obtaining performances consistent with the intrinsic frequency stability capability of an optically pumped cesium beam resonator. This electronic system is described and the technical choices made are discussed.

The 10 MHz quartz crystal oscillator has been frequency locked to the cesium hyperfine transition with a 0.2 s time constant. Its frequency stability has been measured with respect to a hydrogen maser. The result is $\sigma_y(\tau) = 2 \times 10^{-12}/\sqrt{\tau}$ for $1\ s < \tau < 500\ s$. This result confirms the potential interest of optically pumped cesium beam frequency standards.

Prospects for the improvement of frequency stability are discussed.

I. INTRODUCTION

Studies of the frequency stability and accuracy capabilities of the optically pumped cesium beam frequency standard has been started in a number of laboratories and firms [1-8].

The work that we are reporting is the continuation of that presented last year [9], in which the factor of merit of an optically pumped cesium beam tube was measured, in a configuration where the same laser is used for the optical pumping and for the optical detection of the atoms. We have built an electronic system which enables to frequency lock a quartz crystal oscillator to the atomic transition. The short term frequency stability measured is in agreement with that estimated from the value of the figure of merit, M = 29, which was given.

---

[1] Permanent address : CEPE, 44 Avenue de la Glacière
95100 Argenteuil - France

2. FREQUENCY STABILITY CAPABILITY OF OPTICALLY PUMPED CESIUM BEAM FREQUENCY STANDARDS.

Application of optical methods [10][11] to the state preparation and to the state detection of atoms in a cesium beam tube provides an opportunity of improving both the short term and the long term frequency stabilities of this frequency standard.

2.1. Short term frequency stability

The impact of optical methods on the short term frequency stability is essentially related to the increase of the flux of detected atoms. This proceeds from the removal of the magnet state selectors.

i) These state selectors deflect the atom trajectories with an angle which depends greatly on the atom velocity and, consequently, they attenuate the flux of atoms striking the hot wire detector by a factor of approximately one hundred.

ii) The need of using a very narrow detector for avoiding the detection of atoms in an undesired state disappears. It follows that the cross section of the detection zone can be increased noticeably when optical detection is implemented.

iii) Considering atoms in the hyperfine sub-levels F = 3, $m_F = 0$ and F = 4, $m_F = 0$, magnetic state selection eliminates from the beam atoms in one of these states. On the contrary, as saturated optical pumping with a single laser transfers all atoms of one of the hyperfine levels F = 3 or F = 4 to the other one, optical state preparation provides a gain of a factor of about two in the number of detected atoms. In a more involved configuration, two-laser optical pumping is able to provide a gain almost equal to 16 [12].

However, optical detection may bring drawbacks, compared to surface ionization detection. Detrimental effects are due to the additional shot noise of the stray light and, mainly, to the intrinsic noise of the optical detector. In the present state of the art, the beam shot noise would dominate the detection additional noise for a flux of detected atoms (i.e. atoms in one of the two states F = 3, $m_F = 0$ or F = 4, $m_F = 0$) larger than approximately 10^{10} s^{-1} [13]. This condition is not fulfilled presently and the frequency stability is limited by the optical detection noise. In that case, the measurement of the factor of merit of an experimental set-up gave an estimated short therm frequency stability* of $2.6 \times 10^{-12}/\sqrt{\tau}$ [9] [14]

* In reference (9), equation (5) from which the estimated frequency stability is derived should be : $\sigma_y(\tau) = (\sqrt{2}\ M\ \nu_{00})^{-1}\ \tau^{-1/2}$

which is one order of magnitude better than the short term frequency stability of standard performance cesium beam frequency standards.

Fig. 1 represents schematically the experimental set-up used in the present work. Fig. 2 shows the Ramsey pattern obtained. The signal to noise ratio defined as the peak to valley current difference divided by the standard deviation of the peak current is equal to 10,000. The 500 Hz wide central line of the resonance pattern is used to frequency control a quartz crystal oscillator.

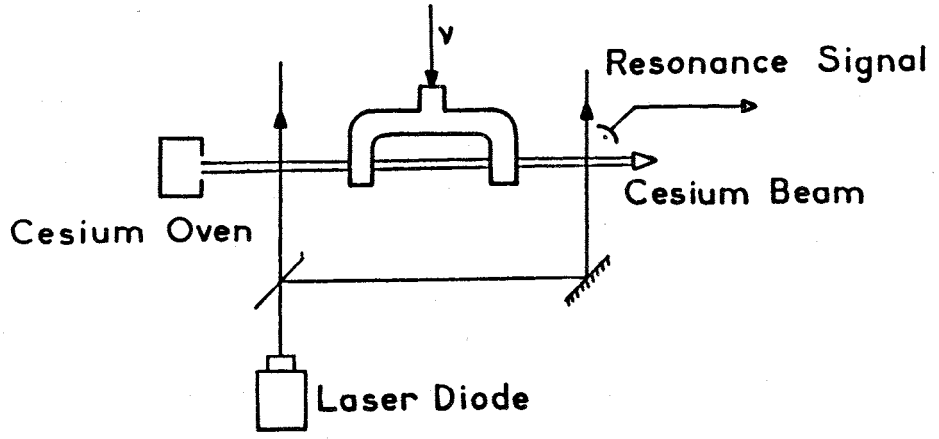


Fig. 1. Schematic representation of the optically pumped cesium beam tube. The separation between the two oscillatory fields is equal to 21 cm. The laser diode is tuned to the F = 3, F' = 3 component of the cesium D_2 line at 0.85 μm. The light is σ polarized. The laser frequency is frequency locked using the fluorescence light in the optical pumping region.

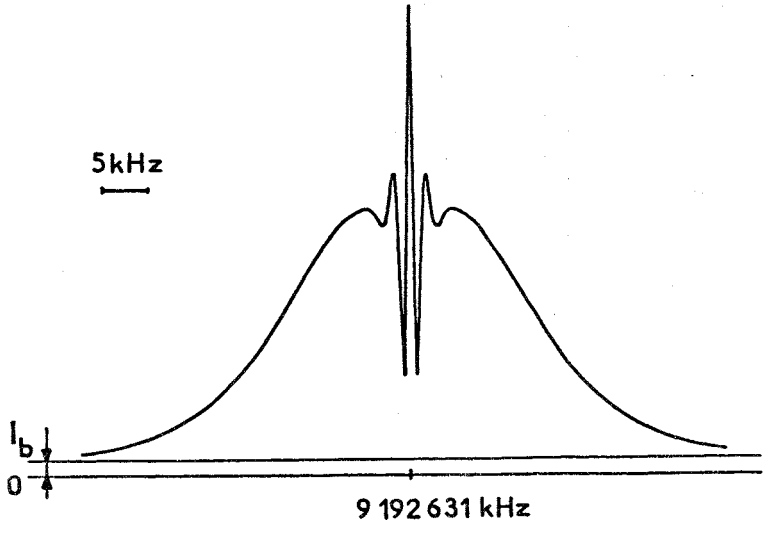


Fig. 2. Ramsey pattern to which a quartz crystal oscillator has been frequency locked. The linewidth of the central fringe is 500 Hz. The 3 dB bandwidth of the x-y plotter is 1 Hz. I_b represents the background of stray light.

The main purpose of this report is to confirm the preceding results by the measurement of the frequency stability of the quartz crystal oscillator frequency-locked to the optically pumped cesium beam tube.

2.2. Long term frequency stability

There are good reasons to believe that optical methods are able to improve the long term frequency stability as well as the accuracy of cesium beam frequency standards.

i) Optical pumping with linearly polarized light populates equally the hyperfine levels having opposite values of the magnetic quantum number m_F. It follows that the amplitude of the seven $\Delta F = \pm 1, \Delta m_F = 0$ lines of the microwave spectrum of the cesium atom varies symmetrically around the central $F = 4, m_F = 0 \longleftrightarrow F = 3, m_F = 0$ line as shown in Fig. 3. In that case, the Rabi pulling frequency shift [15] [16] should disappear along with its associated inaccuracy and with the frequency change related to a variation of the value of the static magnetic induction.

ii) Another benefit of the absence of Rabi pulling relies on the possibility of decreasing the value of the magnetic induction applied to the atoms without introducing a harmful effect. This would reduce still more the sensitivity to ambiant magnetic field changes.

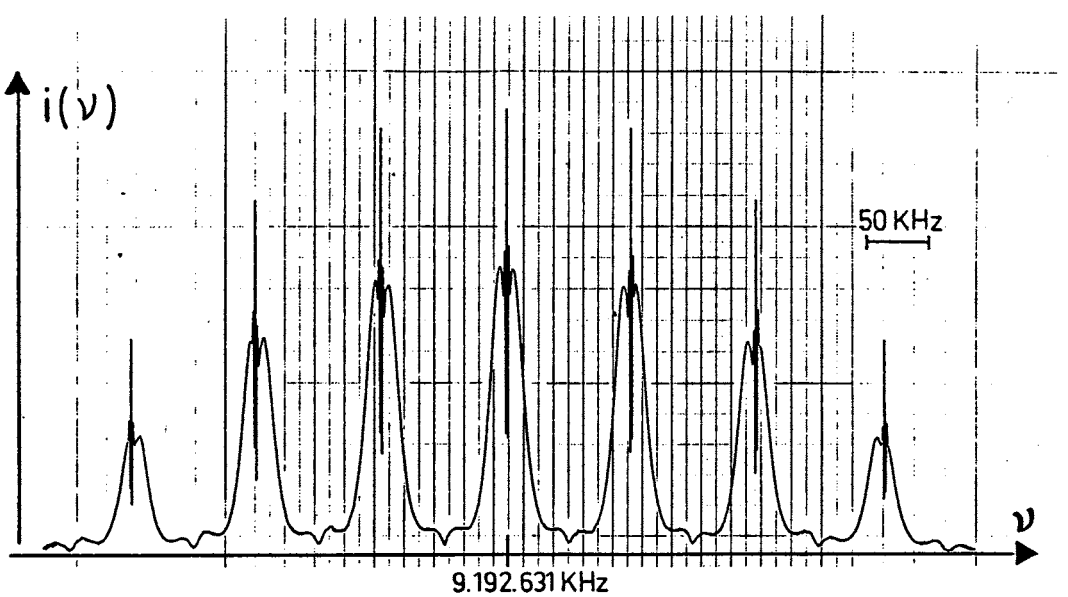


Fig. 3. Microwave spectrum of the optically pumped cesium atoms. Note the symmetry of the spectrum around the central line.

iii) Removal of state selector magnets creating strong magnetic fields will improve the magnetic field homogeneity in the whole region including the two microwave interaction regions and the drift region between

them. This will decrease greatly the related frequency shifts.

iv) Indirect frequency shifts due to Majorana transitions should disappear, as a consequence of the better magnetic field homogeneity.

v) The deflection angle of the atomic trajectories is approximately two orders of magnitude smaller in an optical pumping region than in a magnet state selector (10^{-4} radian instead of 10^{-2} rad.). This is favorable to a better estimate of the distributed cavity phase shift or to its measurement by beam reversal.

Consequently, we have started the study of possible designs for the generation of the interrogation signal and for the processing of the beam tube response. The goal is to investigate means to take advantage of the frequency stability capability of optically pumped cesium beam tubes.

In this paper we shall describe a generator of the resonance line interrogation signal which includes innovative features. It has been implemented to measure the frequency stability of our optically pumped device. However, the poor quality of the balance of the two arms of the microwave cavity and of the magnetic shield of the presently operated set-up did not enable yet to characterize the long term properties of the electronic system.

3. GENERATION OF THE INTERROGATION SIGNAL

A block-diagram of the microwave interrogation signal generator is shown in Fig. 4. The interrogation signal is square-wave frequency modulated. The instantaneous frequency is thus equal to :

$$\nu(t) = \nu_{Cs} + \Delta\nu_z + \nu_m g(t) \qquad (1)$$

where ν_{Cs} is the hyperfine transition frequency of the isolated cesium atom, $\Delta\nu_z$ is the second order Zeeman frequency shift and ν_m is the amplitude of the frequency deviation. $g(t)$ is a modulation function, equal to + 1 during one half of the modulation period and to - 1 during the other half period. The amplitude of the microwave signal is regulated and the undesired spectral components created in the frequency synthesis chain are very well rejected.

3.1. Nine GHz frequency generator

The quartz crystal oscillator to be frequency locked to the cesium hyperfine transition is a high quality 10 MHz QAS quartz crystal oscillator. There is no frequency multiplier used in the 10 MHz to 1 GHz part of the chain. The 1 GHz signal is generated by a temperature controlled

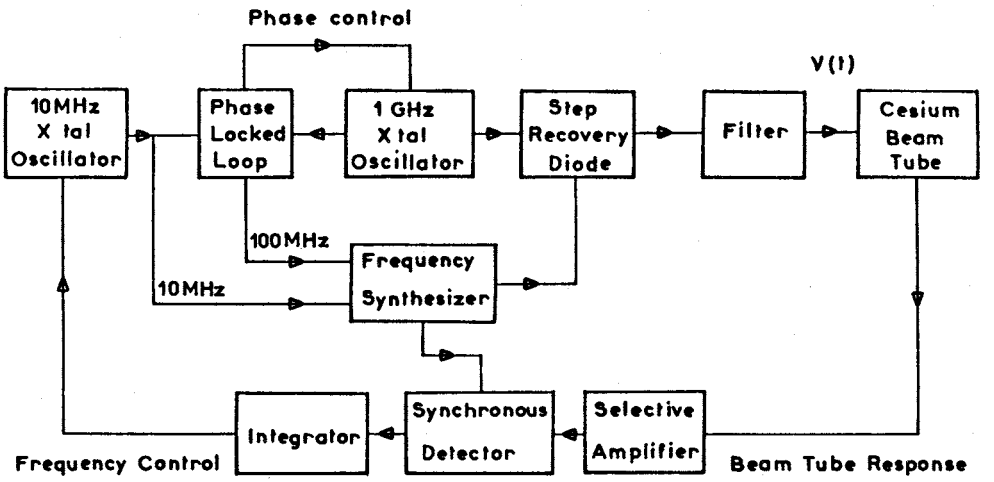


Fig. 4. Simplified block-diagram of the interrogation signal generator and of the frequency control of the 10 MHz quartz crystal oscillator.

crystal oscillator which is phase locked to the 10 MHz oscillator. The loop includes an ECL frequency divider which produces a 100 MHz signal at its output. A second divider per 10 gives the signal which enables the phase comparison with the 10 MHz quartz crystal oscillator output. The bandwidth of the first order phase locked loop is $f_c \sim 4$ kHz.

For $f < f_c$, the phase fluctuation of the 1 GHz output is determined by that of the 10 MHz frequency source. For $f > f_c$, it is identical to that of the free running 1 GHz source. It has been verified that the frequency dividers do not contribute significantly to phase noise.

Fig. 5 shows the power spectral density of the phase noise at the 1 GHz output. One sees that for $f \gtrsim 4$ kHz, the phase spectrum is much improved compared to that which would be obtained after frequency multiplication of the 10 MHz quartz crystal oscillator output.

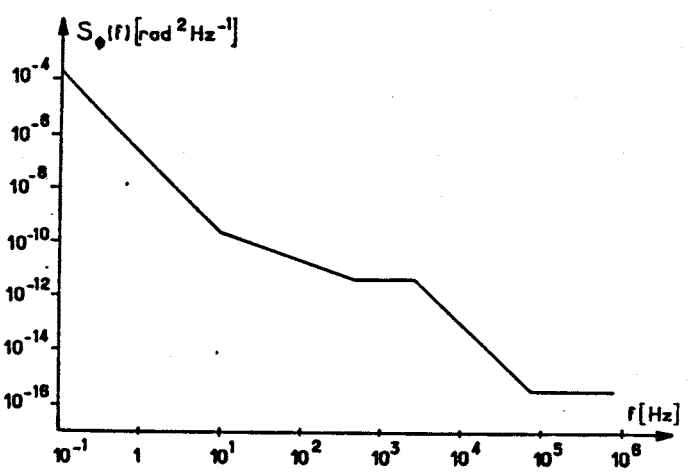


Fig. 5. Power spectral density of the phase fluctuations of the phase locked 1 GHz frequency source (asymptotic representation).

The important advantages of the use of a phase locked high frequency source are i) the absence of any sideband at 10 MHz and at its harmonics and ii) the easiness of implementation, which is mainly due to the absence of tuned circuits.

The frequency of the 1 GHz signal is multiplied by a factor of 9 by means of a step recovery diode.

3.2. Frequency synthesizer

A signal at the frequency $\nu'(t)$ must be mixed to the 9 GHz microwave output in order to generate the interrogation signal at frequency $\nu(t)$. In our case we thus have :

$$\nu'(t) = 192\ 631\ 770\ \text{Hz} + \Delta\nu_z + \nu_m g(t) \quad . \tag{2}$$

We have deliberately chosen a design of the frequency synthesis chain which ensures a large value of $\nu'(t)$. The reason is the following. It is known that a spurious spectral component at angular fequency ω' produces an angular frequency shift $\Delta\omega$, which originates in the quantum properties of the atoms and which is approximately given by [17]:

$$\Delta\omega \sim \frac{\ell}{L} \frac{b_1^2}{(\omega_o - \omega')} \quad , \tag{3}$$

where ω_o is the atomic transition angular frequency, ℓ the length of each interaction region and L the separation between the two interaction regions. b_1 is the Rabi angular frequency of the spurious spectral component. b_1^2 is proportional to the power of this component. In usual designs one often has $|\omega_o - \omega'|/2\pi = 12.6$ MHz or 7.4 MHz and the power of the spurious component may be as large as 1/4 that of the carrier. This practically gives $\Delta\omega/\omega_o \sim 10^{-12}$ for $\ell = 1$ cm and L = 15 cm. This is a large effect which may be the source of long term frequency drifts related to a microwave power change. To avoid such an effect we have chosen $|\omega_o - \omega'|/2\pi = 192$ MHz and the spurious spectral components are very well rejected so that the related fractional frequency shift is smaller than 1×10^{-14}.

Square wave frequency modulation has been chosen since it is favorable to the achievement of a good frequency stability of the controlled source [18(19)] and since it avoids the use of an analog phase modulator which might introduce frequency shifts in the case of a second harmonic distorsion of the modulation waveform[20].

For a given velocity of the atoms, for a given separation between the two interaction regions and between the second one and the detection region,

and for a given frequency of modulation, it exists a value of the amplitude of the frequency deviation and of the microwave power which optimizes the frequency stability [18]. In the present case, the velocity distribution is Maxwellian and the modulation angular frequency, $\omega_M = 2\pi \times 95$ rad. s^{-1}, is such that we have $\omega_M T_o = 0.6$, where T_o is the transit time between the two interaction regions for atoms having the most probable velocity in the oven. For the geometry of our tube, the optimum value of the angular frequency deviation is then ω_m given by $\omega_m T_o \sim 1.6$. This leads to $\omega_m/2\pi \sim 260$ Hz, a value close to half the linewidth. However, the optimum of stability is not very sharp and we have to take into account the fact that frequency shifts such as the second order frequency shift and the cavity phase shift are increasing functions of $\omega_m T_o$. Consequently, we have chosen $\omega_m/2\pi = 190$ Hz as a practical value. The optimum value of the microwave power is given by $b\tau_o \sim 1.6$, where we have $\tau_o = \ell T_o/L$ and where b is the Rabi angular frequency of the atoms. The corresponding value of the microwave power can be adjusted referring to the value which provides the optimum transition probability, for which we have $b\tau_o = 1.88$ for the velocity distribution considered.

One can separate the frequency synthesizer into two parts. One of them generates a square-wave frequency modulated signal at the frequency :

$$\nu''(t) = 7368230 \text{ Hz} - \Delta\nu_z \mp \nu_m \, g(t) \qquad (4)$$

and the second one mixes this signal with another one at 200 MHz. We thus have $\nu''(t) = 2 \times 10^8 - \nu'(t)$, where the frequencies are expressed in Hertz.

Fig. 6 represents schematically the generator of the signal whose frequency is $\nu''(t)$. Two quartz crystal oscillators produce a signal at fixed frequencies, ν_1 and ν_2, respectively. They are multiplexed in a logic gate so that the output frequency is equal to ν_1 during one half period of modulation and to ν_2 during the second half period. The two quartz crystal oscillators are frequency controlled in order to fulfil the two following conditions :

$$\frac{1}{2}(\nu_1 + \nu_2) = 7368230 \text{ Hz} - \Delta\nu_z \qquad (5)$$

and

$$\frac{1}{2}(\nu_1 - \nu_2) = \nu_m \qquad (6)$$

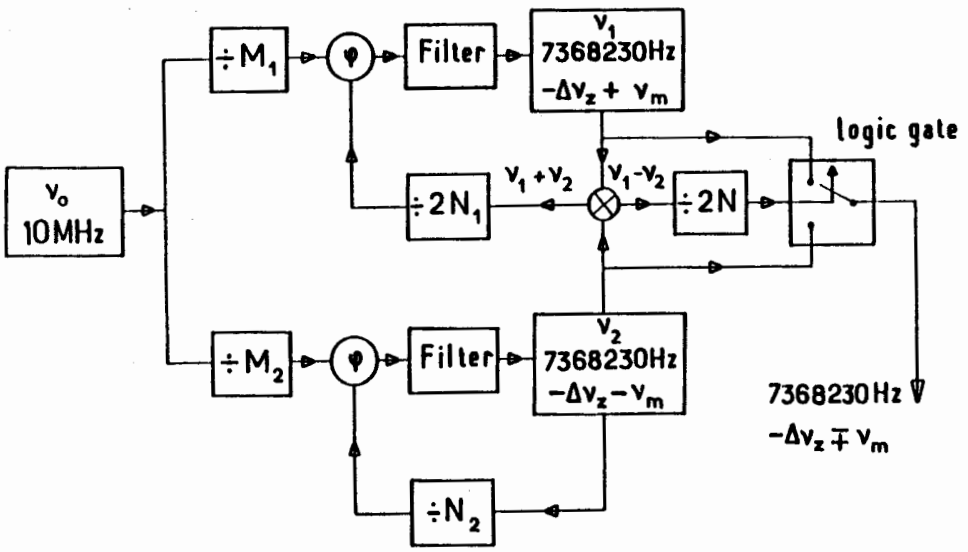


Fig. 6. Block-diagram of the generator of the frequency modulated signal.

For the purpose of their frequency control the two signals at frequency ν_1 and ν_2 are mixed and two phase locked loops of the M/N type are operated. The first one makes use of the signal at the sum frequency $\nu_1 + \nu_2$. This signal is filtered at the mixer ouptut. Its frequency is divided by the factor $2N_1$. A phase locked loop ensures that we have :

$$\frac{1}{2}(\nu_1 + \nu_2) = \frac{N_1}{M_1}\nu_o \quad , \tag{7}$$

where ν_o is the frequency of the 10 MHz quartz crystal oscillator. The choice of the factors N_1 and M_1 sets the value of $\Delta\nu_z$. For example, we have $\Delta\nu_z = 1.596$ Hz with $N_1 = 2013$ and $M_1 = 2732$. The corresponding value of the magnetic induction applied to the atoms is 61.1×10^{-7} Tesla. The second phase locked loop fixes the value of ν_2.

We have :

$$\nu_2 = \frac{N_2}{M_2}\nu_o \quad . \tag{8}$$

From Equations (6), (7) and (8) we obtain :

$$\nu_m = \frac{\nu_1 - \nu_2}{2} = \left(\frac{N_1}{M_1} - \frac{N_2}{M_2}\right)\nu_o \quad . \tag{9}$$

With $N_2 = 2024$ and $M_2 = 2747$, we have $\nu_m = 190$ Hz. The signal at the difference frequency $\nu_1 - \nu_2$, also available at the mixer output, is filtered, amplified and divided by the factor $2N$. The resulting signal drives the logic gate. The frequency, ν_M, of the modulation of the output signal is thus equal to :

$$\nu_M = \frac{\nu_1 - \nu_2}{2N} = \frac{\nu_m}{N} \tag{10}$$

With N = 2, we have ν_M = 95 Hz. It may be noted that the modulation frequency being an even submultiple of the amplitude of the frequency deviation, the carrier of the signal is suppressed [21] (see Fig. 8).

As shown in Fig. 7, the signal at frequency $\nu''(t)$ is mixed with another one at 200 MHz which is issued from the 1 GHz to 10 MHz frequency division chain. The lower sideband, at frequency ν' is selectively

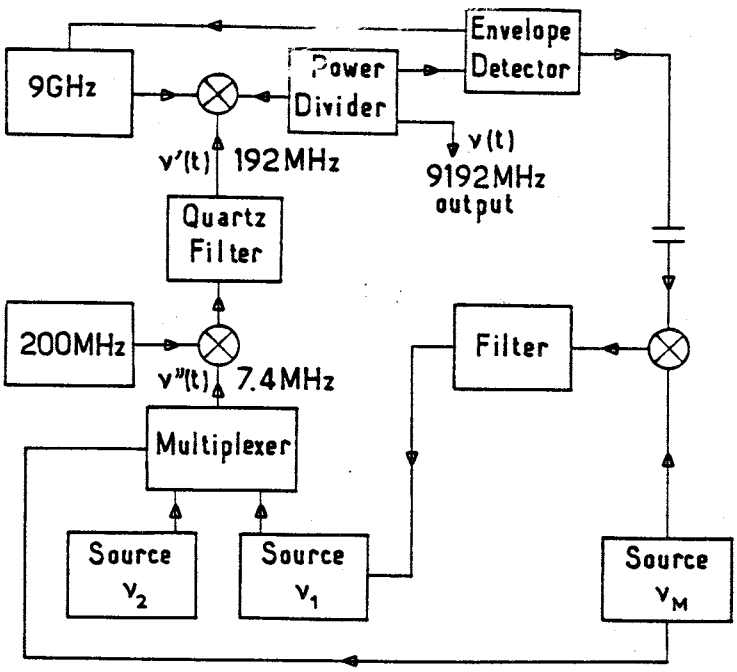


Fig. 7. Simplified block diagram showing i) the relation between the signals at frequencies $\nu''(t)$, $\nu'(t)$ and $\nu(t)$, ii) the control of the 9192 GHz microwave power level and iii) the control of the equality of the amplitude of the interrogation signal during the two half periods of modulation. For simplicity, the generation of the 9 GHz signal and the mixing with the 192 MHz signal are represented separately. Actually, the two functions are accomplished in the same step recovery diode.

filtered in a quartz crystal filter having a bandwidth of 100 kHz.Undesired spectral components are at least 27 dB below the desired one. Then, the signal at frequency $\nu'(t)$ is applied to the step recovery diode, where it is mixed to the harmonics of the 1 GHz signal. A filter made of two coupled coaxial cavities, having a bandwidth of approximately 50 MHz, selects the spectral components around 9.2 GHz. The more selective microwave Ramsey cavity attenuates sufficiently the remaining undesired spectral components.

3.3. Control of the microwave level

A number of frequency offsets such as, for instance, those related to the second order Doppler effect, the cavity phase shift, the cavity mistu-

ning, the spurious spectral components and the imperfections in the electronics depend on the microwave interrogation level. A 1 % variation of this level gives a fractional frequency offset change of the order of 10^{-14}. A control of the microwave level is thus desirable. As shown in Fig. 7, it is implemented by detecting the level of the 9.2 GHz signal and by acting upon the polarization current of the step recovery diode. A level stability of 0.2 % per degree Celsius is achieved.

Moreover, it is known that amplitude modulation of the interrogation signal at odd harmonics of the modulation frequency may be the source of a false error signal in the frequency control loop [20]. In order to avoid such an effect, it is necessary to ensure that the amplitude of the interrogation signal is the same during the two half periods of modulation. For that purpose, as shown in Fig. 7, the amplitude changes of the interrogation signal are synchronously detected, using a signal at frequency ν_M as a reference. Then, the error signal controls the amplitude of the output of the source at frequency ν_1. The residual fractional amplitude modulation is approximately equal to 2×10^{-6}.

The power available at frequency $\nu(t)$ is 0.3 mW. Fig. 8 shows the spectrum of the interrogation signal. Spurious components due to electronic imperfections are at least 60 dB below the signal power. The related fractional frequency shift is smaller than 1×10^{-14}.

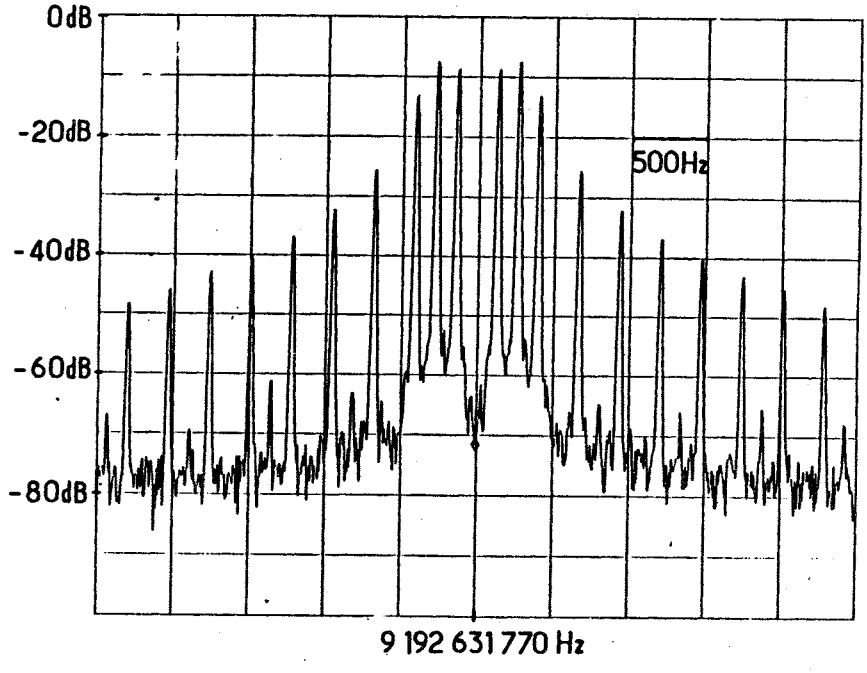


Fig. 8. Power spectrum of the square wave frequency modulated microwave interrogation signal. The level referred as 0 dB is that of the signal without modulation.

4. PROCESSING OF THE BEAM TUBE RESPONSE

At present, the processing of the cesium beam response is made classically. The first harmonic of the response is selectively amplified and synchronously detected at 95 Hz. The error signal feeds an integrator whose output voltage is applied to the frequency control varactor of the 10 MHz quartz crystal oscillator. The value of the loop time constant is set to 0.2 s.

5. MEASURED SHORT TERM FREQUENCY STABILITY

The frequency stability of the frequency controlled 10 MHz quartz crystal oscillator has been measured with respect to a hydrogen maser. Fig. 9 shows the Allan standard deviation obtained. The result is :

$$\sigma_y(\tau) = 2 \times 10^{-12}/\sqrt{\tau} \quad , \qquad (11)$$

for $1\ s < \tau \lesssim 500\ s$. For $\tau > 500$ s, systematic effects due to imperfections of the set up in its present version limit the achievable frequency stability.

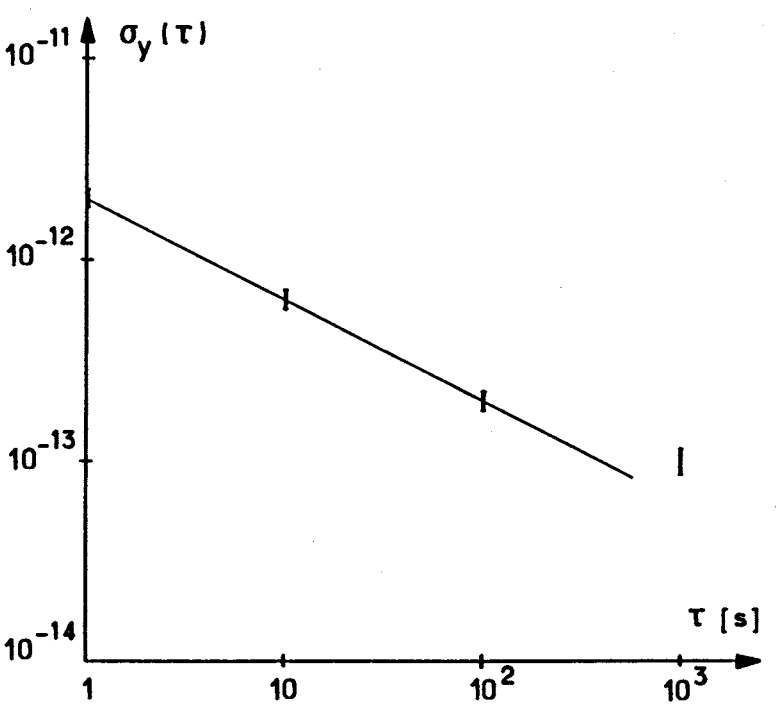


Fig. 9. Measured short term frequency stability of an experimental optically pumped cesium beam frequency standard.

Referring to the current of a hot wire detector placed in this experimental tube, the axial beam intensity was 6.1×10^{14} atoms.sr^{-1}.s^{-1}.

To our knowledge this measured short term frequency stability is the best obtained to date with an optically pumped cesium beam frequency stan-

dard. It confirms the interest of this kind of frequency standard.

6. PROSPECTS FOR IMPROVEMENT OF THE FREQUENCY STABILITY

It can be shown that the frequency stability is given by [22] :

$$\sigma_y(\tau) \sim F^{1/2} \frac{(2I'_b + I_{pv})^{1/2}}{2Q_\ell I_{pv} \tau^{1/2}} \quad . \tag{12}$$

where I_{pv} is the peak to valley flux of detected atoms and Q_ℓ is the line quality factor. F is the noise factor of the optical detection of atoms. It is given by [13] :

$$F = 1 + \frac{1}{\beta\gamma\eta} \quad , \tag{13}$$

where β is the photon yield per atom, γ the photon collection efficiency and η the quantum efficiency of the photocell. Assuming that the shot noise of the stray light collected is negligible, I'_b is a fictitious flux of detected atoms whose shot noise is equivalent to the photocell thermal noise. We have :

$$I'_b = \frac{F_A (P_n R)^2}{2e^2(\beta^2\gamma^2\eta^2 + \beta\gamma\eta)} \quad , \tag{14}$$

where F_A is the noise factor of the current to voltage converter, P_n the noise equivalent power of the photocell, R its responsitivity and e the electron charge. In our experimental set-up, we have $\beta = 3.9$, $\gamma = 0.14$, $\eta = 0.85$, $I_{pv} = 6.4 \times 10^9\ s^{-1}$, $I'_b = 3.9 \times 10^{10}\ s^{-1}$ and $Q_\ell = 1.8 \times 10^7$. This gives $\sigma_y(\tau) \sim 2.2 \times 10^{-12}/\sqrt{\tau}$ in agreement with the measured value.

Equations (12) to (14) enable to discuss the possible improvement of frequency stability.

1°) In the present set-up, the noise factor of the current to voltage converter connected at the photocell output is 3 times its theoretical value. It follows that a gain of a factor 1.6 on $\sigma_y(\tau)$ can be achieved.

2°) If the detector noise were negligible, i.e. if the frequency stability were limited by the beam shot noise, the frequency stability would be improved by a factor of 3.6. The condition assumed can, in principle, be fulfilled using a cycling transition for the optical detection, provided that the laser spectral purity is sufficiently enhanced. It should be noted that the noise factor F of the optical detection would be improved at the same time since the value of β is larger for a cycling transitions than of the presently used pumping one. Assuming $\beta = 30$ instead of 3.9,

the value of $F^{1/2}$ is decreased by a factor 1.6. Obviously, an increase of the collection efficiency γ is also beneficial.

3°) Assuming two-laser optical pumping, the flux of atoms in state $F = 3$, $m_F = 0$ or $F = 4$, $m_F = 0$ is increased by a factor of almost 8. It follows an improvement of the frequency stability by a factor almost equal to 2.8 in the case $I'_b \ll I_{pv}$.

We thus see that improvement of the frequency stability by a factor of the order of 10 is theoretically possible.

Equations (12) to (14) can also be used to look at scale effects.

1°) If the flux of detected atoms is changed, $\sigma_y(\tau)$ varies as I_{pv}^{-1} for $I'_b \gg I_{pv}$ and as $I_{pv}^{-1/2}$ for $I'_b \ll I_{pv}$.

2°) If the separation L between the two microwave interaction regions and if the separation between the oven and the detection zone are scaled by the same factor, whereas the axial beam intensity is kept constant, then $\sigma_y(\tau)$ does not depend on L in the case $I'_b \ll I_{pv}$, but varies as L in the case $I'_b \gg I_{pv}$. The latter result means that, in the present state of the art, where I'_b is larger than I_{pv}, the frequency stability of an optically pumped cesium beam tube is improved when its overall length is decreased, the beam intensity beeing kept constant.

Acknowledgements.

The authors are grateful to CEPE who supplied the quartz crystal oscillators and filters used in this work.

REFERENCES

(1). G. Avila, E. de Clercq, M. de Labachelerie and P. Cérez, IEEE Trans. on Instr. and Meas. IM-34 (1985), 139

(2). Shang Song-quan, Wu Xin-xin, Yao Shu-tong, Xie Lin-zhen and Wan Yi qiu, Chinese Physics Letters 2 (1985), 557

(3). A. Derbyshire, R.E. Drullinger, M. Feldman, D.J. Glaze, D. Hilliard, D.A. Howe, L.L. Lewis, J.H. Shirley, I. Pascaru and D. Stanciulescu, Proc. of the 39th Annual Symposium on Frequency Control, Philadelphia USA (1985), p. 18

(4). S. Oshima, Y. Nakadan and Y. Koga, private communication (1987)

(5). R.E. Drullinger, J. Shirley, D.J. Glaze and L.W. Hollberg, Proc. of the 40th Annual Symposium on Frequency Control, Philadelphia, USA (1986), p. 428

(6). T. Mc Clelland, I. Pascaru, J. Zacharski, N.H. Tran and M. Meirs, Proc. of the 41st Annual Symposium on Frequency Control, Philadelphia USA (1987). To be published

(7). E. de Clercq, A. Clairon and B. Dahmani, Proc. of the 2nd European Frequency and Time Forum, Neuchâtel, Suisse (1988), this issue

(8). P. Thomann, H. Schweda and G. Busca, Proc. of the 2nd European Frequency and Time Forum, Neuchâtel, Suisse (1988), this issue

(9). V. Giordano, A. Hamel, G. Théobald, P. Cérez, C. Audoin and V. Candelier, Proc. of the 1st European Frequency and Time Forum, Besançon, France (1987), p. 294

(10). A. Kastler, Journal de Physique et le Radium, 11 (1950), 255

(11). M. Arditi and J.L. Picqué, Journal de Physique-Lettres 41 (1980), L-379

(12). G. Avila, V. Giordano, V. Candelier, E. de Clercq, G. Théobald and P. Cérez, Physical Review A 36 (1987), 3719

(13). V. Giordano, V. Candelier, A. Hamel, C. Audoin, G. Théobald and P. Cérez. To be published

(14). V. Giordano, A. Hamel, G. Théobald, P. Cérez, C. Audoin and V. Candelier, Metrologia, 25 (1988), in press

(15). A. de Marchi, G.D. Rovera and A. Premoli, Metrologia, 20 (1984), 27

(16). A. de Marchi, IEEE Trans. on Ultrasonics, Ferroelectronics and Frequency Control, UFFC-34 (1987), 598

(17). C. Audoin, M. Jardino, L.S. Cutler and R.F. Lacey, IEEE Trans. on Instr. and Meas. IM-27 (1978), 325

(18). C. Audoin, V. Candelier and J. Vanier, IEEE Transactions on Ultrasonics, Ferroelectronics and Frequency Control, UFFC-34 (1987), 573

(19). A. de Marchi, G.D. Rovera and A. Premoli, IEEE Transactions on Ultrasonics, Ferroelectronics and Frequency Control, UFFC-34 (1987), 582

(20). F.L. Walls, IEEE Transactions on Ultrasonics, Ferroelectronics and Frequency Control, UFFC-34 (1987), 592

(21). C. Audoin, REvue de Physique Appliquée 17 (1982), 273

(22). J. Vanier and C. Audoin in "The Quantum Physics of Atomic Frequency Standards". Adam Hilger, Bristol, to appear, 1989

2 ND EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 16 - 17 - 18 MARCH 1988

Session AFS-03 Number 2

Light Pressure effects on an Optically Pumped Cesium Beam Frequency Standard

A. Clairon, E. de Clercq, B. Dahmani, A. H. Gérard

Laboratoire Primaire du Temps et des Frequences
61, avenue de l'Observatoire
75014 PARIS - FRANCE

and C. Thomas

B.I.P.M. - Pavillon de Breteuil
92312 SEVRES Cédex - FRANCE

Abstract

This paper describes the light pressure effects on an atomic beam and presents our results about travelling wave deflection of a cesium beam and its standing wave narrowing or broadening. Deflection introduces a new cavity phase shift and a velocity selection which should be considered in order to estimate frequency bias in a primary standard.

Introduction

In an optically pumped clock, laser light interacts with an atomic beam of cesium in order to achieve the population difference between the clock levels and to detect the microwave transition. Beyond these typically spectroscopic purposes, such interactions induce spurious effects such as modifications of the shape and velocity distribution of the beam by transfer of momentum from the light to the atoms. The elaboration of an optically pumped frequency standard needs of course to strictly control radiation pressure phenomena and perhaps even to use them in order to get monocinetic and parallel beams.

The aim of this paper is to describe the experimental and numerical study of the deflection, collimation and decollimation of a cesium beam by laser interaction and to evaluate their quantitative impact on the accuracy of an optically pumped atomic clock.

This work is supported by Bureau National de Métrologie - FRANCE

I - Theoretical model

The transverse deflection of an atomic beam by radiation pressure, first observed by FRISCH in 1933 [1] then by other groups in 1972 [2,3], has been the subject of considerable investigation in recent years. GORDON and ASHKIN [4] discussed this problem for a two - level atom travelling with a momentum $m\vec{v}$ in a region illuminated by a laser radiation of wave-vector $\vec{k}$. After absorption of a laser photon, the atomic momentum varies by $\vec{p} = \hbar\vec{k}$. If the atom then emits a stimulated photon, the atomic momentum varies by $-\vec{p}$ so that this pair of events results in no change. But, if the atom emits N spontaneous photons during its travel in the interaction region, the average of the spontaneous emitted photons momenta becomes negligible with large N while the laser transfers to the atom the total momentum $N\hbar\vec{k}$.

The atom is then deflected by the averaged force of radiation pressure [5]

$$\vec{F}(\vec{v}) = \hbar\vec{k}\,\frac{\Gamma}{2}\cdot\frac{s\,(\vec{k}.\vec{v})}{1 + s\,(\vec{k}.\vec{v})} \qquad (1)$$

where $\hbar\vec{k}$ is the momentum of one photon and $\frac{\Gamma}{2}\cdot\frac{s\,(\vec{k}.\vec{v})}{1 + s\,(\vec{k}.\vec{v})}$ is the number of spontaneous photons emitted during one second. Γ is the natural width of the excited state ($\Gamma/2\pi$ = 5 MHz for Cs) and s is the saturation parameter given by :

$$s\,(\vec{k}.\vec{v}) = \frac{w_1^2 \,/\, 2}{(\delta - \vec{k}.\vec{v})^2 + \Gamma^2/4} \qquad (2)$$

w_1^2 is the square of the Rabi frequency associated to the transition and proportional to the laser beam intensity. $\delta = w - w_0$ is the detuning of the laser from the resonant frequency w_0 of the atom and the scalar product $\vec{k}.\vec{v}$ represents the first order Doppler effect.

Let's precise that expression (1) is only valid in the hypothesis of an interacting progressive monochromatic plane wave and a two-level atom, i.e. with a transition which does not allow optical pumping. If one uses a "pumping" transition, one can no more define an averaged force over a large number of photons.

From the preceeding expressions (1) and (2), it follows that the parameters of interest in the deflection of an atomic beam by light pressure are :

- δ the frequency detuning (typically a few MHz)

- s the saturation parameter (typically a few unities for laser intensities of a few mW/cm²)

- L the length of interaction (5 to 30 mm) which determines the interaction duration and then the total transfer of momentum.

- α the angle between the laser and the atomic beam. In our experiment, a great deal of attention is paid to achieve a perpendicular interaction within a few milliradians.

For very small lengths of interaction, the deflection of the atom is maximum when the laser crosses the beam orthogonally and when its frequency is tuned on resonance. As soon as the atom is deviated, its own frequency is shifted by first order Doppler effect and the laser is out of resonance after a thousand of absorption. Thus a general pattern of the angular deviation versus the quantity δ cannot be drawn because it depends at least of the initial velocity distribution, of the length of interaction L and of the spatial energy distribution of the laser beam.

Let's give a few orders of magnitude. The mean longitudinal velocity of the cesium atoms is about 300 m/s, which leads to 500 spontaneous photons for L = 1 cm, with large values of s and a two-level atom. A "one photon absorption" induces a change in the transverse velocity $\hbar k/m$ = 0,35 cm/s (m is the cesium atom mass) and then of 1,7 m/s for L = 1 cm, allowing an angular beam deviation of 5 milliradians.

Now, let's add a second progressive plane wave, similar to the previous one but propagating in the opposite way so that $\vec{k}$ is replaced by $-\vec{k}$. It can be due to the precedent laser beam reflected orthogonally on a mirror of reflection coefficient R. The resulting light pressure force on an atom of velocity $\vec{v}$ can be written [5].

$$\vec{F'}(\vec{v}) = \hbar\vec{k}\,\frac{\Gamma}{2}\left[\frac{s\,(\vec{k}.\vec{v})}{1 + s\,(\vec{k}.\vec{v})} - \frac{R.s\,(-\vec{k}.\vec{v})}{1 + R.s\,(-\vec{k}.\vec{v})}\right] \qquad (3)$$

The expression (3) supposes the parameter s weak enough to add independently the effects of both waves and to neglect the laser intensity gradient force (called dipole force).

For a particular transverse velocity $\vec{v}_t$ and for a non zero detuning δ, the atom "sees" different Doppler shifts from both counterpropagating waves. Thus the absorption coefficient is different for both waves. As an example, if $\delta < 0$ and $\vec{k}.\vec{v}_t < 0$, we have $s(-\vec{k}.\vec{v}) < s(\vec{k}.\vec{v})$ and the mean force $\vec{F}'$ is antiparallel to $\vec{v}_t$; for $\vec{k}.\vec{v}_t > 0$, the result is the same. So in both cases, the transverse velocity is damped and the cesium beam is collimated [6,7]. On the contrary a positive detuning leads to a decollimation of the atomic beam. These two effects are maximum for values of δ approximately equal to $\pm\Gamma/2$ (there is no general expression of the optimal detuning because its value strongly depends of the initial transverse velocity distribution which is given by experimental conditions).

Furthermore, let's notice that two perpendicular standing waves are necessary to provide a complete collimation of the atomic beam.

Of course, the random nature of the light - atom interaction produces fluctuations of the number and of the direction of the spontaneously emitted photons. These fluctuations limit, for cesium, the collimation capability to a width of the order of 10 cm/s for the transverse velocity distribution [7,5].

A numerical calculation of the deflection and collimation processes (for some values of the experimental parameters) is performed by a Monte-Carlo simulation using equations (1) and (3). Particularly, the fluctuations of the momentum transfer due to the scattering of the spontaneous emission are introduced. However, in our preliminary numerical simulation, this diffusion effect is taken into account in the laser direction but not yet in the perpendicular one. Such computations allow comparisons with the experimental features of the beam profile. It gives also transverse velovity distribution, we are unable to measure in our experiment.

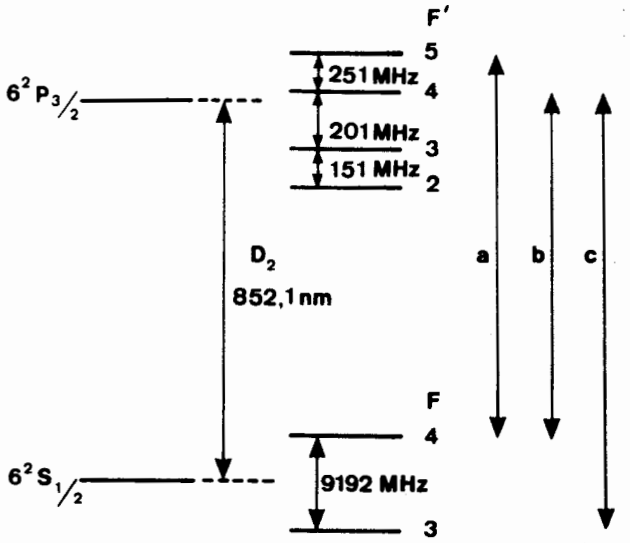


Fig. 1 : Levels scheme involved in the D_2 Cs line

The three transitions used in the experiment are :
a) F = 4 - F'= 5 (deflecting or collimating light)
b) F = 4 - F'= 4 (for frequency calibration)
c) F = 3 - F'= 4 (for optical pumping)

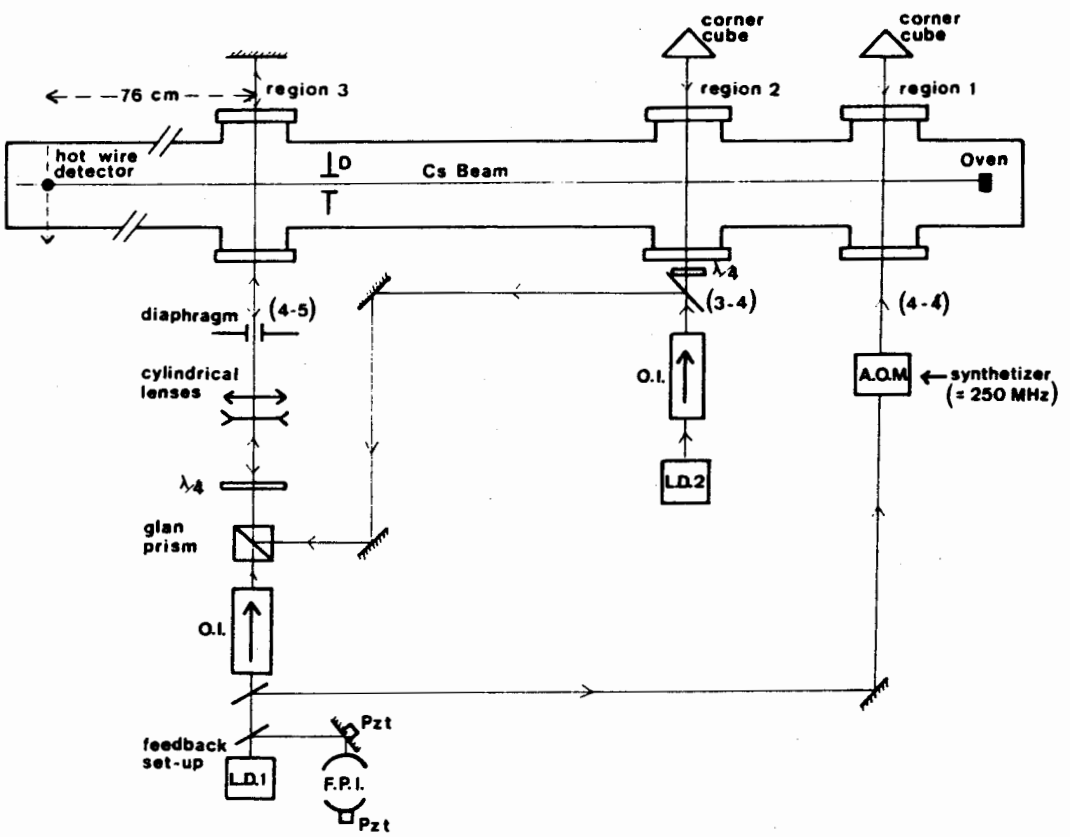


Fig. 2 : Experimental set-up

L.D. Laser Diode, A.O.M Acousto Optic Modulator,
O.I. Optical Isolator, F.P.I. Fabry Perot Interferometer,
(4-5) frequency of deflecting beam (a), (4-4) frequency of calibrating beam (b), (3-4) frequency of pumping beam (c)

II - The experiment

The study of deflection and collimation of an atomic beam of cesium implies the transition between the two levels $6\,^2S1/2\ F = 4$ and $6\,^2P3/2\ F' = 5$ (transition a in figure 1). This transition is chosen because it does not induce any optical pumping. However it is necessary to prepare the atom in the initial state $F = 4$. For this purpose, the fundamental state $6\,^2S1/2\ F = 3$ is emptied by optical pumping on the transition $F = 3 - F' = 4$ (transition c in figure 1). The two corresponding laser beams (wavelength around 0,85µm) are delivered by diodes laser LD1 and LD2 shown on the experimental set up of figure 2.

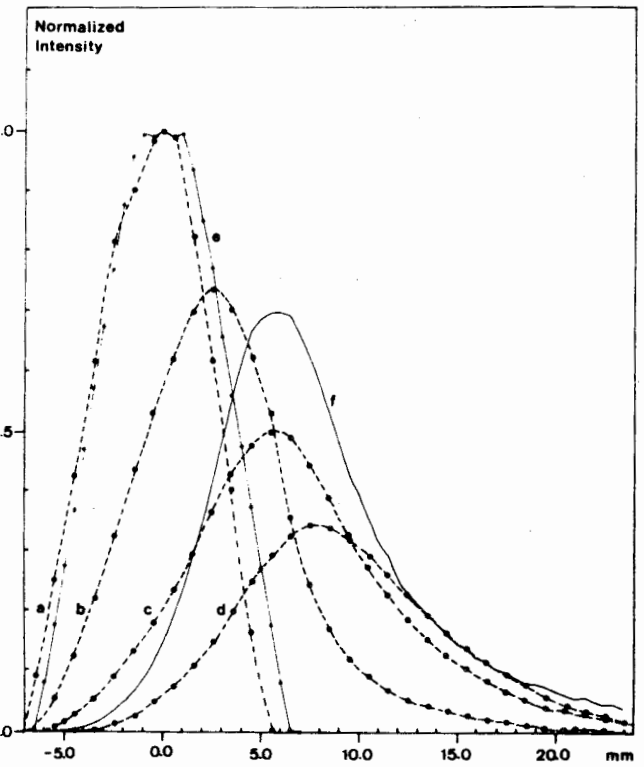
The circular polarized light corresponding to transition a interacts with the atomic beam in region 3. It's spectral purity (less than 1 MHz) is achieved by optical feedback stabilization [8]. A set of cylindrical lenses is used to spread the laser beam and the spot size, which determines the length of interaction L, is controlled by a diaphragm. The frequency of LD1 must be tunable over a few MHz and must be precisely known. For this purpose, a small part of the beam issued from LD1 passes through an acousto-optic modulator driven by a synthesizer output around 250 MHz. The obtained shifted frequency corresponds to the transition b $F = 4 - F' = 4$ and is locked on the atomic beam fluorescence in region 1. By scanning the frequency of the synthetizer output, one can vary the detuning δ of the deflecting light in region 3 and measure it within a few hundred kilohertz.

The saturation parameter s is accessible by the measure of the power of the LD1 beam and the estimation of the area of the light spot.

The pumping frequency, tuned on the transition $F = 3 - F' = 4$, delivered by the laser diode LD2 is injected in region 2 and 3.

The recordings of the tranverse atomic beam profiles is achieved by a hot wire detector movable in the direction perpendicular to the beam, and situated at a distance of 76 cm from the interaction region 3.

The dimensions of diaphragm D are 3 mm (horizontal) x 2 mm (vertical), the laser beams are in the horizontal plane. The whole horizontal beam divergence is 13 milliradians and the hot wire detection distance from D is 78 cm. The initial non deflected beam presents a slight asymmetry (curve a on figure 3) probably caused by a small misalignement of the diaphragm D.



g. 3 : Deflected atomic beam profiles (δ = 0)

Dashed lines = experimental curves :
a) without laser b) s = 1, L = 10 mm
c) s = 1, L = 30 mm d) s = 5, L = 30 mm
Solid lines = calculated curves :
e) without laser f) s = 1, L = 30 mm

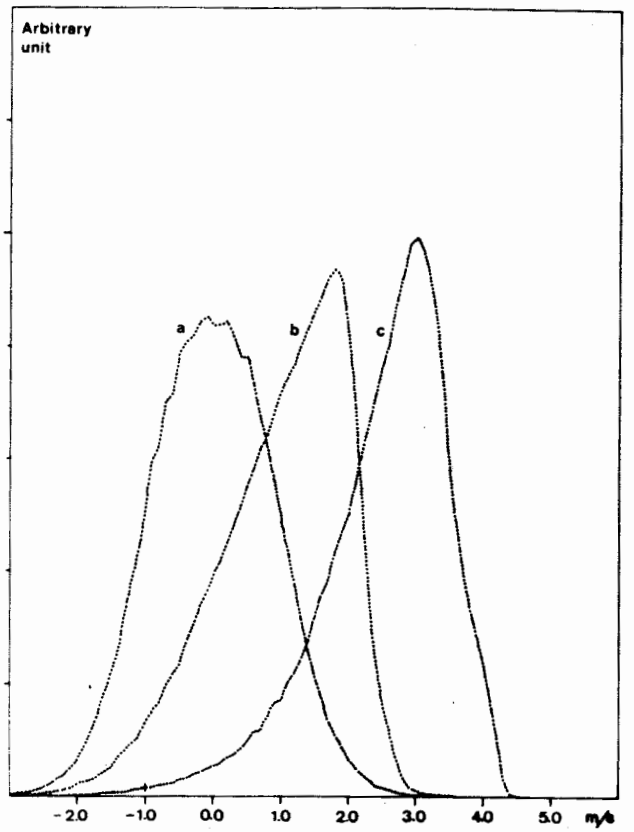


Fig. 4 : Calculated transverse velocity distribution of the deflected beam

a) b) c) are the same as a) b) c) in fig. 3

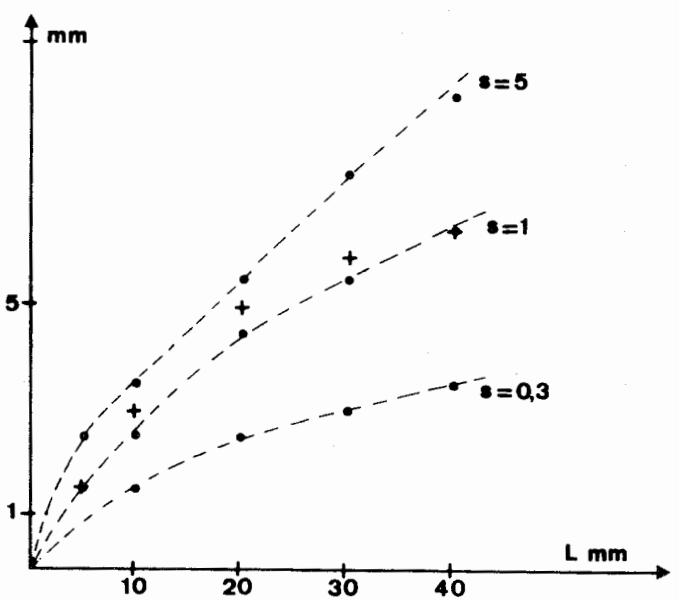


Fig. 5 : Peak displacement of the deflected atomic beam versus L (δ = 0)

+ calculated points (s = 1)

III - Experimental results

1. Deflection of the atomic beam

Deflected beam profiles recorded for a null detuning δ are shown on figure 3 (curves a to d experimental, curves e and f Monte-Carlo simulation). As expected the atomic beam is shifted and spreaded with increasing values of parameters s and L. The disagreement between theoretical and experimental curves (f and c) could be explained by an uncomplete pumping [9] (about 9 % of the atoms remain in the F = 3 ground state sublevel and are not deflected) and by diffusion in the vertical direction which is not taken into account in our simulation.

Figure 4 shows the deflected beam transverse velocity distribution deduced from our numerical simulation; a,b,c, curves correspond to the same experimental conditions as a,b,c, curves in figure 3. As we can see the initial velocity distribution is strongly shifted and distorded even for relatively low laser intensity (s = 1 for a laser intensity of 1 mW/cm^2). We show in figure 5 the measured peak displacement versus the interaction length L for three s values.

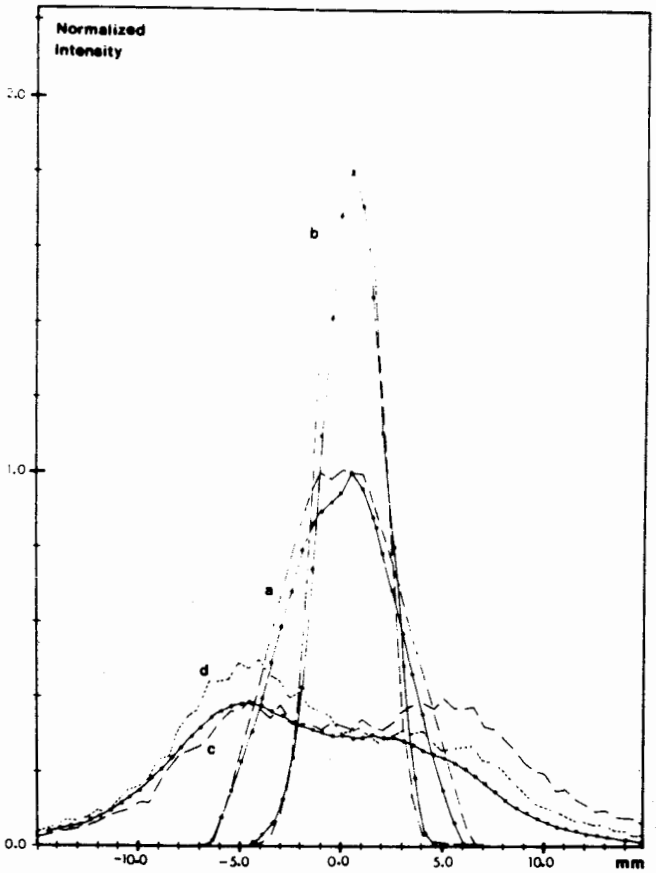
2. Collimation and decollimation of the atomic beam

Figure 6 gives the experimental evidence for the collimation and spreading of an atomic beam by interaction with a standing wave. Both counterpropagating laser beams are out of resonance and their intensities ratio R is equal to 0,85.

The maximum collimation, obtained for a negative detuning of - 2,5 MHz ($\approx -\Gamma/2$) divides by a factor 2 the beam width and increases by nearly the same factor the maximum intensity (curve b). Here also the fluctuations due to the scattering in vertical direction, which are not taken into account in the numerical simulation, and an uncomplete optical pumping could explain the disagreement between experimental and theoretical curves.

An important spreading is observed for δ = + 2 MHz (less than + $\Gamma/2$). The shape of the expanded beam drastically depends on the interaction angle. A tilt from normal incidence of 2 mrad can explain the asymmetry observed on the experimental curve c.

The computed curves of figure 7 show an almost flat distribution for the transverse velocities of the spreaded beam. On the opposite the width of the transverse velocity distribution is reduced by a factor 5 in the collimated case.



ig. 6 : Collimation and decollimation of the atomic beam
(L = 30 mm, s = 1, R = 0,85)

Solid lines : experimental curves,
dashed lines : theoretical curves,
a) without laser b) δ = - 2,5 MHz
c) δ = + 2 MHz d) δ = + 2 MHz and a tilt from normal incidence of 2 mrad

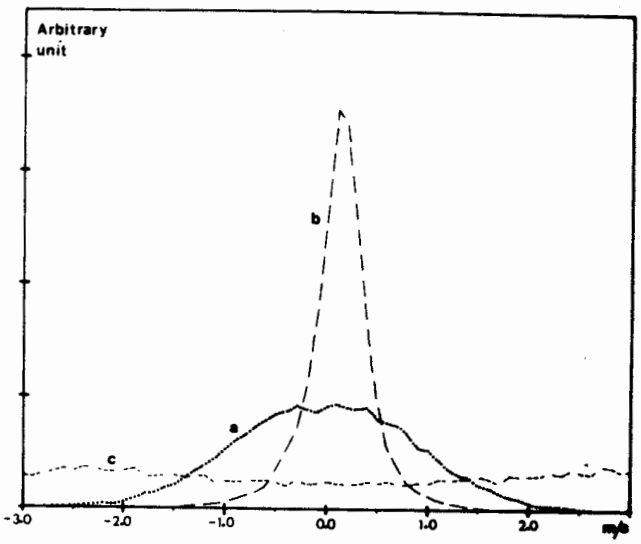


Fig. 7 : Calculated transverse velocity distribution in the atomic beam with a standing wave (L = 30 mm, s = 1, R = 0,85)

a) without laser b) δ = - 2,5 MHz

c) δ = + 2 MHz

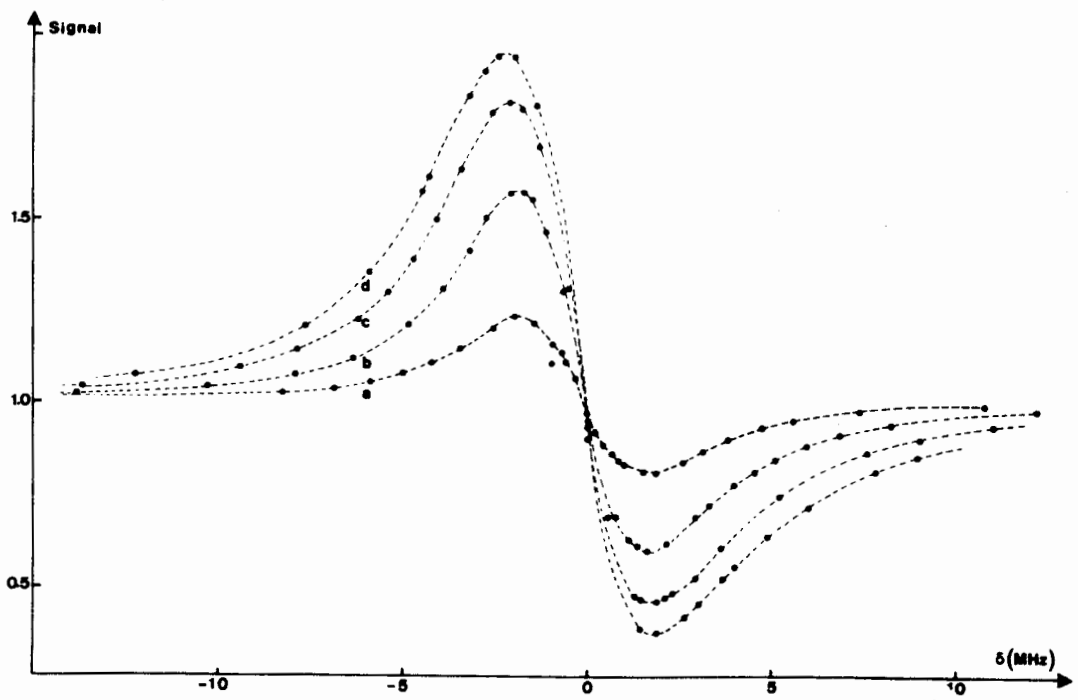


Fig. 8 : Experimental hot wire detector signal versus the detuning δ

The signal is normalized by the observed signal without laser interaction
(L = 30 mm, R = 0,85)
a) s = 0,2 b) s = 0,6 c) s = 1,2 d) s = 2

A more complete and detailed experimental study leads to the curves of figure 8 : the hot wire detector being placed at the maximum of the initial atomic beam, one records its signal versus the detuning δ and for several values of s. The essential features of collimation and decollimation processes are observed here : collimation (resp. decollimation) for $\delta < 0$ (resp > 0) and maximum effect at $\delta = -\Gamma/2$ (resp. $+\Gamma/2$) for weak values of s.

The next section is devoted to the analysis of the consequences that would induce such phenomena for an optically pumped cesium frequency standard.

IV - Applications to an optically pumped cesium frequency standard

A very simplified scheme of an optically pumped cesium frequency standard is given on figure 9 : the cesium atoms are first prepared in the fundamental state $6\ ^2S1/2\ F = 3$ by the pumping transition c $F = 4 - F' = 4$, before crossing the Ramsey cavity. The clock transition is then detected by interrogation with a laser beam issued from LD1 and tuned on the transition a $F = 4 - F' = 5$ which maximises the photon number. This scheme seems to be a good candidate in order to obtain a high signal to noise ratio in the detection region [10]. However the locking of this frequency needs a high band width photodiode to detect the Cs beam fluorescence. This characteristic is incompatible with the required signal to noise ratio in the detection region and the locking is then performed in the first region.

Therefore the atomic beam interacts with two laser beams before entering the cavity and the resulting angular deflection induces a phase shift $\Delta\Phi \approx a\theta L_c$ (a = constant) of the microwave seen by the atoms between both arms of the Ramsey cavity of length L_c [11]. The angular deviation θ and in consequence the phase shift can be numerically estimated by a Monte-Carlo simulation so long as a geometrical configuration is given. The results presented below have been obtained with the configuration of figure 9, for a 50 mrd initial angular divergence of the atomic beam and when the Ramsey cavity is closed by two shot-circuits placed at the distance $\lambda g/2$ (λg= guided wavelength) of the centers of the slots parallel to the laser beam wave-vector $\vec{k}$. The knowledge of $\Delta\Phi$ gives access to the estimation of the contribution $\Delta f/f$ of the studied effect on the accuracy of the clock. $\Delta f/f$ being proportional to $\Delta\Phi/L_c$, it results that the length of the Ramsey cavity has no influence on the clock accuracy as far as deflection of the beam is concerned.

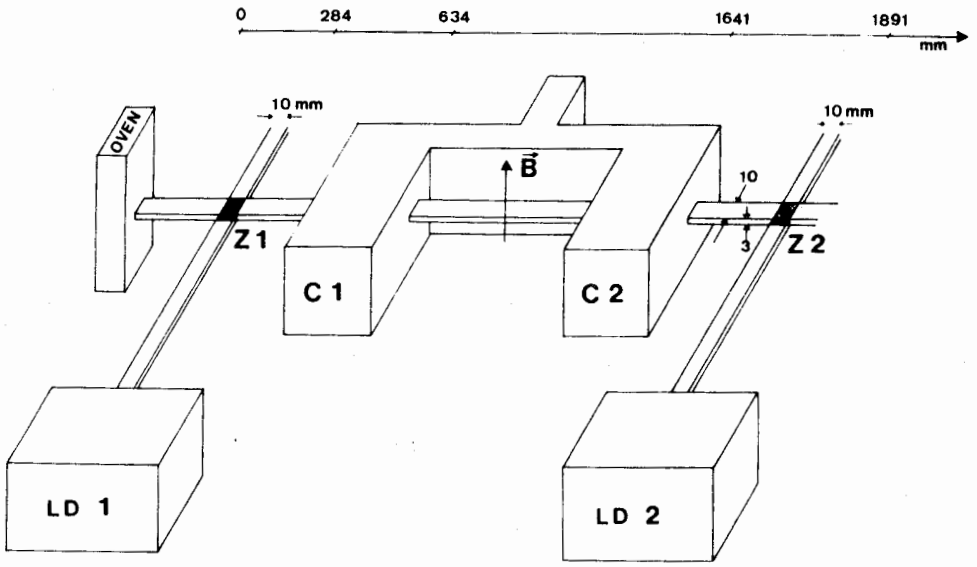


Fig. 9 : Scheme of an optically pumped frequency standard

LD1 and LD2 are laser diodes, B is the transverse static magnetic field,
the parameters used in our numerical simulations are :
10 mm long slots in the Ramsey cavity and 15 mm diaphragms in the laser interaction regions Z1 and Z2.

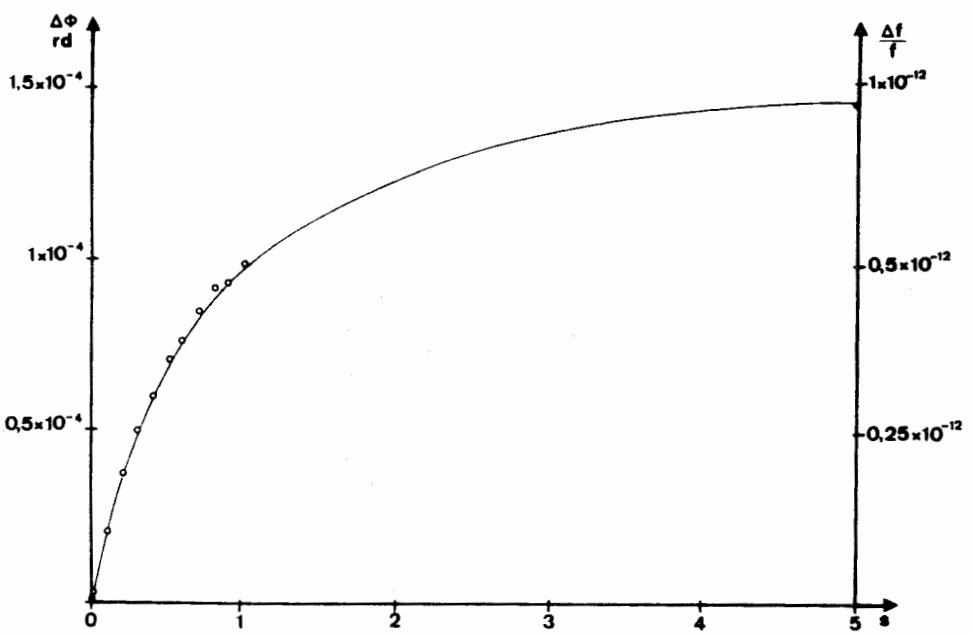


Fig. 10 : Phase shift $\Delta\Phi$ and evaluation of the relative frequency shift $\Delta f/f$ versus the s parameter for a deflecting travelling wave tuned on the 4-5 transition ($\delta = 0$)

The calculation is performed with L = 10mm, $\delta = 0$ and $\Delta f/f$ is evaluated with a Ramsey fringe width equal to 140 Hz.

The figure 10 gives the calculated curves of $\Delta\Phi$ and $\Delta f/f$ versus the s parameter for a travelling wave in first region, tuned on the 4-5 transition. For a large initial angular divergence of the atomic beam, the relative frequency shift $\Delta f/f$ reduces from a few 10^{-12}, for a non pumping travelling wave, to a few 10^{-14} when the counterpropagating wave is added (R = 0,9). For the chosen geometry, it appears that this effect is not simply proportional to the averaged deflection of the Cs beam, as naïvely expected. In fact, the rapid decrease of the frequency shift is closely tied to the fact that the averaged deflection remains much less than the initial beam divergence. In particular, for a pumping travelling wave, $\Delta f/f$ becomes negligible (smaller than 10^{-15}).

As a consequence, for a configuration where the atomic beam is first collimated by two perpendicular standing waves, to increase the detected signal, and is then optically pumped, the frequency shift could become much larger (note that the contribution of the beam perpendicular to the slots of the cavity is negligible). As a conclusion, the effect drastically depends on the particular geometry of the clock.

Until now, we have only considered the atomic trajectories modifications in the microwave cavity. This effect comes with a change in the longitudinal velocity distribution of the detected atoms. According to the experimental configuration, slow atoms can be lost (deflection) or on the contrary, favourized (collimation). The induced variations of the second order Doppler and the phase shift effects can then be of the same order of magnitude.

Conclusion

This paper raises the attention on the importance of the light pressure phenomena for Cs clock operation. The consequences of such effects on the relative frequency shift of an optically pumped atomic clock have been rapidly reviewed. It results that they present an important contribution to the clock error budget and that they must carefully be taken into account in a clock design as well as in a small size commercial standard as in a long primary standard. Furthermore such effects, already important in the actual thermal beam frequency standards, will become crucial in the emerging generation of cooled beam frequency standards.

Références

[1] R. Frisch, Z. Phys. 86,42 (1933)

[2] J.L. Picque and J.L. Vialle, Opt. Comm. 5,402 (1972)

[3] R. Schieder, H. Walther and L. Wöste, Opt. Comm. 5,337 (1972)

[4] J.P. Gordon and A. Ashkin, Phys. Rev. A21,1606 (1980)

[5] C. Cohen-Tannoudji, cours au Collège de France (1983 et 1984)
J. Dalibard, Thèse de Doctorat d'Etat, Université Paris VI (1986) and references therein

[6] T.W. Hänsch, A.L. Schawlow, Opt. Comm. 13,68 (1975)

[7] V.I. Balykin, V.S. Letokhov, V.G. Minogin and T.V. Zueva, Appl. Phys. B35,149 (1984)

[8] A. Clairon, B. Dahmani, Ph. Laurent, Ch. Bréant, this forum AFS04 session and references therein

[9] B. Jaduszliwer, G.F. Shen, and B. Bederson, Phys. Rev. A33,3792 (1986)

[10] G. Avila, V. Giordano, V. Candelier, E. de Clercq, G. Theobald and P. Cerez, Phys. Rev. A36,3719 (1987)

[11] A. de Marchi, J. Shirley, D. Glaze and R. Drullinger to be published (IEEE Trans. Ins. Meas.) and references therein.

2 ND EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 16 - 17 - 18 MARCH 1988

Session AFS-03 Number 3

COMMERCIAL SIZE OPTICALLY PUMPED CESIUM BEAM RESONATOR

P. Thomann, H. Schweda, G. Busca

OSCILLOQUARTZ S.A. - Neuchâtel

Abstract

We have constructed a commercial size optically pumped cesium resonator. Optical pumping of the atomic beam and detection of the Ramsey signal are both achieved with a single laser. The linewidth of the Ramsey signal is 950 Hz and the present signal-to-noise ratio is 1500 $Hz^{-1/2}$. Based on these measured values, the calculated short-term frequency stability is given by $\sigma_y(\tau) \cong 5\ 10^{-11}\ \tau^{-1/2}$. The main present limitation is the noise contribution of the scattered light which can be drastically reduced without significant changes in the design.

Introduction

Optical pumping is a promising alternative to magnetic deflection as a state-selecting technique in atomic beam resonators. Several laboratories have already used this technique [1-4] and short-term stabilities as high as $\sigma_y(\tau) \cong 3\ 10^{-12}\ \tau^{-1/2}$ have been obtained [1]. We report here on a prototype of a 30 cm long optically pumped resonator which is designed to also achieve an order of magnitude improvement in short term stability over presently available conventional resonators with comparable operating parameters.

Description of the resonator

The principle of operation of an optically pumped tube is illustrated in fig. 1. the atomic beam produced by the oven first intersects the pump laser beam in interaction region A. The laser frequency is tuned to one of the hyperfine components of the D_2 resonance line. A combination of hyperfine and/or Zeeman pumping leads to a population difference between the two clock states (F = 3, m = 0 and F' = 4, m' = 0). The fluorescent light emitted by the atoms in the optical pumping region is directed by the "A" collecting optics towards the A detector. Since the fluorescent intensity is maximum at resonance, a weak modulation of the laser frequency induces a modulation of the collected intensity, which is proportional to the detuning and can be used, after synchronous detection, as an errror signal to lock the laser frequency to the line center. The fluorescent intensity is also dependent on the population difference between the various Zeeman

substates and, in particular, between the two clock states. This property is used in region B, where the atomic beam intersects a probe laser beam, to detect any change in the relative population of the clock states induced by the interaction with the microwave field in the Ramsey cavity.

Several pumping and probing schemes can be envisaged, involving the use of one, two or even three lasers [5]. Two-laser or three-laser schemes may improve the optical pumping efficiency and/or probe signal level which is an advantage for the S/N ratio and the corresponding short-term stability. However, these lasers must operate at different wavelengths, which means either that additional interaction regions must be provided along the atomic beam (or in an auxiliary cell) thereby increasing its length and complexity, or that an elaborate locking system must be provided in order to slave the frequency of additional laser(s) to the first one. For reasons of simplicity and reliability, we consider here a single-laser set-up.

As shown in fig. 1, a parallel beam is obtained from the strongly diverging diode laser after a suitable collimator (spherical doublet). A half-wave plate followed by a polarizing beam-splitter make up the lossless variable beamsplitter needed to divide the available power between pump and probe beam. The polarization of each beam can be independently controlled using additional wave-plates. The wavevector k of both pump and probe beams is perpendicular to the C-field. This particular geometry is chosen such that the two possible polarizations (π and σ) can be obtained by the adequate orientation of $\vec{e}_A$ of $\vec{e}_B$ (fig. 1). The fluorescent light is collected perpendicularly to both the laser beam and the atomic beam.

Ramsey cavity length	13 cm
Total tube length	25 cm
Atomic density in B region	$2\ 10^{7}\ cm^{-3}$
Interaction volume V_B	$10\ mm^{3}$
Total atom flux through V_B	$3\ 10^{10}\ s^{-1}$
Cs consumption	0.2 g/year
Ramsey signal amplitude (20 mm^2 detector)	0.5 nW/0.3 nA
Background light level	16 nA
Measured noise current (at 500 Hz)	$200\ fA/Hz^{1/2}$
S/N ratio	$1500/Hz^{1/2}$
Ramsey signal FWHM	950 Hz
Calculated short term stability	$5\ 10^{-11}\tau^{-1/2}$

Table 1 : Operating parameters of
Optically pumped Cs beam resonator

The performance of the resonator depends critically on the amount of fluorescent light received by the detector. The light collection system was designed to maximize the solid angle of light collection. It is made of two concave mirrors (fig. 2), one elliptical and one spherical, facing each other. The interaction region is centered at one focus of the elliptical mirror while the hole in the spherical mirror and in the magnetic shield are at the second focus. A detector placed there would collect approximately 50% of the emitted light. In the measurements reported here, however, the detector was placed outside the vacuum vessel, and some light was lost because of the additional distance introduced by the window.

A traditional problem of conventional cesium beam tubes is the abrupt transition between the high inhomogeneous field of the two state-selecting magnets and the low, homogeneous field in the microwave interaction region. An important advantage of replacing the strong state-selecting magnets by optical pumping is that the C-field region can now extend over both pumping and detection regions. (The Zeeman effect associated wiht the weak C-field - less than 10^{-5} T - has a negligible incidence on the optical pumping process). This configuration allows for a better homogeneity of the C-field over the microwave interaction region.

A summary of relevant operating parameters is listed in Table 1.

Preliminary results

A recording of the B photodetector signal as a function of RF frequency (Ramsey signal) is shown in fig. 3. Both traces were recorded with the laser temperature stabilized. In the lower trace, in addition, the laser frequency was locked to the atomic fluorescence. The scan duration is 100s.

The mesasured signal-to-noise ration, S/N, measured at the center of the Ramsey resonance in a 1 Hz bandwidth around 500 Hz, is S/N = 1500 in both cases. The linewidth (FWHM) is 950 Hz, as expected from the length of the Ramsey cavity. From these measured values, we calculate the Allan variance $\sigma_y(\tau)$ of relative frequency fluctuations of an oscillator locked to such a resonator:

$$\sigma_y(\tau) = \frac{W}{\sqrt{2}\ (S/N)\,\nu_o}\ \tau^{-1/2} = 5\ 10^{-11}\tau^{-1/2}$$

(ν_o = 9.2 GHz). This is comparable to the short-term stability of conventional resonators of the same size where, however, both linewidth and S/N ratio are smaller by a factor of two.

We have carried out an analysis of the contributions to the S/N ratio. The Ramsey signal amplitude is 0.3 nA (photodetector current, responsivity 0.6 A/W) ; the background signal due to unwanted scattered light is 16 nA. Among the various contributions to the total noise, laser intensity noise associated with the background light dominate largely. By reducing the amount of background light,which we have already achieved in separate experiments by a factor of 10, and by moving the detector closer to its ideal position, we expect to improve the S/N ratio by an order of magnitude.

A similar scan to that of fig. 3 but over a larger RF frequency range (fig. 4) shows the Ramsey patterns corresponding to all the field dependent transitions ($m_F = \pm 1, \pm 2, \pm 3$). The symmetry of the overall spectrum contrasts with the generally asymmetric spectra of magnetic deflection resonators; this symmetry is a consequence of optical pumping with linearly polarized light. It is a desirable property since Rabi pulling and Majorana effects are reduced. This spectrum was recorded with a single magnetic shield (outer shield of fig. 2 omitted). The good visibility of the Ramsey patterns on the outermost resonances is proof of the C-field homogeneity (B = 53 mG, homogeneity 1 %) mentioned above.

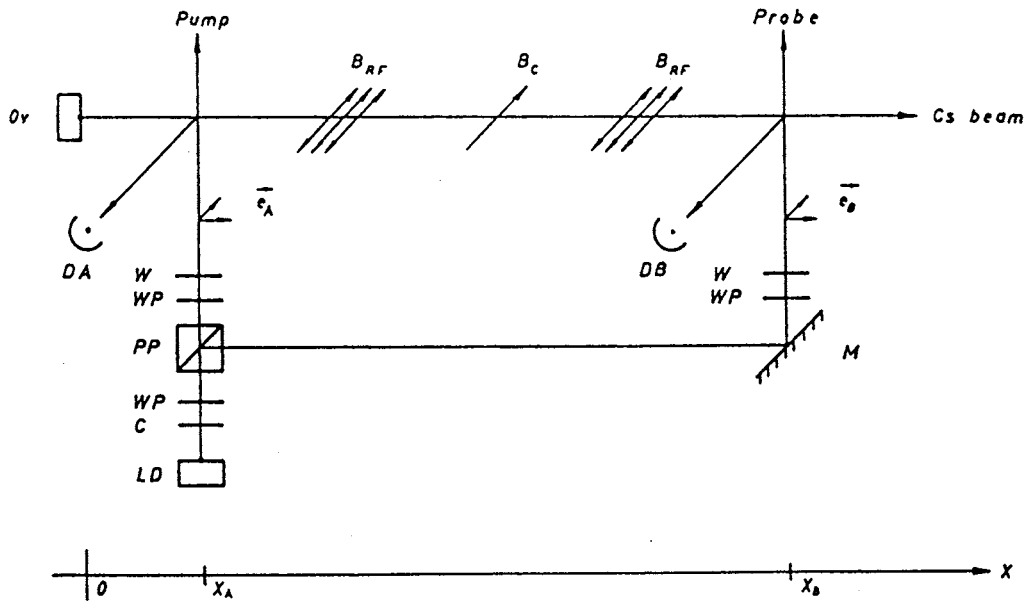
Conclusion

Preliminary measurements on a prototype of commercial-size optically pumped cesium beam tube show that the short-term stability, already equivalent to that of a conventional resonator of same size, may be improved further by a factor of ten.

From the practical point of view, optically pumped tubes have several advantages:

a) the ease of alignment should allow significant weight savings
b) the optical pumping technique opens up new possibilities in terms of miniaturization because the compromise between linewidth and beam flux does not impose the same condition on the tube length as in a magnetic deflection tube
c) the potentially least reliable element of the optically pumped tube - the laser diode - can be kept outside the sealed tube and easily replaced if necessary.
d) magnetic sensitivity should be inherently lower because of the absence of strong magnetic fields near the RF cavity.

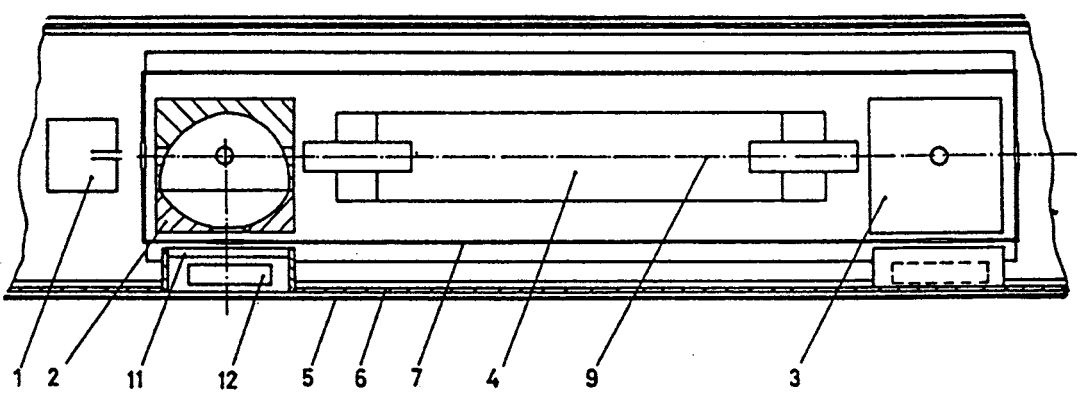
On the other hand, new problems associated with optical pumping (such as laser diode lifetime, light-shift, etc) need to be carefully evaluated.



Ov : Cs oven
Pump : pump laser beam (A)
B_{RF} : microwave interaction regions
B_C : C-field
Probe : probe laser beam (B)
DA, DB : photodetectors
$e_{A,B}$: possible polarization of laser beams
W : windows in vacuum envelope
WP : waveplate
PP : polarizing prism
M : mirror
C : collimating optics
LD : laser diode

F I G U R E 1

Schematic diagram of optically pumped Cs tube (one laser)



1. Oven
2. Collecting optics A
3. Collecting optics B
4. Ramsey cavity
5. Outer magnetic shield
6. Vacuum envelope
7. Inner magnetic shield
9. Atomic beam axis
11. Photodetector window
12. Photodetector

F I G U R E 2

Schematic drawing of the prototype, seen from above

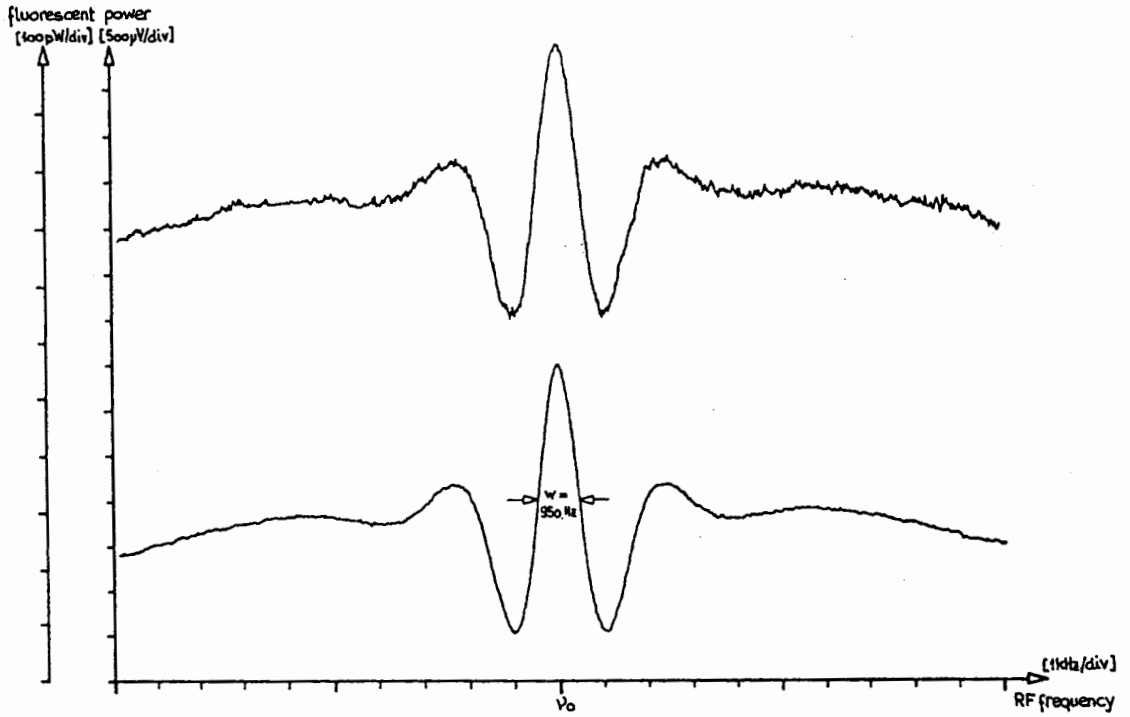


Fig. 3 : Ramsey pattern, 4 - 4 ☞ transition detector UV250B/10M ⏦
upper curve : laser free running
lower curve : laser frequency locked

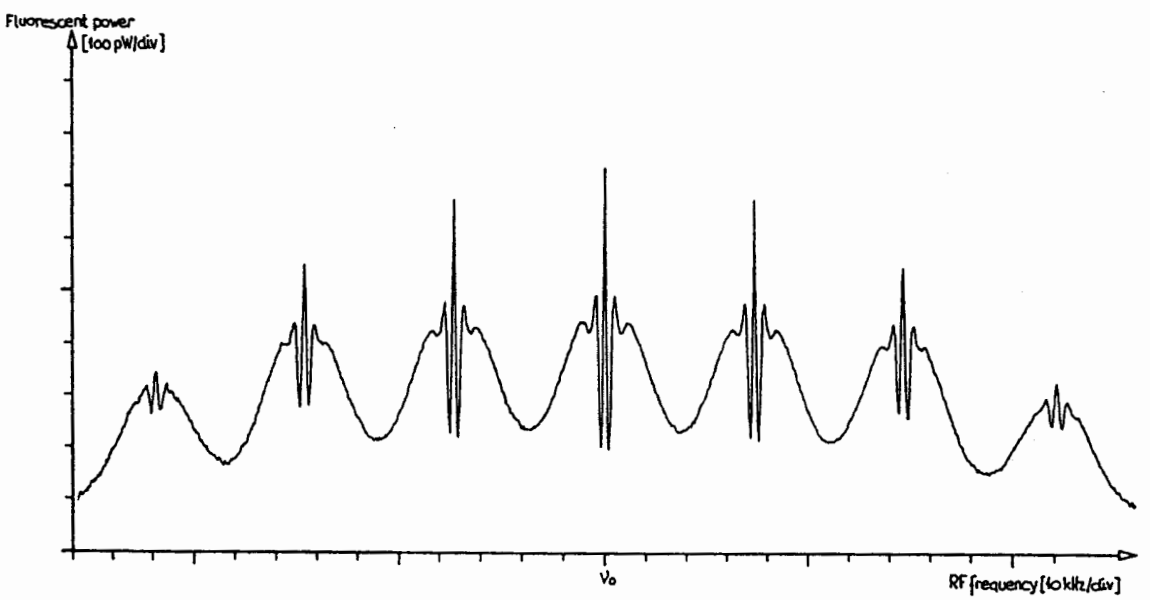


Fig. 4 : RF spectrum (4 - 4 ☞ pumping, détector UV250B/10M ⏦)

References

[1] V. Candelier, V. Giordano, A. Hamel, G. Theobald, P. Cerez, C. Audoin, these proceedings.
[2] J. Umezu, B. Komiyama, M. Saitho, Y. Ohta, these proceedings.
[3] SHANG Song-quan, WU Xin-Xin, YAO Shu-Tong, XIE Lin-Zhen, WANG Yi-qiu, Chinese Phys. Lett. 2, 12(1985) 557-560
[4] T.M. Clelland et al., Proc. of the 41st Annual frequency Control Symposium, 1987, p. 59
[5] G. Avila, V.Giordano, V. Candelier, E. de Clercq, G. Theobald, P. Cerez: Phays. Rev. A 36, 8(1987) 3719-28

Session AFS 03 Number 4

Experiments on optically pumped cesium beam frequency standard

Jun UMEZU, Bokuji KOMIYAMA, Haruo SAITHO, Yasusada OHTA

Radio Research Laboratory, Toky JAPAN

The improuvement on cesium beam frequency standard is anticipated by replacing the conventional magnetic state selection with optical state selection and detection. The RRL's experimental optically pumped cesium frequency standard is constructed to study a medium size potential performance with the frequency stabilized laser diode as a pumping source. The Ramsey cavity is E-bend shaped and it's interaction distance is 40cm long.

We report some experimental results related to the study and development of our new setup.

Session AFS-03 Number 5

A FM-stablized semiconductor laser for optical pumping of Cesium atoms and length measurements

H. R. Telle
Physikalisch-Technische Bundesanstalt
Lab. 4.42
Bundesallee 100
D-3300 Braunschweig, F. R. Germany

Abstract

A GaAlAs laser diode has been locked to different saturated hyperfine lines of the Cesium D2 transition at 852.1 nm by using a FM stabilisation scheme.
The influence of several operation parameters, e. g. modulation index, saturation intensity, cell length on the emission frequency has been studied.
In addition, high frequency FM noise components were reduced substantially by using a fast servo loop. The output spetrum was investigated both in the Fourier- and optical frequency regime.

1. Introduction

The poor short term frequency stability of solitary laser diodes is one of the limiting factors in many high resolution applications, like sub-doppler spectroscopy or optical preparation of atoms in atomic clocks.
On the other hand, light sources having a long term frequency stability of the order of 10 E-8, are required for many technical applications, e. g. interferometric length measurents.
In the following we will see, that both the long and short term FM noise of LDs can be improved by simple electronic frequency control to meet the requirements of most of these applications.

2. Long term frequency control

The set-up used for long term frequency stabilisation is shown schematically in Fig. 1. Doppler-free spectral substructures of the Cesium D2 transitions at 852 nm were induced by applying nearly counterpropagating waves to a cell containing Cs vapor at romm temperature.
In order to obtain error signals of odd symmetry, which are needed for proper operation of the electronic control loop, a small sinusoidal ac current was added to the dc injection current of the laser diode. The resulting FM width was kept much smaller than the modulation frequency of about 30 MHz, i. e. the FM-index was much less than 1. Narrow spectral features due to saturated absorption are expected at the center frequency of each hyperfine transitions and at the frequency positions of the crossover lines.

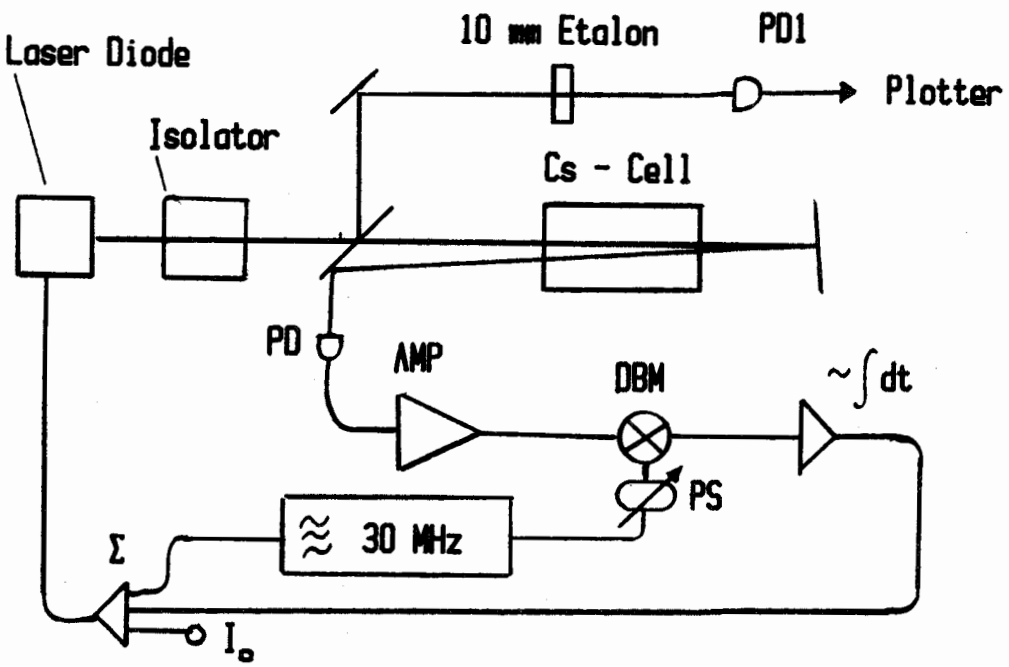


Fig. 1 : Long term stabilisation set-up

These signals were homodyne-detected by using a fast photodiode PD, a rf preamplifier AMP, a double-balanced mixer DBM and a variable rf phase shifter PS. The loop was closed by feeding the DBM output signal back to the laser diode. A typical open loop frequency scan for transitions starting from the F=4 ground state is shown in Fig. 2. The lower curve corresponds to the dc output signal of PD, showing the dips due to saturated absorption on a doppler-broadened background. The upper curve was taken simultaneously at the DBM output. The PS phase was set to quadrature (with respect to the applied FM) during this scan in order to detect the induced dispersion steps. Sharp dispersion profiles are found at the hyperfine transition frequencies and at the correponding crossover positions which provide nearly ideal error signals for the servo loop.

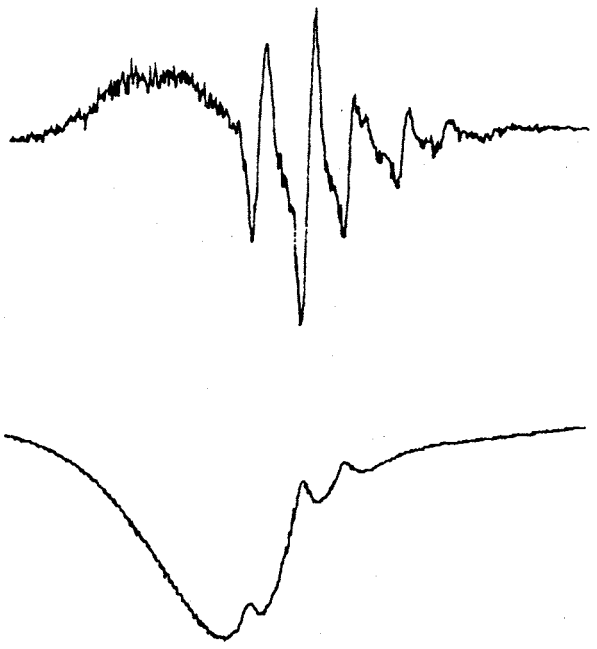


Fig. 2: Cs-D2 (F=4)-spectral scan,
FM-signal (upper curve), DC-signal (lower curve)

Since the spectral widths of the steep dispersion steps seen in Fig. 2 is only about 10E-7 of the optical frequency, one expects a emission frequency under closed loop conditions, which is only weakly affected by changes of the operation conditions. In order to investigate this point in more detail, a temperature stabilized solid state etalon, seen in the upper part of Fig. 1, was used as a slope discriminator to monitor the frequency changes resulting from well defined variations of the operation parameters, like saturation

intensity, cell temperature etc. The results are summarized in Tab. 1. One clearly sees, that even unrealistic strong parameter varations result in relative frequency offsets of much less than 1E-7. From these values one might expect a frequency reproduceability of the order of 1E-8 for a well designed system which is sufficient for most practical applications like length measurements.

Initial conditions:

Cesium cell: L = 5 cm , T = 18 °C
Cs transition: $6^2P_{3/2}$ (F=4) - $6^2S_{1/2}$ (F=4)
Modulationfrequency: 25.2 MHz
Modulation index: 0.2
Saturation beam: r = 2 mm, P = 6 mW, I=50 mW/cm²
Long. magnetic field: zero

Parameter dependence of laser frequency:		Δf/f
Saturation intensity:	50 mW/cm² to 10 mW/cm²	$1.2 \cdot 10^{-8}$
Mod. frequency:	25.2 MHz to 25.8 MHz	$<.5 \cdot 10^{-8}$
Mod. index:	0.2 to 0.1	$<.5 \cdot 10^{-8}$
Cell temperature:	18 °C to 25 °C	$4.0 \cdot 10^{-8}$
Cell length:	5 cm to 2 cm	$2.0 \cdot 10^{-8}$
Laser current:	I_0 to I_0 +.5 mA	$<.5 \cdot 10^{-8}$
optical signal att.:	1 to .2	$1.0 \cdot 10^{-8}$
Long. magnetic field:	0 to 100 G	$1.9 \cdot 10^{-8}$

Tab. 1 : Parameter dependence of laser frequency

3. Fast frequency control

The investigations described above have been carried out using an electronic control bandwidth of about 10 kHz. Therefore only the very slow frequency fluctuations are reduced leading to an almost unchanged emission spectrum. It can be shown that loop bandwidths of at least 10 MHz are required for substantial linewidth reduction.

Attemps to increase the bandwidth of the servo loop considered above failed, although the discriminator linewidths of about 700 MHz for the Doppler line and about 30 MHz for the saturation dips should allow much wider bandwidths. The reason for the observed instabilities is the long ground state population relaxation time which is actually given by the atomic transit time across the laser beam. The resulting additional phase lag of the error signal may exceed the servo loop phase margin at high Fourier frequencies which results in instabilities. In fact, loop oscillations at about 50 kHz were observed under this conditions.
To overcome this problem, the slope of a Fabry-Perot interferometer (temperature stabilized solid state etalon) was used as a fast frequency discriminator. A reference path was added to eliminate AM noise and to stabilize the frequency at the steepest slope of the FPI transmission curve, see Fig. 3. The servo amplifier consists of a fast transimpedance amplifier for the Fourier frequency band from 1 to 35 MHz while a low noise op-amp was used to suppress slow FM noise components.

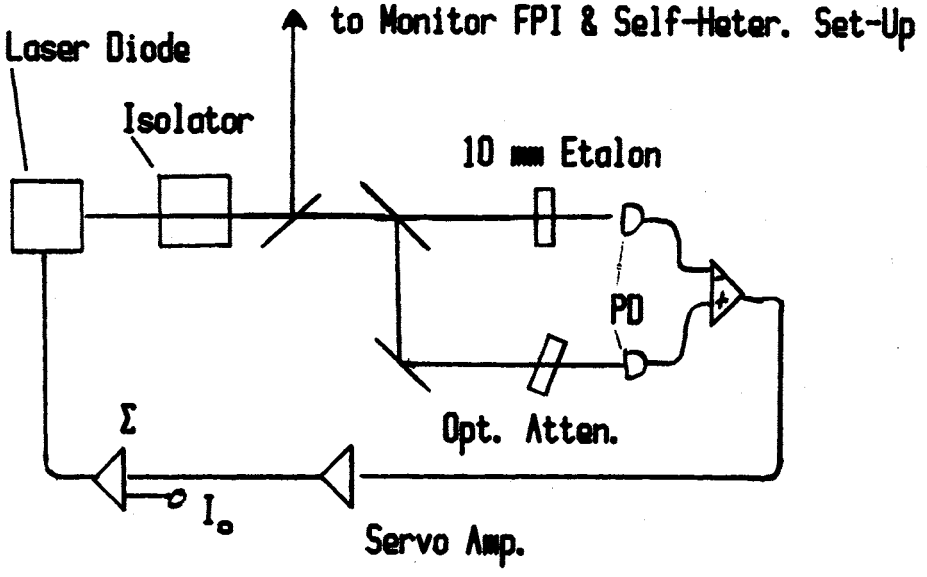


Fig. 3 : Fast servo loop

The emission characteristics of the solitary and stabilized laser where investigated both in the in the Fourier- and optical frequency regime.
Measurements of the FM noise power were carried out with the help of a two monitor FPIs having transmission linewidths of 3 GHz and 400 MHz, respectively, and a rf spectrum analyzer.

The measured FM noise power spectral density of the free running laser shows the well known behavior: white frequency noise for Fourier frequencies above 10 MHz and an additional 1/f component below, see Fig. 4.
Under closed loop conditions, one clearly sees substantial noise suppression for Fourier frequencies below 35 MHz (Fig. 4). For frequencies below 1 kHz, the noise reduction is shot noise limited.

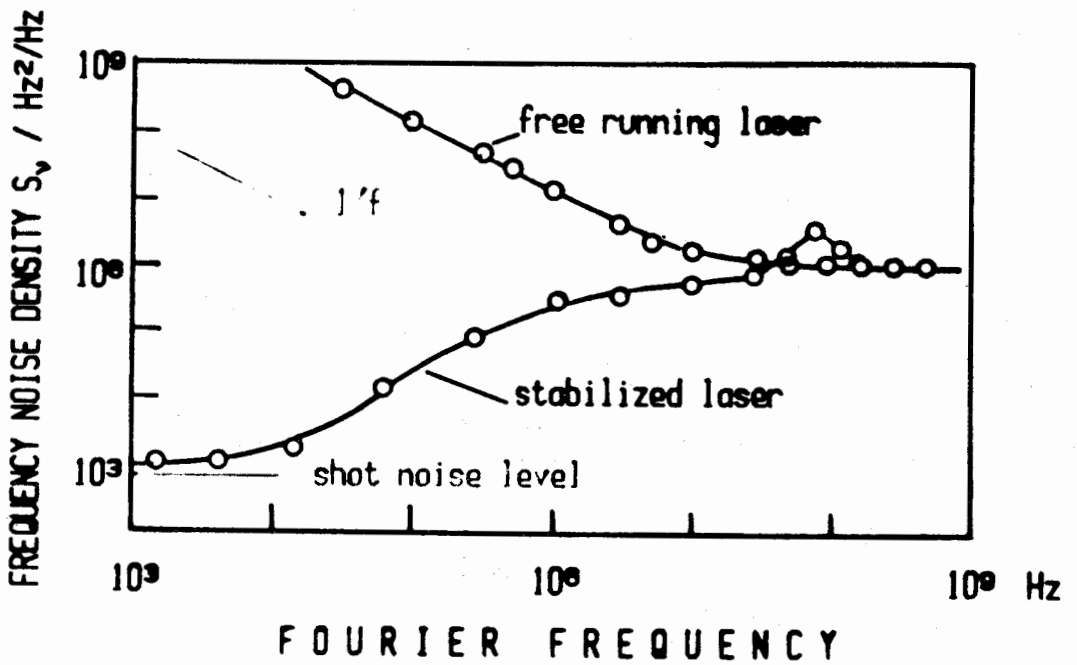


Fig. 4: Measured FM noise spectrum

The optical output spectrum of the laser was investigated by delayed self-heterodyning. Using 200 m of single mode fiber as the optical delay line, we obtained a self heterodyning beat note width of 27.5 MHz (FWHM) for a free running laser and 5 MHz for stabilized system, see Fig. 5. Assuming a gaussian lineshape for the free running laser and a Lorentzian under locked conditions, we estimate linewidths of 20 MHz for the free running system and ~ 2.5 MHz under servo control. It should be noted, that the ~ 1 µsec delay time provided by the fiber is sufficient to measure linewidths of ≥ 1 MHz (FWHM) directly, i. e. without correcting for the finite delay time.

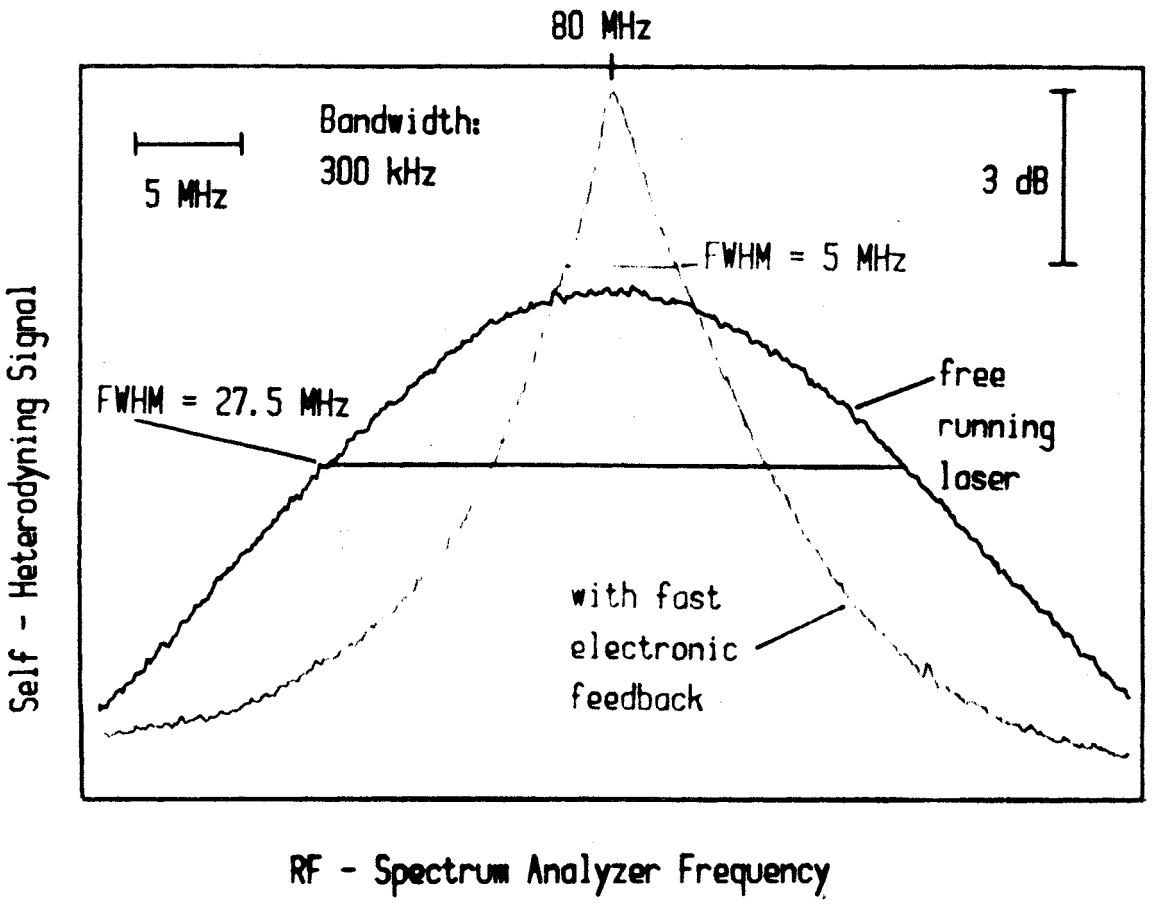


Fig. 5: Self-heterodyning spectrum

4. Conclusion

Both long and short term frequency stabilisation of laser diodes can be accomplished by using electronic servo control. A long term frequency stability of the order of 10E-8 can be expected for a well designed system using saturated atomic absorption as the frequency reference while emission linewidths of the order of 1 MHz can be easily obtained with sufficiently fast discriminators. Linewidths of 100 kHz or less should be possible in the latter case with optimised servo loop transfer function. Both methods can be combined, for example by controlling the temperature of the reference etalon by the homodyne detector output signal.

2 ND EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 16 - 17 - 18 MARCH 1988

Session AFS-03 Number 6

THE EFFECT OF LASER LINE-NARROWING ON THE PERFORMANCE OF OPTICALLY PUMPED CESIUM ATOMIC BEAM FREQUENCY STANDARDS

S. Ohshima, Y. Koga and Y. Nakadan
National Research Laboratory of Metrology
Tsukuba-shi, Ibaraki-ken 305, Japan
and
L. Hollberg and R. Drullinger
National Bureau of Standards
Boulder, Colorado USA

ABSTRACT

An experiment has been conducted to investigate the effect of laser line-narrowing on the performance of an optically pumped cesium beam frequency standard. A two laser scheme ($4 \rightarrow 4' \sigma$ for optical pumping and $4 \rightarrow 5' \pi$ for detection) was used and the lasers were either frequency locked to Cs saturated absorption features in an external cell or line-narrowed by optical feedback from an external Fabry-Perot cavity [1]. Relative to the frequency stabilization to the saturated absorption the line-narrowing on the detection laser produced a large improvement (+14dB) in the signal-to-noise ratio (S/N). However, line-narrowing on both pump and detector lasers was necessary to achieve beam shot noise limited performance. The S/N and line Q achieved in these experiments should result in a clock stability of $\sigma_y(\tau)=3\text{-}5\times10^{-13}\tau^{-1/2}$ [2].

INTRODUCTION

The use of diode lasers for optical state preparation and detection in atomic beam frequency standards holds great promise for improvement in both the short term and long term performance of these standards. The potential for improved short term performance lies in the improved utilization of the atomic beam flux. A number of laboratories have begun experiments on optically pumped standards and the improved signal levels have been demonstrated. However, the potential signal-to-noise ratio (S/N) available has not previously been achieved when semiconductor lasers have been used as the optical source. This seems to be due to residual FM noise in the laser sources

even when they are frequency stabilized to either atomic reference signals or optical cavities with servo systems of moderate speed. We have used an optical feedback scheme with an equivalent servo bandwidth of about a 200 MHz to reduce the laser operating linewidth. This in turn reduces the FM noise and the conversion of FM noise to AM noise by the Cs atoms.

EXPERIMENT

The experiments were conducted on the large optically pumped standard at the NRLM [3]. We used a two laser pumping scheme. The state preparation used the $4 \rightarrow 4'\sigma$ transition while the detection used the $4 \rightarrow 5'\pi$ transition. The optical power densities were 10's of mW/cm_2.

Two different laser frequency stabilization schemes were employed. A small portion of each lasers output was used to produce a saturated absorption spectrum in a room temperature cesium cell. The lasers had a small FM modulation at 2KHz and the saturated absorption signal was synchronously detected to produce a derivative type error signal for frequency reference. In the first lock scheme, this signal was used directly to control the laser center frequency by electronic feedback to the injection current. In the second case, the laser output linewidth was reduced by an optical feedback technique [1,4]. In this case, the saturated absorption signal was used to provide the long term frequency reference for the optical cavity. The laser output spectrum was not measured in these experiments but linewidths of much less than a MHz are to be expected [4].

RESULTS AND CONCLUSIONS

The results are given in table 1. The experiments labeled 1,2 and 3 are: 1) with broadband lasers locked to the saturated absorption, 2) with only the detector laser narrowed and 3) with both lasers narrowed. The measurements were all made in a static way with no modulation on the microwave radiation. The signal levels are given in arbitrary units and represent the difference between the detected signal level with the microwave radiation tuned to the center of the Ramsey resonance and well outside the hyperfine spectrum. The noise is measured at the peak of the Ramsey resonance in a 1Hz bandwidth centered at 75Hz.

Table 1

Experiment	1	2	3
Signal level	1	1.53	1.40
S/N	450	2300	6500

The increased signal produced by narrowing the detection laser was the result of increased power spectral density on the cycling transition while the increased S/N was the result of the reduced FM noise on the laser. The slight loss of signal when the pump laser was narrowed may have resulted from an offset in the cavity frequency reference servo and a corresponding reduction in the pumping efficiency. In-spite-of this error, the detected S/N was further improved. The interpretation of this would seem to be that in the absence of line-narrowing, some laser frequency excursions are large enough and last long enough that some atoms pass the pumping region

without interacting with the laser. This flux of unpumped atoms, while small, can contribute a detectable background signal which contains the added noise. When the pump laser was line-narrowed, even though it was locked slightly off resonance and some flux of atoms passed unpumped, the background they produced at the detector contained no added noise. The total background level which was a composite of scattered lights, fluorescence from background cesium atoms and fluorescence from unpumped atoms in the beam was equal in magnitude to the Ramsey signal.

In the case with both lasers line-narrowed, the S/N ratio on the detected signal was essentially beam shot noise limited. When combined with the Ramsey resonance line Q we observed ($1x10^8$) and a standard clock servo analysis [4], this S/N should result in a clock stability of $\sigma_y(\tau) = 3\text{-}5 \times 10^{-13}\ \tau^{-1/2}$ depending on the type of modulation scheme chosen. This demonstrates that the full potential for short term statistics offered by optically pumped cesium is available if lasers with proper noise performance are used.

REFERENCES

1. B. Dahmani, L. Hollberg and R. Drullinger, Optics Letters 12, 875 (1987).
2. A. De Marchi, G.D. Rovera and A. Premoli, IEEE Trans. Ultra son. Ferro Elec. Freq. Contr. 34 582 (1987).
3. S. Ohshima, Y. Nakadan and Y. Koga, "The Optically Pumped Cs Frequency Standard at the NRLM" proceedings of the 1981 PTTI meeting.
4. A. Clairon, B. Dahmani, Ph. Laurent and Ch. Breant "Ultra-Narrow Semiconductor laser linewidth for Optically Pumped Cesium Clock," in these proceedings.

AFS-04

Atomic Frequency Standards

1. Ultra-Narrow Semiconductor Laser Linewidth for Optically Pumped Cesium Clock
   A. Clairon, B. Dahmani
   Laboratoire Primaire du Temps et des Fréquences, Paris - France
   Ph. Laurent, Ch. Breant
   Laboratoire Lasers Ultra-Stables, E.T.C.A., Arcueil - France

2. Extended-Cavity Semiconductor Lasers for Optical Pumping
   M. de Labachelerie, K. Diomande, N. Dimarcq
   Laboratoire de l'Horloge Atomique, Orsay - France

3. Frequency-Stabilized Laser for Atomic Frequency Standards Using Saturated
   J. Deng, Y. Tan, Y. Xiang, X. Zhu
   Wuhan Institute of Physics, Academia Sinica, China

4. Phase-lock design of a 600 GHz BWO for the Mg frequency standard
   E. Bava, A. Godone, C. Novero
   Istituto Elettrotecnico Nazionale Galileo Ferraris, Torino - Italy

2 ND EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 16 - 17 - 18 MARCH 1988

Session AFS-04 Number 1

ULTRA-NARROW SEMICONDUCTOR LASER LINEWIDTH FOR OPTICALLY PUMPED CESIUM CLOCK

A. CLAIRON, B. DAHMANI
Laboratoire Primaire du Temps et des Fréquences, Observatoire de PARIS
61, avenue de l'Observatoire 75014 PARIS
Ph. LAURENT and Ch. BREANT
Laboratoire Lasers Ultra-Stables, E.T.C.A.
16 bis, avenue Prieur de la Cote d'Or 94114 ARCUEIL Cédex

Abstract: Semiconductor laser frequency noise will shortly be a limiting factor in the performances of optically pumped Cesium beam frequency standard. Here we report our recent progress obtained with an optical self-locking technique to stabilize the frequency of laser diodes operating near 850 nm. The semiconductor laser beam is fed through an external Fabry-Perot cavity and the reflected beam is weakly coupled into the laser to achieve optical feedback stabilization. With this method the laser frequency is locked to the cavity resonance and the linewidth is dramatically reduced by a factor 5000 from 20 MHz to less than 4 kHz. The spectral characteristics of the frequency noise and the beat note spectra of two such stabilized lasers are analyzed both theoretically and experimentally.

INTRODUCTION

The development of monomode semiconductor lasers emitting in the near infra-red at 852 nm quickly opened the field of optically pumped Cesium clock [0]. The most efficient preparation and detection schemes for an optically pumped Cesium beam clock are limited by FM noise. The poor spectral purity and the high frequency noise of diode lasers, mainly due to the low Q cavity and to the strong coupling between phase and amplitude noise, requires special techniques for frequency stabilization. Most techniques use some kind of optical feedback to achieve a short enough response time.

DETECTION SCHEMES

Various optical pumping and detection schemes have been proposed for the Cesium clock. The simplest one uses only one laser for pumping and detection. Since the pumping time is relatively short (3 µs) compared to the interaction duration (30 µs), the number of pumped atoms is not very sensitive to the diode laser frequency noise. Nevertheless the signal to noise ratio of the clock signal is usually limited by detector noise. As shown on fig. 1 to increase the signal to noise ratio and brings it close to the shot noise limit of the Cesium beam it is necessary to use more sophisticated pumping and detection schemes.

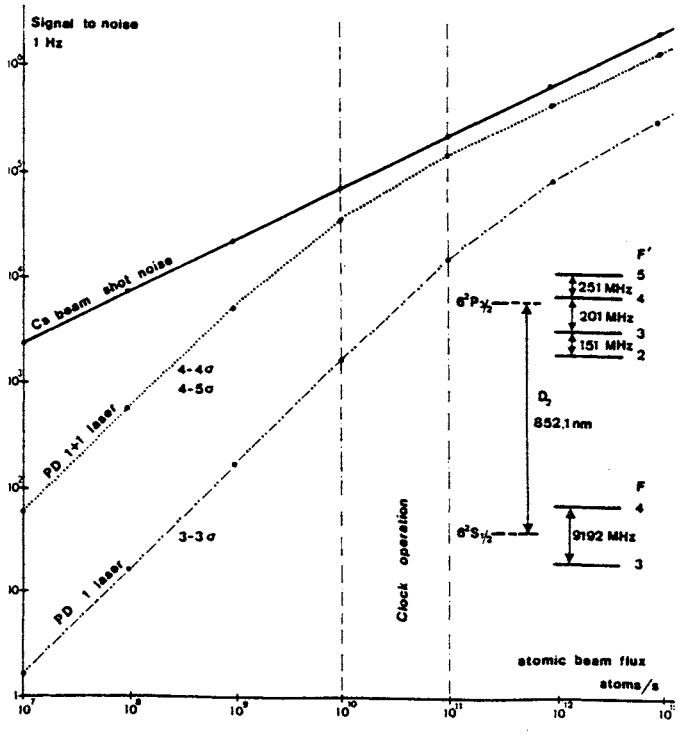


Fig. 1 : Theoritical signal to noise ratio of the detection signal as a function of the atomic beam flux for various detection and pumping schemes.

For instance a two lasers scheme which uses the 4-4 σ transition for preparation and the 4-5 σ transition for detection can be used to approach shot noise limited operation. The 4-5 σ is a "cycling" transition and the signal to noise obtained is therefore very sensitive to the diode laser frequency noise. Under normal conditions the linewidth of the diode laser is much wider (20-25 MHz) than the natural linewidth (5 MHz) of the Cesium D_2 line. Conventional electronic stabilization techniques are not fast enough to improve significantly the linewidth of the diode laser and the signal to noise ratio of the fluorescence signal.

TECHNIQUES FOR FM NOISE REDUCTION

The experimental set-up used for frequency stabilization of the diode lasers is shown on fig. 2, it is similar to the one of Dahmani and Hollberg [2] and uses the transmission like fringe of a tilted Confocal Fabry-Perot (CFP) interferometer to achieve optical feedback.

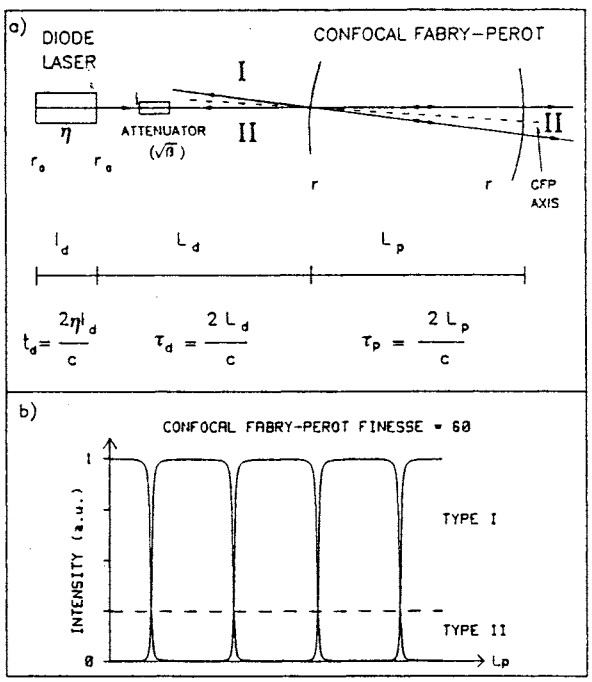


Fig. 2 : Schematic of the optical feedback locking system :

- a) represents a diagram of the optically self-locked laser coupled to a tilted C.F.P.
- b) shows the transmission (II) and reflection (I) fringes obtained with a tilted C.F.P.

For low feedback level the eigen frequency ω of the coupled cavity laser can be written as

$$\omega_N = \omega + K \frac{\sin[\omega(\tau_d + \tau_p) + \theta] - r^4 \sin[\omega(\tau_d - \tau_p) + \theta]}{1 + F^2 \sin^2 \omega \tau_p} \qquad (I)$$

with

ω_N : the free running diode laser frequency

ω : laser frequency with feedback

$$K = (1 + \alpha^2)^{1/2} \sqrt{\beta} \frac{c}{2 \eta l_d} \frac{1}{2} \frac{\mathcal{F}_{cfp}}{\mathcal{F}_d}$$

α : the phase-amplitude coupling factor [3]

$\sqrt{\beta}$: the amplitude coupling factor CFP-diode laser

η : the diode laser refraction index

l_d : the diode laser length

r_0 : the amplitude diode mirror reflectivity

r : the amplitude CFP mirror reflectivity

$\mathcal{F}_{CFP}$: the CFP finesse

$\mathcal{F}_d$: the diode laser finesse

τ_P : the round-trip time in the CFP

τ_d : the round-trip time between the CFP and the diode laser

$\theta = tg^{-1}\alpha$

and

$$F = \frac{2r^2}{(1 - r^4)}$$

The graphic representation of expression (I) is shown on fig. 3. The stable solutions are those with a positive slope. For a given $\Delta\omega_N$ frequency excursion of the free-running diode the frequency excursion of the coupled cavity laser is reduced by the slope factor P defined by :

$$P = \frac{d\omega_N}{d\omega} = 1 + 2K\tau_p$$

$$= 1 + (1+\alpha^2)^{1/2} \sqrt{\beta} \frac{L_p}{\eta l_i} \frac{\mathcal{F}_{cfp}}{\mathcal{F}_d}$$

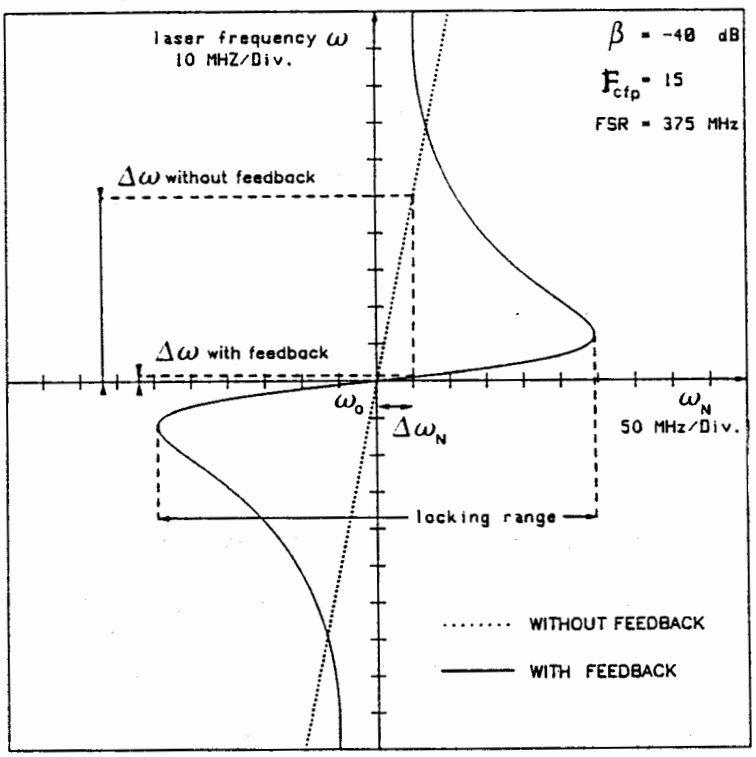


Fig. 3 : Theoritical representation of the laser frequency behaviour ω around frequency ω_0 as a function of the unperturbed laser frequency ω_N. The dotted line represents the free-running diode laser ($\omega = \omega_N$) and the solid line the feedback coupled cavity laser. This figure illustrates the dramatic frequency excursion reduction from $\Delta\omega$ without feedback to $\Delta\omega$ with feedback, for a given $\Delta\omega_N$. This reduction is given by the slope P and the locking range of the coupled cavity laser by the local extrema of the function $\omega(\omega_N)$.

For a high value of the product $\sqrt{\beta}\ L_p\ \mathcal{F}_{CFP}$, the frequency of the coupled cavity laser is completely determined by the resonance frequency of the CFP cavity. For instance with the experimentalparameters $\mathcal{F}_d \equiv 1$, $\eta l_d \equiv 10^{-3}$, $\mathcal{F}_{CFP} \equiv 100$, $L_p \equiv 0{,}2$ m (the CFP length) and $\sqrt{\beta} \equiv 10^{-2}$ the P value is $\equiv 500$.

The theoritical analysis of a coupled cavity laser [4] shows a relative static reduction of frequency noise spectral density between the locked state $S^L_{\dot{\phi}}(0)$ and the free-running state $S^F_{\dot{\phi}}(0)$ given by

$$\frac{S^L_{\dot{\phi}}(0)}{S^F_{\dot{\phi}}(0)} = \frac{1}{(1+2K\tau_p)^2}$$

A full dynamic analysis of the frequency noise reduction has been calculated and is illustrated on fig. 4. With high finesse cavity and high feedback level frequency noise reduction of 50 to 60 dB can be achieved and should lead, if spontaneous emission is the main noise source, to a lorentzian lineshape with a linewidth below 100 Hz given by:

$$\text{HWHM} = \pi\ S^L_{\dot{\phi}}(0) = \frac{\Delta\nu_0}{(1+2K\tau_p)^2}$$

where $\Delta\nu_0$ is the linewidth of the free-running diode laser.

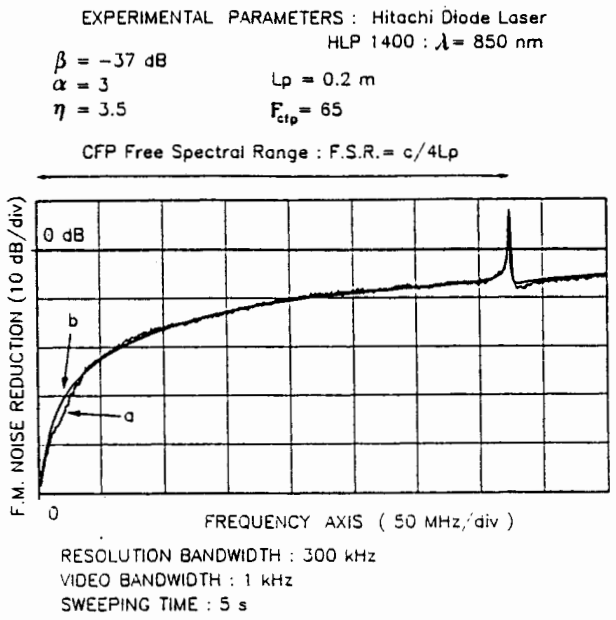


Fig. 4 : Frequency noise reduction of a diode laser coupled to a CFP . Line a) shows the experimental data and line b) represents the theoritical prediction calculated from the experimental parameters. The static noise reduction is of the order of 46 dB in this case. In this particular case we observe an increase of the frequency noise reduction at 375 MHz which corresponds to the next peak of the CFP. But depending on the ratio L_d/L_p, this resonance could be either a decrease or an increase of the noise.

For different feedback levels, we recorded and analyzed the RF power spectra of the beat note between two such coupled cavity lasers and found no lorentzian lineshapes. A careful noise analysis shows that at high feedback level the frequency noise spectrum is mainly due to Flicker frequency noise ($S^L_{\dot{\phi}}(f)=1/f$) and leads to a linewidth which varies as ($S^L_{\dot{\phi}}(1))^{1/2}$. Hence the linewidth is much wider than expected, of the order of 4 kHz for 50-60 dB noise reduction as illustrated on fig. 5.

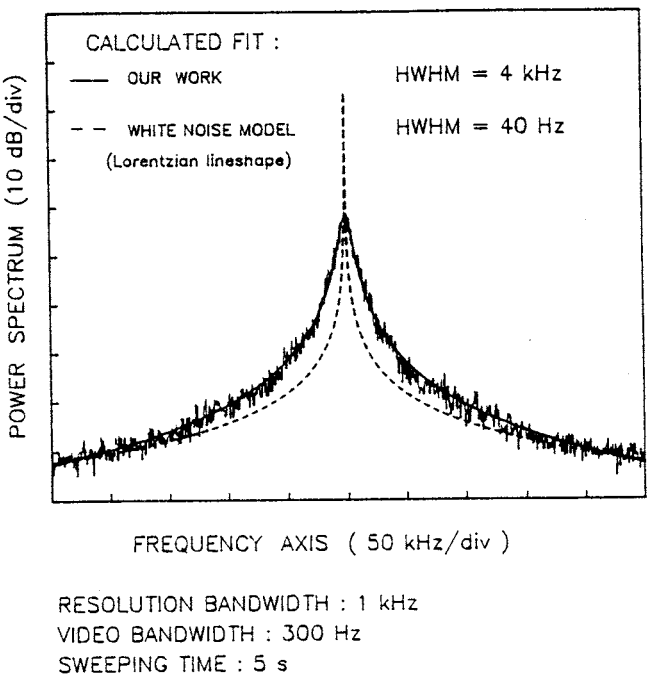


Fig. 5 : Beat-note spectrum of two independently locked diode lasers with frequency noise reduction of 60 dB for diode 1 and 55 dB for diode 2. The dashed line shows the calculated spectrum if we assumed a white frequency noise model. The solid line represents our best fitted spectrum (lorentzian lineshape to the power 3/2) and can be interpretated as 1/f frequency noise.

Another possibility to improve the linewidth of the diode laser is to use an extended cavity laser [5] (also presented at this forum).

To study the frequency stability of the coupled cavity lasers we have measured the Allan variance (fig.6) and found constant levels around 10 kHz (3 10^{-11}) in agreement with the previous frequency noise levels.

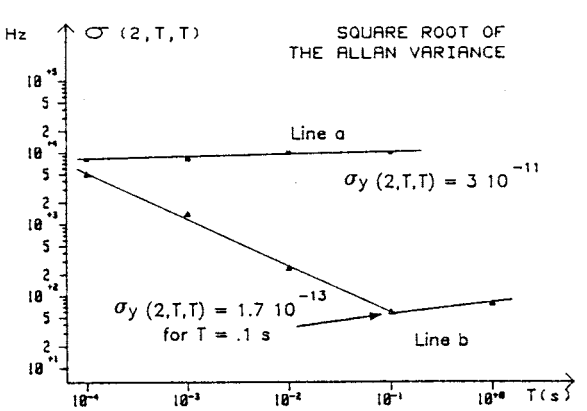


Fig. 6 : Line a represents the frequency fluctuations of two independently optically coupled diode lasers cavities and line b shows the Allan variance when one of the lasers is frequency offset-locked to the other.

Finally it is interesting to note that the coupled cavity lasers can easily be frequency locked to the Cesium beam fluorescence by traditionnal techniques. It is important to notice that for the linewidth and frequency stability now obtained the laser noise contribution to the clock signal detection noise should be completely negligible.

This research is supported by Bureau Nation al de Metrologie and Direction des Recherches et Etudes Techniques.

references

[0] M. Arditi and J.L. Piqué, J. Phys. Letters 41,379 (1980)

[1] V. Candelier, J. Giordano, A. Hamel, G. Theobald, P. Cerez and C. Audoin, this issue and referencies therein.

[2] B. Dahmani, L. Hollberg and R. Drullinger," Frequency stabilization of semiconductor lasers by resonant optical feedback", Optics Letters vol. 12, n° 11, pp 876-878 , 1987

[3] C.H. Henry, "Theory of the linewidth of semiconductor lasers",IEEE J. Quantum Electro., QE-18, n°2, pp 259-264, 1982

[4] Ph. Laurent, A. Clairon and Ch. Bréant to be published

[5] M. De La Bachelerie, K. Diomande and N. Dimarcq, " Extended cavity semiconductor lasers for optical pumping", this issue and referencies therein.

2 ND EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 16 - 17 - 18 MARCH 1988

Session AFS-04 Number 2

EXTENDED-CAVITY SEMICONDUCTOR LASERS FOR OPTICAL PUMPING .

M. de LABACHELERIE , K . DIOMANDE , N. DIMARCQ

Laboratoire de l'Horloge Atomique
Equipe de recherche du CNRS associée à l'université Paris-Sud
Bat. 221 - Université Paris-Sud
91405 ORSAY - France .

Abstract :

As they are compact , costless , and wavelength tunable , laser diodes are very interesting coherent optical sources for applications involving light-atom interactions , such as optical pumping for Cesium beam frequency standards . However , their tunability is not so easy , and their linewidth is often too large to be optimally used for these applications . Using anti-reflection coated laser diodes , we have studied extended-cavity semiconductor lasers which are more suitable for optical pumping . In this paper , we describe the main features of these lasers and show that they improve considerably the tunability and linewidth of laser diodes .

To confirm their suitability , we have shown experimentally that using them leads to a great frequency stability improvement for optically pumped Cesium frequency standards and long-term stabilized lasers .

1 . Introduction :

The achievement of industrial applications using light-atom resonant interactions (pollutants monitoring , optically pumped atomic clocks ...) is bounded to the existence of a compact , easy to operate , tunable laser source . Up to now , only semiconductor lasers seem to be suitable for such applications : they have all the qualities for industrial requirements (compactness , reliability , efficiency ...) and may be spectrally tuned to reach atomic resonant frequencies .

However , they have still a few drawbacks that may limit their usefulness for light-atom interactions :

— They are usually tuned by varying the diode temperature and current, which alters the semiconductor refractive index , but the resulting frequency variation is a continuous sweep over a limited range followed by abrupt frequency jumps to other cavity modes generally located at unpredictible places . This behaviour is probably related to the absence of a strong intra-cavity spectral selection : the diode is then very sensitive to slight defects of the diode which impose the lasing frequency in a somewhat random way .Tuning a laser diode to a predefined frequency is thus rather difficult and . even if the original lasing frequency is very close

to the desired frequency , it is often impossible to attain because the frequency will always jump to other modes before reaching it .
For a set of ten diodes selected at ±3 nm of the desired wavelength , we usually consider that only three of them will reach the fixed atom resonant frequency if their temperature and current is varied . Moreover , we have noticed that the diode behaviour changes on long time scales : we observe a drift of the temperature and current required to obtain a given frequency and sometimes , the desired frequency becomes impossible to attain after aging . It is then absolutely necessary to improve the frequency control of semiconductor lasers for all the applications where the lasing frequency value is important , mainly for light-atom resonant interactions .
— Another feature which is often undesirable for light-atom interactions is the very large linewidth of laser diodes . It is well established [1] , that the linewidth of single-frequency laser diodes is generally greater than10 MHz which is much broader than any other single-frequency laser linewidth and may be on the same order of magnitude than Doppler-free atomic lines . This large linewidth , mainly related to the low cavity length , leads to a broadening of spectral atomic lines and a large noise of detected signals due to the diode frequency noise . For instance , this frequency noise is a major noise source in optically pumped Cesium beam atomic clocks and long-term stabilized lasers .
— Laser diodes are also very sensitive to external parasitic reflections of the laser beam into the diode cavity which perturb the lasing frequency and usually enhance frequency noise . It is thus important to reduce this high sensitivity to "optical feedback" , to limit the frequency noise for optical pumping applications .

To overcome these drawbacks , we have studied extended-cavity semiconductor lasers (ECL) which are designed to improve the spectral characteristics of laser diodes . After a description of these lasers , we will show that using an ECL instead of an ordinary laser diode actually permits to get better results in two applications involving light-atom interaction : optical detection in Cesium beam atomic clocks , and long-term stabilized semiconductor lasers .

2 . Extended cavity lasers :

The linewidth of semiconductor lasers , mainly limited by spontaneous emission phase noise , is inversely proportionnal to the square of the laser optical cavity length L . It is thus highly desirable to increase this length in order to obtain a linewidth reduction that attains 10^4 for a 10 cm long optical cavity. However , increasing the length of the semiconductor crystal is not possible : internal distributed losses would limit drastically the linewidth reduction and would lead to a very poor laser differential efficiency. It is then necessary to extend the laser diode cavity by a long passive section that can be obtained by suppressing one of the diode cleaved mirrors with an anti-reflection coating , and replacing it by an

external mirror located far from the diode .

In our experiments , the light emerging from the diode AR-coated facet was collimated by a large aperture microscope objective (0.6 numerical aperture) , and directed to a Littrow-mounted diffraction grating which ends the extended cavity (fig. 1). This grating acts as a wavelength selective mirror which allows the laser to oscillate only in the very narrow spectral band reflected exactly on the incident direction .Moreover, a broad tuning of the lasing wavelength can be obtained when the grating is rotated about an axis parallel to its grooves .

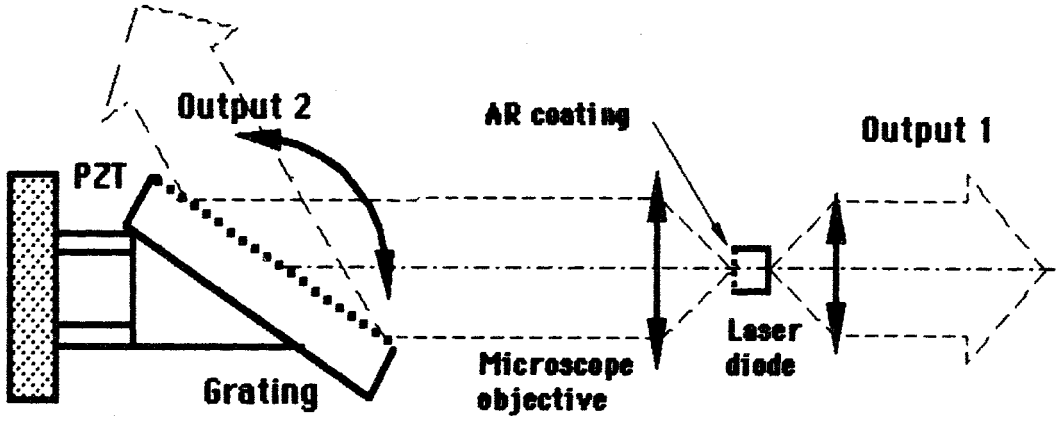


Fig. 1 : Typical extended-cavity laser

Such extended-cavity lasers have been studied with either GaInAsP diodes (1.5 µm lasing wavelength) or GaAlAs diodes (0.85 µm wavelength) which had low residual reflectivities on the AR-coated facet ($R_{residual} < 1\ \%$). Their main features will be described below :

Tunability :

Thanks to the grating , the cavity losses are high , except for a narrow spectral band ; the lasing wavelength is thus imposed by the grating and oscillation will occur provided there is enough gain at the selected wavelength . The grating rotation thus controls completely the lasing wavelength and allows to get any wavelength in the diode gain spectral band . With AlGaAs lasers , we have obtained a tuning range larger than 30 nanometers around 850 nm which is consistent with simple gain bandwidth estimations . The theoretical gain frequency bandwith $\Delta\nu_G$ is almost identical for InGaAsP lasers , however , the corresponding gain wavelength bandwidth $\Delta\lambda_G$ is proportionnal to the square of the wavelength . This leads to a 100 nm tuning range around 1.5 µm which has been well verified experimentally .

This large tuning range is very attractive for light-atom interactions because any roughly selected diode can be used to get the desired

wavelength and moreover , the laser frequency being fixed by the external grating , it is no longer subject to mode hops after aging of the semiconductor material .

Single frequency operation :

A simple estimation of the grating spectral selectivity shows that the wavelength band selected by the grating is generally not narrow enough to ensure single frequency operation of the laser . Even with very dispersive gratings and short extended cavities , there are always several longitudinal cavity modes within the selected wavelength band . However it does not necessarily mean that the ECL will be multimode because ordinary laser diodes actually does not need any spectrally selective element to be single frequency ; it is thus difficult to predict the behaviour of the ECL and we must limit our analysis to experimental observations .

Experimentally we have found that the behaviour of the CSP-AlGaAs lasers we used was very different from that of the BH-GaInAsP lasers . While the 1.5 µm lasers were always single frequency in an extended-cavity configuration , we observed frequently a multi-longitudinal mode operation of the 0.85 µm lasers in the same conditions .
When the laser had many longitudinal modes it was still possible to insert a suitable intra-cavity Fabry Perot selective element to select only one cavity mode . This technique was very efficient to obtain single-frequency operation , however , the laser frequency tuning became difficult and we generally preferred to avoid this solution .

Fine frequency tuning :

The grating rotation is only used to select one extended-cavity mode but cannot change the mode frequency which depends only on the cavity length . When a precise frequency tuning is needed , it is then necessary to control the cavity length in order to tune the laser frequency from one cavity mode to the next one . The grating was thus mounted on a piezoelectric translator (PZT) which allowed a fine electric control of the laser frequency .

Linewidth :

In order to measure the linewidth of extended-cavity lasers , we have made two identical lasers and mixed their beams on a wide bandwidth detector to get the beat note between the optical frequencies . The spectral analysis of the photocurrent , obtained with a RF spectrum analyser , gave the convolution of the laser lines and lead to a good estimation of the laser linewidth . Depending on the laser and the operating conditions , we obtained an instantaneous laser linewidth between 10 and 50 kilohertz . However , the lasing frequency was sensitive to acoustical and mechanical perturbations of the cavity length which broadens the optical spectrum for longer observation times ; for a 1 second observation time , the optical energy was thus distributed over a 1 MHz frequency band .

Sensitivity to optical feedback :

The sensitivity of a laser source to optical feedback is inversely proportionnal to the laser optical cavity length ; an extended-cavity laser is thus far less perturbed by optical feedback than an ordinary diode , for the same power level reflected in the laser . Although it is difficult to measure quantitatively the improvement , we have verified that optical feedback did not perturb experiments that usually require an optical isolator when an ordinary diode was used . This is of great practical interest for optical pumping because residual feedback is a major frequency noise source in this kind of experiment .

3 . Improvement of optical detection in Cs beam frequency standards :

In an optically pumped Cesium beam atomic frequency standard , the error signal used to lock the microwave clock frequency is detected with a laser diode tuned to the top of a Cesium line . When the laser frequency is locked to the top of the line , the variations of the resonant fluorescence , detected by a silicon photodiode , are only related to microwave frequency variations .The fluorescence signal is then used to lock the clock frequency The microwave frequency being modulated at 95 Hz , the clock fractionnal frequency stabilty $\sigma_y(\tau)$ for an observation time τ , is directly related to the signal to noise ratio S/N of the error signal , measured in a 1 Hz bandwidth at the modulation frequency , by the well known formula :

$$\sigma_y(\tau) \approx \left(Q \cdot \frac{S}{N} \cdot \sqrt{\tau} \right)^{-1} \qquad (1)$$

where Q is the microwave resonance quality factor .

Whatever the detection noise comes from , it has to be minimized to reach the best possible clock stability . In most of the interesting experimental setups , it has been found [2] that the laser diode frequency noise was the major noise source in the optical detection .

In order to verify this hypothesis and to evaluate the improvement expected with a noiseless laser, we have compared the detection noise obtained with an ordinary laser diode and with an external-cavity laser .

The results are presented in fig. 2 which shows a spectral analysis of the detection noise with an ordinary laser diode (fig. 2.a) and with an extended-cavity laser (fig. 2.b) when the laser frequency was locked to the top of the atomic line . The signal level was identical in both cases and the detector noise has been measured on the same graph .

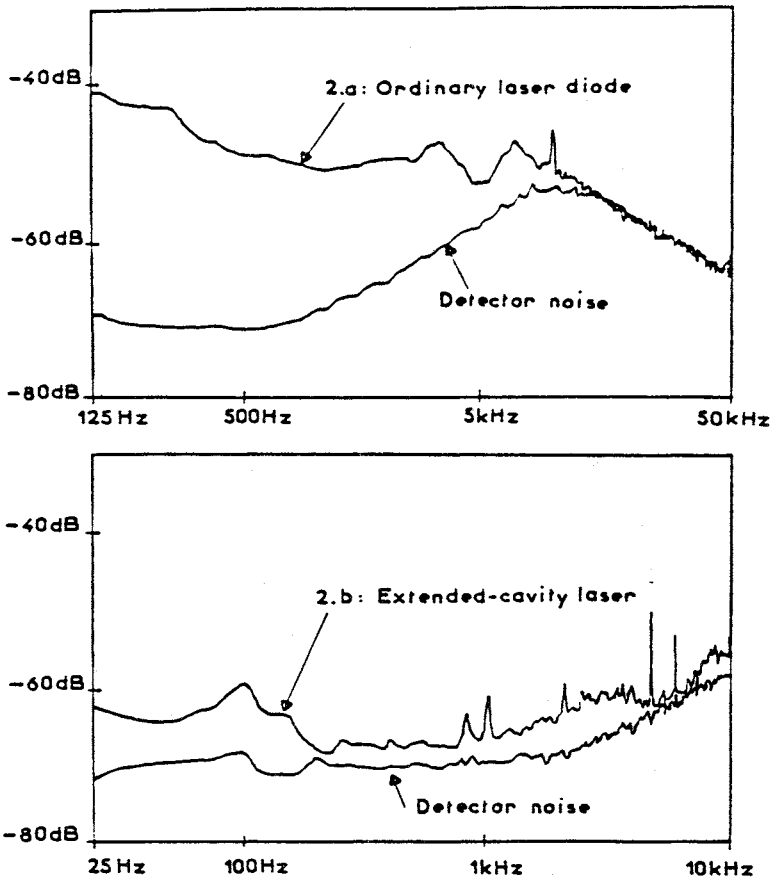


Fig.2 : Spectral analysis of the fluorescence noise for a laser locked on the top of a Cesium beam line :
a) Ordinary laser diode .
b) Extended-cavity laser.

These results show clearly that for a 100 Hz modulation frequency , the detection noise is 30 dB above the detector noise in the first case and only 10 dB above the detector noise in the second case . The signal to noise ratio of the detection signal is then multiplied by 10 when an extended-cavity laser is used instead of an ordinary laser diode .

This demonstrates that the use of extended-cavity lasers may improve the clock stability by at least one order of magnitude .

3 . Long-term frequency stabilization :

Numerous attempts have been devoted to the long-term frequency stabilization of semiconductor lasers using an atomic line as a frequency reference [3 , 4]. The best results have been obtained with AlGaAs laser

diodes locked to Doppler-free absorption lines of a Cesium atomic vapour [4]. However , even with sophisticated detection methods , the laser stability was always greater than 10^{-12} for unitary observation times . In fact , this stability also depends , as in equation 1 , on the detected absorption signal to noise ratio , which is again strongly related to the diode frequency noise . It is then possible to improve the laser long-term stability if an extended-cavity laser is used instead of an ordinary laser diode .

In order to check this hypothesis , we have stabilized the frequency of extended-cavity lasers with a very simple saturated absorption scheme that is described in fig. 3 . A 5 cm long glass cell **C** containing a Cesium atomic vapour at room temperature , is crossed by a saturating beam $\mathbf{B_S}$ coming directly from the laser ; a small fraction of this beam is then reflected back and constitutes a counter-propagating probe beam $\mathbf{B_D}$ which is finally detected by the photodiode **D** . When the lasing frequency is tuned to the 852 nm Cesium line selected , the detected signal is proportionnal to the probe beam absorption saturated by the direct beam . The resulting line is a Doppler profile containing narrow ($\approx$ 15 MHz wide) saturated absorption peaks which are used as frequency references for the laser stabilization .

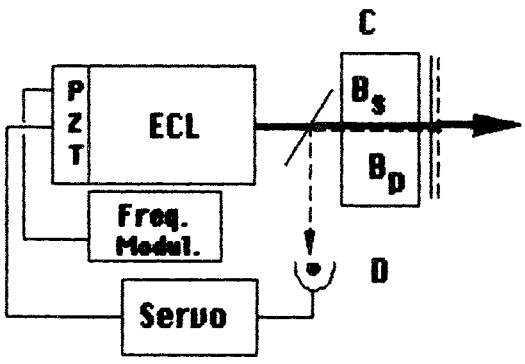


Fig. 3 : **Saturated absorption setup for ECL stabilization .**

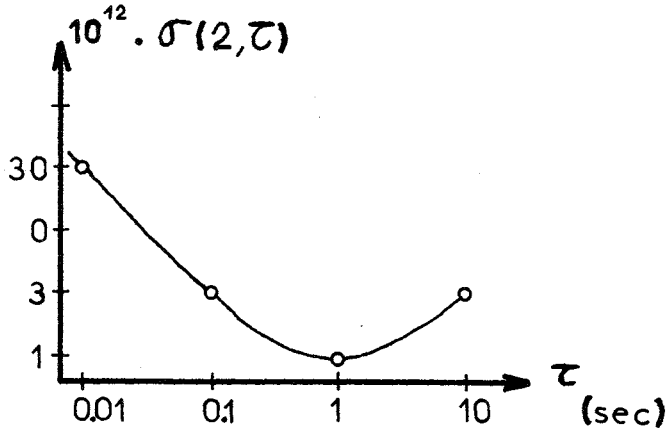


Fig. 4 : Locked laser frequency stability for observation time τ .

The laser frequency was modulated at 5 kHz , using the PZT , around the top of the reference lines , and detected with a lock-in amplifier to obtain an error signal which is fed in a 9 dB/octave integrator used to control the PZT . To measure the laser stability , the beat frequency between two separately locked identical lasers was counted and its Allan variance $\sigma(2,\tau)$ was computed to obtain the frequency stability curve presented in fig. 4 .

With this very simple experimental method , it is thus possible to obtain a 10^{-12} frequency stability for a 1 sec. observation time which demonstrates that the laser long-term stability can actually be improved

when the laser frequency noise is reduced . The stability degradation that is observed for longer times was explained by a bad mechanical stability of one laser which induced slight displacements of the reference lines through laser power variations . It should thus be improved with a better mechanical design and with detection techniques that are less sensitive to power fluctuations .

4 . Conclusion :

It has been shown that extended-cavity semiconductor lasers are very well adapted to optical pumping and light-atom interactions because of their good tunability and low frequency noise which improves detection signals . For instance , we have shown experimentally that using them leads to a one order of magnitude improvement for the frequency stability of optically pumped Cesium atomic clocks . For the same reasons , the frequency stabilization of extended-cavity lasers onto an atomic line leads easily to a 10^{-12} fractionnal frequency stability that has not yet been attained with ordinary laser diodes .

Although other methods can be employed to reduce laser diodes frequency noise [5], we believe that this one can lead to a compact , tunable , easy to operate laser source which would be ideal for industrial requirements .

References :

[1] C.H. HENRY , " Theory of the linewidth of Semiconductor lasers " , IEEE Journ. of Quantum Electron. , Vol. **QE-18** (1982) , p.259

[2] V.CANDELIER et al. , " Frequency stability of an optically pumped Cesium beam frequency standard : preliminary results ", Proc. of 1st European Frequency and Time Forum , Neuchâtel, Suisse (1988) ,this issue.

[3] M.OHTSU , H.TSUCHIDA , T.TAKO , "Frequency stabilization of AlGaAs lasers ", Proc. of 36th Ann. Freq. Control Symposium , Philadelphie (1982)

[4] H.HORI et Al. , " Frequency stabilization of GaAlAs lasers using a Doppler-free spectrum of Cs-D2 line " ,IEEE Journ. of Quantum Electron. , Vol. **QE-19**(1983) , pp. 169-175 .

[5] A.CLAIRON , B.DAHMANI , P.LAURENT , C.BREANT , " Ultra-narrow semiconductor laser linewidth for optically pumped Cesium clock " , Proc. of 1st European Frequency and Time Forum , Neuchâtel, Suisse (1988) ,this issue.

2 ND EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 16 - 17 - 18 MARCH 1988

Session AFS-04 Number 3

Frequency-Stabilized Laser for Atomic Frequency Standards Using Saturated Absorption by Magnetic Modulation

Deng Jingquan, Tan Yongfang, Xiang Yuanhai and Zhu Xiwen

Laboratory of Magnetic Resonance and Atomic and Molecular Physics,

Wuhan Institute of Physics, Academia Sinica, PR China

Optical pumping by laser in place of spectral lamp in atomic frequency standards will increase the output power(active mode) or the intensity of signal(passive mode), and therefore improve the performence of the standards. The frequency and intensity of the laser as a pumping surce should be so stable that the pumping rate and light shift would not vary significantly. In order to meet the requirement mentioned, the frequency of the commercially available laser diode(LD) has to be stabilized. Among various frequercy-stabilization scheme of LD, Doppler-free saturated absorption method has given the best frequency stability due to the narrowness and stability of the absorption line. In previous saturated absorption set-up frequency scanning was used[1], which would cause additional periodical variation of laser frequency. In this report, the signal of saturation absorption has been obtained by applying an oscillating magnetic field to a Rb cell[2] and good frequency stability of an LD(7800A) has been realized.

We firstly regulated the temperature of the GaAlAs laser diode within 10^{-4} °C by the use of thermoelectric cooler and used current regulated power supply with high performence to provide the injection current. A circularly polarized laser beam is sent to a Rb cell and reflected back to the cell as a probe beam, which is detected by a photodetector. The output of the photodetector is fed to a lock-in amplifier. A series of rectangular pulses is applied to a coil around the Rb cell, and the Rb atoms in the cell will experience an oscillating longitudinal magnetic field, or equivalently, be irradiated a right-handed or left-handed circularly polarized beam alternatively. If the laser frequency does not coincide with the transition frequency of atoms, the intensity of the probe beam when right-handed pumping beam applied will differ the intensity in the left-handed case, then the lock-in amplifier will output an error signal, which will be fed back to the injection current of LD to lock the laser frequency to the atomic line. We have observed the signal of saturation absorption, examined the characteristics of laser frequency for various modulation frequency and magnetic field strength, and got preliminary values of laser frequency stability. It is shown that the frequency stability of LD in magnetic modulation scheme is at least one order of magnitude better than that in free running case.

References

[1] H. Hori, Y. Kitayama, T. Yabuzaki and T. Ogawa, IEEE J. QE-19, 169(1983).

[2] T. Yabuzaki, T. Kawamura and T. Ogawa, Proc. 10th Int'l Conf. Atomic Physics, (Tokyo, Japan, 1986), p. 184

2 ND EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 16 - 17 - 18 MARCH 1988

Session AFS-04 Number 4

Phase-lock design of a 600 GHz BWO for the Mg frequency standard

E. Bava, A. Godone, C. Novero

Istituto Elettrotecnico Nazionale Galileo Ferraris, Torino, Italy.

Abstract

The frequency corresponding to the $^3P_1 - {}^3P_0$ ^{24}Mg transition has been measured with respect to a Cs standard with an accuracy of $\pm 3\times10^{-11}$ mainly limited by two technical problems: the uncertainty about the C-field intensity, due to the absence of magnetic shielding, and the cycle slipping of the BWO phase-lock. The latter of these is discussed in this communication and in particular the design of an improved phase-lock loop, which could guarantee an accuracy of the submillimetric synthesized frequency better than 1×10^{-13} for counting times of 0.1-1 s, is reported. The main task in this case is to control the high level of flicker frequency noise of the 600 GHz oscillator, due to its intrinsecally low Q-factor. The theoretical predictions will be compared with the experimental results as far as it regards the IF and control signal spectra, in the case of a closed loop bandwidth of 400 kHz.

1 - Introduction

In the 200 GHz - 1 THz range backward-wave-oscillators (BWO) are among the

Work supported by the Consiglio Nazionale delle Ricerche of Italy.

most interesting sources for metrological and spectroscopic research and applications: tunability makes BWO's more versatile than molecular lasers and power level is much larger than the output from frequency multipliers. They have been used in high resolution spectroscopy experiments such as atomic frequency standards [1], frequency synthesis up to the infrared [2], Rydberg atoms excitation [3] and radioastronomy in general. Phase stabilization is mandatory in all the above fields. Even if phase-locking techniques have been widely developed in the past both theoretically and experimentally, in the BWO's case further considerations need to be taken into account, related to the high operation frequency and the very low Q-factor.

These typical features imply a relatively high flicker-frequency noise which has to be taken into account in the design of the phase-lock loop.

The phase control of a submillimetric source requires also the synthesis of a reference signal in the same frequency range whose phase-coherence can be strongly impaired by the high multiplication order; this problem has been considered elsewhere [4] and will not be discussed in this paper.

We shall focus our attention in the characterization of the frequency noise spectrum of a BWO and then in the proper choice of the control-loop parameters.

The experimental results referring to the phase-stabilization of the 600 GHz carcinotron used in the Mg frequency standard [1] are reported and compared with the theory developed.

2 - BWO frequency noise spectrum

As a first step towards the phase-stabilization of the BWO oscillator it is necessary to characterize its frequency noise spectrum.

Taking into account that this kind of oscillator is a very-low Q device ($Q \sim 10$), it is expected in general to be quite unstable; moreover the simple theory reported in [5] foresees that the dominant noise contribution to the fluctuation spectrum is mainly flicker-frequency noise.

The above considerations are confirmed by the experimental measurements

reported in figs. 1 and 2, referring to the free-running 600 GHz BWO used in our experiments. In fig. 1 the Allan standard deviation $\sigma_y(\tau)$ is reported, where the value $\sigma_y(\tau) \simeq 4 \times 10^{-7}$ may be assumed as the flicker level, which ranges from τ = 10 ms to τ = 100 s. The pattern of the one-sided frequency noise spectral density $S_y(f)$ is shown in fig. 2; the spectrum appears dominated by flicker noise up to about 100 kHz from the carrier with a slope of 15 dB/dec ($f^{-1.5}$).

From the above experimental data it turns out that for a wide range of Fourier frequencies the following expression for $S_y(f)$ may be assumed to characterize the BWO frequency noise instabilities:

$$S_y(f) = h_{-1} f^{-1} \qquad h_{-1} \sim 10^{-13} \tag{1}$$

or, for the one-sided phase spectral density $S_\phi(f)$, the expression:

$$S_\phi(f) = h_{-1} \nu_0^2 f^{-3} \tag{2}$$

being ν_0 the carcinotron operation frequency.

The above assumption is also confirmed by the observed fast-linewidth $W_f \simeq 400$ kHz which agrees with the calculated value $W_f = 2.2\sqrt{h_{-1}}\,\nu_0$, when flicker-frequency noise is assumed.

3 - Heterodyne phase-lock loop

The block diagram of the system adopted for the phase stabilization of a sub-millimetric source is shown in fig. 3. A beam-splitter divides the output beam in two parts in such a way that a fraction of the BWO power is sent to a harmonic mixer (xN_2); the non-linear element is usually a μm-size Pt-GaAs Schottky barrier diode. The output of a K-band Gunn oscillator, locked by injection to a reference quartz oscillator (ν_t) via a low-noise frequency multiplier chain (xN_1), is also coupled to the junction via a triplexer which separates the local oscillator signal, the IF signal and the bias. The IF signal from the harmonic mixer, after low-noise amplification, is down-converted to the new IF frequency value

ν_{IF} and sent to a limiting amplifier.

The phase-detector PD is driven by the IF limiting amplifier and by the reference signal with frequency ν_i; the d.c. output signal (error signal) is sent to the BWO control element through a loop-filter with transfer function $F(j\omega)$ and a buffer amplifier.

From the general linear theory of the phase-lock loops [6] the following expressions for the phase spectral densities of the IF ($S_{\phi\,IF}$) and output ($S_{\phi\,0}$) signals are obtained under the condition $N^2 S_{\phi\,r} \ll S_{\phi\,c}$ [4].

$$S_{\phi_{IF}} = |H(j\omega)|^2 S_{\phi_n} + |1 - H(j\omega)|^2 S_{\phi_c} \tag{3}$$

$$S_{\phi_0} = |H(j\omega)|^2 N^2 S_{\phi_r} + S_{\phi_{IF}} \tag{4}$$

In the expressions (3) and (4) $S_{\phi\,n}$ and $S_{\phi\,r}$ are the phase spectral densities of the additive noise and of the reference quartz (ν_r), $N = N_1 N_2$ the total multiplication factor and $S_{\phi\,c}$ is given by eq. (2). $H(j\omega)$ is the closed loop transfer function; in the case of a 2nd order loop the following expression holds:

$$|H(j\omega)|^2 = \frac{1 + 4\zeta^2 f^2 / f_n^2}{\left(f^2/f_n^2 - 1\right)^2 + 4\zeta^2 f^2 / f_n^2} \tag{5}$$

where f_n and ζ are the natural frequency and the damping factor respectively [6].

In the time-domain the frequency stability of the phase-locked BWO, measured via the two-sample standard deviation $\sigma_y(\tau)_0$ is from eq. 4:

$$\sigma_y(\tau)_0 = \sigma_y(\tau)_r + \frac{\nu_{IF}}{\nu_c}\,\sigma_y(\tau)_{IF} \tag{6}$$

The controlled oscillator output follows the reference signal if:

$$\frac{\nu_{IF}}{\nu_c}\,\sigma_y(\tau)_{IF} \ll \sigma_y(\tau)_r$$

that means, for very high precision frequency measurements, that the following condition is usually required:

$$\frac{\nu_{IF}}{\nu_c}\,\sigma_y(\tau)_{IF} \ll 10^{-12} \quad \forall \tau \geq 1\,s \tag{7}$$

In order to reduce strongly the unlocking probability and to allow for a linear operation of the control loop, the mean-square value of phase fluctuations at the discriminator output $\overline{\varphi_e^2}$ must satisfy the following condition:

$$\overline{\varphi_e^2} = \overline{\varphi_{IF}^2} \ll 1 \tag{8}$$

The expressions of the quantities implied in (7) and (8), for the 2nd order loop considered, are:

$$\sigma_y^2(\tau)_{IF} = \frac{3}{4\pi^2\nu_{IF}^2}\left(S_{\varphi_m}^{w} f_h + \frac{1}{2}\frac{\nu_c^2 h_{-1}}{f_n^4} f_h^2\right)\tau^{-2} \tag{9}$$

$$\overline{\varphi_e^2} = \frac{1}{2}\left(\zeta + \frac{1}{4\zeta}\right) 2\pi f_m S_{\varphi_m}^{w} + \frac{1}{2} f(\zeta)\frac{h_{-1}\nu_c^2}{f_m^2} \tag{10}$$

where $S_{\varphi_m}^{w}$ is the white-phase additive noise, f_h the bandwidth of the low-pass filter in the measuring arm and $f(\zeta)$ a flicker-variance factor reported in [6].

4 - Experimental results

Following the theory outlined above, the phase-lock of the 600 GHz carcinotron has been realized implementing the circuit of fig. 3.

Assuming for the various parameters of eq. (9) and (10) the following experimental values: ν_{IF} = 30 MHz, $S_{\varphi_n}^{w}$ = 1×10^{-9} rad^2/Hz, ν_c = 600 GHz, h_{-1} = 10^{-13}, f_h = 12 kHz, ζ = 1.5 and $f(\zeta) \sim 1$, it turns out that the two conditions (7) and (8) are satisfied when f_n > 100 kHz or a closed-loop control bandwidth f_{-3dB} > 300 kHz ($f_{-3dB} \simeq 2\zeta f_n$ [6]). The system realized for the Mg standard allows a control bandwidth $f_{-3dB} \simeq$ 400 kHz; the spectrum of the IF signal ($S_{\varphi\,IF}$), observed in this condition, is reported in fig. 4a and 4b for two observation bandwidths (10 kHz and 100 Hz). The signal-to-noise ratio near the IF carrier is due to residual BWO flicker noise, not fully controlled, as it can be argued by observing its

dependence from the open loop-gain (fig. 5). A decrease of 8 dB in the open loop gain is completely transferred to the IF spectrum (5a, 5c), whereas an increase in the additive IF noise is nearly compensated by the control loop (Fig. 5a, 5b).

The white-phase additive noise S_{ϕ_n} is at least 30 dB below: an increase of f_n should be therefore useful.

The Allan standard deviation of the IF signal referred to 600 GHz is

$$\frac{\nu_{IF}}{\nu_c}\,\sigma_y(\tau)_{IF} \simeq 1\times10^{-13}\,\tau^{-1}$$

in good agreement with the value estimated from eq. (9).

Even if the measurement of the fluctuations spectrum of the phase-locked BWO should require a second indipendent system, we can estimate it from eq. 4: the measured $S_{\phi_{IF}}$ is lower than $N^2 S_{\phi_t}$ [4] so that the RF spectrum is just limited by the unavoidable multiplied phase noise of the reference oscillator.

Still possible improvements in the phase-lock system, mainly a wider control bandwidth, don't improve the RF output spectrum, but may be useful to extend the pull-out range $\Delta\nu_{po}$ [6], which has not been taken into account in our linear theory:

$$\Delta\nu_{po} \simeq 1.8\,(1+\zeta)\,f_m = 600\ \text{kHz}$$

If the probability of peak frequency deviations of the free-running oscillator higher than 600 kHz is relevant in a time interval comparable to the counting time, an increase of f_n is mandatory.

Conclusions

The experimental results demonstrate the possibility of extending the frequency synthesis up to 600 GHz at the accuracy level required for the realization of a frequency standard operating in the submillimetric region of the electromagnetic spectrum.

References

1 A. Godone, E. Bava, C. Novero
High-resolution frequency measurement of the $^{3}P_{1} - {}^{3}P_{0}$ ^{24}Mg transition
Metrologia 24, pp. 133-138 (1987)

2 C.O. Weiss
Frequency measurement chain to 30 THz using FIR Schottky diodes and a submillimeter backward wave oscillator
Appl. Phys. B34, pp. 63-67 (1984)

3 J. Lian, M. Gross, P. Goy, S. Haroche
Circular Rydberg-state spectroscopy
Phys. Rev. A., vol. 33, n. 6, pp. 4437-4439 (1986)

4 A. De Marchi, A. Godone, E. Bava
Phase-coherent synthesis and precision frequency measurements in the far-infrared
IEEE Trans. Instr. Meas., IM-30, n. 2, pp. 132-138 (1981)

5 D.B. Leeson
A simple model of feedback oscillator noise spectrum
Proc. IEEE, vol. 54, pp. 329-330 (1966)

6 F.M. Gardner
Phase-lock Techniques, 2nd ed., New York: Wiley (1979)

Captions to figures

Fig. 1 - Measured Allan standard deviation of a 600 GHz carcinotron.

Fig. 2 - Experimental patterns of $S_y(f)$ of a 600 GHz carcinotron.

Fig. 3 - Block diagram of the considered phase-lock system.

Fig. 4 - RF spectrum of the IF signal in closed-loop conditions;
a) vert.: 10 dB/div, hor.: 100 kHz/div, BW 10 kHz;
b) vert.: 10 dB/div, hor.: 2 kHz/div, BW 100 Hz.

Fig. 5 - RF spectrum of the IF signal (vert. 10 dB/div., hor. 10 kHz/div., BW 1 kHz).
a) the best result obtained in closed loop conditions;
b) with an increase of S_{ϕ_n} of 8 dB;
c) open loop gain lowered by 8 dB.

Fig. 1

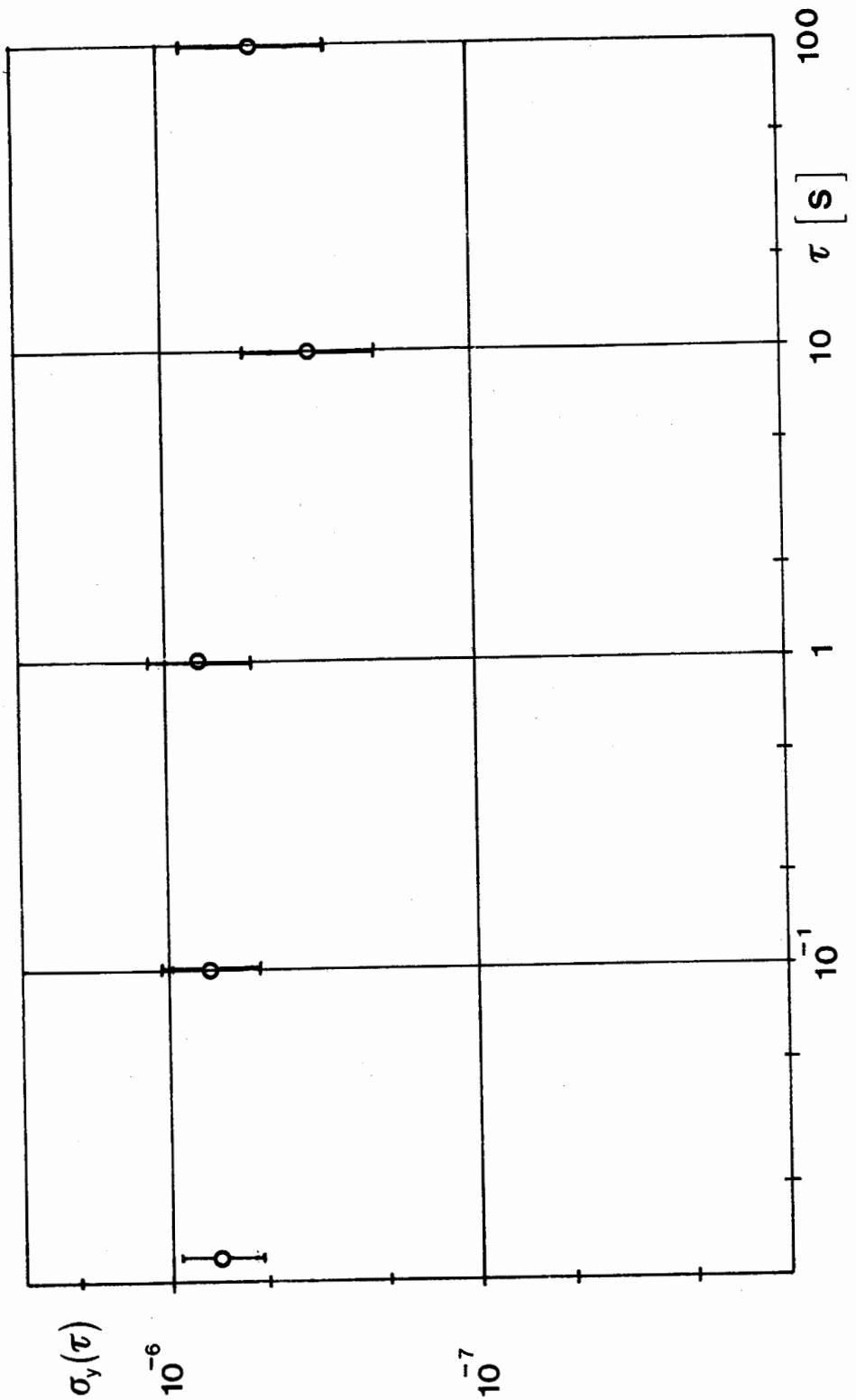


Fig 2

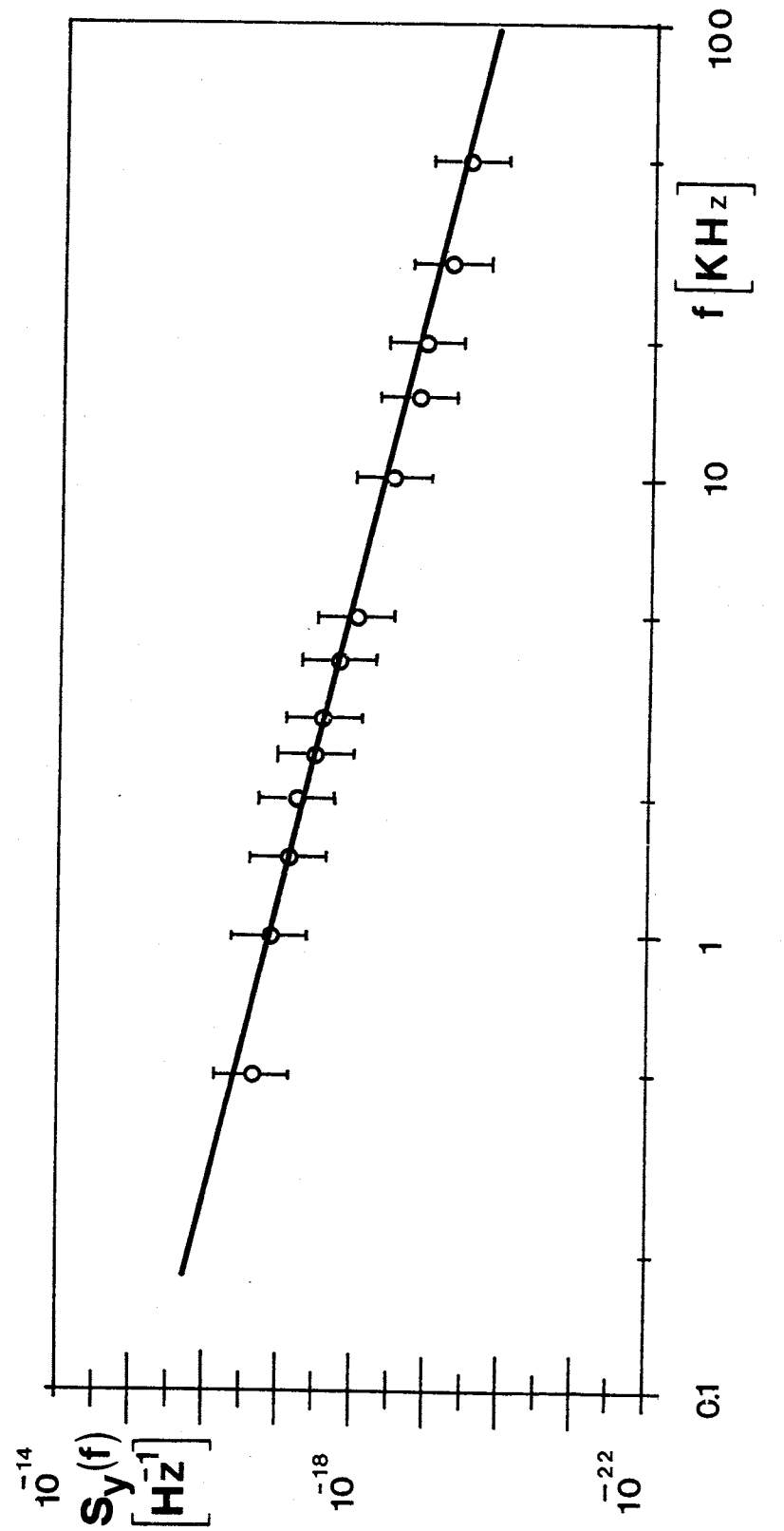
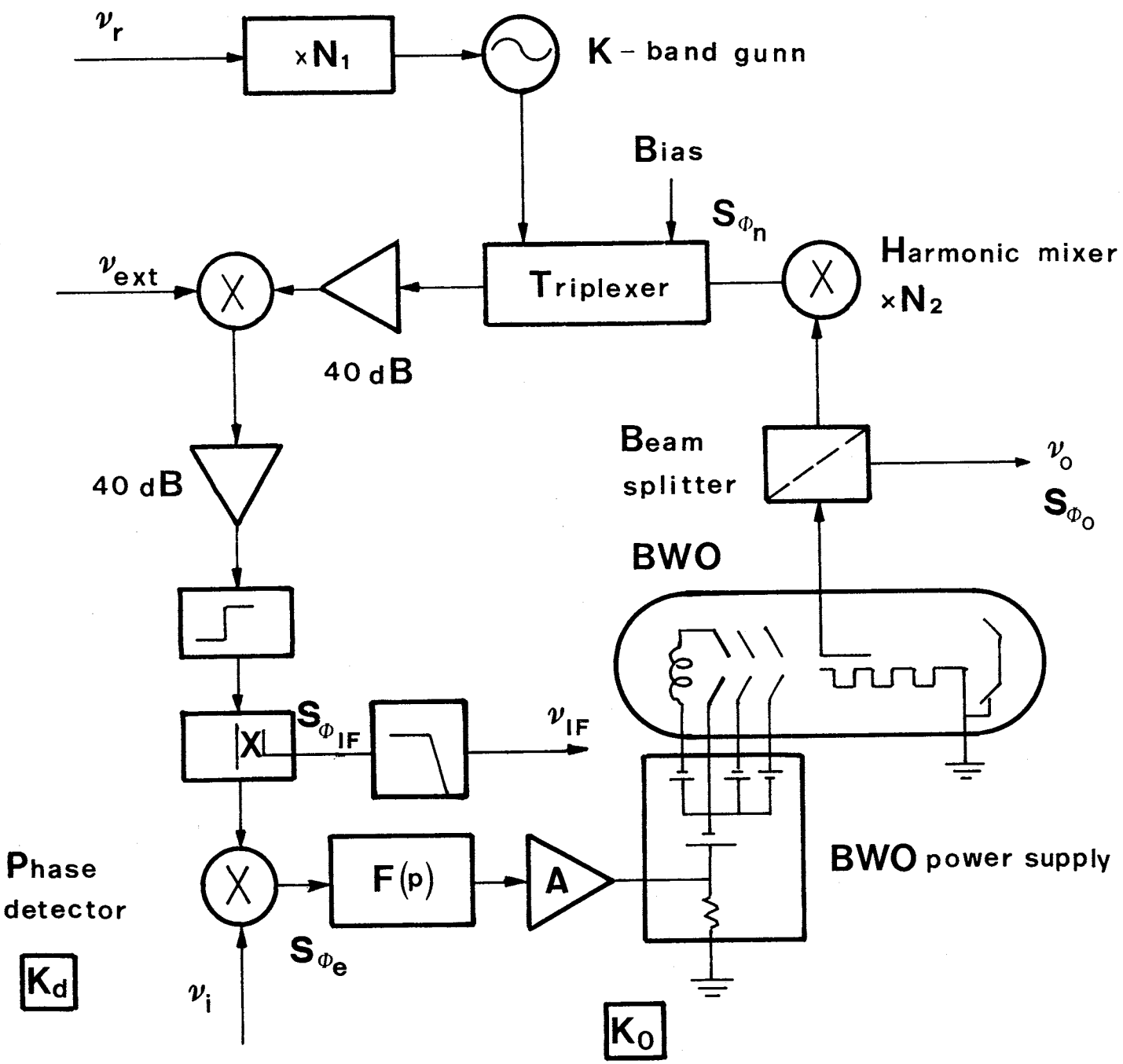


Fig. 3


ν_r
×N1
K - band gunn
Bias
S_{Φ_n}
ν_{ext}
Triplexer
Harmonic mixer
×N2
40 dB
40 dB
Beam
splitter
ν_o
S_{Φ_o}
BWO
$S_{\Phi_{IF}}$
ν_{IF}
Phase
detector
F(p)
A
BWO power supply
S_{Φ_e}
K_d
ν_i
K_o

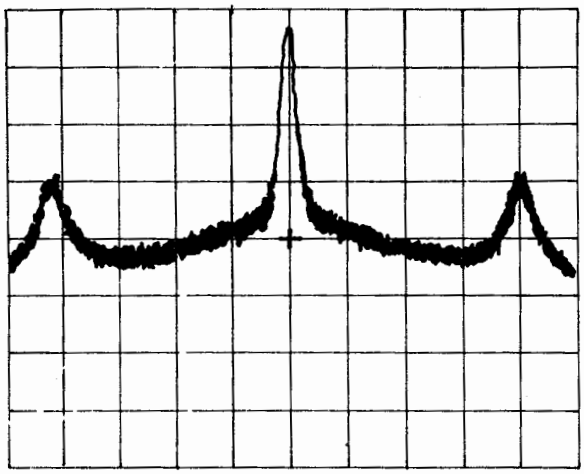


Fig. 4a)

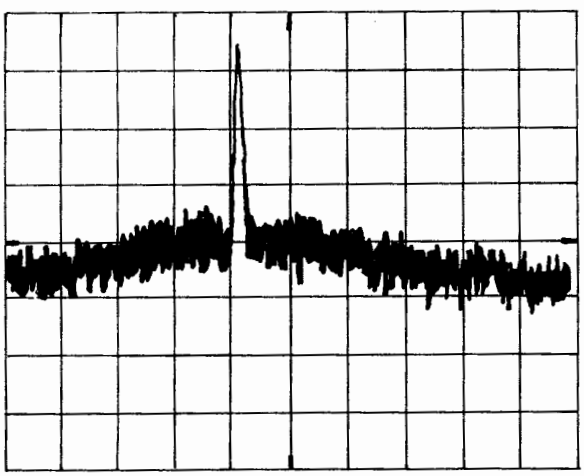


Fig. 4b)

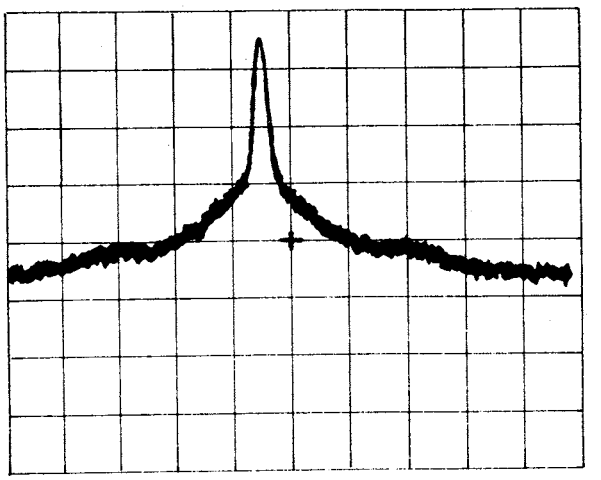


Fig. 5a)

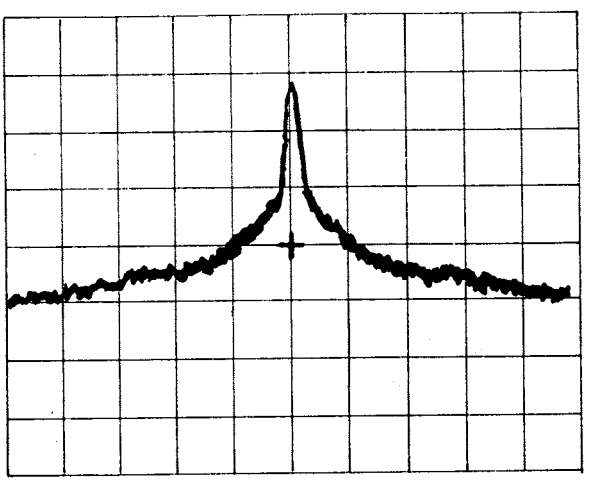


Fig. 5b)

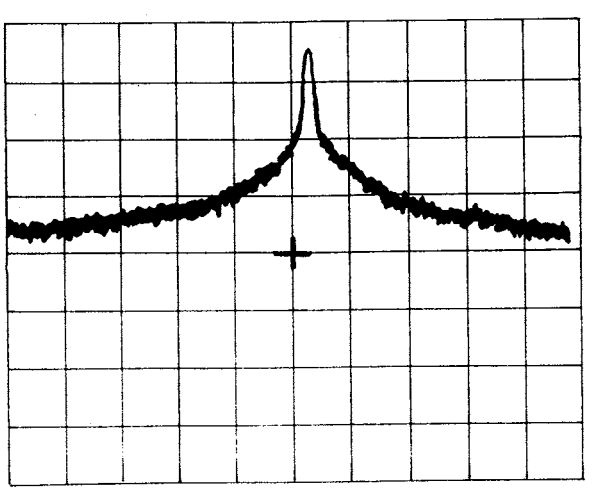


Fig. 5c)

Resonators and Oscillators

		Chairman
RO-01	Resonators 1	H.F. Tiersten
RO-02	Resonators 2	W. Zingg
RO-03	BVA	R. Besson
RO-04	SAW	C. Maerfeld
RO-05	Oscillators	N. Yannoni

RO-01

Resonators 1

1. Vibration Modes of Contoured Piezoelectric Resonators with Additional Film Layers
   H. Nowotny,
   Institut für Theoretische Physik, Wien - Austria
   E. Benes, M. Schmid
   Institut für Allgemeine Physik, Wien - Austria

2. Analysis of Partially or totally Contoured Resonators with Finite Lateral Dimensions
   J. Detaint, H. Carru, J. Schwartzel, C. Joly,
   CNET PAB/MCT/CMM, Bagneux - France

3. Lateral DC Electric Field in Quartz Plate Resonators
   C.K. Hruska
   Piezoeletricity Resarch Laboratory, Toronto - Canada

4. Non Piezoelectric Crystal Resonators
   D. Mazerolle, E. Dieulesaint, D. Royer
   Université P. & M. Curie, Paris - France

2 ND EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 16 - 17 - 18 MARCH 1988

Session RO-01 Number 1

Vibration Modes of Contoured Piezoelectric Resonators with Additional Film Layers

H. Nowotny *, E. Benes #, M. Schmid #

* Institut für Theoretische Physik, Technische Universität Wien
# Institut für Allgemeine Physik, Technische Universität Wien

Abstract

In a previous analysis of layered piezoelectric resonators with two electrodes we have introduced a theoretical solution which is rigorous within the one-dimensional description. This model is especially useful for the analysis of the heavily loaded quartz crystal microbalance. In order to reduce further the deviation between the theoretical description and the properties of the real geometry (e.g. a plano-convex crystal as resonator core) we have extended the one-dimensional treatment of the layered structure. Using the general solution of the Fourier transformed three-dimensional equations of linear piezoelectricity, it is possible to allow variations of all physical variables along the surface directions. Such a physical variable dependency on the directions perpendicular to the thickness dimension is necessary for the description of energy-trapped vibration modes. The resulting integral equations for the determination of the quasiharmonic and anharmonic frequency spectrum and the corresponding modes of vibration of the layered structure can be solved only approximately. For the special case of practical importance, where a layered composite resonator containing a resonator core with slowly varying thickness is investigated, a Taylor series approximation up to second order terms can be used, which yields second order differential equations for the determination of the resonance frequency spectrum. It is shown that known results for contoured resonators without films are incorporated in this more general treatment as special cases.

1. Introduction

The first rigorous and within the one-dimensional concept most general treatment of arbitrarely layered piezoelectric films was introduced by *Ballato* /1,2/ et al. This attempt uses an equivalent circuit representation which - at least from the viewpoint of an electrical engineer - is pictorial. With this concept the interaction of an arbitrary number of stacked layers can be described, but, with increasing number of layers, the necessary electrical network analysis becomes rather complex. Besides this fundamental work, several models for special applications have been developed. Examples of such dedicated approaches are, among others, the work of *Newell* /3/, which is concerned with filter applications, where the piezoelectric resonator plate is face-mounted by use of quarter wavelength acoustic layers, and the work of *Miller* and *Bolef* /4,5/ which has been adopted by *Lewis* and *Lu* /6/ for the heavily loaded quartz crystal microbalance, where the deposited films on the sensor crystal form an acoustic sandwich structure. All these examples are based upon a purely mechanical acoustic model.

In a more recent approach /7/ a general 8x8 transfer matrix description for the representation of every layer was introduced. This model is not restricted to a single displacement direction and covers the general case of multimode excitations and its coupling results. It allows the calculation of the electrical admittance appearing between the two

electrodes for any frequency, as well as the determination of the entire resonance frequency spectrum using only simple matrix multiplications. In contrast to *Ballatos* concept, it is extendable to the case where a slight spatial dependency of all physical quantities on the locus upon the surface is taken into account. This extension presented in the next sections for the first time allows the description of energy trapped vibration modes of sandwich-resonators.

2. Derivation of the theoretical model

2.1. Three-dimensional treatment of the piezoelectric continuum

According to the quasielectrostatic approximation, we start with the 9 linearily independent piezoelectric constitutive material equations /8/ with excluded rigid rotations

$$T_{ij} = c_{ijkl}\partial_l u_k + e_{kij}\partial_k \varphi \ , \tag{1}$$

$$D_i = e_{ikl}\partial_k u_l - \varepsilon_{ik}\partial_k \varphi \ , \tag{2}$$

the 3 stress equations for motion

$$\partial_i T_{ij} = \rho \ddot{u}_j \ , \tag{3}$$

and the electrostatic equation for an insulator

$$\partial_i D_i = 0 \ , \tag{4}$$

where T_{ij} are the stress components, c_{ijkl} the elastic stiffness constants for constant electric field, $\partial_i = \partial/\partial x_i$, u_k are the displacement vector components, e_{ikl} the piezoelectric constants, φ is the electrical potential, D_i are the electrical displacement components, ε_{ik} the dielectric constants for constant strain, ρ is the material density and $\ddot{u}_i = \partial^2 u_i/\partial t^2$. The summation convention for repeated indices is employed throughout this paper.

Substitution of Eqs. (1) and (2) into (3) and (4) yields the basic four three-dimensional differential equations for the linear piezoelectric continuum in the four variables u_k and φ

$$c_{ijkl}\partial_i\partial_l u_k + e_{kij}\partial_k\partial_i \varphi = \rho \ddot{u}_j \ , \tag{5}$$

$$e_{ikl}\partial_i\partial_k u_l - \varepsilon_{ik}\partial_i\partial_k \varphi = 0 \ . \tag{6}$$

The boundary conditions for the interfacing surface between two piezoelectric continua I and II are

$$u_i^{\,I} = u_i^{\,II} \ , \tag{7}$$

$$n_i T_{ij}^{\,I} = n_i T_{ij}^{\,II} \ , \tag{8}$$

$$\varphi^{I} = \varphi^{II} \ , \tag{9}$$

$$n_i D_i^{I} = n_i D_i^{II} \quad , \tag{10}$$

where n_i are the components of the orientation vector perpendicular to the surface.

2.2. Fourier transforms of the constitutive equations

The functions $u_i(\mathbf{r},t)$, $\varphi(\mathbf{r},t)$ can be expressed by the Fourier transforms $u_i(\mathbf{p},\omega)$, $\varphi(\mathbf{p},\omega)$ using the Fourier integrals

$$u_i(\mathbf{r},t) = \int d^3p\, d\omega\, e^{j\mathbf{pr}}\, e^{-j\omega t}\, u_i(\mathbf{p},\omega) \quad , \tag{11}$$

$$\varphi(\mathbf{r},t) = \int d^3p\, d\omega\, e^{j\mathbf{pr}}\, e^{-j\omega t}\, \varphi(\mathbf{p},\omega) \quad , \tag{12}$$

where the inverse transformation is given by

$$u_i(\mathbf{p},\omega) = \frac{1}{(2\pi)^4}\int d^3r\, dt\, e^{-j\mathbf{pr}}\, e^{j\omega t}\, u_i(\mathbf{r},t) \quad , \tag{13}$$

$$\varphi(\mathbf{p},\omega) = \frac{1}{(2\pi)^4}\int d^3r\, dt\, e^{-j\mathbf{pr}}\, e^{j\omega t}\, \varphi(\mathbf{r},t) \quad . \tag{14}$$

Subtituting Eqs. (11), (12) in (5), (6) and in (1), (2) yields the stress equations for motion, the electrostatic equation for an insulator, and the material equations, respectively, for the Fourier transforms

$$c_{ijkl}p_i p_l u_k(\mathbf{p},\omega) + e_{kij}p_k p_i \varphi(\mathbf{p},\omega) = \rho\omega^2 u_j(\mathbf{p},\omega), \tag{15}$$

$$e_{ikl}p_i p_k u_l(\mathbf{p},\omega) - \varepsilon_{ik}p_i p_k \varphi(\mathbf{p},\omega) = 0 \quad , \tag{16}$$

$$-jT_{ik}(\mathbf{p},\omega) = c_{iklm}p_m u_l(\mathbf{p},\omega) + e_{lik}p_l \varphi(\mathbf{p},\omega) \quad , \tag{17}$$

$$-jD_i(\mathbf{p},\omega) = e_{ikl}p_k u_l(\mathbf{p},\omega) - \varepsilon_{ik}p_k \varphi(\mathbf{p},\omega) \quad . \tag{18}$$

Because our physical variables are real functions of $\mathbf{r}$ and t, the following symmetry relations are valid for their Fourier transforms:

$$u_i(-\mathbf{p},-\omega) = u_i(\mathbf{p},\omega)^* \quad , \tag{19}$$

$$\varphi(-\mathbf{p},-\omega) = \varphi(\mathbf{p},\omega)^* \quad , \tag{20}$$

$$T_{ij}(-\mathbf{p},-\omega) = T_{ij}(\mathbf{p},\omega)^* \quad , \tag{21}$$

$$D_i(-\mathbf{p},-\omega) = D_i(\mathbf{p},\omega)^* \; ; \tag{22}$$

where * denotes the complex conjugate of the indicated function.

2.3. Fourier transformed boundary conditions for a plane between two continua

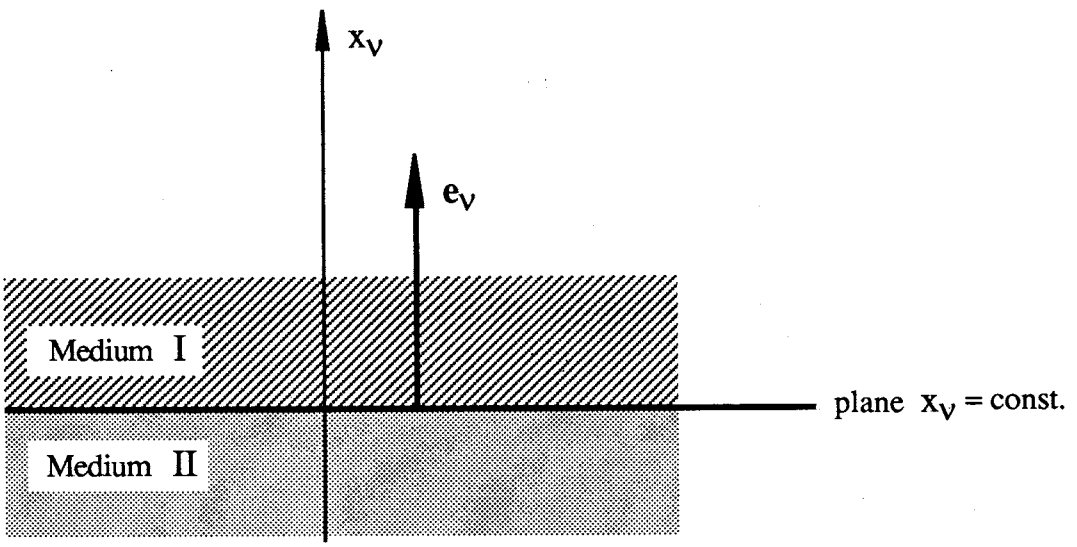


FIG.1. Two arbitrarily oriented piezoelectric media I and II separated by a plane. The direction ν is perpendicular to that plane, $\mathbf{e}_\nu$ being the unit orientation vector and x_ν a coordinate axis chosen in direction ν.

We regard two piezoelectric continua separated by a plane and designate with x_ν the coordinate axis perpendicular to that plane (see Fig.1). Adequate to the problem is the splitting of the spatial vector $\mathbf{r}$ in a spatial vector $\mathbf{r}_s$ within the plane and in a vector $x_\nu \mathbf{e}_\nu$ perpendicular to the plane according to the relation

$$\mathbf{r} = \mathbf{r}_s + x_\nu \mathbf{e}_\nu \,, \quad \text{whereby} \quad \mathbf{r}_s \mathbf{e}_\nu = 0 \,. \tag{23}$$

The corresponding Fourier variable $\mathbf{p}$ can be written in a perfectly analogous form

$$\mathbf{p} = \mathbf{q} + p_\nu \mathbf{e}_\nu \,, \quad \text{whereby} \quad \mathbf{q}\mathbf{e}_\nu = 0 \,. \tag{24}$$

Introducing this splitting of the spatial vectors according to Eq. (23) into the boundary condition (7) and into the Fourier integral (11), we obtain

$$u_i^{\mathrm{I}}(\mathbf{r}_s, x_\nu, t) = u_i^{\mathrm{II}}(\mathbf{r}_s, x_\nu, t) \,, \tag{25}$$

or, written with the Fourier transforms

$$\int d\omega\, e^{-j\omega t} \int d^2q\, e^{jqr_s} \int dp_\nu\, e^{jp_\nu x_\nu}\, u_i^{I}(\mathbf{q},p_\nu,\omega) =$$

$$= \int d\omega\, e^{-j\omega t} \int d^2q\, e^{jqr_s} \int dp_\nu\, e^{jp_\nu x_\nu}\, u_i^{II}(\mathbf{q},p_\nu,\omega)\,. \qquad (26)$$

Since these equations must be fulfilled for any $\mathbf{r}_s$ and t, we obtain with the useful abbreviation

$$u_i^{I}(\mathbf{q},\omega;\, x_\nu) \triangleq \int dp_\nu\, e^{jp_\nu x_\nu}\, u_i^{I}(\mathbf{q},p_\nu,\omega) \qquad (27)$$

the boundary conditions for the Fourier transforms of the mechanical displacement components

$$u_i^{I}(\mathbf{q},\omega) = u_i^{II}(\mathbf{q},\omega)\ . \qquad (28)$$

The boundary conditions for the remaining Fourier transforms are obtained analogously as

$$T_{\nu j}^{I}(\mathbf{q},\omega) = T_{\nu j}^{II}(\mathbf{q},\omega)\ , \qquad (29)$$

$$\varphi^{I}(\mathbf{q},\omega) = \varphi^{II}(\mathbf{q},\omega)\ , \qquad (30)$$

$$D_\nu^{I}(\mathbf{q},\omega) = D_\nu^{II}(\mathbf{q},\omega)\ . \qquad (31)$$

2.4. General solution of the basic equations in the Fourier transform domain

The Fourier transform $\varphi(\mathbf{p},\omega)$ of the electrical potential can be immediately determined from Eq.(16) as

$$\varphi(\mathbf{p},\omega) = \frac{e_{ikl}p_ip_k}{\varepsilon_{mn}p_mp_n}u_l(\mathbf{p},\omega) + \varphi_{hom}(\mathbf{p},\omega)\ , \qquad (32)$$

with $\varphi_{hom}(\mathbf{p},\omega)$ being solution of the homogeneous equation

$$\varepsilon_{ik}p_ip_k\,\varphi_{hom}(\mathbf{p},\omega) = 0\ . \qquad (33)$$

Substituting Eq.(32) into (15) we obtain the general equation for the determination of the $u_i(\mathbf{p},\omega)$

$$\Lambda_{ik}(\mathbf{p})\,u_k(\mathbf{p},\omega) = \rho\omega^2 u_i(\mathbf{p},\omega), \qquad (34)$$

where we have introduced the abbreviation

$$\Lambda_{jk}(\mathbf{p}) = c_{ijkl}p_ip_l + \frac{e_{igj}p_g e_{lhk}p_h}{\varepsilon_{mn}p_mp_n}p_ip_l = c_{ijkl}(\mathbf{p})\,p_ip_l \tag{35}$$

with $c_{ijkl}(\mathbf{p})$ being the piezoelectrically stiffened elastic constant according to the definition /8/ established in the $(\mathbf{r},t)$-domain.

2.4.1. Solution of the homogeneous equation for the Fourier transform of the electrical potential

Eq.(33) shows that a nontrivial solution for $\varphi_{hom}(\mathbf{p},\omega)$ can be obtained only for such $\mathbf{p}$-vectors, which fulfill the equation

$$\varepsilon_{ik}p_ip_k = 0\ . \tag{36}$$

Considering the splitting of the $\mathbf{p}$-vectors in a component in direction ν and a component perpendicular to that direction according to Eq.(24), Eq.(36) is a quadratic equation for p_ν

$$\varepsilon_{ik}p_ip_k = \varepsilon_{\nu\nu}[p_\nu - p_\nu^{(\varepsilon)}(\mathbf{q})][p_\nu + p_\nu^{(\varepsilon)}(-\mathbf{q})^*] = 0\ , \tag{37}$$

where the two solutions $p_\nu^{(\varepsilon)}(\mathbf{q})$ are - as indicated by their argument - dependent on $\mathbf{q}$. Equation (37) can be written in the form shown, because the solutions of the quadratic equation are complex conjugates

$$p_\nu^{(2)}(\mathbf{q}) = p_\nu^{(1)}(\mathbf{q})^* \tag{38}$$

as a consequence of ε_{ik} being a positive definite matrix and because

$$p_\nu^{(\varepsilon)}(\mathbf{q})^* = -p_\nu^{(\varepsilon)}(-\mathbf{q})^* \tag{39}$$

follows from the form of the solution obtained from equation (37). Considering this and taking the symmetry relation Eq.(20) into account, we can write the solution of the homogeneous equation (33) in the form

$$\varphi_{hom}(\mathbf{p},\omega) = \Phi(\mathbf{q},\omega)\,\delta(p_\nu - p_\nu^{(\varepsilon)}(\mathbf{q})) + \Phi(-\mathbf{q},-\omega)^*\,\delta(p_\nu + p_\nu^{(\varepsilon)}(-\mathbf{q})^*) \tag{40}$$

where Φ is an arbitrary chooseable function, which is only determined by the boundary conditions.

2.4.2. General solution for the Fourier transform of the mechanical displacement vector

We introduce the eigenvalue equation

$$\Lambda_{jk}(\mathbf{p})\beta_k^{(n)}(\mathbf{p}) = \Lambda^{(n)}(\mathbf{p})\beta_j^{(n)}(\mathbf{p})\ , \tag{41}$$

where $\Lambda^{(n)}(\mathbf{p})$ correspond to the three (n = 1, 2, 3) eigenvalues of the piezoelectrically stiffened elastic constants with regard to the direction of the vector $\mathbf{p}$, and $\beta_j^{(n)}(\mathbf{p})$ are the respective eigenvectors, which represent an orthogonal and complete set. This is a direct consequence of the fact that the piezoelectrically stiffened elastic constants form a symmetric tensor (compare the defining Eq.(35))

$$\Lambda_{kj}(\mathbf{p}) = \Lambda_{jk}(\mathbf{p}) = \Lambda_{jk}(-\mathbf{p}) = \Lambda_{jk}(\mathbf{p})^* \ . \tag{42}$$

The orthogonality and completeness relations in normalized form are

$$\beta_k^{(m)}(\mathbf{p})\,\beta_k^{(n)}(\mathbf{p}) = \delta_{mn} \ , \tag{43a}$$

$$\sum_{n=1}^{3} \beta_j^{(n)}(\mathbf{p})\,\beta_k^{(n)}(\mathbf{p}) = \delta_{jk} \ , \tag{43b}$$

respectively. Additionally follows

$$\Lambda^{(n)}(\mathbf{p}) = \Lambda^{(n)}(-\mathbf{p}) = \Lambda^{(n)}(\mathbf{p})^* \ , \tag{44}$$

and the eigenvectors can be choosen to be real and to satisfy

$$\beta_k^{(n)}(\mathbf{p}) = \beta_k^{(n)}(-\mathbf{p}) \ . \tag{45}$$

Introducing the substitution of the Fourier transform of the mechanical displacement vector $u_i(\mathbf{p},\omega)$ by its displacements components $u^{(n)}(\mathbf{p},\omega)$ in direction of the corresponding eigenvectors $\beta_i^{(n)}(\mathbf{p})$

$$u_i(\mathbf{p},\omega) = \sum_{n=1}^{3} \beta_i^{(n)}(\mathbf{p})\, u^{(n)}(\mathbf{p},\omega) \ , \tag{46}$$

from Eq. (34) we obtain

$$[\,\Lambda^{(n)}(\mathbf{p}) - \rho\omega^2]\, u^{(n)}(\mathbf{p},\omega) = 0 \ . \tag{47}$$

This means that

$$u^{(n)}(\mathbf{p},\omega) = 0 \tag{48}$$

except for $\mathbf{p}$-vectors satisfying

$$\Lambda^{(n)}(\mathbf{p}) = \Lambda^{(n)}(\mathbf{q},p_\nu) = \rho\omega^2 \ . \tag{49}$$

Equation (49) determines the p_ν -values for any given $\mathbf{q}$-vector. There are two solutions ($s = \pm 1$; see Fig.2) for every n-value ($n = 1,2,3$) ,

$$p_\nu = p_\nu^{(ns)}(\mathbf{q},\omega) \ , \qquad \mathbf{p}^{(ns)}(\mathbf{q},\omega) = \mathbf{q} + p_\nu^{(ns)}(\mathbf{q},\omega)\,\mathbf{e}_\nu \ . \tag{50}$$

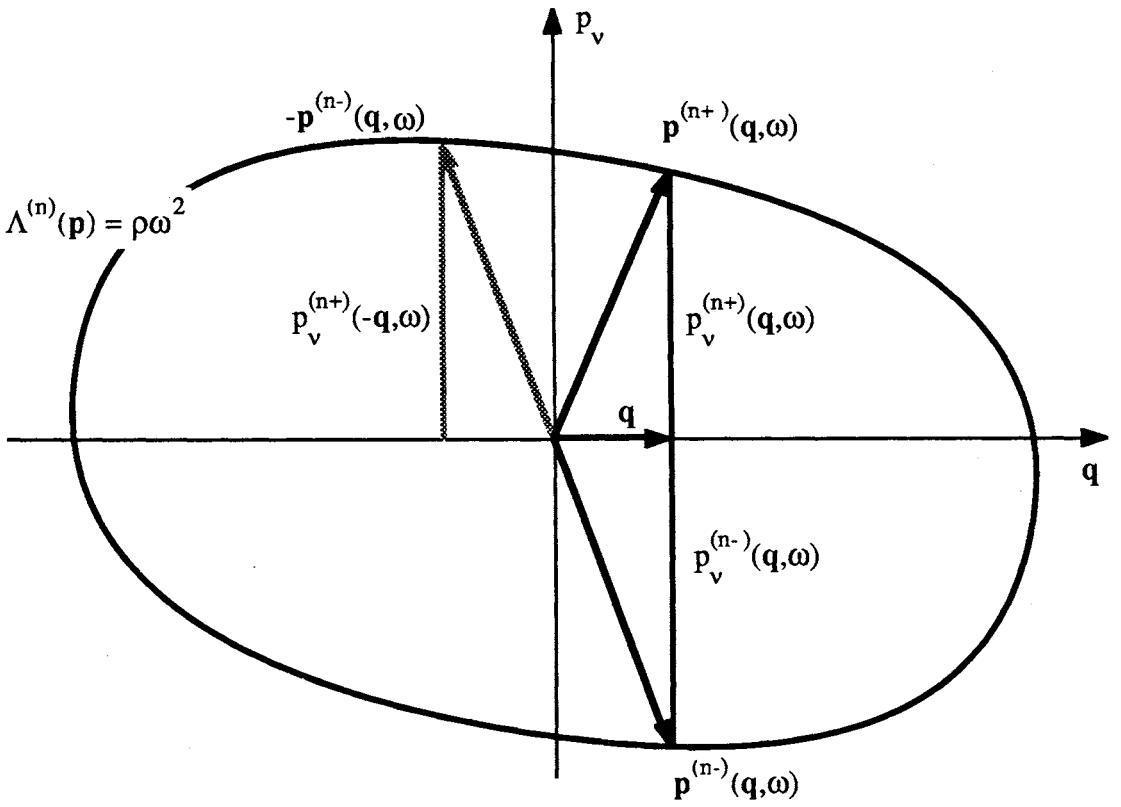


FIG.2. Possible $\mathbf{p}_\nu^{(ns)}$-vectors for non-vanishing displacements $u^{(n)}(\mathbf{p},\omega)$ and for a given circular frequency ω and the $\mathbf{q}$-vector regarded. The direction of the abscissa is choosen in coincidence with the $\mathbf{q}$-vector. The curve is analogous to the well known cross-section curves of the slowness surfaces in the $(\mathbf{r},t)$-domain.

These solutions are connected by the following equations

$$p_\nu^{(n-)}(\mathbf{q},\omega) = -p_\nu^{(n+)}(-\mathbf{q},\omega) = -p_\nu^{(n+)}(-\mathbf{q},-\omega) \ , \tag{51}$$

$$\beta_k^{(n-)}(\mathbf{q},\omega) = \beta_k^{(n+)}(-\mathbf{q},\omega) = \beta_k^{(n+)}(-\mathbf{q},-\omega) \ , \tag{52}$$

where the fact has been utilized that the $p_\nu^{(ns)}$- and $\beta_k^{(ns)}$-dependency on ω is exclusively in form of a function of ω^2 . In the following we omit the superscript s , $p_\nu^{(n)}$ and $\beta_k^{(n)}$ in each case express one of the two solutions and we take advantage of the symmetry of the two solutions for their representation.

Thus, we can write the general solution for the Fourier transforms of the displacement

vector components in the following form:

$$u_i(\mathbf{p},\omega) = \sum_{n=1}^{3} \beta_i^{(n)}(\mathbf{q},\omega)\, u^{(n)}(\mathbf{q},\omega)\, \delta(p_\nu - p_\nu^{(n)}(\mathbf{q},\omega)) + \\ + \sum_{n=1}^{3} \beta_i^{(n)}(-\mathbf{q},-\omega)\, u^{(n)}(-\mathbf{q},-\omega)^*\, \delta(p_\nu + p_\nu^{(n)}(-\mathbf{q},-\omega)) \,. \qquad (53)$$

2.5. Solution for a plane x_ν

Assuming all variables are known at a starting plane $x_\nu = 0$, we can easily calculate all respective variables in a parallel plane in a certain distance $x_\nu = \ell$ of the starting plane. If we utilize Eqs.(24), (27), (32), (40) and (53), we obtain

$$\varphi(\mathbf{q},\omega;x_\nu) = \sum_{n=1}^{3} \bar{e}_1(\mathbf{p}^{(n)}(\mathbf{q},\omega))\, \beta_1^{(n)}(\mathbf{q},\omega)\, e^{jx_\nu p_\nu^{(n)}(\mathbf{q},\omega)}\, u^{(n)}(\mathbf{q},\omega) + \\ + \sum_{n=1}^{3} \bar{e}_1(-\mathbf{p}^{(n)}(-\mathbf{q},-\omega))\, \beta_1^{(n)}(-\mathbf{q},-\omega)\, e^{-jx_\nu p_\nu^{(n)}(-\mathbf{q},-\omega)}\, u^{(n)}(-\mathbf{q},-\omega)^* + \\ + e^{jx_\nu p_\nu^{(\varepsilon)}(\mathbf{q})} \Phi(\mathbf{q},\omega) + e^{-jx_\nu p_\nu^{(\varepsilon)}(-\mathbf{q})^*} \Phi(-\mathbf{q},-\omega)^* \,, \qquad (54)$$

$$u_i(\mathbf{q},\omega;x_\nu) = \sum_{n=1}^{3} \beta_i^{(n)}(\mathbf{q},\omega)\, e^{jx_\nu p_\nu^{(n)}(\mathbf{q},\omega)}\, u^{(n)}(\mathbf{q},\omega) + \\ + \sum_{n=1}^{3} \beta_i^{(n)}(-\mathbf{q},-\omega)\, e^{-jx_\nu p_\nu^{(n)}(-\mathbf{q},-\omega)}\, u^{(n)}(-\mathbf{q},-\omega)^* \,, \qquad (55)$$

where we have used the abbreviation

$$\bar{e}_1(\mathbf{p}) = \frac{e_{ikl} p_i p_k}{\varepsilon_{mn} p_m p_n} \,. \qquad (56)$$

Finally, the stress tensor components $T_{ik}(\mathbf{q},\omega;x_\nu)$ and the dielectric displacement vector components $D_i(\mathbf{q},\omega;x_\nu)$ can be obtained by substitution of Eqs.(54) and (55) into the material equations (17) and (18). Introducing the following abbreviations

$$\gamma_i^{(n)}(\mathbf{q},\omega) = j\, c_{vikl}(\mathbf{p}^{(n)}(\mathbf{q},\omega))\, p_l^{(n)}(\mathbf{q},\omega)\, \beta_k^{(n)}(\mathbf{q},\omega) \quad , \tag{57}$$

$$k_i(\mathbf{q}) = j\, e_{kvi}\, p_k^{(\varepsilon)}(\mathbf{q}) \quad , \tag{58}$$

$$k^{(n)}(\mathbf{q},\omega) = \bar{e}_k(\mathbf{p}^{(n)}(\mathbf{q},\omega))\, \beta_k^{(n)}(\mathbf{q},\omega) \quad , \tag{59}$$

$$\bar{k}^{(n)}(\mathbf{q},\omega) = j\, e_{vlk}(\mathbf{p}^{(n)}(\mathbf{q},\omega))\, p_l^{(n)}(\mathbf{q},\omega)\, \beta_k^{(n)}(\mathbf{q},\omega) \quad , \tag{60}$$

$$\bar{\varepsilon}(\mathbf{q}) = -j\, \varepsilon_{vl} p_l^{(\varepsilon)}(\mathbf{q}) \quad , \tag{61}$$

where we used the notation

$$e_{ikl}(\mathbf{p}) = e_{ikl} - \frac{\varepsilon_{ig} p_g e_{hkl} p_h}{\varepsilon_{mn} p_m p_n} \tag{62}$$

and the analogous notation defined by Eqs. (35) and (56), we can write

$$\begin{pmatrix} u_i(\mathbf{q},\omega) \\ T_{vi}(\mathbf{q},\omega) \\ \varphi(\mathbf{q},\omega) \\ D_v(\mathbf{q},\omega) \end{pmatrix}_{x_v} = \mathcal{B}(\mathbf{q},\omega) \begin{pmatrix} e^{jx_v p_v^{(n)}(\mathbf{q},\omega)}\, u^{(n)}(\mathbf{q},\omega) \\ e^{-jx_v p_v^{(n)}(-\mathbf{q},-\omega)}\, u^{(n)}(-\mathbf{q},-\omega)^* \\ e^{jx_v p_v^{(\varepsilon)}(\mathbf{q})}\, \Phi(\mathbf{q},\omega) \\ e^{-jx_v p_v^{(\varepsilon)}(-\mathbf{q})^*}\, \Phi(-\mathbf{q},-\omega)^* \end{pmatrix} , \tag{63a}$$

$$\mathcal{B}(\mathbf{q},\omega) = \begin{pmatrix} \beta_i^{(n)}(\mathbf{q},\omega) & \beta_i^{(n)}(-\mathbf{q},-\omega)^* & \mathbf{0} & \mathbf{0} \\ \gamma_i^{(n)}(\mathbf{q},\omega) & \gamma_i^{(n)}(-\mathbf{q},-\omega)^* & k_i(\mathbf{q}) & k_i(-\mathbf{q})^* \\ k^{(n)}(\mathbf{q},\omega) & k^{(n)}(-\mathbf{q},-\omega)^* & 1 & 1 \\ \bar{k}^{(n)}(\mathbf{q},\omega) & \bar{k}^{(n)}(-\mathbf{q},-\omega)^* & \bar{\varepsilon}(\mathbf{q}) & \bar{\varepsilon}(-\mathbf{q})^* \end{pmatrix} . \tag{63b}$$

2.6. Transfer matrix description

The eight boundary values $u_i(q,\omega;x_\nu=0)$, $T_{\nu i}(q,\omega;x_\nu=0)$, $\varphi(q,\omega;x_\nu=0)$, $D_\nu(q,\omega;x_\nu=0)$, at $x_\nu = 0$ determine the eight $(n = 1,2,3)$ unknown functions

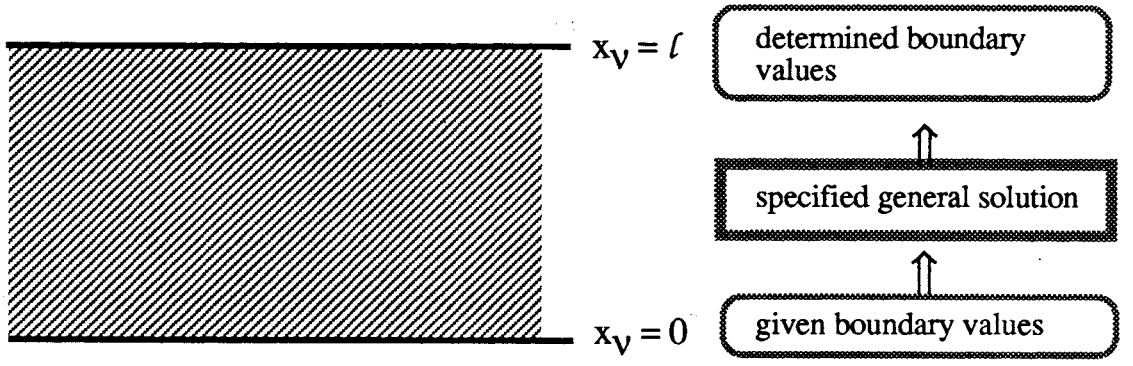


FIG.3. Concept of determining the boundary values in the plane $x_\nu = \ell$ by specifying the general solution by the boundary values in the plane $x_\nu = 0$.

$u^{(n)}(q,\omega)$, $u^{(n)}(-q,-\omega)^*$, $\Phi(q,\omega)$, $\Phi(-q,-\omega)^*$ of the general solution. This specified solution can be used to obtain the values of u_i , $T_{\nu i}$, φ , D_ν at the surface $x_\nu = \ell$. Utilizing this, we obtain a linear connection of the eight surface values at $x_\nu = 0$ and $x_\nu = \ell$

$$\begin{pmatrix} \mathbf{u}(q,\omega) \\ T_\nu(q,\omega) \\ \varphi(q,\omega) \\ D_\nu(q,\omega) \end{pmatrix}_{x_\nu = \ell} = \mathcal{M}(q,\omega;\ell) \begin{pmatrix} \mathbf{u}(q,\omega) \\ T_\nu(q,\omega) \\ \varphi(q,\omega) \\ D_\nu(q,\omega) \end{pmatrix}_{x_\nu = 0} , \tag{64}$$

$$\mathcal{M}(\mathbf{q},\omega;\ell) = \begin{pmatrix} \mathbb{M}^{uu} & \mathbb{M}^{uT} & \mathbf{M}^{u\varphi} & \mathbf{M}^{uD} \\ \mathbb{M}^{Tu} & \mathbb{M}^{TT} & \mathbf{M}^{T\varphi} & \mathbf{M}^{TD} \\ \mathbf{M}^{\varphi u} & \mathbf{M}^{\varphi T} & M^{\varphi\varphi} & M^{\varphi D} \\ \mathbf{M}^{Du} & \mathbf{M}^{DT} & M^{D\varphi} & M^{DD} \end{pmatrix}, \tag{65}$$

where the font-styles $\mathbb{M}$, $\mathbf{M}$, M , have been introduced in order to discriminate between second rank tensors (3x3 matrices), vectors (3 element columns or rows), and scalars, respectively. The 8x8 transfer matrix (65) is given by

$$\mathcal{M}(\mathbf{q},\omega;\ell) = \mathcal{B}(\mathbf{q},\omega)\mathcal{L}(\mathbf{q},\omega;\ell)\mathcal{B}(\mathbf{q},\omega)^{-1}, \tag{66}$$

where the ℓ-independent matrix $\mathcal{B}(\mathbf{q},\omega)$ is given by Eq.(63) and the ℓ-dependent matrix $\mathcal{L}(\mathbf{q},\omega;\ell)$ is a diagonal matrix of the explicite form

$$\mathcal{L}(\mathbf{q},\omega;\ell) = \begin{pmatrix} e^{j\ell p_v^{(n)}(\mathbf{q},\omega)} & 0 & 0 & 0 \\ 0 & e^{-j\ell p_v^{(n)}(-\mathbf{q},-\omega)} & 0 & 0 \\ 0 & 0 & e^{j\ell p_v^{(\varepsilon)}(\mathbf{q})} & 0 \\ 0 & 0 & 0 & e^{-j\ell p_v^{(\varepsilon)}(-\mathbf{q})^*} \end{pmatrix}. \tag{67}$$

Considering this, the following properties of the matrix $\mathcal{M}(\mathbf{q},\omega;\ell)$ can be derived immediately

$$\lim_{\ell \to 0} \mathcal{M}(\mathbf{q},\omega;\ell) = 1 , \tag{68}$$

$$\mathcal{M}(\mathbf{q},\omega;\ell)^{-1} = \mathcal{M}(\mathbf{q},\omega;-\ell) . \tag{69}$$

3. Description of a layered resonator structure

The physical boundary condition for the interfacing surfaces of two layers in a plane

x_ν = const. is the identity of the eight chosen variables $\mathbf{u}$, $\mathbf{T}_\nu$, φ, D_ν, for the transfer matrix description. Therefore, the total transfer matrix for a layered structure can be obtained

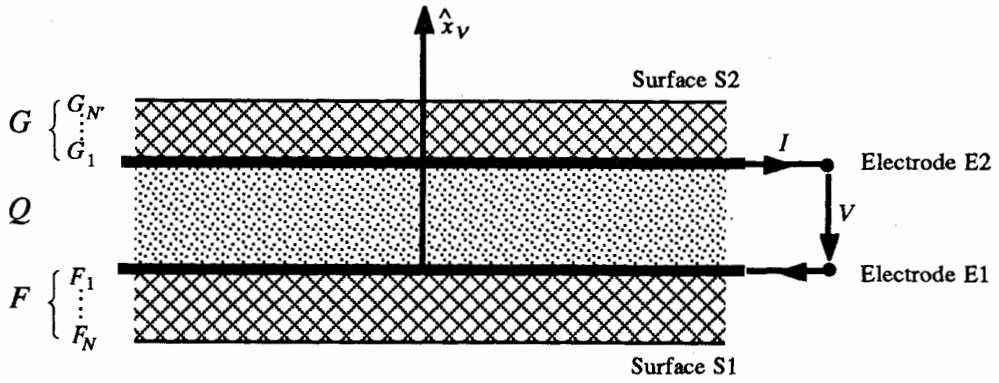


FIG.4. Structural division of an arbitrary layer configuration with two electrodes, where Q represents the film stack between the electrodes and F as well as G represents the film stacks outside the electrodes.

directly by matrix multiplications of the transfer matrices of each film, e.g. for the lower film stack F in Fig.4

$$\mathcal{M}_F = \mathcal{M}_{F1}\mathcal{M}_{F2}\ldots\mathcal{M}_{FN}\,. \tag{70}$$

For the analysis of the total arrangement shown in Fig.4, we have to consider the influence of the two electrodes carefully. The electrodes provide a constant electrical potential, the electric surface charge of these electrodes, however, is determined by the vicinity of the electrodes, not by the electrodes itself. This means, we are not able to define a transfer matrix description for the electrodes, as it was possible in the one-dimensional analysis /7/. We can only use seperate total transfer matrix descriptions for the two film stacks F and G outside the electrodes (compare Fig.4)

$$\underset{S2}{\begin{pmatrix}\mathbf{u}\\ \mathbf{T}_\nu\\ \varphi\\ D_\nu\end{pmatrix}} = \mathcal{M}_G \underset{E2}{\begin{pmatrix}\mathbf{u}\\ \mathbf{T}_\nu\\ \varphi\\ D_\nu^{+}\end{pmatrix}}, \qquad \underset{E1}{\begin{pmatrix}\mathbf{u}\\ \mathbf{T}_\nu\\ \varphi\\ D_\nu^{-}\end{pmatrix}} = \mathcal{M}_F \underset{S1}{\begin{pmatrix}\mathbf{u}\\ \mathbf{T}_\nu\\ \varphi\\ D_\nu\end{pmatrix}} \tag{71}$$

as well as for the film stack Q between the electrodes

$$\begin{pmatrix} \mathbf{u} \\ T_\nu \\ \varphi \\ D_\nu^- \end{pmatrix}_{E2} = \mathcal{M}_Q \begin{pmatrix} \mathbf{u} \\ T_\nu \\ \varphi \\ D_\nu^+ \end{pmatrix}_{E1}, \tag{72}$$

where we have used the notation D_ν^+ and D_ν^- in order to discriminate between the different values of D_ν on the upper and lower side of each electrode. In contrast to that, we have made no distinction of the mechanical variables $\mathbf{u}$ and T_ν on both sides of an electrode, because for sufficient thin electrodes these values should be the same (see Eq.(68)). For thick electrodes one has to distinct between the values on both electrode surfaces, the connection between these values can than be provided simply by the mechanical part of the transfer matrix.

3.1. Resonator analysis in the Fourier transform domain

For the connection between the surface values of Eqs. (71) and (72) we restrict to non-piezoelectric materials F and G outside the electrodes with free surfaces $S1$ and $S2$, which means

$$T_\nu(\mathbf{q},\omega) = 0 \ , \quad D_\nu(\mathbf{q},\omega) = 0 \ , \ \text{on surfaces } S1 \text{ and } S2 . \tag{73}$$

Furthermore, the electric potential on the electrodes is chosen as

$$\varphi(\mathbf{q},\omega) = \frac{1}{2} V(\omega)\, \delta(\mathbf{q}) \ , \ \text{for electrode } E1 , \tag{74}$$

$$\varphi(\mathbf{q},\omega) = -\frac{1}{2} V(\omega)\, \delta(\mathbf{q}) \ , \ \text{for electrode } E2 . \tag{75}$$

Hence, the interesting parts of Eqs. (71) and (72) can be written as

$$\mathbf{u}_{S2} = \mathbb{M}_G^{uu}\, \mathbf{u}_{E2} + \mathbb{M}_G^{uT}\, T_{\nu E2} \ , \tag{76}$$

$$0 = \mathbb{M}_G^{Tu}\, \mathbf{u}_{E2} + \mathbb{M}_G^{TT}\, T_{\nu E2} \ , \tag{77}$$

$$\mathbf{u}_{E2} = \mathbb{M}_Q^{uu}\, \mathbf{u}_{E1} + \mathbb{M}_Q^{uT}\, T_{\nu E1} + \mathbb{M}_Q^{uD}\, D_{E1}^+ \ , \tag{78}$$

$$T_{\nu E2} = \mathbb{M}_Q^{Tu}\, \mathbf{u}_{E1} + \mathbb{M}_Q^{TT}\, T_{\nu E1} + \mathbb{M}_Q^{TD}\, D_{E1}^+ \ , \tag{79}$$

$$-V(\omega)\, \delta(\mathbf{q}) = \mathbb{M}_Q^{\varphi u}\, \mathbf{u}_{E1} + \mathbb{M}_Q^{\varphi T}\, T_{\nu E1} + \mathbb{M}_Q^{\varphi D}\, D_{E1}^+ \ , \tag{80}$$

$$\mathbf{u}_{E1} = \mathbb{M}_F^{uu}\,\mathbf{u}_{S1} \,, \tag{81}$$

$$T_{\nu E1} = \mathbb{M}_F^{Tu}\,\mathbf{u}_{S1} \,. \tag{82}$$

Thereby we could use the transfer matrix given in Ref./7/ which is valid for $\mathbf{q} = \mathbf{0}$, since the terms with the electrode potential φ contain a factor $\delta(\mathbf{q})$. Eliminating D_{E1}^{+} in Eqs. (78) and (79) by substitution of the explicite expression for D_{E1}^{+} obtained from Eq. (80) and utilizing for $\mathbf{u}_{E1}$ and $T_{\nu E1}$ the expressions given by Eqs. (81), (82), we obtain $\mathbf{u}_{E2}$ and $T_{\nu E2}$ as functions of $\mathbf{u}_{S1}$ and $V(\omega)$. Introducing these functions in Eq. (77) yields

$$\underline{\mathbb{M}}_{GQF}^{Tu}(\mathbf{q},\omega)\,\mathbf{u}(\mathbf{q},\omega;x_\nu{=}S1) = \underline{\mathbf{M}}_{GQ}^{T\varphi}(\mathbf{q},\omega)\,V(\omega)\,\delta(\mathbf{q}) \,, \tag{83}$$

with

$$\underline{\mathbb{M}}_{GQF}^{Tu} = \mathbb{M}_G^{Tu}\,\underline{\mathbb{M}}_Q^{uu}\,\mathbb{M}_F^{uu} + \mathbb{M}_G^{Tu}\,\underline{\mathbb{M}}_Q^{uT}\,\mathbb{M}_F^{Tu} + \mathbb{M}_G^{TT}\,\underline{\mathbb{M}}_Q^{Tu}\,\mathbb{M}_F^{uu} + \mathbb{M}_G^{TT}\,\underline{\mathbb{M}}_Q^{TT}\,\mathbb{M}_F^{Tu}, \tag{84}$$

$$\underline{\mathbb{M}}_Q^{xy} = \mathbb{M}_Q^{xy} - \mathbf{M}_Q^{xD}\frac{1}{M_Q^{\varphi D}}\mathbf{M}_Q^{\varphi y} \,, \tag{85}$$

$$\underline{\mathbf{M}}_{GQ}^{T\varphi} = (\mathbb{M}_G^{Tu}\,\mathbf{M}_Q^{uD} + \mathbb{M}_G^{TT}\,\mathbf{M}_Q^{TD})\frac{1}{M_Q^{\varphi D}} \,. \tag{86}$$

Equation (83) determines the displacement $\mathbf{u}(\mathbf{q},\omega;x_\nu{=}S1)$ at the free surface $S1$ if an electric voltage $V(\omega)$ is applied on the electrodes. The electric current $I(\omega)$ via the electrodes can be calculated from the change of the surface charge per unit time. The surface charge itself is connected to the difference of the dielectric displacement vector components perpendicular to the electrode surfaces on both sides of the electrode according to

$$D_{E1}^{+} - D_{E1}^{-} = -\underline{\mathbf{M}}_{QF}^{\varphi u}\,\mathbf{u}_{S1} - \frac{1}{M_Q^{\varphi D}}V(\omega)\,\delta(\mathbf{q}) \,, \tag{87}$$

with

$$\underline{\mathbf{M}}_{QF}^{\varphi u} = \frac{1}{M_Q^{\varphi D}}(\mathbf{M}_Q^{\varphi u}\,\mathbb{M}_F^{uu} + \mathbf{M}_Q^{\varphi T}\,\mathbb{M}_F^{Tu}) \,. \tag{88}$$

The homogeneous equation obtained from Eq. (83) for $V(\omega) = 0$ determines the resonance frequencies of the total sandwich arrangement. If we multiply this homogeneous equation by $[\mathbb{M}_G^{TT}]^{-1}$ and use as unknown displacement function

$$\mathbb{M}_F^{uu}\, \mathbf{u}_{S1} = \mathbf{u}_{E1} = \mathbf{u} \ , \tag{89}$$

we obtain

$$\underline{M}_{ij}^{Tu}(q,\omega)\, u_j(q,\omega) = 0 \ , \tag{90}$$

with

$$\underline{\mathbb{M}}^{Tu} = \underline{\mathbb{M}}_Q^{Tu} + \mathbb{M}_G \underline{\mathbb{M}}_Q^{uu} + \underline{\mathbb{M}}_Q^{TT} \mathbb{M}_F + \mathbb{M}_G \underline{\mathbb{M}}_Q^{uT} \mathbb{M}_F \ , \tag{91}$$

$$\mathbb{M}_G = [\mathbb{M}_G^{TT}]^{-1}\, \mathbb{M}_G^{Tu} \ , \tag{92}$$

$$\mathbb{M}_F = \mathbb{M}_F^{Tu}\, [\mathbb{M}_F^{uu}]^{-1} \ . \tag{93}$$

It is emphasized that the whole dependency of the resonance frequencies on a film stack G or F outside the electrodes is described by one 3x3 matrix $\mathbb{M}_G$ or $\mathbb{M}_F$, respectively.

In order to get the solutions of the homogeneous Eq. (90), it is favourable to solve the eigenvalue problem

$$\underline{M}_{ij}^{Tu}(q,\omega^2)\, \mu_j^{(\lambda)}(q,\omega^2) = m^{(\lambda)}(q,\omega^2)\, \mu_i^{(\lambda)}(q,\omega^2) \ , \tag{94}$$

where we have explicitely indicated the fact that the dependency on ω is exclusively in form of a function of ω^2 and the three different solutions are discriminated by $\lambda = 1,2,3$. Equation (94) is directly comparable with Eq. (41). Thus, we can write in analogy to Eq. (46)

$$u_i(q,\omega^2) = \sum_{\lambda=1}^{3} \mu_i^{(\lambda)}(q,\omega^2)\, u^{(\lambda)}(q,\omega^2) \ , \tag{95}$$

and obtain by substitution into Eq. (90) the three equations

$$m^{(\lambda)}(q,\omega^2)\, u^{(\lambda)}(q,\omega^2) = 0 \ , \qquad \lambda = 1,2,3 \ , \tag{96}$$

Hence, the general solution of Eq. (90) can be written as

$$u_i(q,\omega^2) = \sum_{\lambda=1}^{3} \mu_i^{(\lambda)}(q,\omega^2) \sum_n u^{(\lambda n)}(q)\, \delta(\omega^2 - \omega_{\lambda n}^2(q)) \ , \tag{97}$$

where the functions $\omega_{\lambda n}^2(q)$ are the solutions of

$$m^{(\lambda)}(q,\omega_{\lambda n}^2(q)) \equiv 0 \tag{98}$$

for each λ-value and the $u^{(\lambda n)}(\mathbf{q})$ are arbitrary functions of $\mathbf{q}$ (compare Eq. (53)). For every branch of interest we can also write

$$[\omega^2 - \omega_{\lambda n}^2(\mathbf{q})]\, u^{(\lambda n)}(\mathbf{q},\omega^2) = 0 \quad , \tag{99}$$

using the fact that

$$x\delta(x) = 0 \quad . \tag{100}$$

The dispersion relations $\omega_{\lambda n}^2(\mathbf{q})$ following from Eq. (98) represent the entire resonance vibration spectrum. From the general symmetry

$$\mathbb{M}(-\mathbf{q},-\omega)^* = \mathbb{M}(\mathbf{q},\omega) \quad , \tag{101}$$

(compare Eqs. (19) to (22)), together with the fact that the transfer matrix is only a function of ω^2, one obtains

$$\omega_{\lambda n}^2(-\mathbf{q}) = \omega_{\lambda n}^2(\mathbf{q}) \quad , \tag{102}$$

which is a valuable symmetry equation for the dispersion relation.

3.2. Resonator analysis in the space-domain

Using the transformation equation

$$u_i(\mathbf{r}_s,\omega; x_\nu) = \int d^2q\, e^{j\mathbf{q}\mathbf{r}_s}\, u_i(\mathbf{q},\omega; x_\nu) \tag{103}$$

and analogous formulas for $T_{\nu i}$, φ and D_ν we can write Eq. (64) in the integral form

$$\begin{pmatrix} \mathbf{u}(\mathbf{r}_{s2},\omega) \\ T_\nu(\mathbf{r}_{s2},\omega) \\ \varphi(\mathbf{r}_{s2},\omega) \\ D_\nu(\mathbf{r}_{s2},\omega) \end{pmatrix}_{x_\nu = \ell} = \int d^2r_{s1}\, \mathcal{M}(\mathbf{r}_{s2}-\mathbf{r}_{s1},\omega;\ell) \begin{pmatrix} \mathbf{u}(\mathbf{r}_{s1},\omega) \\ T_\nu(\mathbf{r}_{s1},\omega) \\ \varphi(\mathbf{r}_{s1},\omega) \\ D_\nu(\mathbf{r}_{s1},\omega) \end{pmatrix}_{x_\nu = 0} , \tag{104a}$$

with the space domain transfer matrix

$$\mathcal{M}(\mathbf{r}_s,\omega;\ell) = \frac{1}{(2\pi)^2} \int d^2q\, e^{j\mathbf{q}\mathbf{r}_s}\, \mathcal{M}(\mathbf{q},\omega;\ell) \quad . \tag{104b}$$

In order to obtain the total transfer matrix of a layered structure, we have to use the

convolution theorem to transform the matrix multiplication of Eq. (84) into the space domain. Equation (83) can be written as

$$\int d^2 r_{s1}\, \underline{\underline{M}}^{Tu}_{GQF}(\mathbf{r}_{s2}-\mathbf{r}_{s1},\omega)\, \mathbf{u}(\mathbf{r}_{s1},\omega) \;=\; \underline{\underline{M}}^{T\varphi}_{GQ}(\mathbf{q}=\mathbf{0},\omega)\, V(\omega) \qquad (105)$$

and from Eq. (87) follows

$$\frac{Y(\omega)}{j\omega A} V(\omega) = -\langle \underline{\underline{M}}^{\varphi u}_{QF}\, \mathbf{u}_{s1}\rangle - \left\langle \frac{1}{M^{\varphi D}_{Q}} \right\rangle V(\omega) \;, \qquad (106)$$

where we have indicated the mean value within the surface area by $\langle \ldots \rangle$,

$$\langle \underline{\underline{M}}^{\varphi u}_{QF}\, \mathbf{u}_s \rangle = \frac{1}{A} \int_A d^2 r_s\, \underline{\underline{M}}^{\varphi u}_{QF}(\mathbf{q}=\mathbf{0},\omega;\, \ell(\mathbf{r}_s))\mathbf{u}(\mathbf{r}_s) \;, \qquad (107)$$

the integration is made over the area A of the electrodes, and the total electric current $I(\omega)$ is given by

$$I(\omega) = Y(\omega)\, V(\omega) \;. \qquad (108)$$

In principle it is possible to calculate $V(\omega)$ from Eq. (106) as a function of $\mathbf{u}$ and

$$Y(\omega) = -Y_{ext}(\omega) \;, \qquad (109)$$

where $Y_{ext}(\omega)$ is the external admittance between the electrodes, and introduce it in equation (105). This specified equation than determines the resonance frequencies of an arrangement with an external admittance.

From the Fourier transformation of Eq. (99) we obtain

$$\int d^2 q\, e^{j\mathbf{q}\mathbf{r}_s} [\omega^2 - \omega^2_{\lambda n}(\mathbf{q})] \frac{1}{(2\pi)^2} \int d^2 r'_s\, e^{j\mathbf{q}\mathbf{r}'_s}\, u^{(\lambda n)}(\mathbf{r}'_s, \omega^2) = 0 \;, \qquad (110a)$$

or, after interchanging the integrations

$$\frac{1}{(2\pi)^2} \int d^2 r'_s \int d^2 q\, [\omega^2 - \omega^2_{\lambda n}(\mathbf{q})]\, e^{j\mathbf{q}(\mathbf{r}_s-\mathbf{r}'_s)}\, u^{(\lambda n)}(\mathbf{r}'_s, \omega^2) = 0 \;. \qquad (110b)$$

For functions $\mathbf{u}(\mathbf{r}_s,\omega)$ which are changing slowly along the surface, the most important Fourier components are those in the vicinity of $\mathbf{q}=\mathbf{0}$. For such small $\mathbf{q}$-vectors a Taylor series expansion for the dispersion relation $\omega^2_{\lambda n}(\mathbf{q})$ can be used

$$\omega_{\lambda n}^2(\mathbf{q}) = \omega_{\lambda n}^2(\mathbf{0}) + \frac{1}{2}\left(\frac{\partial^2 \omega_{\lambda n}^2}{\partial q_i \partial q_j}\right)_{\mathbf{0}} q_i q_j + \ldots . \qquad (111)$$

It should be noted that there are no first order terms, which can be seen from the symmetry relation (102). Introducing this into Eq. (110b) and using the substitution

$$q_i e^{j\mathbf{q}\mathbf{r}_s} = -j\frac{\partial}{\partial x_i} e^{j\mathbf{q}\mathbf{r}_s} \quad , \qquad (112)$$

we obtain finally

$$\left(\frac{1}{2}\left(\frac{\partial^2 \omega_{\lambda n}^2}{\partial q_i \partial q_j}\right)_{\mathbf{0}} \frac{\partial}{\partial x_i}\frac{\partial}{\partial x_j} - \omega_{\lambda n}^2(\mathbf{0}) + \omega^2\right) u^{(\lambda n)}(\mathbf{r}_s, \omega^2) = 0 . \qquad (113)$$

Keeping in mind the ℓ-dependance of $\omega_{\lambda n}^2(\mathbf{q}; \ell)$, which follows from the ℓ-dependance of the transfer matrices that causes subsequently a thickness dependence of the eigenvalue function $m^{(\lambda)}(\mathbf{q}, \omega^2; \ell)$, the form of Eq. (113) is very advantageous, because it allows the consideration of slowly varying thicknesses $\ell(\mathbf{r}_s)$. The calculation of the second derivative of the dispersion relation $\omega_{\lambda n}^2(\mathbf{q})$ can be done by differentiation of Eq. (98). Taking advantage of the fact that the first derivatives are zero, we obtain

$$\left(\frac{\partial^2 \omega_{\lambda n}^2}{\partial q_i \partial q_j}\right)_{\mathbf{0}} = -\left(\frac{\dfrac{\partial^2 m^{(\lambda)}}{\partial q_i \partial q_j}}{\dfrac{\partial m^{(\lambda)}}{\partial \omega^2}}\right)_{\mathbf{q}=\mathbf{0},\ \omega=\omega(\mathbf{0})} . \qquad (114)$$

4. Treatment of a composite quartz crystal resonator with a convex boundary sphere

Considering a contoured quartz crystal resonator core and an additional non-piezoelectric film (or film stack) at the plane crystal side according to Fig. 5, the $\mathbf{r}_s$-dependence of the quartz thickness ℓ_Q can be approximated after *Wilson* /9/ and *Tiersten* /10,11/ by

$$\ell_0 - \ell_Q(\mathbf{r}_s) = \frac{x^2 + z^2}{2R} \quad , \qquad (115)$$

where ℓ_0 is the maximum thickness of the contoured crystal at $x = z = 0$; x, z are the coordinates within the plane crystal surface, y is the coordinate in the ν-direction, and R is the radius of the convex curvature (see Fig. 5). We use now the approximation

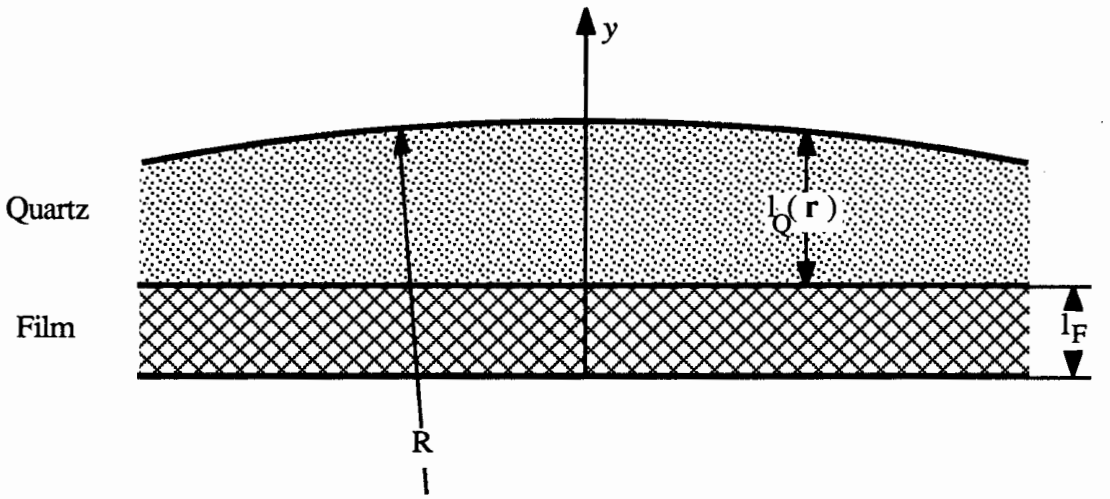


FIG.5. Contoured quartz crystal resonator core and an additional non-piezoelectric film (or film stack)

$$\omega_{\lambda n}^2(\mathbf{q}=\mathbf{0};\, \ell_Q,\ell_F) = \omega_{\lambda n}^2(\mathbf{q}=\mathbf{0};\, \ell_0,\ell_F) + \left(\frac{\partial \omega_{\lambda n}^2}{\partial \ell_Q}\right)_{\ell_0} (\ell_Q - \ell_0) + \ldots\ldots , \qquad (116)$$

and write for the first derivation term in Eq. (116)

$$\left(\frac{\partial \omega_{\lambda n}^2}{\partial \ell_Q}\right)_{\ell_0} = \frac{2}{\underline{\ell}^{(\lambda n)}}\, \omega_{\lambda n}^2(\mathbf{q}=\mathbf{0};\, \ell_0,\ell_F) \quad , \qquad (117)$$

which defines an effective thickness $\underline{\ell}^{(\lambda n)}$. A similar definition is introduced by the equation

$$\frac{1}{2}\left(\frac{\partial^2 \omega_{\lambda n}^2}{\partial q_i \partial q_j}\right)_{\mathbf{0}} = \frac{C_{ij}^{(\lambda n)}}{\underline{\rho}} \quad , \qquad (118)$$

which defines an effective ratio of elastic stiffness constants $C_{ij}^{(\lambda n)}$ and material density $\underline{\rho}$. Utilizing these definitions, we obtain from Eq. (113)

$$\left(- \frac{C_{ij}^{(\lambda n)}}{\varrho} \frac{\partial}{\partial x_i} \frac{\partial}{\partial x_j} + \frac{\omega_{\lambda n}^2(0)}{\ell^{(\lambda n)}} (x^2 + z^2) \right) u^{(\lambda n)}(x,z) =$$

$$= [\omega^2 - \omega_{\lambda n}^2(0)]\, u^{(\lambda n)}(x,z) \quad . \qquad (119)$$

After rotating the x,z-plane in order to obtain a diagnal form of $C_{ij}^{(\lambda n)}$, this is the differential equation for a two-dimensional harmonic oscillator. The eigenfunctions and eigenvalues of this problem are well known /10/.

4.1. Specialization to a plane-convex quartz crystal

In order to solve a concrete example, one has to calculate all material functions and the transfer matrices, solve the eigenvalue problem of Eq. (94) to obtain the function $m^{(\lambda)}(\mathbf{q},\omega^2)$ and then calculate the coefficients for Eq. (119). Approximately we can calculate the transfer matrix $\mathcal{M}(\mathbf{q},\omega^2)$ for small q-values as (see Eq. (66))

$$\mathcal{M}(\mathbf{q},\omega^2) = \mathcal{B}(\mathbf{0},\omega^2)\mathcal{L}(\mathbf{q},\omega^2;\ell)\mathcal{B}(\mathbf{0},\omega^2)^{-1} , \qquad (120)$$

that is, only the q-dependency in connection with the thickness is considered. With this approximation - after a simple algebraic rearrangement - we obtain immediately the following formulas

$$M_{ij}^{uu}(\mathbf{q},\omega^2) = M_{ij}^{TT}(\mathbf{q},\omega^2) = \sum_n \beta_i^{(n)}\beta_j^{(n)}\, \underline{\cos}^{(n)}(\mathbf{q},\omega^2) , \qquad (121)$$

$$M_{ij}^{uT}(\mathbf{q},\omega^2) = \sum_n \beta_i^{(n)}\beta_j^{(n)} \frac{\alpha^{(n)}}{\rho\omega}\, \underline{\sin}^{(n)}(\mathbf{q},\omega^2) , \qquad (122)$$

$$M_{ij}^{Tu}(\mathbf{q},\omega^2) = -\sum_n \beta_i^{(n)}\beta_j^{(n)} \frac{\rho\omega}{\alpha^{(n)}}\, \underline{\sin}^{(n)}(\mathbf{q},\omega^2) , \qquad (123)$$

$$M_i^{\varphi u}(\mathbf{q},\omega^2) = M_i^{TD}(\mathbf{q},\omega^2) = -\sum_n \beta_i^{(n)} k^{(n)} [1 - \underline{\cos}^{(n)}(\mathbf{q},\omega^2)] , \qquad (124)$$

$$M_i^{\varphi T}(\mathbf{q},\omega^2) = M_i^{uD}(\mathbf{q},\omega^2) = \sum_n \beta_i^{(n)} k^{(n)} \frac{\alpha^{(n)}}{\rho\omega}\, \underline{\sin}^{(n)}(\mathbf{q},\omega^2) , \qquad (125)$$

$$M^{\varphi D}(\mathbf{q},\omega^2) = -\frac{\ell}{\varepsilon_{vv}} + \sum_n k^{(n)2} \frac{\alpha^{(n)}}{\rho\omega}\, \underline{\sin}^{(n)}(\mathbf{q},\omega^2) , \qquad (126)$$

$$M^{\varphi\varphi}(\mathbf{q},\omega^2) = M^{DD}(\mathbf{q},\omega^2) = 1 , \qquad (127)$$

all other terms are zero. Hereby we have used the angular functions

$$\underline{\cos}^{(n)}(\mathbf{q},\omega^2) = \frac{1}{2}\left[e^{j\ell p_\nu^{(n)}(\mathbf{q},\omega)} + e^{-j\ell p_\nu^{(n)}(-\mathbf{q},\omega)}\right] , \tag{128a}$$

$$\underline{\sin}^{(n)}(\mathbf{q},\omega^2) = -\frac{1}{2}\left[e^{j\ell p_\nu^{(n)}(\mathbf{q},\omega)} - e^{-j\ell p_\nu^{(n)}(-\mathbf{q},\omega)}\right] , \tag{128b}$$

which specialize for $\mathbf{q}=\mathbf{0}$ to the common cosine and sine functions, and the substitutions

$$p_\nu^{(n)}(\mathbf{q},\omega) = \omega\,\alpha^{(n)}(\mathbf{q}/\omega) \;,\quad \alpha^{(n)}(\mathbf{0}) = \alpha^{(n)} \;. \tag{129}$$

Introducing in analogy to Eq. (92)

$$\mathbb{M}_Q = [\mathbb{M}_Q^{TT}]^{-1}\,\mathbb{M}_Q^{Tu} \;, \tag{130}$$

we obtain for the scalar description of a single shear mode

$$\mathbb{M}_Q = -\frac{\rho_Q\omega}{\alpha_Q}\tan\left(\frac{\ell_Q}{2}[p_\nu(\mathbf{q},\omega^2) + p_\nu(-\mathbf{q},\omega^2)]\right) \;, \tag{131}$$

which, in this case, can be considered directly as the eigenvalue-function $m^{(\lambda)}(\mathbf{q},\omega^2;\ell_Q)$. The solution of Eq. (98) can therefore be written as

$$\frac{\ell_Q}{2}[p_\nu(\mathbf{q},\omega^2) + p_\nu(-\mathbf{q},\omega^2)] = n\pi \;, \tag{132}$$

This equation determines $\omega_n^2(\mathbf{q})$. If we take into account that the functions

$$-\left(\frac{1}{\alpha_Q}\frac{\partial^2\alpha_Q}{\partial(q_i/\omega)\partial(q_j/\omega)}\right)_{\mathbf{q}=\mathbf{0}} = \frac{C_{ij}^Q}{\rho_Q} \tag{133}$$

are ω-independent pure material functions, we obtain from Eq. (132)

$$\omega_n^2(\mathbf{0};\ell_Q) = \left(\frac{n\pi}{\alpha_Q\,\ell_Q}\right)^2 , \tag{134}$$

$$\frac{1}{2}\left(\frac{\partial^2\omega_n^2}{\partial q_i\partial q_j}\right)_{\mathbf{0}} = \frac{C_{ij}^Q}{\rho_Q} \;. \tag{135}$$

Using Eq. (134), we obtain from Eq. (117) $\ell = \ell_0$. Assuming that the material coefficients C_{ij}^Q form a diagonal matrix (which can be always achieved by a rotation of the surface coordinates /11/), we can write Eq. (119) in the form

$$\left\{\left(-\frac{C^Q_{xx}}{\rho_Q}\frac{\partial^2}{\partial x^2}+\frac{\omega_n^2(0)}{R\ell_0}x^2\right)+\left(-\frac{C^Q_{zz}}{\rho_Q}\frac{\partial^2}{\partial z^2}+\frac{\omega_n^2(0)}{R\ell_0}z^2\right)\right\}u^{(n)}(x,z) =$$

$$= [\omega^2-\omega_n^2(0)]\,u^{(n)}(x,z)\ , \qquad (136)$$

which is the same result as it was obtained by *Wilson* /9/, basing upon a different treatment:

$$\omega^2_{n,n_x,n_z} = \omega_n^2(0)\left(1+\frac{1}{\omega_n(0)}\sqrt{\frac{C^Q_{xx}}{R\rho_Q\ell_0}}(2n_x+1)+\frac{1}{\omega_n(0)}\sqrt{\frac{C^Q_{zz}}{R\rho_Q\ell_0}}(2n_z+1)\right). \qquad (137)$$

Since the second and third term within the bracket is for a sufficient flat curvature (R large) much smaller than 1 , a binomial expansion can be used to write the resonance frequency equation in a more convenient form

$$\omega_{n,n_x,n_z} = \frac{n\pi}{\alpha_Q\ell_0}+\sqrt{\frac{C^Q_{xx}}{R\rho_Q\ell_0}}(n_x+\tfrac{1}{2})+\sqrt{\frac{C^Q_{zz}}{R\rho_Q\ell_0}}(n_z+\tfrac{1}{2})\ , \qquad (138)$$

with the harmonic mode numbers $n = 1,2,3,\ldots$
and the anharmonic mode numbers $n_x = 0, 1, 2,\ldots$ and $n_z = 0, 1, 2,\ldots$.

4.2. Treatment of a plane-convex crystal with one additional layer

In the simplest approximation we obtain

$$z_Q\tan\left(\frac{\ell_Q}{2}[p_Q(q,\omega^2)+p_Q(-q,\omega^2)]\right)+z_F\tan\left(\frac{\ell_F}{2}[p_F(q,\omega^2)+p_F(-q,\omega^2)]\right)=0\ , \qquad (139)$$

in full analogy to the results /7/ for $q=0$. In Eq. (139) we have introduced the specific acoustic impedances

$$z_Q=\frac{\rho_Q}{\alpha_Q}\ , \qquad z_F=\frac{\rho_F}{\alpha_F}\ . \qquad (140)$$

For thin films F we can write equation (139) in the approximate form

$$\frac{\ell_Q}{2}[p_Q(q,\omega^2)+p_Q(-q,\omega^2)] = n\pi-\frac{z_F}{z_Q}\frac{\ell_F}{2}[p_F(q,\omega^2)+p_F(-q,\omega^2)]\ . \qquad (141)$$

Now we are able to calculate

$$\omega_n^2(\mathbf{q}=\mathbf{0};\, \ell_Q,\ell_F) = \left(\frac{n\pi}{\alpha_Q \ell_0}\right)^2 \left(\frac{1}{\frac{\ell_Q}{\ell_0} + \eta}\right)^2 , \tag{142}$$

where we have used the abbreviation η for the ratio of the areal mass densities

$$\eta = \frac{z_F \alpha_F \ell_F}{z_Q \alpha_Q \ell_0} = \frac{\rho_F \ell_F}{\rho_Q \ell_0} , \tag{143}$$

and the second derivatives

$$\frac{1}{2}\left(\frac{\partial^2 \omega_n^2}{\partial q_i \partial q_j}\right)_0 = \frac{C_{ij}^Q}{\rho_Q}\left\{1 - \eta\left(1 - \frac{C_{ij}^F}{\rho_F}\frac{\rho_Q}{C_{ij}^Q}\right)\right\} . \tag{144}$$

From Eq. (142) we can calculate $\underset{\sim}{\ell}^{(n)}$ as defined by Eq. (117)

$$\underset{\sim}{\ell}^{(n)} = \ell_0(1+\eta) . \tag{145}$$

Substituting the values given by Eqs. (142), (144), (145) into (119), assuming a diagonal form of the C_{ij}^Q and an isotropic film F as well as using a binomial expansion in analogy to Eqs. (137) and (138) yields

$$\omega_{n,n_x,n_z} = \frac{n\pi}{\alpha_Q \ell_0}(1-\eta) + \sqrt{\frac{C_{xx}^Q}{R\rho_Q \ell_0}}\,(n_x+\tfrac{1}{2})\left(1 - \eta + \eta\,\frac{1}{2}\frac{C^F}{\rho_F}\frac{\rho_Q}{C_{xx}^Q}\right) +$$

$$+ \sqrt{\frac{C_{zz}^Q}{R\rho_Q \ell_0}}\,(n_z+\tfrac{1}{2})\left(1 - \eta + \eta\,\frac{1}{2}\frac{C^F}{\rho_F}\frac{\rho_Q}{C_{zz}^Q}\right) . \tag{146}$$

From Eq. (146) we can learn that the difference between the anharmonic overtones should change only half compared with the absolute value of the harmonic if the film has the same ratio C/ρ as quartz. For a Cu- and an Al-film upon an AT-cut quartz crystal from Eq. (146) can be followed that the slope of the reduced frequency versus thickness curve is greater than $\eta/2$. This higher slope seems to be in good agreement with experimental data shown in Fig.6 for plano-convex AT-cut crystals with a fundamental frequency of 6 MHz, radius of curvature 45 cm, plate diameter 14 mm, and a deposition aperture diameter of 12 mm.

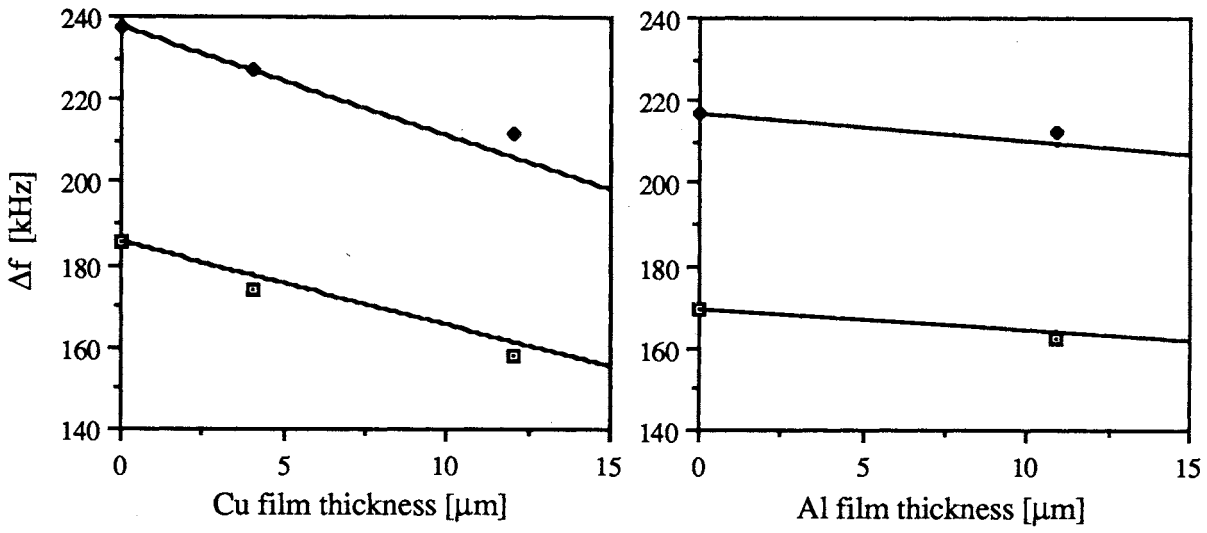


FIG.6. Comparison of measured (dots) and calculated (lines) frequency spacings Δf between the fundamental mode and the first two anharmonics (1,0,2) and (1,2,0) for plano-convex AT-cut crystals loaded with copper (left) and aluminum (right). The curvature of the crystal was adjusted so that the calculated values match the measured frequency spacings for the unloaded crystal.

5. Conclusion

A matrix description for piezoelectric resonators with two electrodes and an arbitrary number of plane parallel layers has been derived utilizing the Fourier transforms of the mechanical and electrical variables. It includes the full set of material constants and covers all the three components of displacement and the electrical field. An approximation considering only vibration modes with small transverse variations and decoupling of the modes was used to obtain a scalar differential equation for the displacement in space domain, which can be applied to the problem of layered contoured resonators. For the case of plano-convex geometry of the piezoelectric resonator core, e.g. a plano-convex quartz crystal, the frequency spectrum was derived and solved for a crystal coated with one thin film. Although the capabilities of the theoretical model are not fully exhausted by this first proofable approximation formula, good agreement with measured data of plano-convex AT-cut crystals loaded with films of copper and aluminum could be shown. The presented theoretical model can be regarded as a fundamental step forward in the rigorous analysis of layered resonator structures, which, in all earlier approaches, was restricted to the one-dimensional treatment.

6. Acknowledgements

The authors wish to thank F. W. Thomas, H. Siegmund and G. Thorn of the Leybold-Heraeus (Inficon) company in Hanau and Alzenau, BRD, for many valuable discussions and support of our research.

This Project was financially supported in part by the "Jubiläumsfonds 150 Jahre Technische Universität Wien" der *Kammer der Gewerblichen Wirtschaft* (Chamber of Commerce, Vienna).

7. Literature

/1/ A. Ballato, H. L. Bertoni, and Th. Tamir, "Systematic Design of Stacked-Crystal Filters by Microwave Network Methods", IEEE Trans. Microwave Theory and Techn. **MTT-22**, 14-25 (1974).

/2/ A. Ballato, "Doubly Rotated Thickness Mode Plate Vibrators", in *Physical Acoustics,* edited by W. P. Mason and R. N. Thurston (Academic, New York, 1977), Vol. XIII, pp.115-181.

/3/ W. E. Newell, "Face-Mounted Piezoelecric Resonators", Proc. IEEE **6**, 575-581 (1965).

/4/ J. G. Miller and D.I. Bolef, "Sensitivity Enhancement by the Use of Acoustic Resonators in cw Ultrasonic Spectroscopy", J.Appl. Phys. **39**, 4589-4593 (1968).

/5/ J. G. Miller and D.I. Bolef, "Acoustic Wave Analysis of the Operation of Quartz-Crystal Film-Thickness Monitors", J.Appl. Phys. **39**, 5815-5816 (1968).

/6/ C. Lu and O. Lewis, "Investigation of Film-Thickness Determination by Oscillating Quartz Resonators with Large Mass Load", J.Appl. Phys. **43**, 4385-4390 (1972).

/7/ H. Nowotny and E. Benes, "General one-dimensional treatment of the layered piezoelectric resonator with two electrodes", J. Acoust. Soc. Am. **82** (2), 513-521 (1987)

/8/ IEEE Standard on Piezoelectricity, No. 176 (1978)

/9/ C. J. Wilson, "Vibration modes of AT-cut convex quartz resonators", J. Phys. D: Appl. Phys. **7**, 2449-2454 (1974)

/10/ H. F. Tiersten and R. C. Smythe, "An Analysis of Contoured Crystal Resonators operating in Overtones of Coupled Thickness Shear and Thickness Twist", J. Acoust. Soc. Am. **65** (6), 1455-1460 (1979)

/11/ D. S. Stevens and H. F. Tiersten, "An analysis of doubly rotated quartz resonators utilizing essentially thickness modes with transverse variation", J. Acoust. Soc. Am. **79** (6), 1811-1826 (1986)

Abstract 1

1. Introduction 1

2. Derivation of the theoretical model 2

2.1. Three-dimensional treatment of the piezoelectric continuum ... 2

2.2. Fourier transforms of the constitutive equations 3

2.3. Fourier transformed boundary conditions for a plane between two continua 4

2.4. General solution of the basic equations in the Fourier transform domain 5

2.4.1. Solution of the homogeneous equation for the Fourier transform of the electrical potential 6

2.4.2. General solution for the Fourier transform of the mechanical displacement vector 6

2.5. Solution for a plane xn 9

2.6. Transfer matrix description 11

3. Description of a layered resonator structure 12

3.1. Resonator analysis in the Fourier transform domain 14

3.2. Resonator analysis in the space-domain 17

4. Treatment of a composite quartz crystal resonator with a convex boundary sphere 19

4.1. Specialization to a plane-convex quartz crystal 21

4.2. Treatment of a plane-convex crystal with one additional layer. 23

5. Conclusion 25

6. Acknowledgements 25

7. Literature 26

2 ND EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 16 - 17 - 18 MARCH 1988

Session RO-01 Number 2

ANALYSIS OF PARTIALLY OR TOTALLY CONTOURED RESONATORS WITH FINITE LATERAL DIMENSIONS

J. DETAINT, H. CARRU, J. SCHWARTZEL, C. JOLY

CNET PAB/BAG/MCT/CMM
196 Rue Henri Ravera 92220 BAGNEUX FRANCE

INTRODUCTION

The totally or partially contoured thickness mode resonators are of very common use for frequency generation or filtering. They present the advantage to have in the frequency range 0,5 to 10 MHz high Q factors together with reasonable lateral dimensions. The partially contoured ones (bevelled resonators) can also be made to present very few unwanted responses, this fact is of particular interest for filter applications. All these favourable properties are the consequence of the occurence in these resonators of a particular type of energy trapping that results mostly from the lateral variations of the thickness.

Mathematical models of these resonators having more and more accuracy were presented in the recent years. Most of them were derived from the recent developments of the theories of thickness vibrations of the piezoelectric plates established by Professors Mindlin, Lee, and Tiersten. In most case, the theory established by Professor Tiersten of the resonators vibrating near one overtone of a thickness mode having a thickness dependance essentially similar to that of the corresponding one dimensional mode of an infinite piezoelectric plate was used (1) (2) (3).

In this theory an approximate dispersion relation is established which is valid for the modes having slow lateral variations. A scalar partial derivative equation is then derived from this dispersion relation. It governs the lateral dependance of these modes. This equation permits to consider the case of contoured resonators provided that the lateral variations of the thickness are also sufficiently slow.

Models of plano-convex resonators of successively increased generality and accuracy were established by Tiersten and coworkers (2) (3) (4). Resonators with cylindrical bevels were considered by Vangheluwe (5) and Sinha (6). Plane and corrugated resonators with arbitrary electrode geometry were then considered (7). Recently Dulmet and Tiersten (8) have treated the case of plano-convex fully electroded rectangular plates (finite dimension).

In this paper, we present semi-algebraical solutions of the approximate equation for several types of partially electroded contoured resonators with finite dimensions. A finite element solution of this equation is then proposed for several case of contoured resonators.

I - THE APPROXIMATE EQUATION

As a basis of this model, we use the scalar equation governing the lateral dependance of the anharmonics in the vicinity of one overtone (number n) of a given thickness mode (noted(1)). This equation was first established for monoclinic plates (2)) then for plates of an arbitrary orientation (3). In this case the equation has, in a coordinate system with x_2 normal to the thickness and x_1 and x_3 choosen to eliminate the mixed derivatives (3) the following form :

$$M_n' \frac{\partial^2 \hat{u}_1^n}{\partial x_1^2} + P_n' \frac{\partial^2 \hat{u}_1^n}{\partial x_3^2} - \frac{n^2\pi^2 c^{*(1)}}{4h^2} \hat{u}_1^n + \rho\omega^2 \hat{u}_1^n = \rho\omega^2 (-1)^{(n-1)/2} \frac{e_{26}\, 4V_0\, e^{j\omega t}}{c^{(1)}_n \pi^2}$$

In this equation, $\hat{u}_1^n(x_1, x_3, t)$ is the lateral dependance of $\hat{u}_1^n(x_1, x_2, x_3, t) = \hat{u}_1^n \sin(n\pi x_2/2h)$.
$\hat{u}_1$ arises from a transformation of the usual component of the displacement u_1 and of the potential ϕ, made to replace the inhomogenous boundary conditions $\varphi = \pm V_0 e^{j\omega t}/2$ at $x_2 = \pm h$ by the homogeneous conditions $\varphi = 0$. The inhomogeneous term of the equation results from this transformation.

$$u_1^n = \hat{u}_1^n - \frac{e_{26}\, V_0\, e^{j\omega t}}{c^{(1)} 2h} \cdot x_2 \; ; \quad \phi = \hat{\phi} + \frac{x_2 V_0 e^{j\omega t}}{2h} \; ; \quad u_1 = \sum_n u_1^n$$

$c^{*(1)}$ is either $\bar{c}^{(1)}$ for the unelectroded part of the resonator or $\hat{c}^{(1)}$ for the electroded part. $\bar{c}^{(1)}$ is the stiffened elastic constant relative to the corresponding one dimensional mode (Eigen-value of the Christoffel Matrix).

$$\hat{c}^{(1)} = \bar{c}^{(1)}(1 - 8\, k^{(1)2}/n^2\pi^2 - 2\,\hat{R})$$

$\hat{c}^{(1)}$ is a constant that includes the electrical and mechanical effects of the metallization. $4k^{(1)2}/n^2\pi^2$ is the relative frequency lowering due to the electrical effect of the metallization ; $\hat{R} = 2\rho'h'/\rho h$ is the relative frequency lowering due to the inertia of the electrodes (figure 1). $c^{(1)} = \bar{c}^{(1)}\,(1-k^{(1)2})$ is a pseudo ordinary elastic constant. $k^{(1)}$ is the coupling coefficient of the corresponding one dimensional mode. M'_n, P'_n are intricate functions of the material constants and of the plate orientation (3).

I.2 - Boundary and continuity conditions : The traction free conditions on the major free surfaces, normal to the thickness are automatically verified as a consequence of the method used to establish the dispersion relation. On the surfaces limiting the electroded and the unelectroded regions of the resonator we have to specify the continuity of u_1 and of its normal derivative (figure 1).

On the external edges of the resonators, the approximate boundary conditions to be applied are not uniquely determined (7) and we have to choose in each case what type of mechanical condition, compatible with the scalar equation is the best approximation of the traction free conditions. However, for a well designed resonator this choice is of small importance since u_1, $\frac{\partial u_1}{\partial x_1}$, $\frac{\partial u_1}{\partial x_3}$ and the corresponding energy are extremely small near the edge.

II - SEMI ALGEBRAICAL MODEL

This model is based upon algebraical solutions of the approximate equation. As for the case of the plane resonator (7) we are looking for solutions, that are, for each region of the resonator, in the form of an infinite serie of eigen functions which are exact solutions of the

equation. As these solutions are obtained by separation in coordinates systems for which the electrode boundary and/or the plate boundary are not, in general, lines with one constant coordinate, the continuity and boundary conditions are expressed at a discrete number of points. This permits to obtain, by truncation of the series, an approximate solution that converges toward the exact solution when the number of points is increased. Most often, the convergence is very fast, so that a limited amount of points (of terms) is required to have a great precision.

II.1 - Separable forms for the thickness variations : The more commoun contoured resonators are the plano-convex, the biconvex and the bevelled plates. In some instances, plano-convex or biconvex resonators that present a second shorter radius of curvature (bevel) near their edges are used (Figure 2). All these resonators have spherical contouring. The most commonly used separable approximation of the term $(1/2h)^2$ that appears in the partial derivative equation is a first order (2) approximation in $r^2 = (x_1^2+x_3^2)$:

$$(\frac{1}{2h})^2 = (\frac{1}{2h_0})^2 \ (1 + \frac{x_1^2+x_3^2}{2\ R_c h_0}) \qquad [1]$$

The same approximation applied to many surfaces tangent to or osculating the sphere on the x_2 axis (figure 2) will lead to the same expression. This is the case for the paraboloid of revolution $-x_2 = \frac{x_1^2+x_3^2}{2p}$ with p = Rc.

A more accurate approximation may be usefull when a precise representation of the thickness variations near the edges of the resonator is needed ; one was proposed in reference (9) :

$$(\frac{1}{2h})^2 = \frac{1}{2h_0})^2 \ (1 + \frac{\pi^2\delta(x_1^2+x_3^2)}{8\ h_0\ l_2^2}) \qquad [2]$$

where δ, l_2 are defined on figure 2.

If we consider $R'_c = \frac{4\, l_2^2}{\pi^2 \delta}$ and, to include the case of all contoured resonators considered in figure 2, if we introduce the quantity h'_0 (defined on figure 2), the form of relation [1] can be conserved for all cases.

$$(\frac{1}{2h})^2 = \frac{1}{2h'_0})^2(1 + \frac{x_1^2+x_3^2}{2R'_c h'_0}) \qquad [3]$$

As can be seen on figure 3 the approximation given by expression [1] is generally better near the center of the resonator whereas the approximation [2] is better near the edges of the plates.

Although they seem to constitute less precise approximations to the spherical contour, it is interesting from a mathematical point of view and probably from a pratical point of view to consider the case of non spherical anisotropic contours that lead to separable expression of $(1/2h)^2$ like :

$$(\frac{1}{2h}) = (\frac{1}{2h'_0})^2 (1 + \frac{x_1^2}{2R'_1 h'_0} + \frac{x_3^2}{2R'_3 h'_0}) \qquad [4]$$

Simple surfaces (elliptical paraboloids, ellipsoids) corresponds to this case.

II.2 - Eigen solutions at V = 0 for the spherical regions : In the following $C^{*(1)}$ here after noted C^*_1 is equal to $\hat{C}^{(1)}$ if the spherical portion considered is electroded or to $\bar{C}^{(1)}$ in the other case.

The homogeneous form of the equation is

$$M'_n \tilde{u}^n_{1,11} + P'_n \tilde{u}^n_{1,33} + 4\pi^2\rho(f^2 - f_{C1}^2(1 + \frac{x_1^2 + x_3^2}{2h'_0 R'_c})) \tilde{u}^n_1 = 0 \qquad [5]$$

where : $f_{C1} = \frac{n}{4h'_0}\sqrt{\frac{C^*_1}{\rho}}$

The transformation of coordinates (5)

$$\begin{cases} X_1 = (C^*/M'_n)^{1/4} . \gamma . x_1 \\ \\ X_3 = (C^*_1/P'_n)^{1/4} . \gamma . x_3 \end{cases} \qquad \text{with } \gamma = \left(\frac{n^2\pi^2}{2R'_c h'^3_o}\right)^{1/4}$$

permits to obtain a solution in the separate form $u_1 = F(X_1)\ G(X_3).\sin(n\pi x_2/2h.)$ where F ang G are solutions of the parabolic cylinder equations

$$\begin{cases} F'' - \left(\frac{X_1}{4} + a_1\right) F = 0 \\ \\ G'' - \left(\frac{X_3}{4} + a_3\right) F = 0 \end{cases} \qquad [6]$$

The separation constants have to verify the expression :

$$4\pi^2\rho\ (f^2 - f^2_{C_1}) = -a_1 M'^{1/2}_n\ C^{*1/2}\gamma^2 \ - \ a_3 P'^{1/2}_n\ C^{*1/2}\gamma^2$$

or

$$-\ a_1 \left(\frac{M'_n}{C^*_1}\right)^{1/2} - a_3 \left(\frac{P'_n}{C^*_1}\right)^{1/2} = A = \frac{n\pi}{2}\left(\frac{R'_c}{2h'_o}\right)^{1/2} \left(\frac{f^2 - f^2_{C_1}}{f^2_{C_1}}\right) \qquad [7,8]$$

Several conventional forms of solution of the parabolic cylinder equations are known (10). Since we are considering the symmetrical eigen modes of a finite plate excited by symmetrical electrodes, we have chosen as Dulmet (8) to use the even serie solutions $y_1\ (a,X)$ which is finite at X = 0. If we set :

$$\begin{cases} a_1^m \left(\frac{M'_n}{C^*_1}\right)^{1/2} = \frac{A}{2} + f(m) \\ \\ a_3^m \left(\frac{P'_n}{C^*_1}\right)^{1/2} = \frac{A}{2} - f(m) \end{cases} \qquad \text{with } m = 1, 2, 3,...$$

a_1^m and a_3^m verify the relation [7] and a solution of [5] is given by :

$$\hat{u}_1 = F(a_1^m, X_1).G(a_3^m, X_3)$$

A series solution can be obtained if f(m) is conveniently choosen :

$$\hat{u}_1 = \sum_{m=0}^{\infty} C^m y_1 (a_1^m, X_1).y_1(a_3^m, X_3) \qquad [9]$$

In our applications we have taken f(m) = 2 m (with m even)

II.3 - Eigen solutions for plane regions (beveled resonators) : The homogeneous equation is now :

$$M_n' \hat{u}_{1,11} + P_n' \hat{u}_{1,33} + 4\pi^2\rho(f^2 - f_{c_1}^2)\hat{u}_1 = 0 \qquad [10]$$

To have common notations with the preceeding paragraph we use the coordinate transformations :

$$\begin{cases} X_{1'} = \left(\frac{C_1^*}{M_n'}\right)^{1/2} . \gamma_2 . X_1 \\ X_3 = \left(\frac{C_1^*}{P_n'}\right)^{1/2} . \gamma_2 . X_3 \end{cases} \qquad \begin{cases} X_1 = r \cos t \\ X_3 = r \sin t \end{cases} \qquad \text{with } \gamma_2^4 = \frac{n^2\pi^2}{2ho^4}$$

which lead to the equation :

$$\hat{u}_{1,rr} + \frac{1}{r}\hat{u}_{1,r} + \frac{1}{r^2}\hat{u}_{1,tt} + A'\hat{u}_1 = 0$$

with $A' = \frac{n\pi}{4}(f^2 - f_{C_1}^2)/f_{C_1}^2$

that can be separated as $u_1 = R(r).\ T(t)$ in :

$$\begin{cases} r^2 R'' + rR' + R\,(\Lambda r^2 - \nu^2) = 0 \\ T'' + \nu^2 T = 0 \end{cases} \qquad [11]$$

The symetrical solutions can be expressed as (7) :

$$\tilde{u}_1 = \sum_{m=0}^{\infty} B_m\, J_m(r\sqrt{\Lambda})\, \cos mt \qquad [12]$$

with $m = \nu$ even integer, J_m Bessel function of 1^{st} kind and order m.

II.4 - Expression of the continuity and boundary conditions : In the following we suppose that the resonator is constituted of only two regions : the central region (plane or spherical), being electroded and the external contoured one unelectroded. Except for the cases of plano-convex or biconvex resonators where the electrode shape can be chosen in a quite arbitrary manner, the electroded region is thus circular.

At all points of the electrode edge we have to specify the continuity of $\tilde{u}_1$ and of its normal derivative. As previously discussed (7) (11), this can be approximated by the expression of these conditions at only a discrete number q_1 of points (figure 4). If the solutions for the two regions are :

$$\begin{cases} \tilde{u}_1^{I} = \left(\sum_{m=0} D_m\, H_m(x_1, x_3) \right) \sin (n\pi x_2/2h) \\ \tilde{u}_1^{II} = \left(\sum_{m=0} E_m\, K_m(x_1, x_3) \right) \sin (n\pi x_2/2h) \end{cases} \qquad [13]$$

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

$$\begin{cases} u_1^{I}(x_1^P, x_2, x_3^P) = u_1^{II}(x_1^P, x_2^P, x_3^P) \\ n_1 \dfrac{\partial u_1^{I}(P)}{\partial x_1} + n_3 \dfrac{\partial u_1^{I}(P)}{\partial x_3} = n_1 \dfrac{\partial u_1^{II}(P)}{\partial x_1} + n_3 \dfrac{\partial u_1^{II}(P)}{\partial x_3} \end{cases} \qquad [14]$$

where ; $\vec{N}(n_1, n_3)$ is the normal to the boundary between region I and region II at P. (Figure 4)

At the resonator edge we have to express a boundary condition compatible with the form of the approximate equation. Since only Dirichet. Neumann or mixed conditions relative to the lateral part of u_1 ie $\hat{u}_1(x_1, x_3)$ are compatibles with the equation, we have to choose among these ones the best approximation to the zero traction conditions. For well designed resonators this choice is of little importance since we have simultaneously at the plate edge

$$u_1 \neq 0 \; ; \; \frac{\partial u_1}{\partial x_1} \neq 0 \quad \frac{\partial u_1}{\partial x_3} \neq 0 \; ; \; \frac{\partial u_1}{\partial n} \neq 0 \text{ and } \alpha u_1 + \beta \frac{\partial u_1}{\partial n} \neq 0$$

If we choose the condition $\hat{u}_1 = 0$, at the q_2 points we obtain q_2 conditions of the form

$$0 = \sum_m E_m K_m (x_1^P, x_3^P) \sin (n\pi x_2/2h) \qquad [15]$$

If the number of points at the electrode edge and at the plate edge are chosen equal ($q_1=q_2$), we have $2q_1+q_2=3q$ conditions so that the series solution for u_1^I must be truncated to $2q$ terms and the series for u_1^{II} has to be truncated to q terms.

The ensemble of the relations [14] and [15] written for the q points at the electrode edge and the q points at the plate edge constitutes thus an homogeneous linear system in the coefficients $D_0 D_2 \ldots D_{4q-2}$, E_0, $E_2, \ldots, E_{2q-2}$. To have non trivial solutions, the determinant of this system must vanish ; this condition constitutes the eigen frequency equation. This equation is solved numerically so that we obtain the eigen frequencies f_n^1, f_n^2, f_n^3, f_n^s.....

For each of the eigen frequencies the homogeneous system is solved and the coefficients D_0^s, $D_2^s, \ldots, D_{4q-2}^s$, $E_0^s, \ldots E_{2q-2}^s$ are obtained as a function of one of them. Thus the eigen modes are obtained in the form

$$\begin{cases} u_1^{nsI}(x_1, x_2, x_3, t) = e^{j\omega t} \sum_{m=0}^{4q-2} D_m^s \; H_m(x_1, x_2, f_n^s) \sin(n\pi x_2/2h) \\ u_1^{nsII}(x_1, x_2, x_3, t) = e^{j\omega t} \sum_{m=0}^{2q-2} E_m^s \; K_m(x_1, x_2, f_n^s) \sin(n\pi x_2/2h) \end{cases} \qquad [16]$$

where : $\begin{cases} m \text{ is even} \\ H_m(x_1, x_2) \;\; ; \;\; K_m(x_1, x_2) \text{ are given by relations [9] or [12]} \\ \text{according to the geometry of the contoured resonator.} \end{cases}$

II.5 - Non spherical contours : An approximate expression of $1/h^2$ was found for contour being elliptical paraboloids :

$$\left(\frac{1}{2h}\right)^2 = \left(\frac{1}{2h_0}\right)^2 \left(1 + \frac{x_1^2}{2ph_0} + \frac{x_3^2}{2qh_0}\right) \qquad [17]$$

where : p and q are the radii of curvature at x_1 and $x_3 = 0$ in the two coordinate planes.

An approximation of similar form can be found for ellipsoidal contours. The eigen modes for such a contour can be obtained by separation in transformed rectangular coordinates of the homogeneous approximate equation :

$$\begin{cases} X_1 = \left(\frac{C_1^*}{M_n'}\right)^{1/4} \gamma_1 x_1 \\ X_3 = \left(\frac{C_1^*}{P_n'}\right)^{1/4} \gamma_3 x_3 \end{cases} \qquad \begin{cases} \gamma_1^4 = \frac{n^2\pi^2}{2ph_0^3} \\ \gamma_3^4 = \frac{n^2\pi^2}{2qh_0^3} \end{cases}$$

$$u_1^n = F(X_1).G(X_3)$$

$$\begin{cases} F'' - \left(\frac{X_1^2}{4} + a_1\right) F = 0 \\ G'' - \left(\frac{X_3^2}{4} + a_3\right) G = 0 \end{cases}$$

with : $- a_1 \, M_n'^{1/2} \, C_1^{*1/2} \, \gamma_1^2 - a_3 p_n'^{1/2} \, \overset{*}{C}_1^{1/2} \, \gamma_3^2 = 4\pi^2 \rho (f^2 - f_{C1}^2)$ [18]

For a totally electroded planoconvex resonator of this kind in the case of the Gauss-Hermite approximation (2) $a_1 = -m-1/2$; $a_3 = -m' - 1/2$ with m ; m' integers (even for the symmetric modes). The parabolic cylinder functions are then :

$$\mathcal{U}(-m - \tfrac{1}{2}, X) = D_m(X) = 2^{-\frac{1}{2}} \, m \, e^{-\frac{1}{4} X^2} \, H_m (X/\sqrt{2}) \qquad [19]$$

The eigen frequency equation is then obtained from [18] :

$$f_{nmm'}^2 = f_{C_1}^2 \left[1 + \frac{2}{n\pi} \left[(m + \tfrac{1}{2}) \left(\frac{M_n' \, 2ho}{C_1^* \, p} \right)^{1/2} + (m' + \tfrac{1}{2}) \left(\frac{P_n' 2h_0}{C_1^* \, q} \right)^{1/2} \right] \right] \qquad [20]$$

II.6 - Forced solutions

The forced modes are expressed as a sum of all the eigen modes (7) :

$$\hat{u}_1^n = \sum_s H^{ns} \hat{u}_1^{ns} \sin(n\pi x_2/2h) \quad , \quad \hat{u}_1 = \Sigma \, \hat{u}_1^n$$

where : $\hat{u}_1^{ns}$ is either $\hat{u}_1^{ns^I}$ or $\hat{u}_1^{ns^{II}}$

Using the orthogonality of the eigen modes

$$\iint_S \hat{t}_1^{ns} \, \tilde{u}_1^{nv} \, dS = \delta_{sv} \, N^{(n)}(s)$$

and the fact that $\hat{u}_1^{ns}$ is a solution of the homogeneous approximate equations we find (7) the values of H^{ns}. Then using the constitutive equation giving D_2, the electrical admittance is found in the form :

$$Y = j\omega\left[\left(C_Z - \sum_n \sum_s C_{ns}\right) - \sum\sum_{ns} C_{ns}\left(\frac{f^2_{ns}}{f^2 - f^2_{ns}}\right)\right]$$

with

$$C_z = \frac{S'}{2h}\,\varepsilon_{22}\,(k^2+1) \qquad \text{(capacitance at } f = 0\text{)}$$

$$C_{ns} = \left(\frac{4k^2\,\varepsilon_{22}}{n^2\pi^2 ho}\right)\frac{\left(\iint_{se} \tilde{u}_1^{ns}\, ds\right)^2}{N^{(n)(s)}} \qquad \text{dynamic capacitance}$$

S' surface corrected for the thickness variations.

III - FINITE ELEMENT SOLUTION

The homogeneous form of the approximate equation is solved by a finite element technique when the parameter h as any slowly variation with x_1 and x_3. The natural boundary condition implies the normal derivative of displacement equal to zero, so that the continuity relations between electroded and unelectroded regions are automatically satisfied and the condition at the edge of the plate is of Neumann type. This later one can also be made of Dirichlet type. The circular domain is divided into triangular elements with the displacement at the corners as unknown. The problem is solved as an eigenvalue one. The matrices resulting from the assemblage of these elements are stored on a "skyline" basis. The dynamic capacitance of the resonator can also by calculated.

The overall accuracy can be established by considering the variation of the frequency when one adds a few elements.

IV - COMPARISON OF COMPUTED AND EXPERIMENTAL RESULTS

Computer programs were set up using the previously described semi-algebraical and finite element method. The computed eigen frequencies and eigen modes are compared with experimental ones in the case of a bevelled resonator and of a plano-convex resonator.

IV.1 - Comparison of eigen frequencies : This comparison is made for a bevelled resonator designed for filter application to have very reduced anharmonic responses.

Table I : Resonance frequency of a bevelled resonator : 2ho = 1,094 mm $2l_1$ = 8,93 ; $2l_2$ = 15.00 mm ; R_c = 60 mm ; R = 0,0028

Fundamental mode	Fr. experimental	1556. KHz
	Fr. finite element method	1557.
3rd overtone	Fr. experimental	4563. KHz
	Fr. finite element method	4558.

In table 1, We can see that a good agreement compatible with the uncertainty on the thickness and mass loading measurements and the material constants is obtained.

A comparison was then made in the case of a plano convex resonator with circular electrodes. In table II we compare the experimental frequency with the eigen frequencies computed for the Gauss Hermite approximations for $1/h^2$ (given (in paragraph II), with the results of the semi algebraical model and with the result of the finite element calculation.

Table II : Resonance frequency of a plano convex resonator : [5^{th} overtone, 2ho = 1.321 ± 1 mm ; R = 0.0017 ; $2l_2$ = 15 mm ; R_c = 175 mm; AT cut ; circular electrode $2l_1$ = 8 mm (Fr in Hz)]

Fr (experimental)	6338.560
Fr (approx. 1 for $1/h^2$; Gauss Hermite)	6333.988
Fr (approx. 2 for $1/h^2$; Gauss Hermite)	6340.312
Fr (approx. 1 for $1/h^2$; Semi Algebraical)	6334.012
Fr (approx. 2 for $1/h^2$; Semi Algebraical)	6339.907
Fr (Exact $1/h^2$; Finite Element Method)	6334.

In table II no thickness adjustments were made to allow for the uncertainty on the resonators parameters and the material constants. During the computation, made in the last three cases with Neumann conditions at the external edge of the plate, it was observed that the displacement near this edge is so small that it is of comparable magnitude as the numerical noise. This is a confirmation of the weak importance of the choice of the edge boundary conditions for well designed plano-convex resonators.

Considering the approximations of $1/h^2$ the first one leads to a minorant of the eigen frequency (figure 3) the second to a majorant, but for this type of resonator the first one is much more accurate.

IV.3 - Comparison of the vibration modes : The vibration modes corresponding to the eigen frequencies of table I and II were computed and are compared to experimental ones in figures 5 and 6. In these figures the variations computed by the finite element method, of u_1 along the x_1 axis of the resonators were obtained by X-ray measurements (14) (15). This method gives a good approximation of the value of $\tilde{u}_1$ (x_1, x_3) except for the small values. This can be observed in figure 7.

In figure 8 to 10 are represented the computed eigen modes and the experimental modes of the plano convex resonator of table II. It can be observed that a good agreement is obtained except for the 1st mode which is strongly coupled to the flexure.

V - CONCLUSION

The semi algebraical method of resolution of the approximate equation established by Professor Tiersten and coworkers permits to compute the eigen frequencies and the modes of partially or totally contoured resonators with finite lateral dimensions and quite arbitrary electrode shape. This method has the advantage to require a moderate amount of computations to obtain a high precision.

The finite element method of solution of the approximate equation permits to consider the exact thickness variations together with arbitrary geometry. This method gives accurate results and has the possibility to analyse the thickness vibration of more complex piezoelectric structures.

The results obtained are a new confirmation of the fundamental interest of the approximate equation in the description of thickness modes of any practical resonators.

Aknowledgments : The authors wish to aknowledge the skillful assistance of R. Chenebault, A. Daniel for experimental work and sample preparation. The authors are indebted to M. E. Caquot, for stimulating discussions. Special thanks are due to M.P. Louis for preparing this manuscript.

REFERENCES

(1) H.F. Tiersten
J. Acoust. Soc. Am. Vol. 59 p 879 (1976)

(2) H.F. Tiersten, R.C. Smythe
J. Acoust. Soc. Am. Vol. 65 n° 6 p (1979)

(3) D.S. Stevens ; H.F. Tiersten
J. Acoust. Soc. Am. Vol. 79 n° 6 p 1811 (1986)

(4) H.F. Tiersten ; D.S. Stevens
Proc. 34th Annual Frequency Control Symp p 37 (1982)

(5) D.C.L. Vangheluwe
Proc. 34th Annual Frequency Control Symp p 412 (1980)

(6) B.K. Sinha ; D.S. Stevens
J. Acoust. Soc. Am. Vol 66 n° 1 p 192 (1979)

(7) J. Détaint, J. Schwartzel, C. Joly, E. Philippot
Proc. 41st Annual Frequency Control Symp. p (1987)

(8) B. Dulmet, H.F. Tiersten
Proc. IEEE Ultrasonics Symp. p 355 (1986)

(9) A.V. Apostolov ; S.H. Slavov
Appl. Phys. Vol. A39 p 33-37 (1982)

(10) M. Abramowitz ; I. Stegun
Handbook of Mathematical functions Dover (1972)

(11) K. Nakamura ; H. Shimizu
Proc. IEEE Ultrasonics Symp. p. 606 (1976)

(12) E.T. Whittaker ; G.N. Watson
A course of Modern Analysis Cambridge Press 1927

(13) O.C. Zienkiewicz
The finite Element Methode Mc Graw Hill 1977

(14) J. Détaint, A. Zarka, B. Capelle, Y. Toudic, J. Schwartzel, E. Philippot, J.C. Jumas, A. Goiffon, J.C. Doukhan
Proc. 40th Annual Frequency Control Symposium (1986)

(15) A. Zarka, B. Capelle, J. Détaint, J. Schwartzel
Stroboscopic Xray topography of Quartz resonators
This proceedings

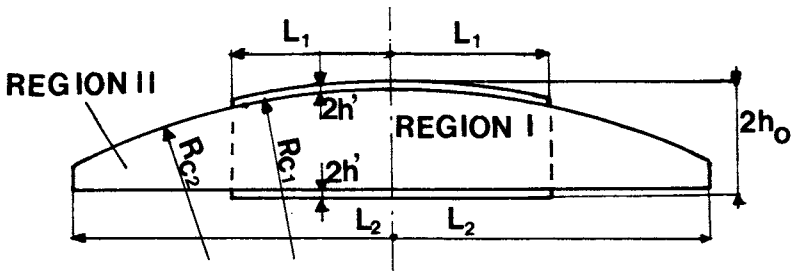
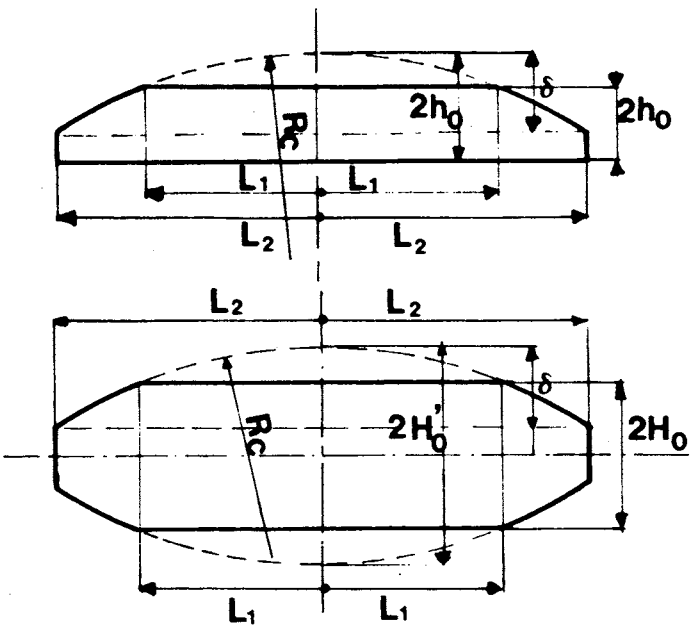
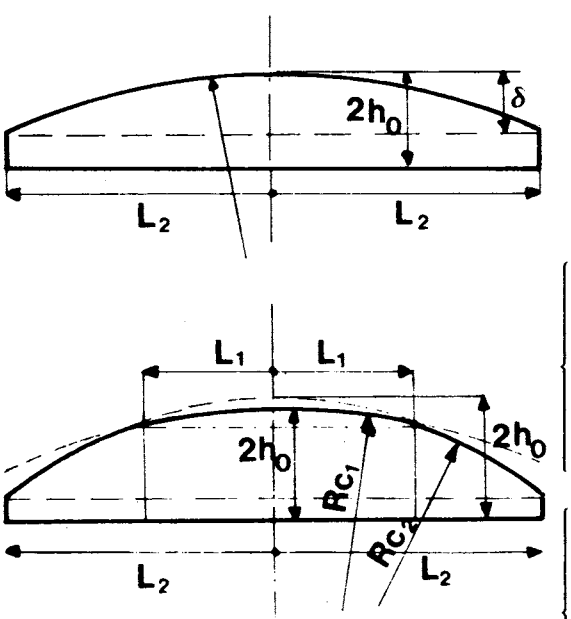


Figure 1 : Contoured resonators .



$$\begin{cases} 2h_o' = 2h_o + R_c - \sqrt{R_c^2 - l_1^2} \\ \delta = R_c - \sqrt{R_c^2 - l_2^2} \\ R_c' = \dfrac{4\, l_2^2}{\pi^2 \delta} \end{cases}$$

$$\begin{cases} 2H_o' = 2H_o + 2R_c - \sqrt{R_c^2 - l_1^2} \\ R_c' = \dfrac{2\, l_2^2}{\pi^2 \delta} \\ \delta = R_c - \sqrt{R_c^2 - l_2^2} \end{cases}$$



$$\begin{cases} 2h_o' = 2h_o \\ R_c' = \dfrac{4\, l_2^2}{\pi^2 \delta} \\ \delta = R_c - \sqrt{R_c^2 - l_2^2} \end{cases}$$

Internal region

$$\begin{cases} 2h_o' = 2h_o \\ R_c' = \dfrac{4\, l_2^2}{\pi^2 \delta} \\ \delta = R_c - \sqrt{R_c^2 - l_1^2} \end{cases}$$

External region

$$\begin{cases} 2h_o' = 2h_o + R_{c2} - R_{c1} + \sqrt{R_{c1}^2 - l_1^2} \\ R_c' = \dfrac{4\, l_2^2}{\pi^2 \delta} \qquad - \sqrt{R_{c2}^2 - l_1^2} \\ \delta = R_{c2} - \sqrt{R_{c2}^2 - l_2^2} \end{cases}$$

Figure 2 : Geometries of contoured resonators .

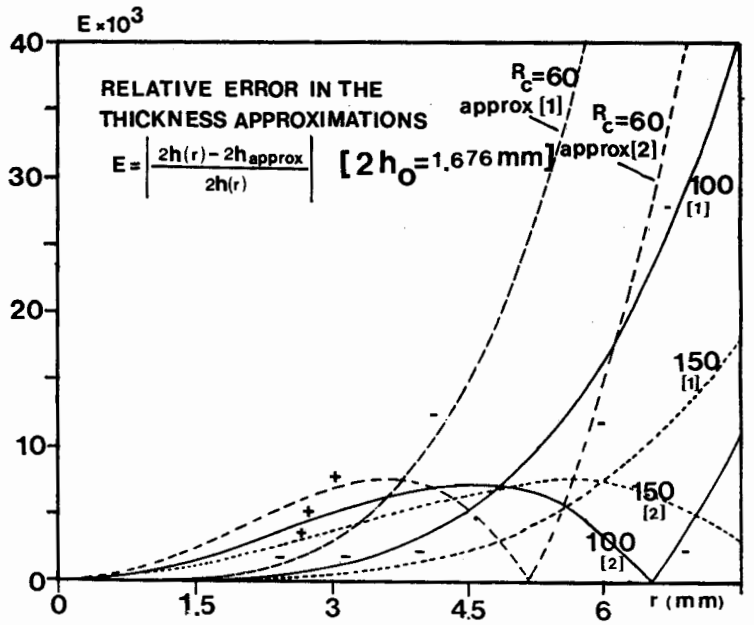


Figure 3 : Approximations of $(1/2h)^2$.

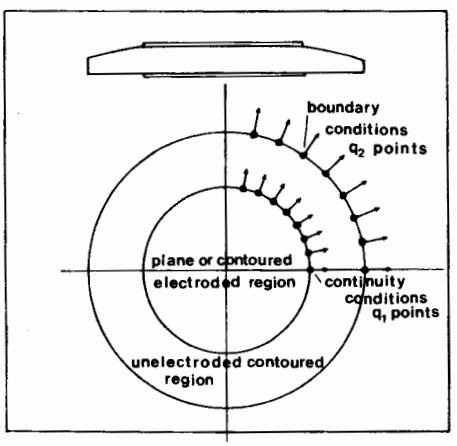


Figure 4 :Discretization of the continuity and boundary conditions .

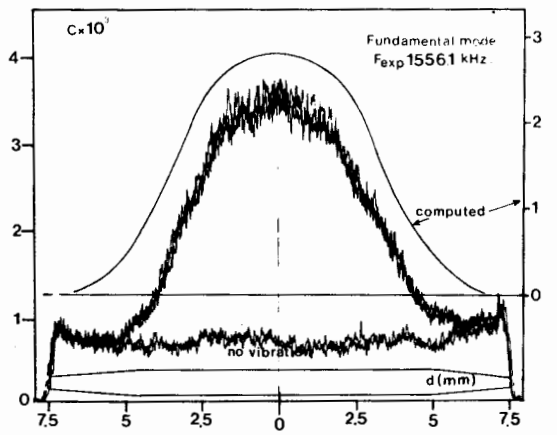


Figure 5 :Fundamental mode of a bevelled resonator .

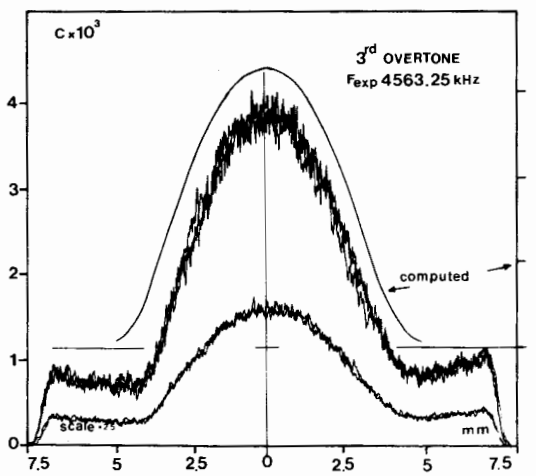


Figure 6 :3rd Overtone of a bevelled resonator .

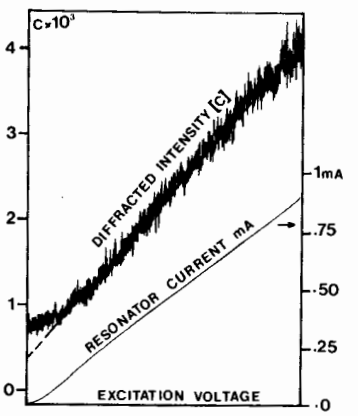


Figure 7 :Linearity of the amplitude detection

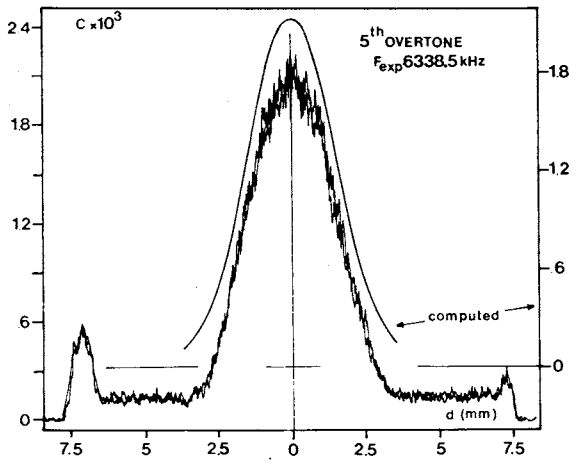


Figure 8 : 5th overtone of a plano convex resonator .

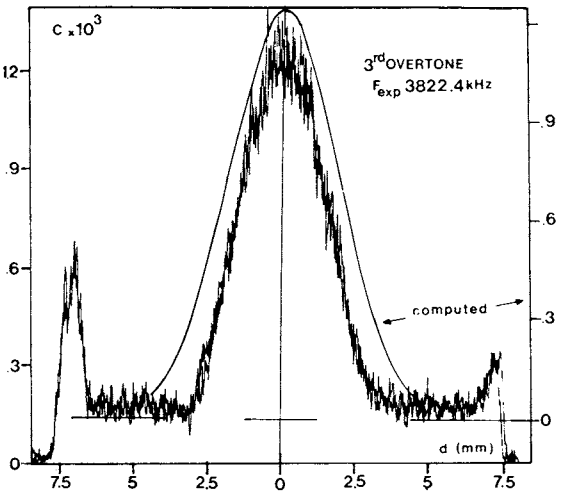


Figure 9 : 3rd overtone of a plano convex resonator .

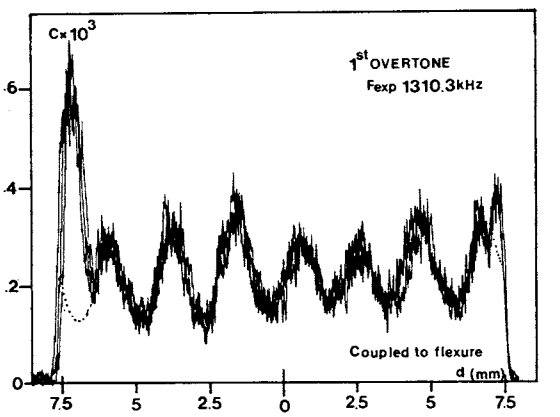


Figure 10 : Fundamental mode of a plano convex resonator .

2 ND EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 16 - 17 - 18 MARCH 1988

Session RO-01 Number 3

LATERAL DC ELECTRIC FIELD IN QUARTZ PLATE RESONATORS

Carl K. Hruska
Piezoelectricity Research Laboratory
Department of Mathematics, York University
4700 Keele Street, Toronto, Ontario M3J 1P3, Canada

Summary

The frequency of doubly rotated quartz plate resonators vibrating in thickness can be affected by a dc electric field in the lateral directions (parallel to the main surfaces of the plates). The frequency change depends on the orientation of the plate and on the direction of the dc electric field in the plate.

The available experimental evidence suggests that the shape of the dependence including the location of its extrema and zero points can be predicted by a simplified model using the effective electroelastic tensor.

Observations of the frequency shifts in the lateral fields of various directions can provide useful information about the nonlinear effects in quartz.

Introduction

Most of the earlier studies of the frequency-dc field dependence of quartz resonators (also known as the polarizing effect) were done for the three thickness modes of vibration of quartz plates subjected to a dc electric field in the direction of the plate thickness. The frequency changes detected were found to be related to all four nonlinear third-order electromechanical effects in quartz [1,2]. They have been recognized as one of the principal sources of information about two nonlinear phenomena of current interest: the electroelastic effect and electrostriction.

The disadvantage of the above data is that they provide only incomplete access to the material constants describing the two mentioned nonlinear phenomena [3,4]. This disadvantage can be overcome if additional experimental values of the frequency change are obtained for dc fields applied in directions other than that of the plate thickness.

The most obvious choice is the lateral dc electric field i.e. an electric field in any direction normal to the plate thickness. This choice also defines the scope of this work which is to study the effect of the lateral dc electric field on the frequency of arbitrarily oriented quartz plate resonators vibrating in any one of the three thickness vibration modes.

The numerical values of all quantities in this paper apply to the left hand quartz at room temperature and the frame of reference defined by the IRE Standard in 1949 [5].

The measured quantity and its model

The dc field-induced frequency change in quartz resonators is predominantly linear with a small quadratic component. The latter has been removed from our observations and our study is thus limited to the linear coefficient of the frequency-dc field dependence defined as

$$L = (1/f).(df/dE)_{E=0} \tag{1}$$

and determined experimentally as

$$L = \Delta f/(f.\Delta E), \tag{2}$$

where Δf is the linear change of the original series resonance frequency f caused by the dc electric field ΔE.

The resonators used were square plates of orientation (xzlw)Φ/Θ according to [5]. The linear coefficient L was studied as a function of the two orientation angles Φ and Θ defining also the direction of the resonator thickness and of the angle α which determines the direction of the lateral dc field vector in the plane of the resonator plate. This angle is regarded as positive when measured in the direction from +Y' to +Z" as marked in Fig.1.

The measurement of the frequency-dc field dependence was done for the fundamental frequency of the three thickness vibration modes under the condition of zero phase shift across the resonator (series resonance). More details concerning the experimental method can be found in [6].

The first indication of the potential of the linear coefficient L as a productive source of information about the nonlinearities in quartz was its remarkable consistency and predictability [3,7]. This was shown by comparing the experimental values of L with an expression for L according to Eq.(1) where the necessary derivative was obtained by ad hoc differentiation of the frequency formula of the linear theory [8] and with the use of the effective electroelastic constants introduced during the differentiation process [9]. These results have been obtained for the thickness vibrations of plates with the dc electric field applied in the direction of their thickness.

The geometry of the present experiment represents a departure from the relative simplicity of the earlier case. It appears to be more difficult to create a lateral dc field of sufficient uniformity acting on the whole vibrating portion of the resonator. Moreover, the lateral dc field can be applied in any direction normal to the resonator thickness which adds a new independent variable to the quantity L. For these reasons it is believed that a demonstration of consistency and predictability (similar to the above) would be of interest.

The derivation of the prediction formula for L needed here follows the steps described in detail for the dc field in thickness [4,9] except for the differentiation in Eq.(1) which has to be made for dc fields E of any

direction normal to the resonator thickness. The derivation presents no difficulty; it is just tedious and the details are unwieldy. For this reason, only the final result is recorded here in a qualitative form

$$(3)\quad L = (1/f).(df/dE)_{E=0}$$

$$= c_1 \cdot d_{11} + c_2 \cdot d_{14} + a_1 \cdot e_{111} + a_2 \cdot e_{113} + a_3 \cdot e_{114}$$

$$+ a_4 \cdot e_{122} + a_5 \cdot e_{124} + a_6 \cdot e_{134} + a_7 \cdot e_{144} + a_8 \cdot e_{315} \cdot$$

The coefficients c_1, c_2, and a_j, $j = 1,2, \ldots,8$, are relatively complicated functions of the three angles Φ, Θ, and α, of the mode of the vibrations of the resonator and of the linear material constants of quartz: elastic stiffnesses, piezoelectric stress constants and dielectric permittivities; d_{11} and d_{14} are the piezoelectric strain constants. For the actual calculations their values have been taken from [10].

The quantities e_{111}, e_{113}, ..., e_{315} are the eight effective electroelastic constants of quartz. The complete set is available from the transit time experiment in bulk quartz. Their most recent values referred to implicitly in [11] have been obtained directly from their authors [12].

Experimental results and their predicted values

Two types of resonators (I and II) were used for the experiment. Both were square plates, about 19 mm in size and between 0.4 and 1.0 mm in thickness.

Each resonator plate of type I had plane main faces. Each plate was completely coated with silver except for a continuous gap of constant width passing through the centre of its main faces and around its whole body. The gap divided the silvered surface of the plate into two electrodes. The resonator and the position of the electrodes are illustrated in Fig.2 (a,b). A dc potential applied to these electrodes created a prevailingly lateral dc electric field. The same electrodes were used to excite the resonator vibrations. In order to change the direction of the dc electric field the electrodes were washed off and repositioned. For the actual measurement the resonator was placed in a grounded metallic envelope.

A representative result pertaining to the resonators of type I is shown in Fig.3. It was obtained for the fundamental frequency of the mode A of a resonator of orientation (xzlw)105.07°/-9.24° and shows the dependence of the linear coefficient L on the direction α of the lateral dc field. A curve fitted to the individual experimental points (Eq.(2)) is compared with the predicted curve obtained from Eq.(3). Characteristic for resonators of this type is the good agreement in the shape of the two curves as well as a consistent difference in the magnitude of L with the absolute value of the observations always smaller than that of the predictions.

Trying to account for the relative smallness of the experimental values of L in Fig.3 the magnitude of the applied electric dc field ΔE and of the measured frequency change Δf used in Eq.(2) was further examined.

The values of ΔE substituted into Eq.(2) to produce the experimental values of L in Fig.3 are obtained as a ratio of the dc potential across the resonator electrodes and the electrode separation. A better approximation obtained by a numerical solution of the Poisson equation yields an average value of ΔE which is only 93% of the value used originally.

Although the use of the corrected value for the lateral dc field would bring the experimental and the predicted curves in Fig.3 closer together the improvement would not be substantial. For a more substantial change one has turn to the measured value of Δf.

While the vibrations are excited in the central unelectroded part of the resonator they tend to propagate into the rest of the resonator body due to the additional mass of the electrodes. This has been confirmed by x-ray topographic observations of the vibrations. Our experiments showed that if more silver is added to the electrodes to make them heavier, the measured frequency change Δf decreases. On the other hand, if the central unelectroded parts of the resonator are loaded with a balast (wax), the frequency change Δf increases considerably. These physical modifications of the resonator result in redistribution of the energy of vibrations and the recorded variations in Δf are consistent with the notion that the observed frequency change Δf depends on the portion of the vibrational energy which stays in the unelectroded part of the resonator, the only part with a nonzero lateral dc field.

This has led to the conclusion that the observed frequency changes Δf used in Eq.(2) to determine the experimental values of L in Fig.3 were only a fraction of the changes predicted by calculations done using Eq.(3) where it is assumed that all resonator vibrations are affected by the lateral dc field. The best way to rectify the situation is to confine the vibrations to a volume of the resonator where they can be exposed to the lateral dc field as fully as possible.

As a result resonators of a new type (II) have been tested. Their main faces were no longer plane. The centre of each main face was worked into a circular mesa intended to confine the vibrations to the central volume of the resonators. The electrodes, positioned in the same way as with the resonators of type I, were reaching to the side of the mesa circles and their separation was defined by the mesa diameter. The resonator shape and electrodes are shown in Fig.2 (c,d).

Comparisons of observations and predictions made for the resonators of type II produced a substantially improved agreement. An example shown in Fig.4 has been obtained for the fundamental frequency of the mode C of a resonator of orientation (xzlw)100°/-40°. It shows the dependence of the linear coefficient L on the lateral dc field direction α. The plotted experimental values of L have not been corrected for the lateral field distortion. This correction would increase the absolute value of all observations by 5% thus further improving the agreement between the experimental and the predicted curves in Fig.4.

Conclusion

If a quartz plate resonator vibrating in thickness is subjected to a dc electric field in its thickness direction the magnitude of the linear

coefficient L of its frequency-dc field dependence is fixed and defined by the resonator orientation.

If the dc electric field is applied laterally, the linear coefficient depends on the electric field vector which may assume any direction normal to the resonator thickness. Compared to its fixed value for the dc field applied in thickness the linear coefficient L may increase very substantially. This is demonstrable most easily with the AT cuts, where the linear coefficient L for the dc field applied in thickness is zero. At the same time, there always exists a lateral dc field direction where the linear coefficient L is zero. This direction may be of interest in designing electrodes for the lateral field excitation of resonator units to maximize their frequency stability.

The shape of the dependence of the linear coefficient L on the resonator orientation and on the direction of the lateral electric field as predicted by means of the current values of the effective electroelastic tensor appears to be sufficiently accurate to provide estimates of frequency-dc field dependence for engineering applications.

If the resonator vibrations are confined (trapped) to a defined location of the resonator body and the nonhomogenous lateral dc electric field in that part of the resonator are corrected (replaced by its space mean value) the observations of the linear coefficient L of the frequency-dc field dependence are likely to provide useful data to study the nonlinear material constants of quartz.

Acknowledgement

The author wishes to acknowledge the assistance with the experimental part of the work received from R. Besson, J. F. Darces, G. Genestier, M. Kucera and H. Merigoux and fruitful discussions with A. Ballato and J. H. Sherman, Jr., and H. F. Tiersten. P. Hruska helped with the final stages of preparation of this paper. This work was supported in part by the NSERC (Grant A4911) and by the Faculty of Arts (MRG Grant), York University.

References

[1] Baumhauer J. C., Tiersten H. F., "Nonlinear electroelastic equations for small fields superimposed on a bias", J. Acoust. Soc Am. vol. 54(4), 1973, pp. 1017-1034.

[2] Brendel R., "Material nonlinear piezoelectric coefficients for quartz", J.Appl.Phys. 54(9), September 1983, pp. 5339-5346.

[3] Hruska C. K., "The electroelastic tensor and second-order phenomena in quasilinear interpretation of the polarizing effect with α-quartz plates", Proc. 31st Annual Frequency Control Symposium, U.S.Army Electronics Command, Ft. Monmouth, June 1977, pp. 159-170.

[4] Hruska C. K., "Independent verification of the values of electroelastic constants of alpha quartz", J. Appl. Phys. 61(4), 1 February (1987), pp. 1127-1129.

[5] IRE Standards on Piezoelectric Crystals, IEEE Standard 176; see also Proc. IRE, vol. 37, Dec. 1949, pp. 1378-1395.

[6] Hruska C. K. and Kucera M., "The thickness modes of a quartz plate resonator in a lateral dc field", IEEE Trans. Sonics Ultrason., vol. SU-32, July 1985, pp. 600-603.

[7] Hruska C. K., "Polarizing effect with doubly rotated alpha-quartz plates vibrating in thickness", IEEE Trans. Sonics and Ultrason., vol. SU-25, November 1978, pp. 390-392.

[8] Tiersten H. F., "Thickness vibrations of piezoelectric plates", J. Acoust. Soc. Amer., vol. 35, 1963, pp. 53-58.

[9] Hruska K., "Polarizing effect with piezoelectric plates and second order effects, IEEE Trans. Sonics and Ultrason., vol. SU-18, January 1971, pp. 1-7.

[10] Bechmann R., "Elastic and piezoelectric constants of alpha-quartz", Phys. Rev., vol. 110, 1958, pp.1060-1061.

[11] Kittinger E., Tichy J., "Mapping of the electroelastic effect in quartz based on fundamental material constants", Proc. 1st European Time and Frequency Forum, Scientific Communities of Besancon and Neuchatel, March 1987, pp. 222-225.

[12] Kittinger E., Tichy J., private communication.

List of captions

Fig.1. Linear coefficient L is a function of three angles: Φ, Θ (defining the orientation of the resonator (xzlw)Φ/Θ and the direction of the resonator plate thickness X") and α (defining the direction of the lateral dc field vector E). The angle α is positive when measured from +Y' to +Z" direction. The axes X, Y, and Z form the basic reference frame in quartz. The notation corresponds to the IRE Standard [5].

Fig.2. Crossection (a) and plan view (b) of the type I resonator. Crossection (c) and plan view (d) of the type II resonator showing the circular mesa on the resonator main faces. Electrodes are shown bordered by dotted lines; α is the direction of the lateral dc field.

Fig.3. Dependence of the linear coefficient L on the lateral field direction angle α. Observations (points and dotted line) are compared with predictions (solid line).Obtained for the fundamental frequency of mode A (4.9 MHz) of type I resonator of orientation (xzlw)105.07°/-9.24°, thickness 0.643 mm, electrode separation of 3 mm and a dc potential across the electrodes of 2 kV.

Fig.4. Dependence of the linear coefficient L on the lateral field direction angle α. Observations (points and dotted line) are compared with predictions (solid line).Obtained for the fundamental frequency of mode C (3.6 MHz) of type II resonator of orientation (xzlw)100°/-40°, central thickness of 0.479 mm, circular mesa diameter and electrode separation of 7 mm and a dc potential across the electrodes of 2 kV. The mesa height is 0.027 mm.

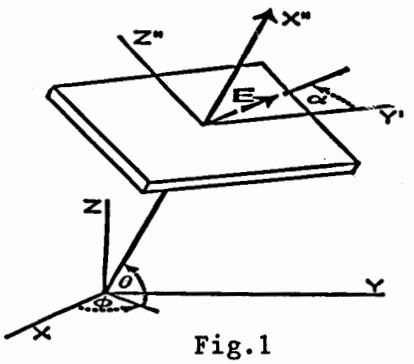


Fig.1

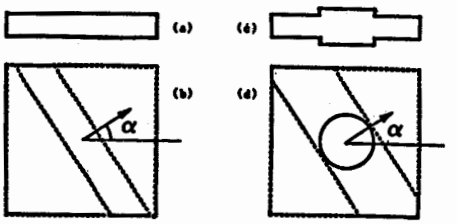


Fig.2

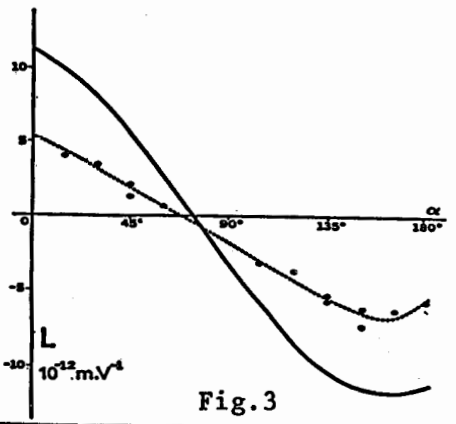


Fig.3

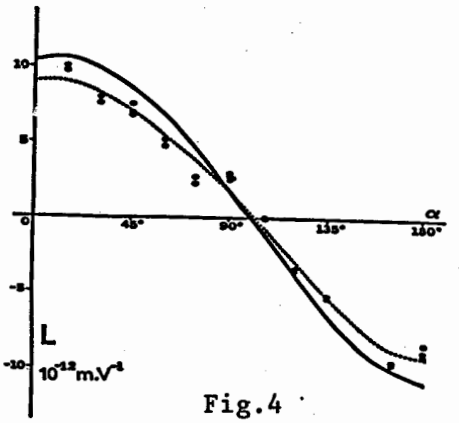


Fig.4

2 ND EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 16 - 17 - 18 MARCH 1988

Session RO-01 Number 4

NON PIEZOELECTRIC CRYSTAL RESONATORS

D. MAZEROLLE, E. DIEULESAINT and D. ROYER

Laboratoire d'Acoustoélectricité de l'Université P. & M.Curie
ESPCI - 10 rue Vauquelin - 75231 Paris Cedex 05

ABSTRACT

As reported at the first Forum, held at Besançon (1987), sapphire resonators can be excited by photothermal or capacitive effect, their vibration being optically detected [1]. The work on these resonators has been continued in order to increase their resonance frequency. On the other hand, similar work has been carried out with YAG resonators. For instance, a YAG bar has been excited by capacitive effect at a frequency of 195.76 kHz in an extensional mode. Its vibration was detected either by the optical wedge method or by capacitive method. A Q-factor of 1.9×10^6 was measured. A sapphire rod has been excited by photothermal effect at a frequency of 483.1 kHz also in an extensional mode. Its vibration was detected with a heterodyne optical probe. The Q-factor measured was 2×10^5.

1 - INTRODUCTION

The small number of piezoelectric crystals exploitable as resonators (SiO_2, $LiTaO_3$) justifies our interest in non piezoelectric crystals able to compete with quartz with regard to intrinsic mechanical properties. The experiments carried out with sapphire crystals (Al_2O_3, class $\bar{3}$m) which we reported last year, at the first European Frequency and Time Forum, showed the quality of this material: resonators at 20 kHz, in flexural mode, and at 260 kHz, in extensional mode, revealed Q-factors of about 2×10^6. The purpose of this paper is to report the results of new experiments performed, on the one hand with a YAG (Yttrium and Aluminum Garnet: $Y_3Al_5O_{12}$, class: m3m) crystal whose acoustic wave attenuation factor is even lower than the one of sapphire (YAG: 0.2 dB/cm [2], sapphire: 0.3 dB/cm [3]), on the other hand, with a sapphire resonator vibrating at a higher frequency.

2 - CAPACITIVE EXCITATION OF A YAG BAR IN AN EXTENSIONAL MODE.

The YAG bar, of length $l = 40.40$ mm oriented along a four-fold axis (100), is held at two points in order to vibrate longitudinally in a second overtone mode (figure 1). This overtone mode is preferred to the fundamental one because holding the bar at two points instead of one suppresses most of the low frequency vibrations and so facilitates the detection. Thus the expected resonance frequency is:

$$f_0 = V/l \qquad \text{where} \qquad V = 1/\sqrt{\rho s_{11}} \tag{1}$$

For $\rho = 4550$ kg/m^3 and $s_{11} = 3.60\times10^{-12}$ m^2/N $\Rightarrow$ $f_0 = 193$ kHz.

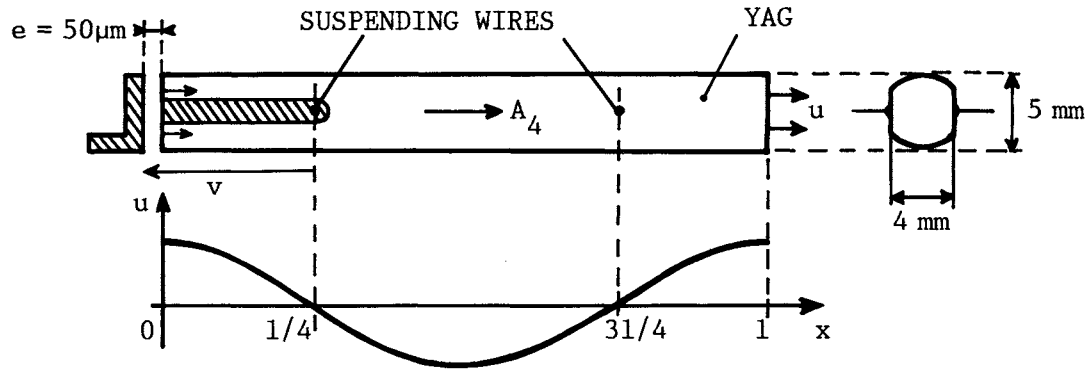


Fig.1. YAG bar. Second overtone extensional mode of vibration excited by capacitive effect. u is the mechanical displacement. l = 40.40 mm.

Capacitive excitation was mounted at one end of the bar: a 1000 Å thick gold layer was deposited onto that end and to the flat part. The end layer connected to a suspending wire forms the moving electrode of a capacitor, whose other electrode is a fixed metal plate facing the bar end at a distance e. When a voltage v is applied across the capacitor, the electrostatic pressure on the end of the rod is:

$$p = \varepsilon_0 \frac{v^2}{2e^2} \cdot \qquad (2)$$

In order to excite the bar at angular frequency ω, we can apply either an alternating voltage at $\omega/2$: $v = v_2 \cos(\omega t/2)$ or an alternating voltage at ω, added to an offset voltage: $v = v_0 + v_1 \cos \omega t$. In this case the pressure at frequency ω is

$$p_1 = \varepsilon_0 \frac{v_0 v_1}{e^2} \cdot \qquad (3)$$

This pressure produces at resonance frequency a displacement amplitude

$$u = \frac{Q}{2\pi^2} s_{11} l\, p_1 \qquad Q: \text{ quality factor} \qquad (4)$$

For $v_0 = v_1 = 30$ V, $e = 50$ μm and $Q = 10^6 \Rightarrow p_1 = 3.2$ N/m^2 and $u \cong 230$ Å.

This displacement is large enough to be detected either by optical wedge technique or by capacitive technique. Results obtained with the first technique have already been given [4]: f_0 and Q measured were respectively 195.760 kHz and 1.9×10^6, the resonator being placed in a primary vacuum enclosure (pressure $\cong 10^{-2}$ Torr).

The assembly for detection by the second technique is shown in figure 2. The capacitive receiver is symmetrical to the capacitive transmitter. The receiver being polarized by a steady voltage v_0', the vibration of the rod generates an alternating current across the capacitor. The main difficulty is to prevent coupling between the detection capacitor and the excitation capacitor. This was overcome by embedding input and output electrodes in grounded metal plates with epoxy resin and by applying an excitation voltage without offset i.e. at a frequency half that of detection. The output signal was then extracted with a selective amplifier.

For an alternating voltage amplitude $v_2 = 30$ V and a steady voltage across the detection capacitor $v_0' = 60$ V, the voltage across the detection capacitor load (27 kΩ) at resonance frequency was about 0.5 mV.

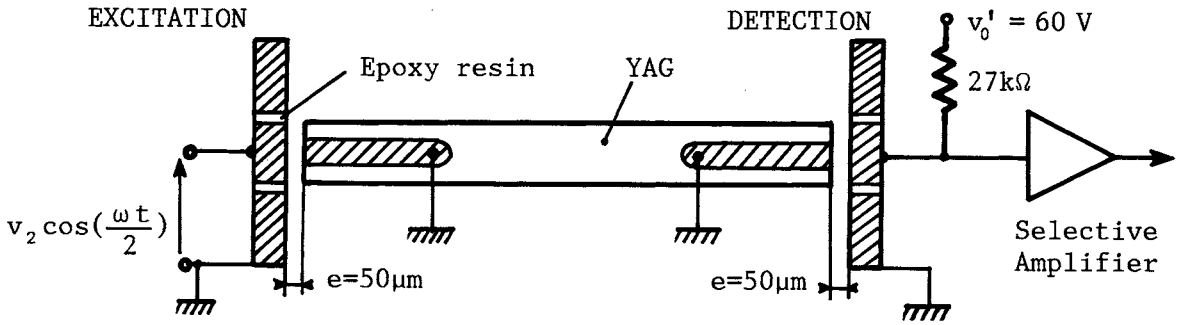


Fig. 2. Capacitive excitation and detection of a YAG bar resonator.

Figure 3 shows the decrease of the signal detected after suddenly turning off the excitation voltage. It reveals a Q-factor of about 700 000, less than half the value in the first experiment. This is probably because the bar was more rigidly mounted.

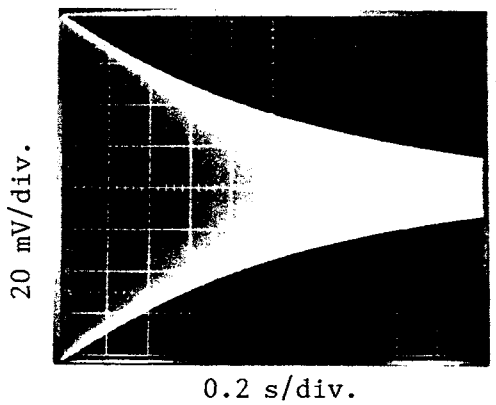


Fig. 3. Decrease of the detection signal after suddenly turning off the voltage applied to the excitation capacitor.

3 - PHOTOTHERMAL EXCITATION OF A SAPPHIRE BAR IN AN EXTENSIONAL MODE.

The sapphire bar, of length 1 = 22.22 mm oriented along the x_3-axis was mounted similarly to the YAG bar so as to vibrate in a second overtone extensional mode. The bar was excited by heating periodically the two nodal zones with a 50 mW He-Ne laser beam, the intensity of which was modulated by a Bragg cell. Direct beam and deflected beam have opposite phase. As the nodal zones also deform in opposite phase, both beams were used, each beam heating one node. A 1000 Å silver layer was deposited onto 2 mm wide zones around nodes ; this layer was then oxidized to absorb the laser beam energy.

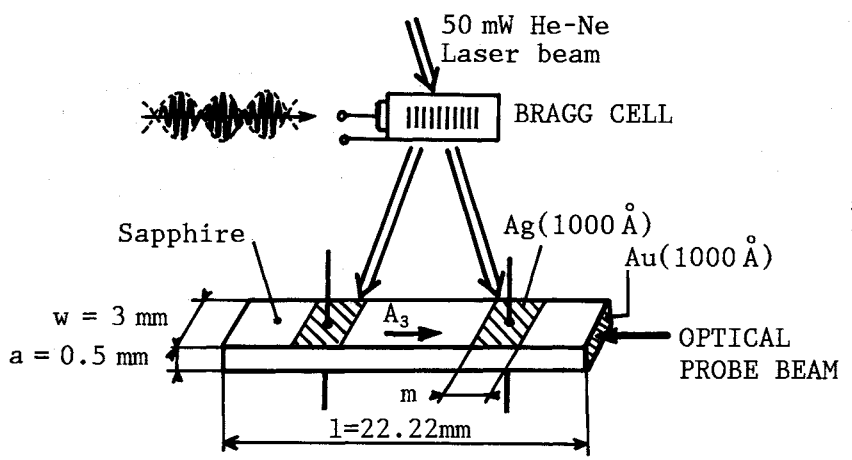


Fig.4. Photothermal excitation of the sapphire rod.

The expected resonance frequency is given by formula 1. For $V = 1/\sqrt{\rho s_{33}} = 10\ 700$ m/s : $f_0 = 482$ kHz.

The heat periodically absorbed in the layer generates a thermal wave in the nodal zones propagating along the thickness of the bar. The equation of conduction of heat provides the field of the periodic variation of temperature

$$\theta(x,t) = \theta_0 e^{-(\frac{1+i}{\mu})x} e^{i\omega t} \quad \text{with} \quad \theta_0 = \frac{1}{1+i} \cdot \frac{\mu}{K} \cdot \frac{P}{wm.}, \quad \mu = \sqrt{\frac{2K}{\rho C_p \omega}} \qquad (5)$$

μ is the thermal wavelength, P the amplitude of the modulated power absorbed by the layer, K the thermal conductivity of sapphire and C_p its specific heat. This thermal wave causes periodic expansion i.e. excitation of the crystal. Considering the average, on the bar cross-section, of stress caused by expansion, the displacement amplitude can be calculated at resonance frequency :

$$u = \frac{2Q}{\pi} \frac{\alpha L}{C} \frac{P}{\omega_0} \qquad (6)$$

where α is the expansion factor of sapphire parallel to x_3 axis, and C the heat capacity of the bar. With $C_p = 782\ Jkg^{-1}\ K^{-1}$, $\alpha = 5.7 \times 10^{-6} K^{-1}$, $Q = 10^5 \Rightarrow u = 2.6$Å.

This displacement amplitude is not easily detected by optical wedge method. So we used a heterodyne optical probe whose resolution reaches 10^{-4} Å/$\sqrt{Hz}$ [5] . The experimental resonance frequency was found to be $f_0 = 483\ 100$ Hz.

Figure 5a shows the spectrum of the output signal of the probe when the bar is at resonance frequency ; the central peak corresponds to the frequency shift ($f_B = 70$ MHz) of the probe, the two lateral components ($f_B \pm f_0$) correspond to, the vibration of the bar. The displacement amplitude measured is 2.5 Å. This signal was demodulated in order to observe directly the vibration frequency ; Figure 5b shows the decrease of demodulated signal amplitude when excitation is suddenly interrupted. These results agree with theoretical calculation if we assume that about 25 % of the incident beam modulated power is absorbed by the layer (P = 5 mW).

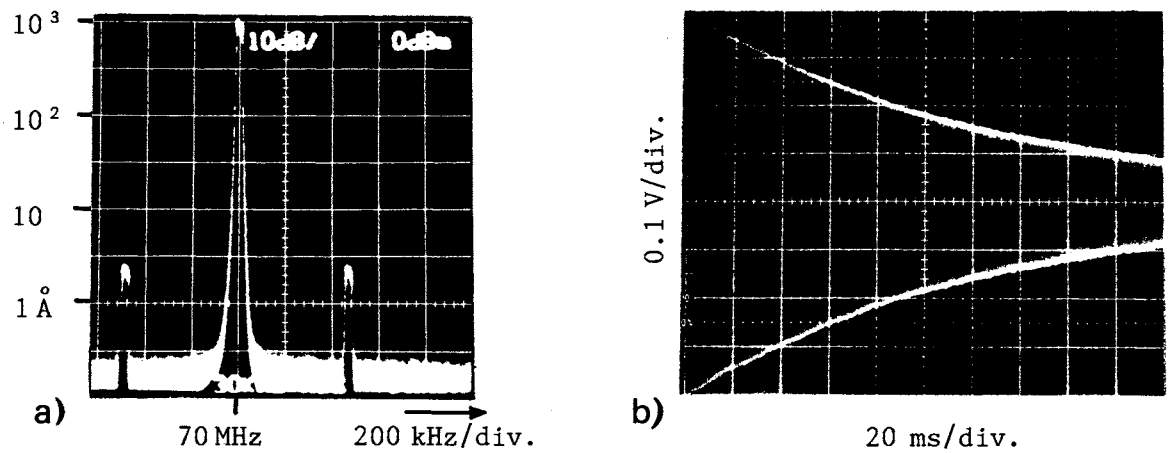


Fig. 5. Vibration of the sapphire bar at 483.100 kHz .
a) Spectrum of the output signal of the probe revealing a displacement amplitude of 2.5 Å.
b) Decrease of the vibration amplitude after suddenly interrupting the excitation ; Q factor : 2 x 10^5.

4 - CONCLUSION

These experiments show that resonators made of non piezoelectric crystals such as sapphire or YAG in the form of bars can be excited by photothermal or capacitive effect at frequencies of several hundreds of kilohertz. The vibrations can be detected by optical or capacitive method. The optical techniques have the advantage to sense at a distance and not to require any electrical connection, however, they are bulkier than the capacitive technique. The capacitive receiver has the disadvantage of requiring, on the one hand a small gap between a part of the resonator used as a moving electrode and a reference electrode and on the other hand, a shielding is necessary if the resonator is also excited by capacitive effect. In any way, further experiments are necessary to increase the frequency, to study the influence of temperature and to explore other optical detection and excitation configurations using, for example, optical fibers.

ACKNOWLEDGEMENTS

The authors are grateful to H. Duchaussoy of DRET (Paris), J.P. Aubry of CEPE, Thomson-CSF (Argenteuil) for helpful discussions, and to P. Kou and L. Michel of CEPE for technical assistance. Work supported by DRET.

REFERENCES

1 - Dieulesaint, E., Mazerolle, D., and Royer, D., "Résonateur à cristal de corindon", First Europ. Freq. and Time Forum Proc., 36 (1987).

2 - Dutoit, M., "Microwave phonon attenuation in yttrium aluminum garnet" J. Appl. Phys., 45, 2836 (1974).

3 - Damon, R.W., IEEE Spectrum, p. 87 (June 1967).

4 - Dieulesaint, E., Mazerolle, D., and Royer, D., "YAG resonator", Electron. Lett., 23, 581 (1987).

5 - Royer, D., Dieulesaint, E., and Martin, Y., "Improved version of a polarized beam heterodyne interferometer", IEEE Ultrason. Symp. Proc., 432 (1985).

RO-02

Resonators 2

1. An LCC Monolithic Crystal Filter
   R.C. Smythe
   Piezo Technology, Orlando - USA

2. Force-Frequency Characteristics of a Plano-Convex Square AT-Cut Plate for Force-Sensor Use
   N. Oura, N. Kuramochi, A. Okada
   Tokyo Institute of Technology,Yokohama - Japan

3. Stroboscopic X-Ray Topography of Quartz Resonators
   A. Zarka, B. Capelle
   Université P. & M. Curie, Paris - France
   J. Detaint, J. Schwartzel
   Centre National d'Etudes des Télécommunications, Bagneux - France

4. Studies on Oscillating Quartz Resonators for Application in Frequency Control and Standardization
   H. Bahadur, R. Parshad, B.S. Mathur
   National Physical Laboratory, New-Delhi - India

5. The Development of a Transportable Hydrogen Maser to be used for the VLBI Network in China*
   Z.C. Zhai, H.X. Huang, C.F. Lin, G.X. Jiang
   W.H. Luo, L. Jia-Fu
   Shanghai Observatory, Shanghai - China

2 ND EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 16 - 17 - 18 MARCH 1988

Session RO-02 Number 1

AN LCC MONOLITHIC CRYSTAL FILTER

R. C. Smythe

Piezo Technology, Inc.
Orlando, Florida 32854
U.S.A.

Abstract

A four-pole monolithic crystal filter has been developed in a ceramic leadless chip carrier (LCC) package 11.4 x 11.4 x 1.6 mm. Center frequency is 21.4 MHz, 3 dB bandwidth is ±7.5 kHz, and ultimate attenuation is 70 dB. The filter achieves a combination of miniaturization and performance not previously available. This paper discusses the acoustical design, packaging, and measured performance of the filter. The potential and limitations of the configuration are considered.

Introduction

The trend toward the use of low-profile surface-mount components in communications and position-locating equipments exposes the crystal filter as one of the largest components. Increasingly often, overall filter dimensions and mounting requirements pose difficult problems in equipment packaging. While no single solution exists, the use of leadless ceramic packages is advantageous in certain applications. Moreover, in cases where such packages are incompatible with the substrate, leads can be added, although with a small penalty in installed height.

To illustrate the possibilities of this approach, this paper describes a four-pole 21.4 Mhz monolithic crystal filter using a standard leadless chip carrier (LCC) package. Of primary importance is the acoustical design. Because of the limited space available, a somewhat novel resonator array

configuration has been combined with an unorthodox wafer mounting method.

Acoustical Design

Present-day monolithic crystal filter technology uses as its primary building block the two-resonator (two-pole) device, or monolithic dual resonator (MDR). Typically, the MDR's are individually packaged. Multi-pole filter characteristics are realized by tandem-connected MDR's. Filter assemblies combine one or more MDR's with additional coupling and impedance-matching components in a single package.

For the present requirement, it was decided to retain a conventional circuit arrangement having two-pole sections, Fig. 1. Two device configurations were considered:

1. two MDR's on separate wafers.
2. two MDR's on the same wafer.

The advantage of the first configuration is good acoustical isolation between the two filter sections. Offsetting this is the requirement to process and mount two wafers. The second configuration offers the advantage of being truly monolithic; further, a particularly simple wafer mounting arrangement is possible. Provided adequate isolation can be achieved, it represents the simpler approach.

In the past we have made extensive use of the second configuration, especially for VHF applications [1,2]. In those devices the two MDR's are situated side-by-side on the wafer, so that good input-to-output acoustical isolation is obtained. For the present requirement, such an in-line array would be too long to fit in the available space; consequently, a folded array was chosen. The photograph of Fig. 2 shows the arrangement. Starting at the upper left and going counterclockwise, resonators 1 and 2 are acoustically coupled to form the first MDR; 3 and 4 form the second. Resonators 2 and 3 are electrically coupled, the coupling being determined by capacitor C_J, Fig. 1, which is shown at the bottom of Fig. 2.

Of particular interest is the cross-coupling between input and output resonators 1 and 4. This coupling produces either a pair of real-frequency or real-axis transmission zeros in the normalized low-pass frequency plane, and may therefore be used to improve either selectivity [3, 4] or phase linearity [5]. For the present application, an all-pole design is appropriate; hence, we wish to minimize the 1-4 coupling. In order to do this on a small wafer, we use 1 mm. square electrodes with a substantial amount of mass loading -- 600 kHz. This allows the use of gold electrodes with a film thickness of 1430 Angstroms. A 0.1 dB ripple Chebyshev design with a 3 dB. bandwidth of 18.5 kHz. was chosen. The corresponding MDR mode spacing is 12.67 kHz., and is achieved with a 0.25 mm. gap.

In order to study the effect of coupling between resonators 1 and 4, the circuit model of Figure 3 was analyzed. From Williams [4] it can be shown that 1-4 coupling produces a pair of transmission zeros whose frequencies, expressed in terms of the normalized low-pass frequency variable, $S = A + j\Omega$, can be written using Dishal's normalized coupling coefficients:

$$S^2 = -k_{23}^2 [1 - k_{12}^2 / k_{23} k_{14}].$$

In the present case, k_{23} and k_{14} are of opposite sign; hence the transmission zeros lie on the $j\Omega$-axis. As a function of separation between resonators, the effect of 1-4 coupling is as follows:

Spacing (mm.)	Ω_s	Min. Atten. (dB.)
1.0	2.3	49
1.25	4.0	68
1.5	7.2	89

Figure 4 shows the calculated attenuation curves for these spacings. Based on these calculations, a spacing of 2.3 mm. was used between MDR's.

Construction and Performance

The general construction was shown in Fig. 2. A standard LCC, .45 x .45 x .062 inch (11.4 x 11.4 x 1.6 mm.) was used. The quartz wafer and the chip coupling capacitor are located in the die cavity, which measures 6.35 mm. square. The wafer is mounted on a central pedestal of conductive epoxy approximately .12 mm. high. The pedestal also provides the ground connection. All other connections are by wire bonds. Small epoxy spots, not shown, along the centerline between the MDR's, are used to suppress whole-plate modes.

Measured attenuation characteristics are shown in Fig. 6. The 3 dB. bandwidth is 18.1 kHz; 40 dB. bandwidth is 49.2 kHz.; flat loss is 1.5 dB., and stability is better than ±30 ppm over the temperature range -55 to +85°C.

Conclusions

The four-pole LCC filter discussed in this paper represents one practical solution to present and future needs for surface-mountable miniature filters. Further improvement can be obtained by obvious design refinements, such as a customized package to allow more efficient use of interior volume.

Devices of the type shown may be used as filter elements as well as by themselves. Higher order filters might be obtained by using two or more LCC's, and these might be packaged together with input and output impedance-matching networks to realize filter assemblies.

References

[1] R. C. Smythe, "Some Recent Advances in Integrated Crystal Filters," Proc. IEEE, v. 67, no. 1, pp119-129; Jan., 1979.

[2] R. C. Smythe, "Bulk Acoustic Wave Filters," in Precision Frequency Control, Vol. 1, E. A. Gerber and A. Ballato, Editors, Academic Press, Orlando, Florida; 1985.

[3] R. M. Kurzrok, "General Four-Resonator Filters at Microwave Frequencies," IEEE Trans. Microwave Theory and Technique, v. MTT-14, no. 6, pp. 295-296; June, 1966.

[4] A. E. Williams, "A Four-Cavity Elliptic Waveguide Filter," IEEE Trans. Microwave Theory and Technique, v. MTT-18, no. 12, pp. 1109-1114; Dec., 1970.

[5] J. D. Rhodes, "A Low-Pass Prototype Network for Microwave Linear-Phase Filters," IEEE Trans. Microwave Theory and Technique, v. MTT-18, no. 6, pp.290-301; June,1970.

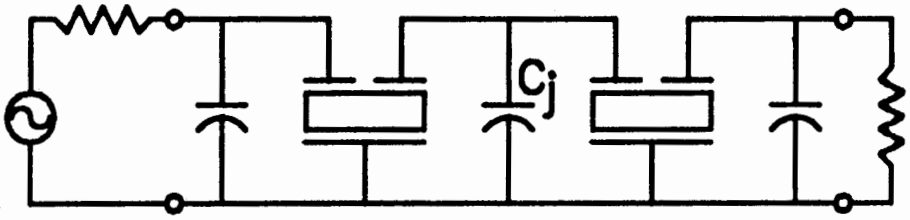


Fig. 1. Circuit Diagram.

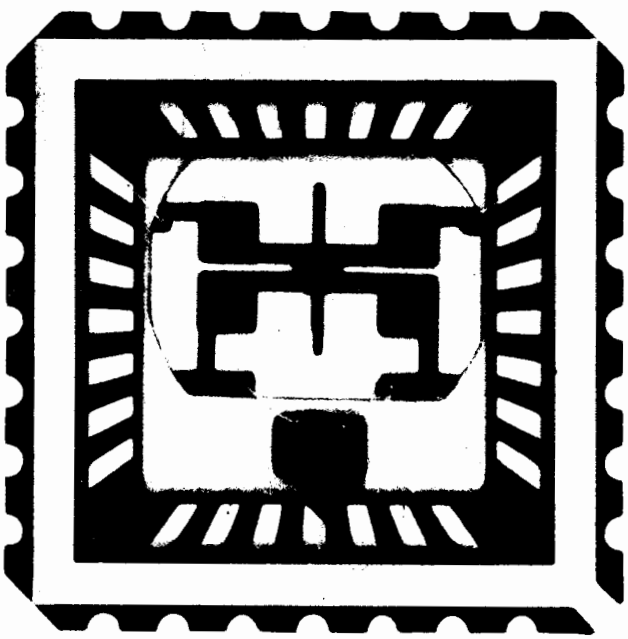


Fig. 2. Internal View of the LCC Monolithic Crystal Filter.

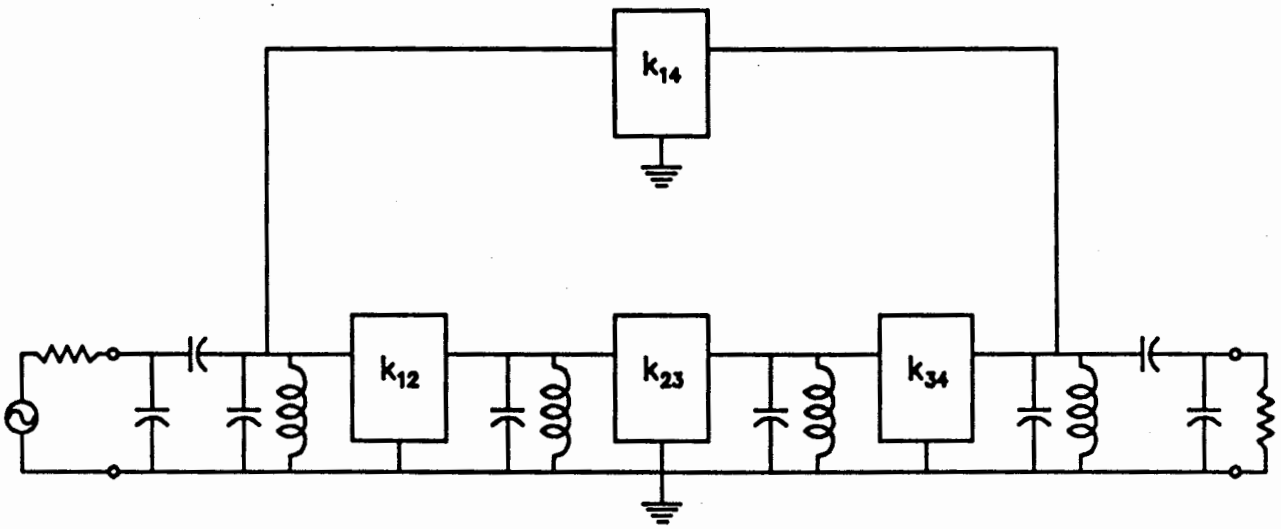


Fig. 3. Equivalent Circuit for Analysis of Cross-Coupling between Resonators 1 & 4; boxes are impedance inverters.

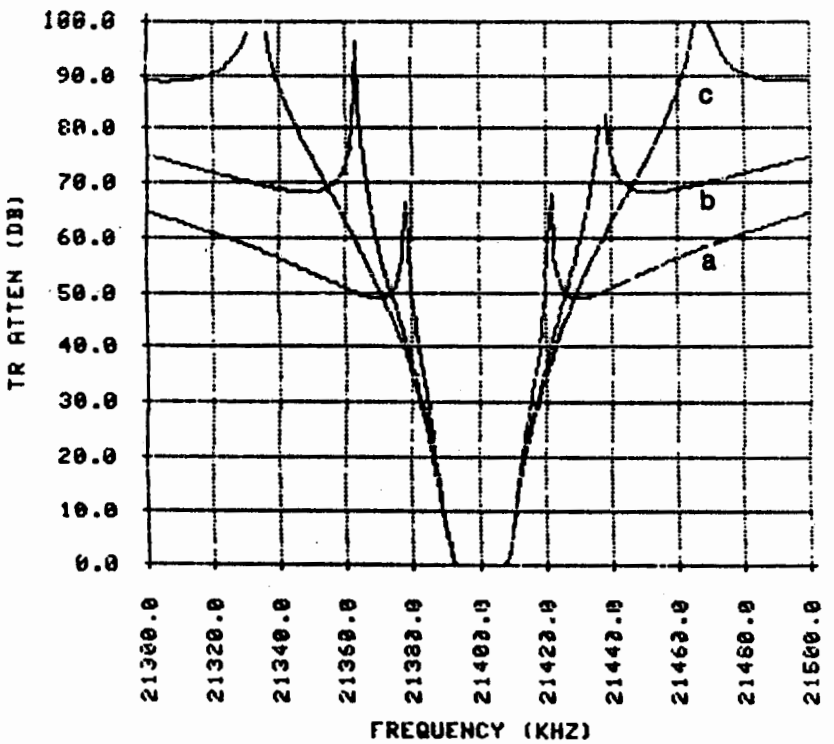
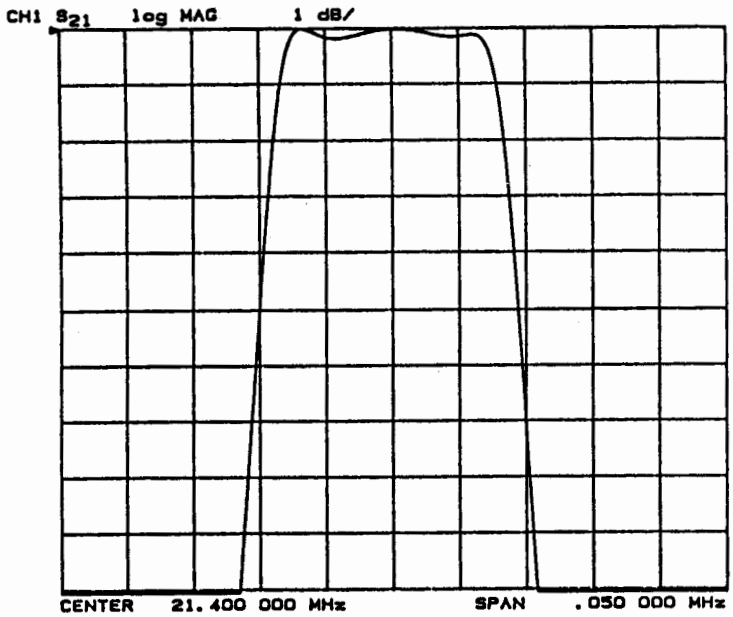
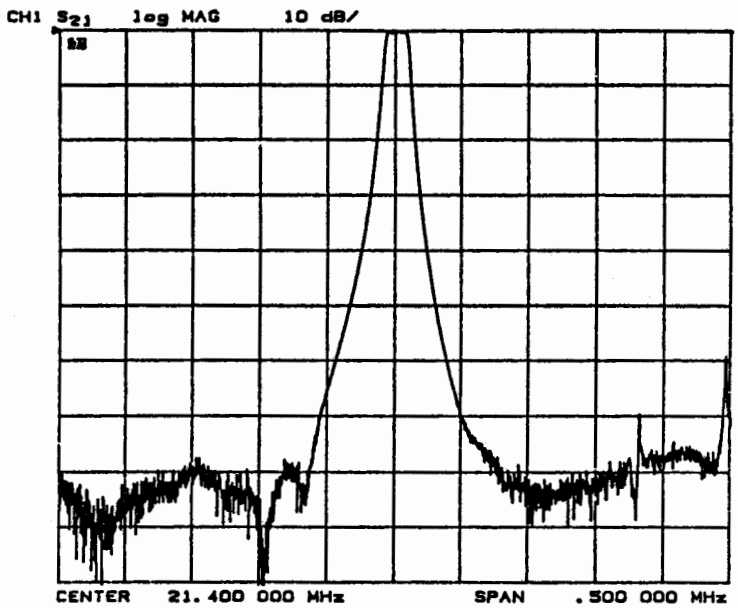


Fig. 4. Calculated Attenuation as a Function of Separation between Resonators 1 & 4: a) 1.0 mm.; b) 1.25 mm.; c) 1.5 mm.



a) Passband.



b) Stopband.

Fig. 5. Measured Attenuation Characteristics.

2 ND EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 16 - 17 - 18 MARCH 1988

Session RO-02 Number 2

Force-Frequency Characteristics of a Plano-Convex Square AT-Cut Plate for Force-Sensor Use

Nobunori OURA, Naimu KURAMOCHI, and Atsushi OKADA

Tokyo Institute of Technology
Research Laboratory of Precision Machinery and Electronics
4259, Nagatsuta, Midori-ku, Yokohama 227, Japan

Summary

Frequency changes in a plano-convex square AT-cut plate due to an opposing pair of external point forces or line forces on the side surfaces of the plate are derived theoretically and the validity of the theory is confirmed by experiment. The temperature and angle-of-cut dependence of the force-frequency characteristics of the plate are studied experimentally from a viewpoint of the application of the plate to sensor. An angle of cut at which the temperature coefficient of the force-frequency characteristics becomes zero is obtained for the application of a line-force pair.

1. Introduction

It is well known that the resonant frequencies of a quartz crystal plate change to some degree in response to external forces applied to the plate[1-5] . In some cases, the external forces may be replaced by internal forces such as thermal stress caused by thermal inhomogeneity due to the temperature fluctuation of an oven in which the plate is placed[6]. Resonators sensitive to the applied external forces are not desirable for use in high stable oscillators. For high stable oscillator use, new cuts of plate such as IT- and SC-cut less immune to the external forces have been devised[7].

Resonators highly stable to the external forces, however, are rather desirable for the application to force sensor. A measure of the magnitude of force-frequency sensitivity is the stress sensitivity coefficient K_f[8]. The coefficient has been studied with respect to circular plates in search of the supporting points of a plate immune to the external forces such as acceleration[9] or gravity. Circular plano-convex plates are not appropriate for force sensor elements. Because the difficulty lies in the fabrication of the sensor element, that is, the difficulty in exact setting of the force applying points on the peripheral side surface of the plate.

The use of plano-convex square plate eliminates the difficulty. Because two opposing sides of a plate are parallel and the azimuth, which is the angle between the electrical axis and a normal to the side of the plate to be subjected to a pair of external forces, can be adjusted beforehand in the processing of the plate. Since the stress sensitivity coefficient K_f of a plate is characteristic to the angle of cut and the azimuth of a plate, the use of a combination of plates makes it possible to increase the force-frequency sensitivity and to implement a sensor element having the zero temperature derivative of the force-frequency characteristics desirable for force sensor use.

In this paper, we discuss the force-frequency characteristics of a plano-convex square AT-cut plate from the following aspects; the experiment and the theory as an extension of the concept of a virtual force pair[10] to two-point support of a square plate and a possibility applicable to force sensor.

2. Extension of the theory of the force-frequency equation for a circular plate to a square plate

Since the specimen plates used are plano-convex square AT-cut plates, it can be regarded that the active region of vibration lies within a circle, which is concentric with the center of the plate and has a diameter of about half the side length of the plate[11]. Therefore, the forces instrumental in the frequency change may be considered force components giving rise to stress in the central part of the plate subjected to an opposing force pair F as shown in Fig.1. Let us take the following two assumptions to obtain the stress generated in the center of the plate.
(1) The frequency change in the plate is equal to that of a circular plate with the same thickness as the square plate and a radius equal to the distance between the center of the plate and one of the points being applied the external forces. This circular plate is named an equivalent circular plate.

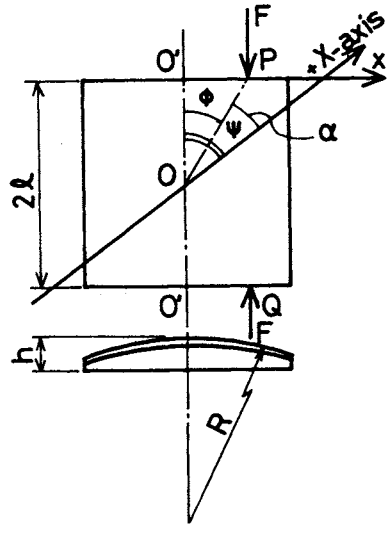


Fig.1. Geometries of a plano-convex square AT-cut plate, the definition of azimuth Ψ and an opposing pair of external forces F.

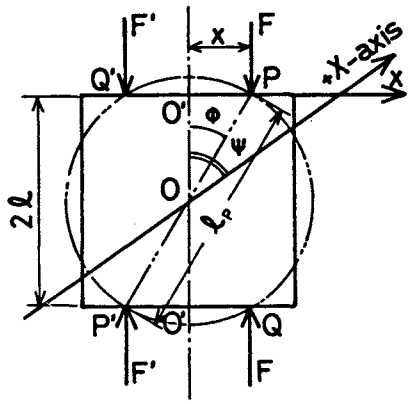


Fig.2. Virtual force F' and a real force-pair F.

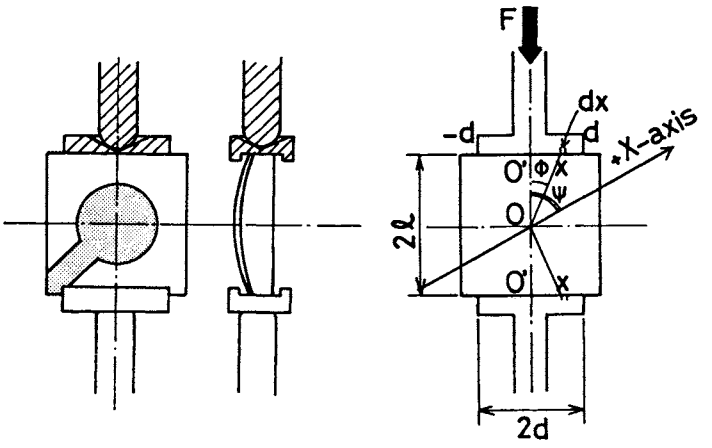


Fig.3. Configuration of application of line-force to the plate.

(2) Only the centripetal component of the external forces results in the frequency change, and the force component tangential to the side of the plate is ignored.

Moreover, the concept of the virtual force pair[10,12] and the superposition of an opposing force pair for the multiple point support are employed.

The geometry of a plano-convex square plate and the configuration for applying an opposing force pair F are shown in Fig.1. The point P and Q are separated at the same distance from the centerline O, O'. A virtual force pair F', whose magnitude is equal to the force of the pair F, is introduced to obtain the centripetal force component of the force F as shown in Fig.2. The radial force component F_p of the applied force at the point P shown in Fig.2 in an equivalent circular plate is give by (1).

$$F_p = F\cos\phi \qquad (1)$$

where ϕ is $\tan^{-1}x/\ell$. The relation between the length of the plate 2ℓ and the diameter ℓ_p of the equivalent circular plate is given by (2).

$$\ell_p = 2\ell(1/\cos\phi) \qquad (2)$$

On the other hand, the normalized frequency deviation ratio of a circular plate, that is, the frequency change of a circular plate due to an external diametric force pair to the resonant frequency is given by (3) [8].

$$\Delta f/f = K_f(\alpha)F_p N/Dh \qquad (3)$$

where Δf is an amount of the frequency change, f is the resonant frequency of the plate in free state, K_f is the stress sensitivity coefficient of the plate, α is an angle between the +X-axis and a line drawn from the center of the plate to the force-applying point P, N is the frequency constant of the plate, D is a diameter of the plate, which is equivalent to ℓ_p, h is the maximum thickness of the plate. The substitution of (1) and (2) into (3) yields (4), which gives the frequency change due to a virtual forace pair F and F' at the position P and P', respectively.

$$\Delta f/f|_p = K_f(\Psi-\phi)F_p N/\ell_p h = K_f(\Psi-\phi)FN\cos^2\phi/2\ell h \quad (4)$$

where Ψ is the azimuth between the +X-axis and a line normal to the side of the plate subjected to the force, ϕ is an angle between the centerline and the line drawn from the center to the point being applied the force.

Similarly the frequency change due to another force pair of F and F' at the position Q and Q', respectively, is given by (5).

$$\Delta f/f|_o = K_f(\Psi+\phi)F_o N/\ell_o h = K_f(\Psi+\phi)FN\cos^2\phi/2\ell h \quad (5)$$

The frequency change due to the force pair F shown in Fig.1 as a whole is given by the superposition of (4) and (5) as follows.

$$\Delta f/f = \frac{1}{2}\{K_f(\Psi+\phi) + K_f(\Psi-\phi)\}FN\cos^2\phi/\ell h \qquad (6)$$

The frequency changes due to the external force pair can be computed as a function of the position being applied forces by using (6).

Next, let us calculate the force-frequency characteristics of the square plate applied external forces from the flat sides of two bars having the same length 2d as shown in Fig.3. Since the force F_x on a finite length dx of the bar is given by (7), the substitution of (7) into (6) yields (8), which gives the frequency change due to external forces F_x on a finite length dx of the bar.

$$F_x = (F/2d)dx \qquad (7)$$

$$(df/f)_x = \frac{1}{2}\{K_f(\Psi+\phi) + K_f(\Psi-\phi)\}F_x N\cos^2\phi/\ell h \quad (8)$$

The total frequency change due to F applied on the bars is obtained by integrating (8) form -d to d. The total length 2d is the force-applying length.

$$\Delta f/f = \int_{-d}^{d}(df/f)_x dx \qquad (9)$$

3. Experimental

3.1 Specimen plate and application of the external force

Specimen plates were plano-convex square AT-cut plates of 5MHz as shown in Fig.1. The plates were processed in the same manner as in reference[10]. The circular exciting electrodes of Cr and Au of 8mm in diameter were deposited on the central regions of the plates one after another. The geometries and electrical characcteristics of the specimen plates are listed in Table 1.

Table 1. Geometries and electrical characteristics of plano-convex square AT-cut plate specimens.

No.	A-1	A-2	B-1	B-2	B-3	B-4
A side length of the square plate 2ℓ (mm)	15.00_5 x15.01_0	15.01_0 x15.00_8	14.95_0 x14.98_0	14.90_8 x14.93_2	14.92_0 x14.93_9	14.91_2 x14.91_0
Maximum thickness h (mm)	1.68_0	1.68_2	1.67_8	1.68_0	1.67_9	1.67_2
Radius of curvature R (mm)	116	116	116	116	116	116
Angle of cut θ (°)	34°55'	34°55'	35°05'±03'	35°05'±03'	35°05'±03'	35°05'±03'
Azimuth from the X-axis Ψ (°)	0 (90)	0 (90)	41.5 (48.5)	30 (60)	0.5 (89.5)	12.5 (77.5)
Q-value Q (x 10^6)	1.12	1.03	1.23	1.05	1.21	0.71
Equivalent resistance R_1 (Ω)	204	184	198	267	200	
Resonant frequency f (Hz)	4968122	4985797	4997230	5001800	5005470	5014780

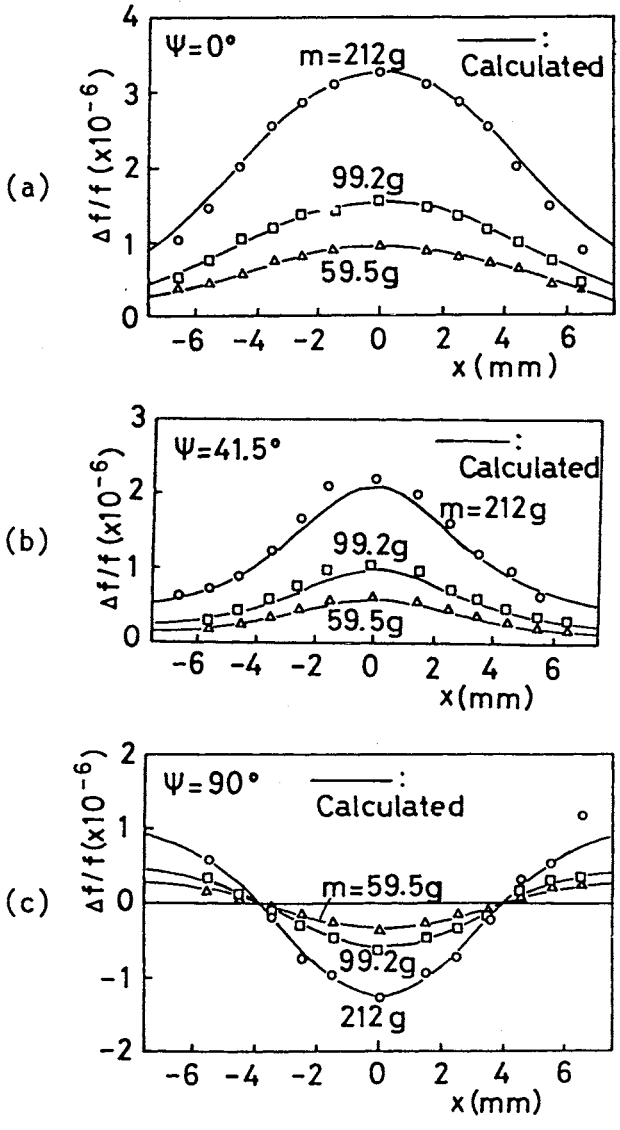


Fig.4. Force-frequency characteristics with respect to the applying position of the point-force. (a) specimen A-1 (Ψ = 0°), (b) specimen B-1 (Ψ = 41.5°), (c) specimen A-1 (Ψ = 90°)

The plate is set in a test holder to apply weights as an external force pair on the sides of the plate. The force was transmitted to the plate in two ways as shown in Figs.1 and 3, that is, point- and line-application of force. Specifically the line-force was applied on the both sides of the plate from two bars through a pivot on which the weight was added. A cone-shaped pit is engraved on the upper surface of the upper bar. The pit bears the pivot and serves to avoid the misfit, which likely happens between the flat side of the upper bar and the plate.

3.2 Measurement of the force-frequency characteristics

A specimen plate set in the holder was connected with an active circuit to compose an oscillator. The oscillating frequency was measured before and after the external force was applied to the plate. The temperature dependence of the force-frequency characteristics was measured in an oven a few to ten few times at each temperature, and measured values were averaged.

3.2.1 Force-frequency characteristics of the plate applied external forces from two opposing points

Since the force-frequency characteristics are symmetrical with respect to the X-axis, it is convenient to represent the force applying angle by the azimuth. The force applying position on the right side of the central line drawn against the convex surface of the plate is expressed as the +X's and the -X's on the left side.

Figure 4 shows the force-frequency characteristics of the plates measured by varying weights as a parameter. Solid lines in the figures are the theoretical lines computed by using (6).

3.2.2 Force-frequency characteristics of the plate applied external forces from two bars

The forces were applied symmetrically with respect to the central line of the plate. Figure 5 shows the force-frequency sensitivity as a function of the azimuth for a force applying length of 8mm. Figure 6 shows the force-frequency sensitivity measured by varying the length of the bar for various azimuth. It is noticeable that the frequency change does not always decrease with the increase in the length of the bar.

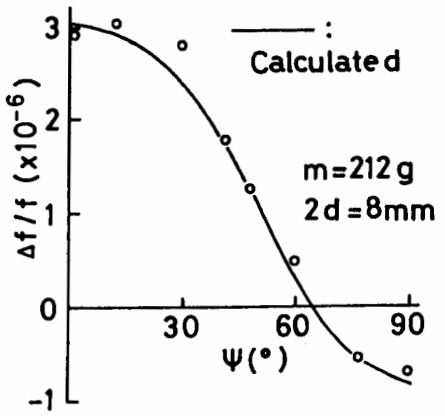


Fig.5. Frequency change vs. azimuth characteristics of the plate applied a line-force pair of a constant weight of 212g and a length of 8mm.

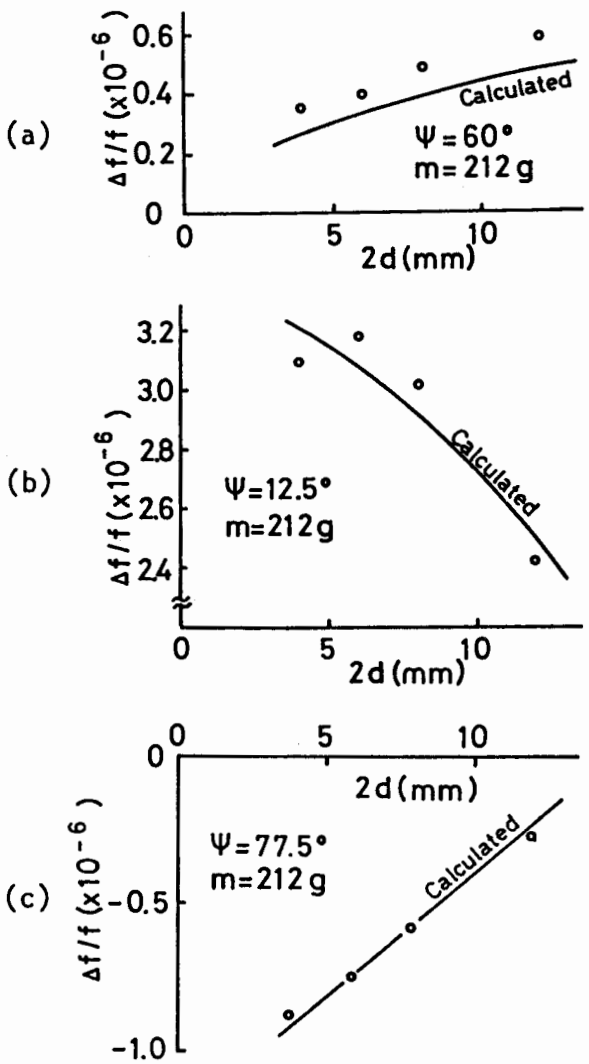


Fig.6. Frequency change vs. force-applying length. (a) Ψ =60° ,(b) Ψ =12.5° ,(c) Ψ =77.5°

3.2.3 Temperature dependence of the force-frequency characteristics

Since the force-frequency sensitivity varies with temperature, the temperature dependence of the force-frequency characteristics must be studied in advance of applying quartz plates to force sensor. Figure 7 shows the temperature dependence of the force-frequency characteristics of a plate having an azimuth angle of Ψ = 90° applied different weights as an external force pair at the center of the side of the plate. On the other hand, Fig.8 shows the angle-of-cut dependence of the force-frequency characteristics of plates applied the same weight approximately at the same azimuth. The solid lines are drawn by the least-square approximation using average values of the frequency deviation ratios measured at each temperature. The temperature coefficient of the characteristics is not only a function of temperature but also the azimuth of the plate. Figure 9 shows the first order temperature derivatives of the force-frequency characteristics of a plate subjected to a point-force pair, which are similar to the results obtained for circular plates[13,14], and to a line-force pair of a length of 8mm. The ordinate on the right side is graduated with magnitude of the derivatives of the stress sensitivity coefficient with respect to temperature.

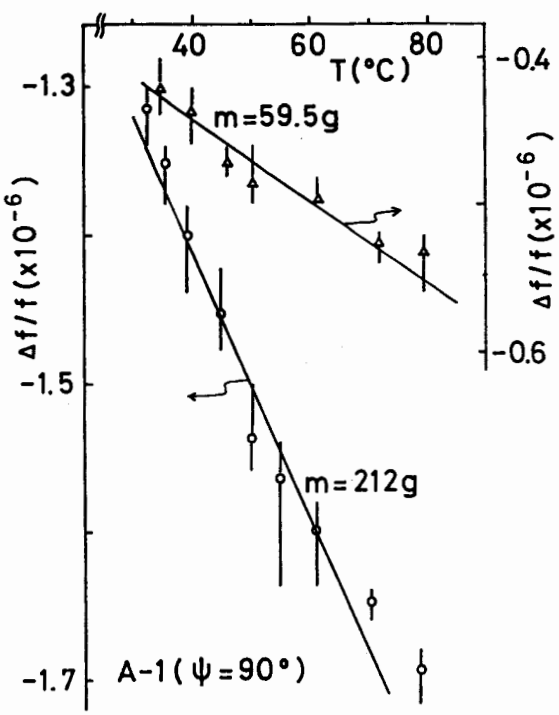


Fig.7. Temperature dependence of the force-frequency characteristics of a plate applied a point force pair at the centers of the sides. Solid lines are drawn by least squares.

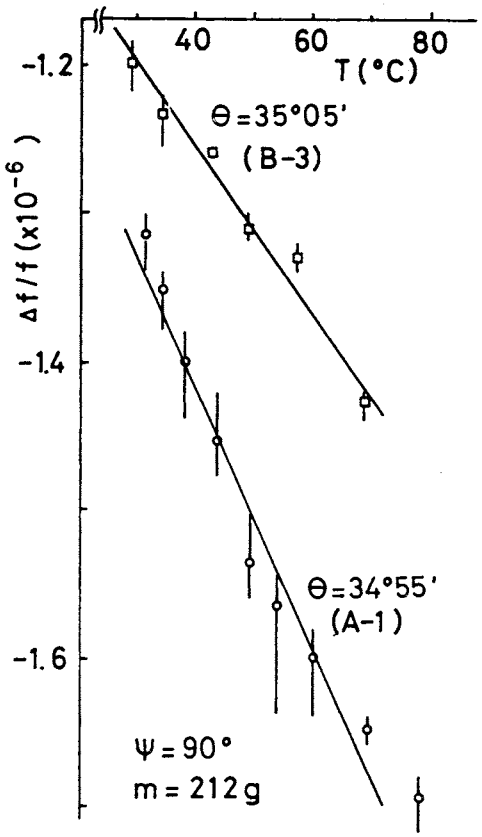


Fig.8. Temperature dependence of the force-frequency characteristics of plates of different angle of cut applied a point force pair at $\phi = 0°$. Solid line is drawn by least squares.

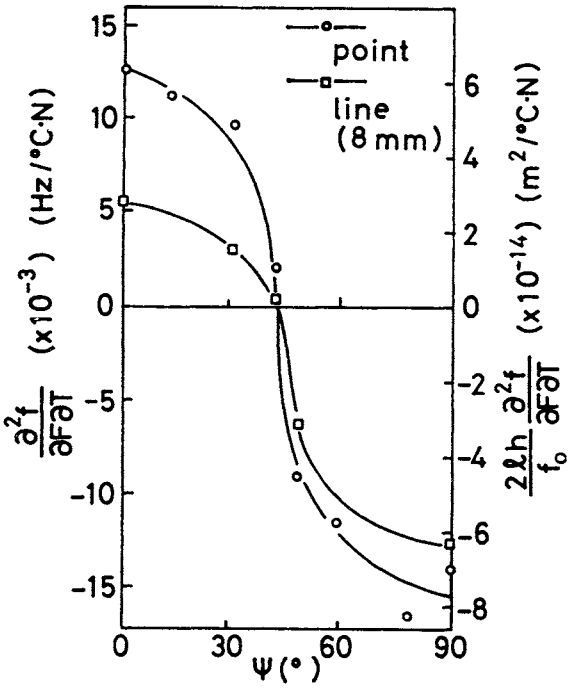


Fig.9. Temperature dependence of the force-frequency characteristics on the azimuth of plates applied a point force pair or a line-force of 8mm.

4. Discussion

Equations expressing the force-frequency characteristics of a plano-convex square AT-cut plate subjected to a point-force pair and a line-force pair are derived from an extension of the concept of the virtual force pair. The theory explains the experimental characteristics to a sufficient degree for the application. This theory may be extended to the calculation of the force-frequency characteristics of a square plate subjected to multiple pairs of point-force or arbitrary length of line-forces by using a combination of (6) and (9) or either of them.

The temperature dependence of the force-frequency characteristics of the plate is not so different from that of a circular plate. Therefore the use of a square plate is superior to a circular plate in the feasibility of exact setting of the force-applying point on the side surface of the plate. The temperature dependence should be derived from the derivative of (3) with respect to temperature in a context where the force-frequency characteristics are in general expressed by using the stress sensitivity coefficient K_f. The theoretical expression of the temperature dependence needs the measurement of the temperature derivatives of K_f.

At a glance it seems that the higher the derivative of the frequency deviation ratio, the greater the force-frequency sensitivity at a temperature is. The force-frequency deviation ratios at a temperature T_1, however, can be expanded into a series of temperature T for a point-force pair and a line-force pair using the first oder derivative of the ratio as follows;

$$\Delta f_P/f = \Delta f_{PT1}/f + [\partial(\Delta f_{PT1}/f)/\partial T]\cdot(T-T_1) \qquad (10)$$

$$\Delta f_F/f = \Delta f_{FT1}/f + [\partial(\Delta f_{FT1}/f)/\partial T]\cdot(T-T_1) \qquad (11)$$

where $\Delta f_{PT1}/f$ and $\Delta f_{FT1}/f$ are the force-frequency deviation ratios of a point-force pair and a line-force pair at a temerature T_1, respectively. It is obvious in comparison with the above equations that the first order derivatives of the force-frequency deviation ratio are not only important, but also the first term of both the equations. Therefore a new criterion is introduced as follows;

$$\xi_P = \int \left|\frac{\partial(\Delta f_P/f)}{\partial T}\right| d\Psi \Big/ \int |\Delta f_{PT1}/f| d\Psi \qquad (12)$$

$$\xi_F = \int \left|\frac{\partial(\Delta f_F/f)}{\partial T}\right| d\Psi \Big/ \int |\Delta f_{FT1}/f| d\Psi \qquad (13)$$

$$A = \xi_F/\xi_P \qquad (14)$$

The values ξ_P and ξ_F mean that which of the first order temperature derivative of the frequency deviation ratio integrated with azimuth Ψ or the linear frequency deviation integrated with respect to the azimuth has a greater effect on the frequency deviation due to the external force. The quantity A given by (14) is a measure of the quality of the force-applying type, that is, if the quantity is small, it means that the application of line-force is superior to the point application of force in the temperature dependence. Incidentally the quantity A for a supporting length of 8mm in this experiment is approximately 0.8.

The use of equilateral polygon plates and their combinations will make it possible to generate a beat frequency between two pieces of plate and also make it easier to support the plate and to compensate the static and the first order temperature dependence of the force-frequency sensitivity compared with the circular plates.

The easy computation of such characteristics shown in Fig.4 will make it possible to design the square resonator supported with the zero frequency shift points to be enclosed in a ceramic flat package[15] without employing a plate of special geometry.

5. Conclusion

The force-frequency characteristics of a plano-convex square AT-cut plate for force sensor were derived theoretically as an extension of the concept of virtual force pair and is expressed by using the stress sensitivity coefficient K_f as a common measure of the force-frequency sensitivity of quartz plate. The validity of the theory and the usefulness of the square plate were confirmed by the experiment. The temperature dependence of the force-frequency characteristics were measured from a view point of application to sensor element and an angle of cut at which the temperature derivative of the force-frequency characteristics becomes zero was obtained for the application of a line-force pair.

References

[1] A.D.Ballato, "Effects of initial stress on quartz plates vibrating in thickness mode," Proc. 14th AFCS, pp.89-114 (1960)

[2] C.R.Mingins et al., "Effects of external force on the frequency of vibrating crystal plates," Proc. 16th AFCS, pp.46-76 (1962)

[3] P.C.Y.Lee et al., "High frequency vibrations of crystal plates under initial stresses," J. Acoust. Soc. Am., Vol.57, No.1, pp.95-105 (1975)

[4] R.W.Keyes and F.W.Blair, "Stress dependence of the frequency of quartz plates," Proc. IEEE, Vol.55, pp.565-566 (1967)

[5] P.C.Y.Lee et al., "Elastic waves and vibrations in deformed crystal plates," Proc. 27th AFCS, pp.1-6 (1973)

[6] E.P.EerNisse, "Quartz resonator frequency shifts arising from electrode stress," Proc. 29th AFCS, pp.1-4 (1975)

[7] A.Ballato, Precision frequency control, Vol.1, (Academic Press, 1985)

[8] J.M.Ratajski et al., "Force-frequency coefficient of singly rotated vibrating quartz crystals," IBM Journal, Vol.12, No.1, pp.92-99 (1968)

[9] R. Besson et al., "Design of a bulk wave quartz resonator insensitive to acceleration," Proc. 33rd AFCS, pp.337-345 (1979)

[10] N. Oura et al., "Frequency changes in fifth-overtone 5-MHz circular plano-convex AT- and IT-cut plates due to radially applied three forces," IEEE Trans. Sonics Ultrason., Vol.SU-30, No.2, pp.111-114 (1983)

[11] H. Fukuyo et al., "Studies on circular planoconvex AT-cut quartz crystal vibrators(1)," Bull. Tokyo Inst. Tech., No.80, pp.18-24 (1967)

[12] N. Oura et al., "Generalized equation for the force-frequency characteristics of circular quartz," IEEE Trans. Sonics Ultrason., Vol.SU-32, No.6, pp.845-849 (1985)

[13] C.R.Dauwalter, "The temperature dependence of the force sensitivity of AT-cut quartz crystals," Proc. 26th AFCS, pp.108-112, (1971)

[14] E.P.EerNisse, "Temperature dependence of the force frequency effect for the AT-, FC-, SC-, and rotated X-cuts," Proc. 34th AFCS, pp.426-430 (1980)

[15] A.D.Ballato, "Force-frequency compensation applied to four-point mounting of AT-cut resonators," IEEE Trans. Sonics Ultrason., Vol.SU-25, pp.223-226 (1978)

2 ND EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 16 - 17 - 18 MARCH 1988

Session RO-02 Number 3

STROBOSCOPIC X-RAY TOPOGRAPHY OF QUARTZ RESONATORS.

by

A. ZARKA & B. CAPELLE.
Laboratoire de Minéralogie- Cristallographie. U.A.09. C.N.R.S.
Université P. et M. CURIE. 4 Place Jussieu. 75252 Paris. France.

and

J. DETAINT & J. SCHWARTZEL.
Centre National d'Etudes des Télécommunications.
PAB/BAG/MCT. 92220. Bagneux. France.

SUMMARY

Stroboscopic X-ray topography has been performed to observe the different states of acoustic vibrations in quartz resonators. These time resolved experiments have been done at L.U.R.E. (Orsay. France) using the pulsed light delivered by the storage ring of D.C.I. The studied resonators have shown that the vibration is affected by the nature, the density and the distribution of the defects.

INTRODUCTION

Time resolved on short time scale (about 1 nanosecond) can be done using the intrinsic structure of synchrotron radiation originating from a Storage Ring. So, the stroboscopic topography could be very well adapted to study the resonators because the frequency of the vibrations in the crystals and the frequency of the pulsed beam, are of the same order (some MHz). Studies of acoustic waves on YZ-LiNbO3 (1,2,3) and on quartz resonators (4,5) have been developed to study the vibration modes with this time resolution.

In this article, we report some results obtained using the pulsed beam delivered at L.U.R.E. by the storage ring D.C.I. concerning the interactions between the B.A.W.(Bulk Acoustic Waves) vibrations and the lattice defects observed in quartz resonators.

EXPERIMENTAL TECHNIQUES

Synchrotron radiation from Storage Rings has a time structure due to the fact that the circulating particules are concentrated in bunches. For D.C.I., which operates in single-bunch mode, the circulation time of an individual bunch is 315 ns and is length is 1ns.

The setting of the experiment is composed of two parts:

- The topographic setting is simple because the topographic station is equiped with a double axis spectrometer. This spectrometer permits, when the quartz sample is placed in the white incident beam to obtain a Laue pattern of topographs (6). Other interesting topographs can be obtained by the use of the monochromatic setting. In this case, the first axis holds a sample of germanium which is adjusted to select one wavelength (0.7 Å). By the use of a very fine slit we have also obtained "section" topographs in the two cases of setting.

- The electronic control. To obtain an exact synchronization of the X-ray pulses and of the piezoelectric vibrations of the resonators, a pulse signal obtained from the positron bunch (by a capacitive pick up) is used to generate the excitation signal of the resonator. This signal is

shaped in sinusoidal form by filtering at the recurrence frequency of the synchrotron (fo = 3,169280 MHz) or at an harmonic of this frequency n.fo (with n = 2,3 ...).
A phase shifter permits to vary the relative phase between the X- ray pulse and the sinusoidal signal used to excite the resonance of the crystal (after a level ajustment).

To have an exact resonance condition (the resonator voltage in phase with the resonator current) the resonators were adjusted by metallization to have frequencies very close to fo (or n.fo). A fine adjustment (a few hertz) is made with a large serial variable capacitance.

The electrical parameters of the resonator are measured, in situ, with a vector voltmeter. Another vector voltmeter is used to measure the relative phase angle between the resonator current and the X-ray pulse. After calibration this relative phase (R.P.C.X. in the following) is known with a precision of about 3°, most of the incertitude resulting of the effects of the temperature fluctuations on the resonator.

The general principle of the set-up is given in Figure 1.

EXPERIMENTS

The purpose of the study was to determine the role of the defects on the acoustical vibrations. The first Q1 sample apart, all the studied resonators were AT-Cut plano-convex (Table1) with the basic design of high Q factor, 5MHz, 5th overtone resonators. Q1 and Q3 samples were made from natural quartz and they were used at two times the synchrotron frequency. Q1 crystal was an AT-Cut plane resonator used at the fundamental mode. Q2 and Q4 resonators were made from premium Q synthetic quartz. All these resonators were deeply etched in saturated ammonium bifluoride and were electroded with Cr/Au metallization (electrode diameter 8mm, mass loading about 1%).

As established by the theory (7), the used vibration modes are more trapped in the case of Q3 and Q4 samples than for Q2.

TABLE 1.

	RESONANCE FREQUENCY	MODE	Rc	Φ	2ho
Q1	6.338560 MHz	Fundamental	none	15mm	0.253mm
Q2	3.169280 MHz	Third Overtone	150mm	15mm	1.592mm
Q3	6.338560 MHz	Fifth Overtone	150mm	15mm	1.322mm
Q4	6.338560 MHz	Fifth Overtone	175mm	15mm	1.319mm

In this table, Rc indicates the curvature radius of the convex part of the resonator, Φ, the total diameter, and 2ho, the thickness at the centre.

The dominating vibration mode in the AT-Cut resonators is the thickness-shear mode which corresponds to a standing transverse wave in the crystal (Figure 2). The planes parallel to the surface are displaced in the X1 direction leading to a sinusoidal shear of the planes perpendicular to the X1 axis.

The components u2 and u3 of the displacement are usually weak in regard to the u1 component, when no coupling to flexure or plane shear occurs.

On the X-ray topographs, and mostly on classical (integrated) topographs, the most important contast is due to the u1 component and depends on the level of the amplitude of the vibration applied on the crystal. This contrast depends also on the diffraction conditions.

Particularly, when the product **g.v** (**g** is the reflection vector and **v** the displacement vector due to the deformation) is zero, the contrast due to the vibration vanishes.

In our case, the most important u1 component of the displacement is along the [100] direction. Then, on classical or on stroboscopic topographs corresponding to the (Okl) reflecting planes the contrast of the deformation should not appear.

OBSERVATIONS

Different types of X-ray topographs have been obtained varying the R.C.P.X., the level of the input signal on the resonator and the topographic setting (Laue, monochromatic or section topographs). We shall only report here the main observations we have obtained for the studied resonators.

Q1 RESONATOR

This resonator was obtained from an excellent natural crystal and any dislocation have been seen in this sample.

In Figure 3, we present two topographs obtained from the same Laue pattern. In this case, the resonator was working at a medium power but was not synchronized with the X-ray pulse. These topographs were obtained by varying continuously the R.P.C.X. during the exposure time. So, the contrasts are similar to those which can be obtained by a classical setting (8) and the result is an integrated image of the deformation during the exposure time.The central black contrast which depends on the deformation is clearly visible for the $2\bar{1}0$ (Fig.3a) and vanishes for the $0\bar{1}3$ reflection(Fig.3b). For this reflection, **g.u1** = 0 and the observed contrast do not reveal other stationary component of the deformation and the crystal seems to be out of vibration.

In Figure 4, we present two topographs obtained in the same conditions but in contrary to the previous experiment the resonator was synchronized with the X-ray pulse (R.P.C.X. = 0). In Figure 4a ($2\bar{1}0$ reflection), it can be note some differences in the black central vibrating zone. This concerns mainly the concentric fringes which are due to X-rays interferences in the deformed zone. In Figure 4b ($0\bar{1}3$ reflection) the broad fringes are due probably to the u2 and/or u3 components of another mode of vibration which is coupled to the principal thickness-shear mode. This suplementary mode seems to be a flexure mode and is only revealed by the stroboscopic topographs. In contrary the contrast observed in Figure 3b is an integrated image for which these fringes are blurred out because their contrast (due to the continuous change of the R.P.C.X.) varies during the time of the exposure(6 minutes).

Since similar observations were made on all the studied plane resonators (about 20 samples of different designs) and since the frequency stability remains very high when the RPCX phase changes continuously at few radians per second, it can be deduce from these experiments that, most probably, other parasitic components of vibration can exist in plane resonators.

So, these 4 topographs show how sensitive the stroboscopic topography is to reveal the instantaneous state of vibration in resonators.

Q2 RESONATOR.

This resonator was produced from a synthetic sample intentionally choosed for its high density of dislocations.

In Figure 5a is shown a topograph of this crystal obtained for $11\bar{2}$ reflection with 90° R.P.C.X. which corresponds to the state of vibration when the deformation of the diffracting planes is equal to zero. The topograph is similar to that obtained without vibration and it can be observed the high density of dislocations oriented parallel to the **c** axis and a faint circular contrast due to stresses at the edge of the electrodes. The effects of the dislocations on the

vibration mode are shown in Figure 5b where it can be obviously seen that the zone of vibration has not the circular symetry expected, and where numerous dislocations unconfine the mode. This was also observed elsewhere by conventional X-ray topography. This last topograph (Figure 5b; $2\bar{1}0$ reflection) obtained with the monochromatic setting shows another interesting contrast in the central resonating zone: fringes of "moiré" type (Z.fringes in the following). Their configuration varies with the R.P.C.X. and they are due to interferences between X-ray wavefields propagating in the crystal. They are located very close to the dislocations and caracterize very locally deformed zones.

Q3 RESONATOR

This crystal is a natural one and the defects present are only growth bands. The topograph presented in Figure 6 was obtained with an 45° R.P.C.X. and with an input power on the resonator of 90 mW. This resonator shows different interesting features:

- The central black zone presents a less contrasted (white) part which depends on the applied R.P.C.X.
- This is also verified for the Z. fringes which appear inside the growth bands.
- The growth bands seem to interact with the vibration mode but this effect is less important tin the case of the dislocations.

Q4 RESONATOR.

This crystal is a synthetic one and is was choosen for its excellent quality. The topograph shown in Figure 7 was obtained with an 0° R.P.C.X. using the monochromatic setting with the $2\bar{1}0$ reflection.In this case the principal features observed are :
- The central vibrating zone presents a quite perfect circular form.
- The Z.fringes, probably due to the important level of vibration, present a quite symetrical configuration disturbed only by the two observed dislocations (indicated by two arrows in Fig.7).

These two principal features are due to the excellent quality of this resonator in contrary to the observations obtained from the Q2 and Q3 resonators.

DISCUSSION

To summarize the different observations done on these AT-Cut resonators, it can be said that the global effect of the defects is to affect the purity of the vibration modes:

- The effect produced by the dislocations is to decrease the lateral confinement of the thickness shear mode when their density is important (10^2 to 10^3 / cm^2). The dislocations seem to modify locally the phase velocity of the shear mode and so to contribute to untrap locally the mode. This hypothesis is renforced by the observation that bunches of dislocations of similar orientations constitute a guiding structure for the shear wave at the periphery of the vibration mode. So, in the case of weakly confined modes, an important untrapping effect can be observed (Q2 resonator).
- The high impurity level in the growth bands may also introduce local variations of the phase velocity of the shear mode and thus in the energy trapping parameters, but due to the complexity of the growth bands structure (specially in the Q3 sample), no simple conclusion about this point can be done until now and further experiments are required.
- The Z.fringes seem to reveal the local deformations and are probably due to modifications of the tranverse wave shape due to the influence of the defects.

CONCLUSION

The time structure of Synchrotron Radiation enables stroboscopic X-ray topography with nanosecond time resolution. By the use of different settings this technique can reveal directly the different modes of the vibration by visualization of all the components of the (instantaneous) deformation. This is a new advantage in comparison with the classical (integrated) topography.

Defects, like growth bands and specially dislocations disturb the B.A.W. propagation in crystals and can induce losses either by untrapping a part of the energy of the mode or by mode conversion.

The stroboscopic method can also permit to obtain other results about the piezoelectric resonators such as information relative to the localization of acoustic dissipations, or to the influence of the imperfect geometry of the surfaces.

However, the topographic contrasts due to the propagation of the X-ray wavefields in the deformation fields of acoustic waves, are not well understood and calculations of the beam trajectories in the crystal and simulations of topographs will be necessary to explain the contrasts (specially the Z.fringes).

ACKNOWLEDGMENTS

The authors wish to acknowledge M. X. BUISSON and his group, S.I.C.N. Annecy, FRANCE, for all the synthetic quartz and particularly dislocation free samples which have been produced by this company .

This work is supported by the D.R.E.T. (France) under the contract N° 53585.

REFERENCES

1. Cerva, H, and Graeff, W. (1984). Phys. Stat. Sol. (a). **82**. 35.

2. Cerva, H, and Graeff, W. (1985). Phys. Stat. Sol. (a). **87**. 507.

3. Cerva, H, and Graeff, W. (1986). Phys. Stat. Sol. (a). **93**. K129.

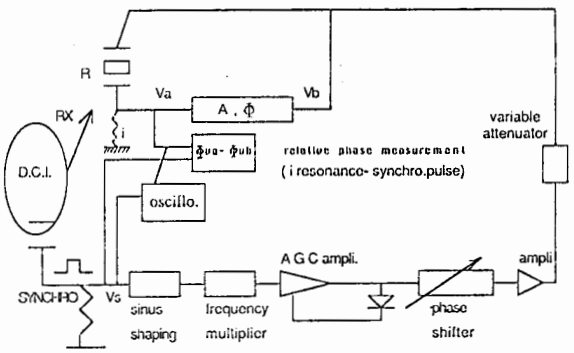
4. Gluer, C.C., Graeff, W. and Moller, H. (1983). Nucl. Inst. & Meth. **208**. 701.

5. Zarka, A., Capelle, B., Detaint, J. and Schwartzel, J. (1987). **41**th Annual Frequency Control Symposium. Philadelphia. U.S.A. p. 236.

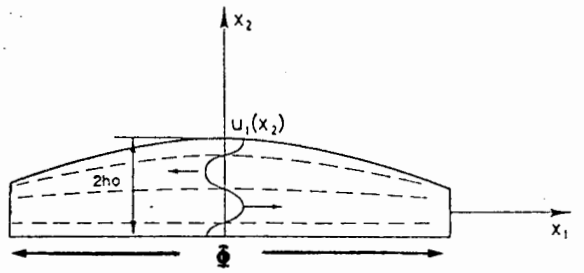
6. Tuomi, T., Naukkarinen, K. and Rabe, P. (1974). Phys. Stat. Sol. **A 25**. 93.

7. Tiersten,H.F. and Smythe, R.C.(1979). J.A.S.A., **65,** 1555

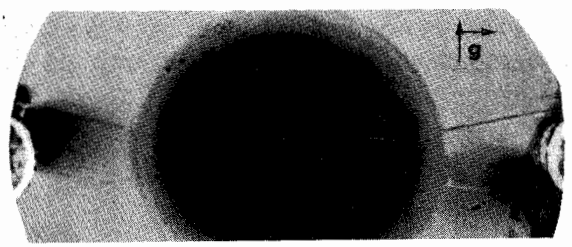
8 Detaint, J., Schwartzel, J., Joly,C. and Philippot,E. (1987). **41**th Annual Frequency Control Symposium. Philadelphia. U.S.A. p. 314.



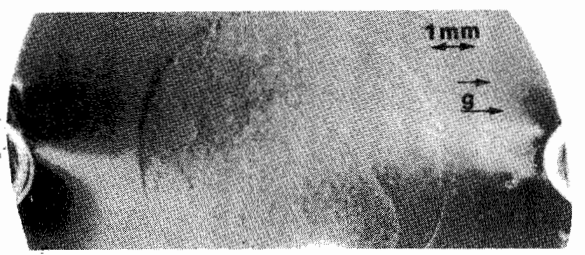
1



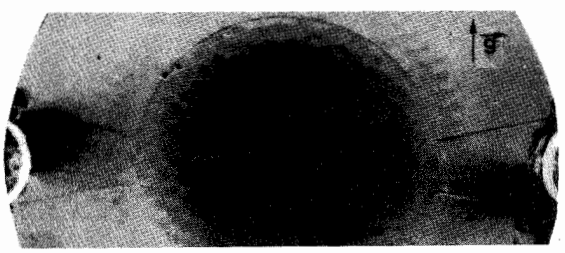
2



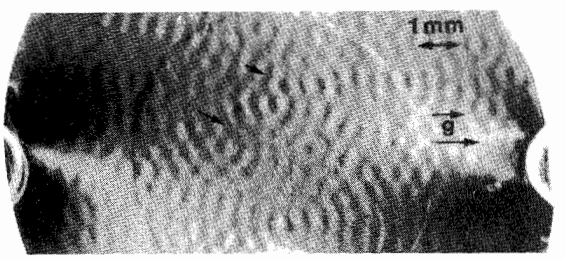
3a



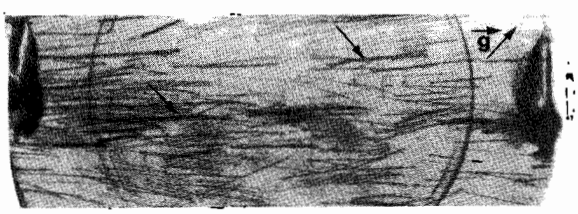
3b



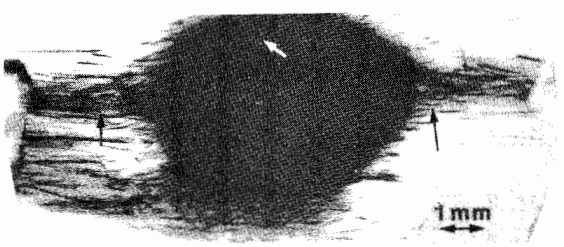
4a



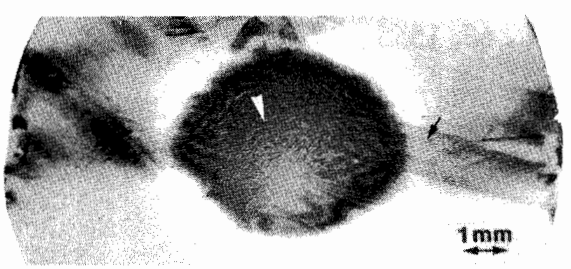
4b



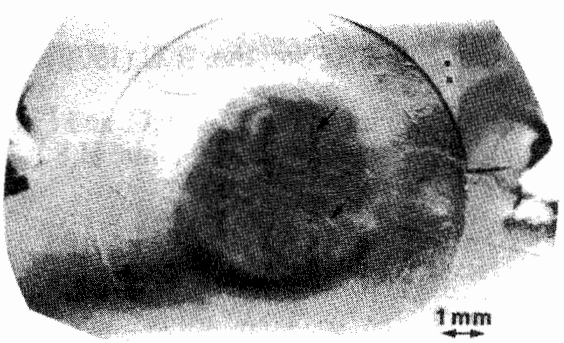
5a



5b



6



7

2 ND EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 16 - 17 - 18 MARCH 1988

Session RO-02 Number 4

STUDIES ON OSCILLATING QUARTZ RESONATORS FOR APPLICATION IN FREQUENCY CONTROL AND STANDARDIZATION

Harish Bahadur, R. Parshad and B.S. Mathur
National Physical Laboratory
Dr. K.S. Krishnan Road,
New Delhi-110012
INDIA

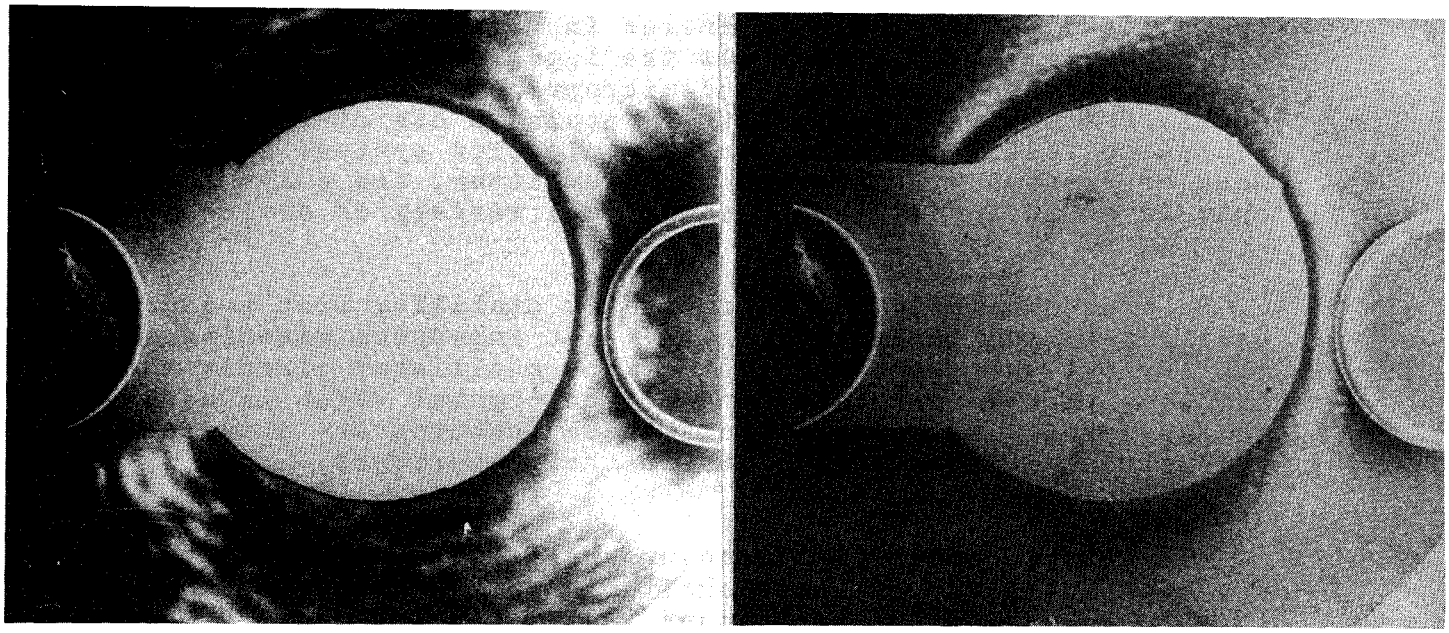
In the modern electronic technology the quartz crystal unit is an indispensable component. Not only, it is used to control and manage the frequencies in virtually all communication systems but it provides the isochronous element in most clocks, watches, computers and microprocessors. To obtain stable frequencies from quartz crystals is one of the most desirable features in various communication systems and development of Frequency Standards. In the latter, the quartz crystal unit is an integral part in almost all variety of devices (1,2).

For obtaining a high frequency stability from a quartz oscillator what is desired is the best impedance matching of the crystal unit with the rest of the oscillator circuit compounded with the increased mass of the crystal which is equivalent to its series inductance. Both these features are obtainable by exciting crystal resonators in their overtones.

In the recent past Bahadur and Parshad developed two new techniques for overtone excitation of quartz resonators (3,4). In one of the techniques (3), a heavier forward dc bias in a transistorized self-oscillating crystal-controlled oscillator circuit resulted in generation of increasing number of harmonic overtones of quartz resonators. The other techniques consisted of coupling a tunable L-C resonant network across the quartz crystal tuned to the desired overtone number (4). This technique also resulted in determination of the entire frequency spectrum of quartz crystals (5).

In the present work we have supplemented the studies of overtone excitation of quartz crystals by observing various mode shapes of quartz resonators in a Scanning Electron Microscope.

Quartz crystals when set into vibrations develop polarization charges spread onto their surfaces. The manner in which these charges are distributed on the surface vary with the type of mode of oscillations of the crystal. SEM is a very powerful tool to investigate the nature of electric potential distribution on the surfaces. In the technique of using SEM for our purpose the electron beam is usually accelerated to 2 kV. The other SEM parameters were :tilt = 0^{o}, beam current = 175 μA. The SEM was operated in the back scattered voltage contrast mode. Figure 1 depicts a typical micrograph of a 10 MHz flat fundamental TS frequency crystal unit and Figure 2 represents its mode shape in the third harmonic overtone. From the comparison of these Figures it can be easily observed that the band dimensions in the two vibrational modes are in proportion of their oscillation frequencies. On the other hand, 5 MHz resonator units (relatively thicker) exhibited a mode shape pattern very different from those presented in Figures 1 and 2.



SEM MICROGRAPHS

FIGURE 1 FIGURE 2

Fig.1 Generation of circular flexural waves in a disk resonator of 10 MHz fundamental TS frequency.

Fig.2 Third harmonic overtone excitation of crystal resonator of Figure 1.

Earlier (6,7) it has been shown that SEM micrographs in some rectangular resonator units exhibited appropriate overtones of flexural vibrations coupling with the fundamental TS vibrations. Flexural vibrations were found to be propagating both parallel and perpendicular to X-axis. The above micrographs (Figures 1 and 2) show that the flexural vibrations both propagating parallel and perpendicular to X-axis are coupled with a phase difference of $n\pi/2$ where n is an odd integer; thus resulting in generation of circular flexural waves.

The present studies indicate that for obtaining higher frequencies relatively thicker crystal units may be excited at their appropriate overtones to avoid generation of flexural vibrations and hence decrease the elastic coupling of flexural and TS modes of vibrations. Thicker units may also be driven at high drive levels without the risk of fracture as contrasted to the use of thin crystal units which due to their reduced thickness would exhibit flexural vibrations along with the risk of fracture at high drive levels.

The studies reported would find a direct and immediate application in obtaining higher frequency stability from quartz crystal units than the conventional techniques.

REFERENCES

1. H.Bahadur and R.Parshad, J.Sc.Ind.Res. 33, 561(1974).

2. E.A.Gerber and A.Ballato (Eds), Precision Frequency Control Vols. 1 & 2. Academic Press Inc., U.S.A. (1985).

3. H.Bahadur and R.Parshad, Proc. IEEE, 68, 1345 (1980).

4. H.Bahadur and R.Parshad, Proc. 7th Precise Time & Time Interval Planning Meeting, NASA Goddard Space Flight Center, Greenbelt, MD, U.S.A. pp 215-229 (1975).

5. H.Bahadur and R.Parshad, IEEE Trans. Sonics and Ultrason, SU-25, 24 (1978).

6. H.Bahadur, V.K.Lall and R.Parshad, Scanning Electron Microscopy/1981/I, Chicago, IL. 333 (1981).

7. H.Bahadur and R.Parshad, Physical Acoustics (Eds. W.P.Mason and R.N.Thurston), Vol. 16, Ch.2 pp 37-171, Academic Press Inc., N.Y., U.S.A. 1982.

2 ND EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 16 - 17 - 18 MARCH 1988

Session RO-02 Number 5

THE DEVELOPMENT OF A TRANSPORTABLE HYDROGEN MASER TO BE USED FOR THE VLBI NETWORK IN CHINA

Z.C.ZHAI, H.X.HUANG, C.F.LIN, G.X.JIANG, W.H.LUO AND LU JIA-FU

SHANGHAI OBSERVATORY, ACADEMIA SINICA,
SHANGHAI, CHINA

ABSTRACT

Shanghai observatory has developed hydrogen masers Since 1970, and six hydrogen masers, so far, were made in Shanghai Observatory. These hydrogen masers are being used in China for establishment of atomic time scale and for research of astronomy, especially for VLBI experments.

Three new hydrogen masers for VLBI network in China are being developed in Shanghai Observatory Since 1985, and one of them is finished. The new generation is a readily transportable, rugged and completely integrated frequency standard. The entire system is engineered as a unit, and all electronics, including the receiver and the synthesizer, were mechanically and electrically integrated with the physical portion into the overall design. This paper describes the mechanical and electronics design philosophy of the new maser, as well as the preliminary tested data.

RO-03

B V A

1. Systematic Characterization of a High Stability Resonator in Pre-Serie Production
   R. Delaite, Th. Migeot
   E.N.S.M.M., Besançon - France

2. The Radiation Susceptibility of BVA-AT Resonators to Simulated Medium Earth Orbit Radiation
   J.J. Suter, A.G. Bates, J.M. Cloeren
   The Johns Hopkins University, Laurel - USA

3. In Plane Orientation Optimization of QVA Resonators
   J.P. Aubry, B. Bonnotte
   CEPE, Argenteuil - France
   J.C. Craveur, F. Deyzac, D. Janiaud
   Onera, Châtillon - France

4. Space qualified 100 MHz BVA-VHF-Resonator
   U.R. Peier, G. Jenni
   Oscilloquartz SA, Neuchâtel - Switzerland

5. Results of the Qualification Programme of a BVA-Type Crystal Oscillator for Satellite use
   W. Schäfer
   University of Stuttgart, Stuttgart - FRG

6. Environmental Effects on the Medium and Long Term Frequency Stability of quartz Oscillators
   F.L. Walls
   National Bureau of Standards, Boulder - USA

2 ND EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 16 - 17 - 18 MARCH 1988

Session RO-03 Number 1

SYSTEMATIC CHARACTERIZATION OF A HIGH STABILITY RESONATOR IN PRE-SERIE PRODUCTION

Richard DELAITE and Thierry MIGEOT

Laboratoire de Chronométrie, Electronique et Piézoélectricité
E.N.S.M.M. - Route de Gray - La Bouloie - 25030 BESANCON CEDEX - FRANCE
Tél : 81.50.36.55

ABSTRACT

Miniaturized piezoelectric high stability resonators are mandatory components for various applications such as localization, space communications and so on. Their specifications are determined by a range of tests, electrical, mechanical or climatical tests for example, which enable us to verify if they match various requirements.

A few tests carried out on a prototype are not sufficient to obtain a cheap but reliable resonator. Thus, a set of other conditions must be taken into account like : pre-series results, reproductibility of performance and reliability of the product. Under these conditions a pre-serie production is suitable in which resonators are systematically tested in order to gather statistics and define how the performances can be reproduced according to dispersion. Moreover, these tests point out any defect in the resonator construction and allow technical changes and industrial production.

From pre-serie production of the new generation miniaturized high stability resonator, first statistical results are given for turn over point, quality coefficient, aging. This is also performed for environments tests (acceleration, external pressure and vibrations).

All these tests increase the cost of the resonator. Yet, this overcost is offset by a very low rate of defect, a fast after-sales and warranty service and a greater reliability.

RESUME

Les résonateurs piézoélectriques miniaturisés de haute stabilité sont des composants fabriqués industriellement pour différentes applications comme les localisations ou les communications. Leurs spécifications passent par une série de tests comme les essais climatiques ou mécaniques pour vérifier la conformité au cahier des charges.

Quelques essais sur prototype ne suffisent pas pour obtenir à la fois la qualité et un faible coût du résonateur ; interviennent alors des notions de présérie, de reproductibilité des performances, de coeffient de sécurité ou de fiabilité du produit. C'est la raison pour laquelle une présérie est fabriquée industriellement, puis testée systématiquement pour obtenir des statistiques et définir la reproductibilité des performances du résonateur en fonction de la dispersion. De plus, ces essais ont l'avantage de mettre en évidence une défaillance du résonateur et la possibilité d'y remédier avant toute production.

A partir d'une fabrication en pré-série d'un résonateur miniaturisé de haute stabilité de nouvelle génération, les premiers résultats statistiques sont donnés pour la température du point d'inversion de la courbe fréquence-température, pour le coefficient de qualité, pour la stabilité à long terme (vieillissement) ainsi que pour les tests en environnement (accélération, pression extérieure et vibration).

L'ensemble de ces essais contribuent à augmenter le coût du résonateur. Cependant, ce surcoût est limité par un taux de défaillance très faible ainsi que par un service après vente et un retour en garantie très réduits.

1 - INTRODUCTION

For various spatial and earth applications of localization and communication, the miniaturized high stability resonators[1] equip boarded instruments.

At the time of the new resonator conception, climatical or mechanical prototype tests allow to verify the conformity of the requirements.

These different prototype tests are not sufficient to obtain the required quality for this kind of component, therefore notions of reproduction of performances, security coefficient, stability time and reliability of the product or specific tests on pre-serie, have to be studied.

2 - SYSTEMATIC TESTS

These systematic tests cover two distinct types specifications of resonators. The first tests generaly implemented, allow to define the turnover point of the frequency-temperature curve, the quality coefficient, as well as the short or long term stability (called aging). Then, climatical and mechanical tests are performed.

The resonator must sustain some attacks during his life time and that's why, in the conception, the whole parameters linked to the environment have to be known.

Then, some solutions are proposed, according to the data, to improve the resonator performance, reduce the fabrication and the assembly times as well as the cost of the product. When the resonator is completely defined, the prototype is put under severe environment.

However, these first results are only pieces of information and to caracterize the resonator correctly, it is necessary to put them in an industrial product environment. The pre-serie is manufactured and then systematically tested to obtain statistics and define the reproduction of the resonator performance according to the scattering.

Moreover, these tests have the benefit to underline a weakness of the resonator, and give us the possibility to remedy before launching production. The conditions between the simulated and the real environment can be different.

Today, a security coefficient can be applied arbitrary or with comparison between the test results and those which are defined in real environment.

With this correction mistakes arising from the imprecision of the measures, have also to be taken into account to obtain a guarantee on the resonator.

At last, when all conditions are gathered for the manufacture, it is important to follow the achievement of the industrial process with notably some control tests. This procedure aims at eliminating the failing resonators.

At this stage, the defects of the manufacture are detected, those of the conception are detected in the pre-serie tests.

All these tests increase the cost of the resonator. Yet, this overcost is offset by a very low rate of defect, a fast after-sales and a warranty service, and a greater reliability.

3 - PRE-SERIE RESONATOR DESCRIPTION

Further to some patents[2][3][4][5] and to the achievement of some prototypes of the new miniaturized high stability resonator[1], a loan[6] of A.N.V.A.R. (Agence Nationale de Valorisation de la Recherche) has been granted to the L.C.E.P. laboratory (Laboratoire de Chronométrie, Electronique et Piézoélectricité) to allow the pre-serie development and the statistic study of the resonator. A detailed description has been given in a precedent communication[1].

We shall simply quote that the system (making a global stability when the whole perturbations are applied), is composed of a HC 40 holder, an eachhanger plan-convex, AT or SC cut quartz crystal which vibrates in a thickness mode at 10 MHz P3, two non adherent electrodes[3] and

peripheral structure[5] which link the crystal to the holder. All these element were placed in an improved aging holder[7] to obtain a long term stability.

The peripherical structure is realized with four rigid nippers associated to a symetrical suspension with springs which have three degrees of liberty in translation with equal stiffness. At last, the influence of the outside pression is stopped.

The resonator assembly is mounted in a cold-weld enclosure with chimney, to obtain 10^{-7} mbar at 250°C (Figure 1).

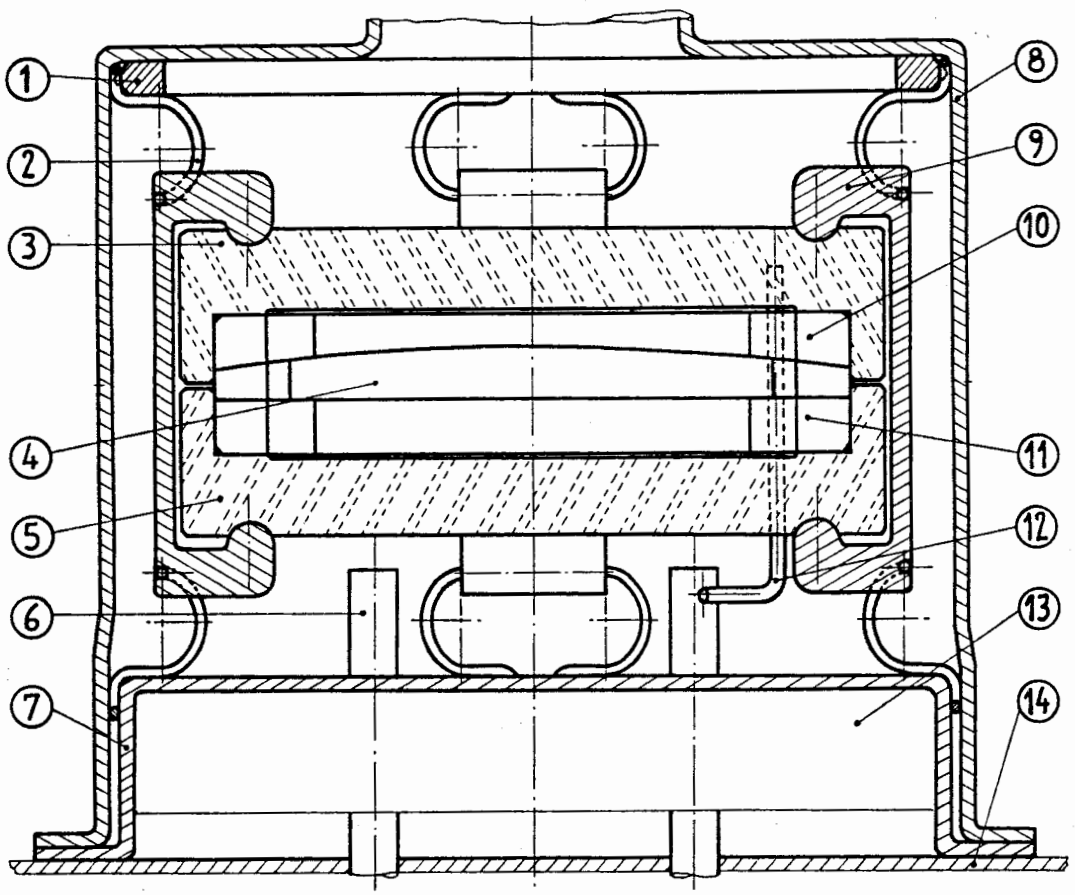


FIGURE 1 — Detail of symmetric internal structure of miniaturized piezoelectric high stability resonator : the plan-convexe cristal with two non adherent electrodes, the peripheral structure with two rigid quartz half-supports, four rigid nippers, a symetrical spring and a ring mounted in a cold-weld enclosure with chimney.

4 - TESTS AND EXPERIMENTAL RESULTS

4.1/ General characteristics

The thirty tested resonators in pre-serie development are made with AT-cut quartz crystal and with the symetrical syspension of peripheral structure.

The turn over point temperature of the frequency-temperature curve and the quality coefficient are measured for each resonator. The long term stability (on one week, one month, two months and three months) is measured for the first four resonators. These tests are realized in the L.C.E.P. laboratory.

We know that the temperature of the turnover point is determined with the frequency-temperature cubic extremum and that the variation of the crystal frequency is linked to the variation of the temperature around the resonator.

For the thirty tested resonators, the statistic curve (Figure 2) can be assimilated to a gaussian curve, the average value of which

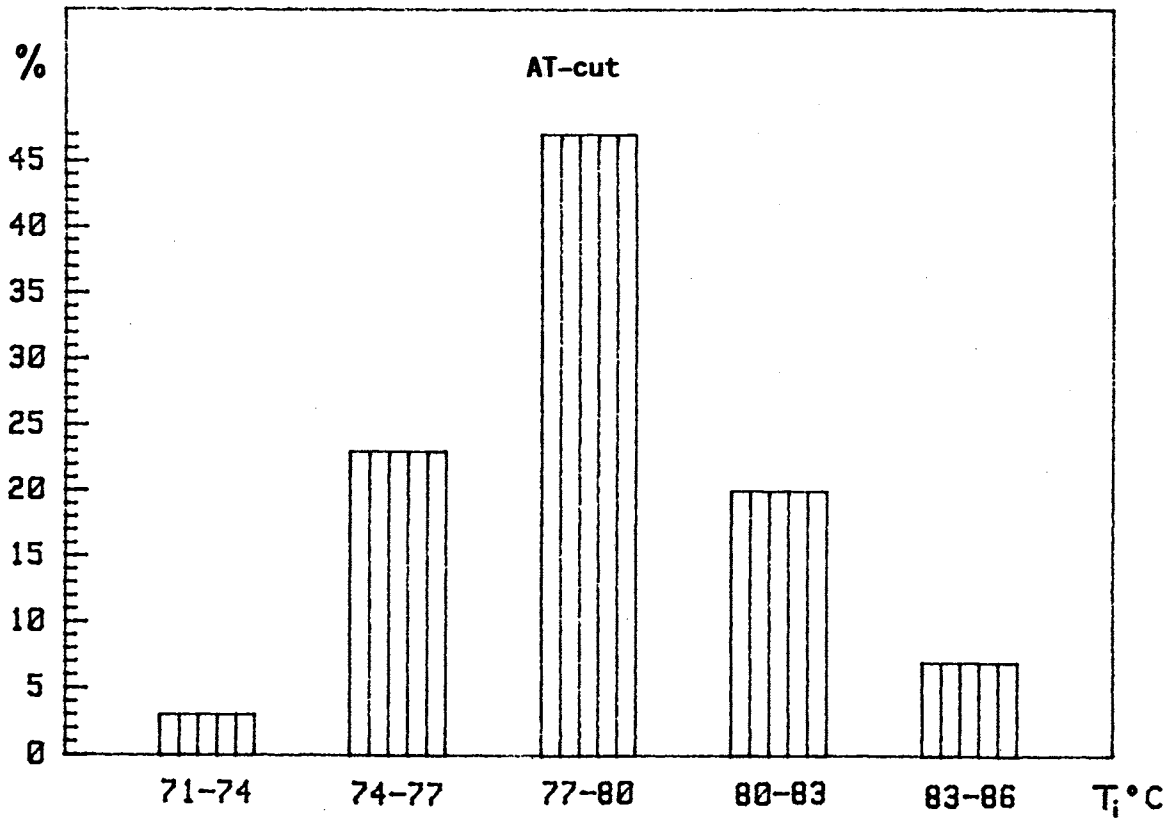
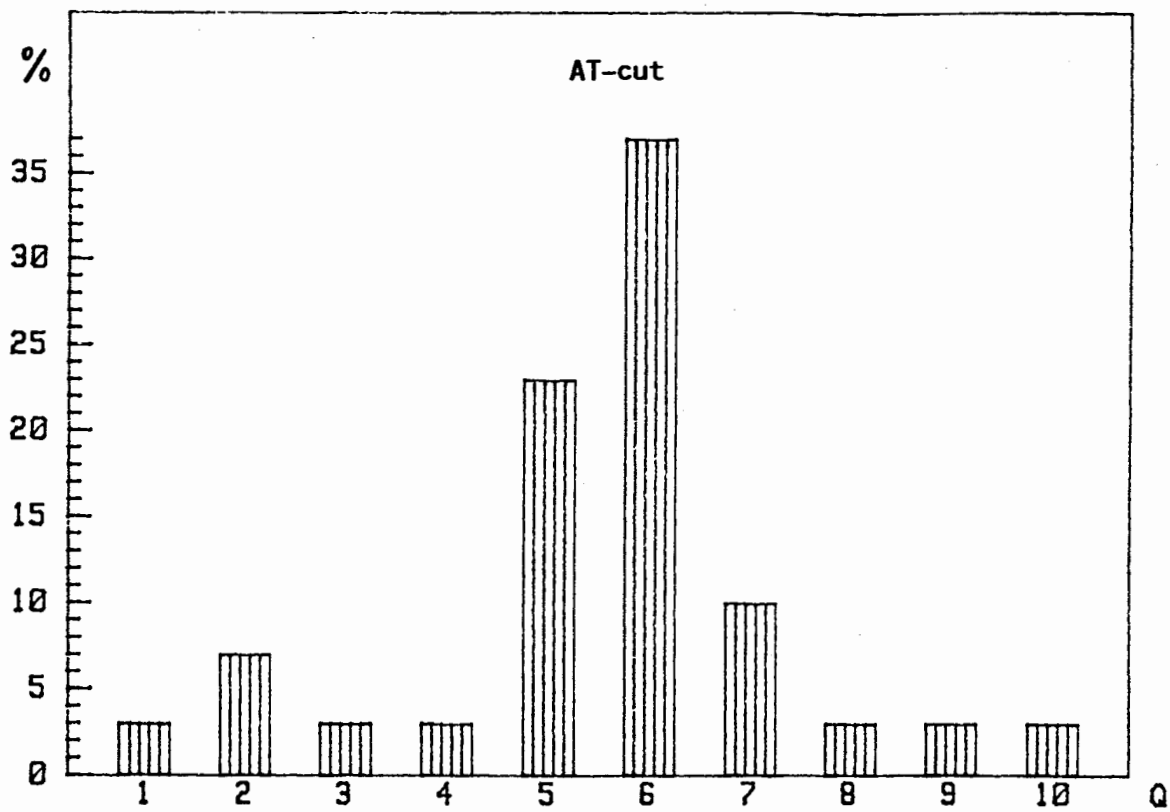


FIGURE 2 - Temperatures of the turn-over point of thirty resonators.

being 78.5°C. The respectively maximum and minimum value are 71°C and 86°C. A reduced scattering is obtained with 47 % of the resonators for some turn over temperatures contained between 77°C and 80°C.

About quality coefficient for the thirty tested resonators, the representative statistic curve of the quality coefficient Q can be linked to a gaussian curve with an average value of 1150000. The minimum value is 1038000 and the maximum value is 1258000 (Figure 3).

The dispersion is still small since 70 % of the resonators have a quality coefficient contained between 1126000 and 1192000.



1: 1.038.000-1.060.000
2: 1.060.000-1.082.000
3: 1.082.000-1.104.000
4: 1.104.000-1.126.000
5: 1.126.000-1.148.000
6: 1.148.000-1.170.000
7: 1.170.000-1.192.000
8: 1.192.000-1.214.000
9: 1.214.000-1.236.000
10: 1.236.000-1.258.000

FIGURE 3 - Quality coefficients of thirty resonators.

The thermostated resonators are placed in an isolated room for the long term stability measures (aging). A frequencymeter records simultaneously the frequency of each resonator and gives the frequency variation according to the long term time. This resonator has preliminary a sustain stability treatment of fiveteen days with 10^{-7} mbar at 250°C. For the first four resonators, long term stability (on three months) is contained between 2.10^{-11} and 5.10^{-11} (Figure 4).

AT-cut

long term / Resonator	1 week	1 month	2 months	3 months
1	$3,2 \cdot 10^{-10}$	$1,5 \cdot 10^{-10}$	$5 \cdot 10^{-11}$	$2 \cdot 10^{-11}$
2	$4 \cdot 10^{-10}$	$1,7 \cdot 10^{-10}$	$6 \cdot 10^{-11}$	$3 \cdot 10^{-11}$
3	$3 \cdot 10^{-10}$	$1 \cdot 10^{-10}$	$9 \cdot 10^{-11}$	$5 \cdot 10^{-11}$
4	$3,8 \cdot 10^{-10}$	$1,6 \cdot 10^{-10}$	$5 \cdot 10^{-11}$	$2 \cdot 10^{-11}$

FIGURE 4 - Long term frequency stability (per day) of the first four resonators.

4.2/ Mechanical sensitivities

The mechanical sensitivities of acceleration and pressure are measured for the first seven resonators, but only the first two resonators are tested in vibration environment.

The acceleration frequency variation is determined directly by the turn round resonator principle[8] which correspond to twice g (gravity acceleration). The resonator is put into a thermostat and is tested at his turn over temperature. The first seven AT-cut resonators have been tested in the L.C.E.P. laboratory and in the aerospatial agency O.N.E.R.A. (Office National d'Etudes et de Recherches Aérospatiales). The minimum sensibility is $0,8.10^{-10}$ for an axial acceleration of 1 g (Figure 5). The greater amplitude alone is taken into account for the radial sensitivity, smaller values than $0,6.10^{-10}$ are recorded for 1 g in the crystal plane (Figure 6).

In the twice case (Figures 5 and 6) a scattering of $1.10^{-10}/g$ or $0,5.10^{-10}/g$ is obtained for 55 % of the resonator, respectively in radial or axial acceleration.

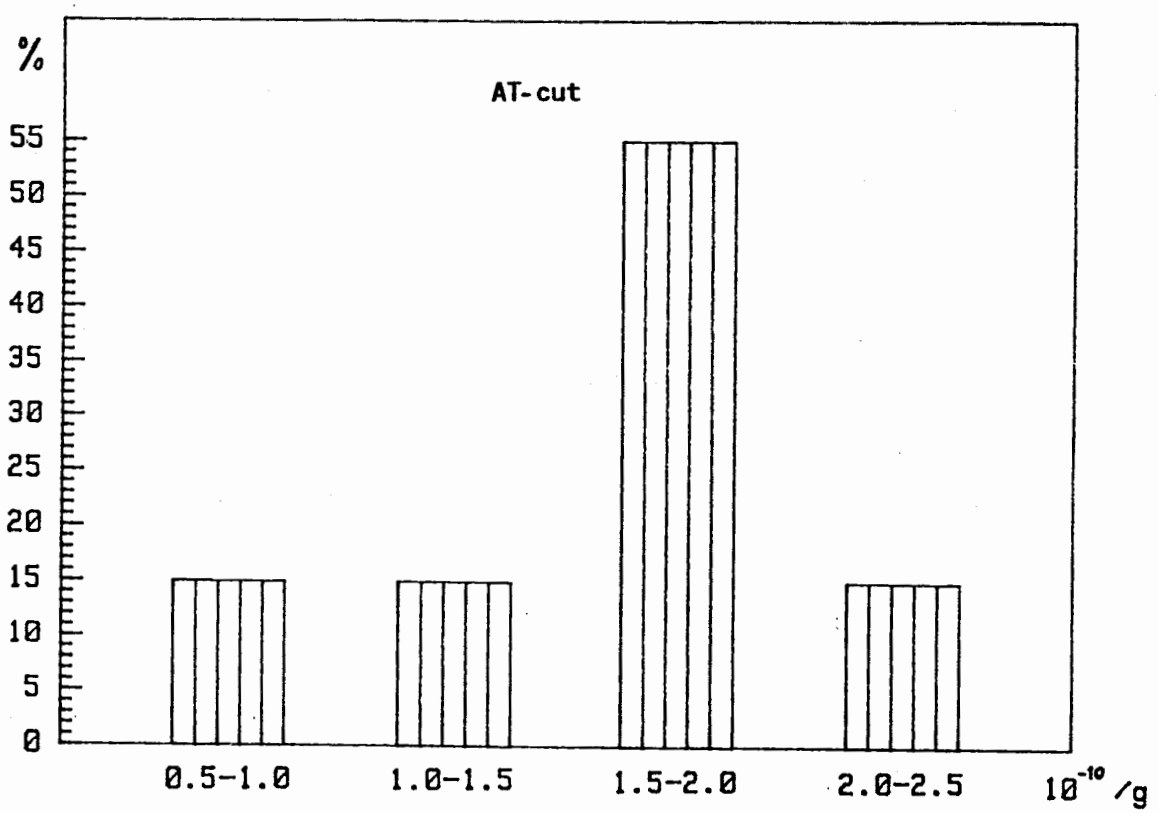


FIGURE 5 - Axial acceleration sensitivity of the first seven resonators.

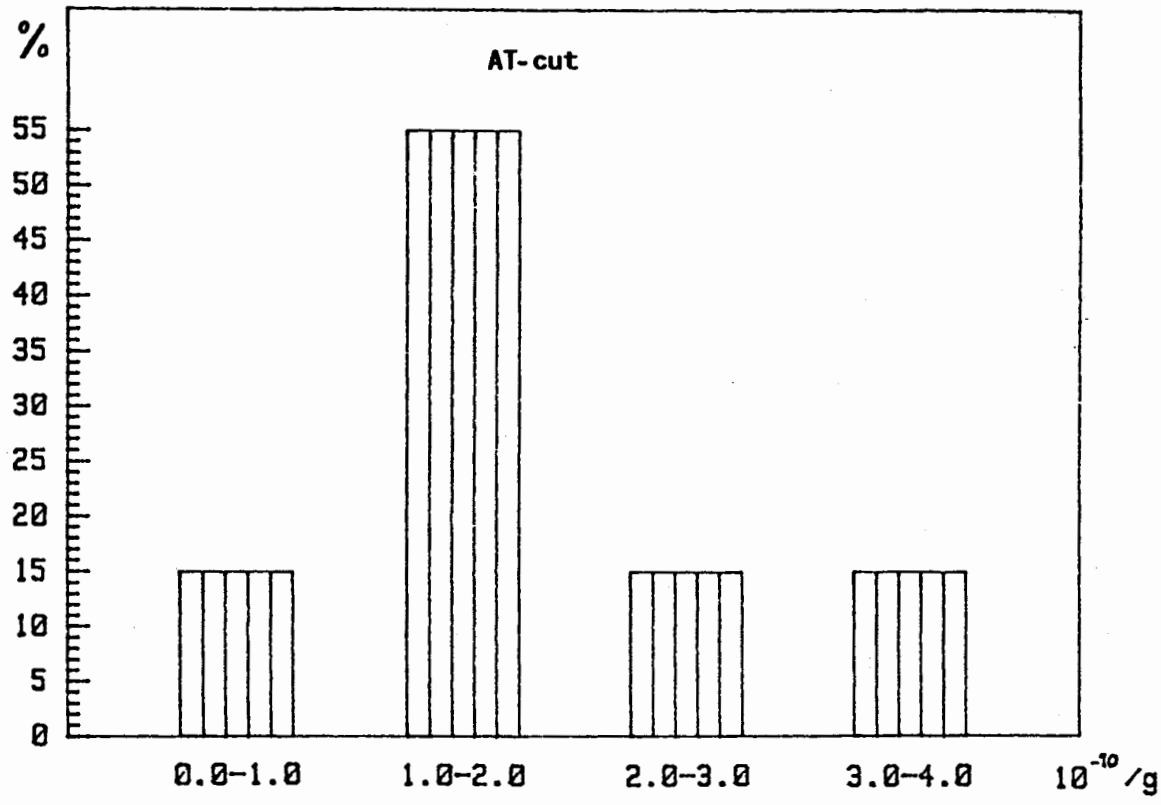


FIGURE 6 - Radial acceleration sensitivity of the first seven resonators.

The influence of the pressure[9][10] outside the resonator is tested at O.N.E.R.A. It begins at 2 bars down to the atmospheric pressions and return, within two hours. The sensibility value is determined by taking into account the middle term stability resonator and the thermic effect associated to the pressure variation. The first seven resonators have been tested with pression (Figure 7). The lowest sensitivity was $0{,}5.10^{-10}$ for 1 bar variation and a scattering of 2.10^{-10}/bar is obtained with 70 % of the tested resonators.

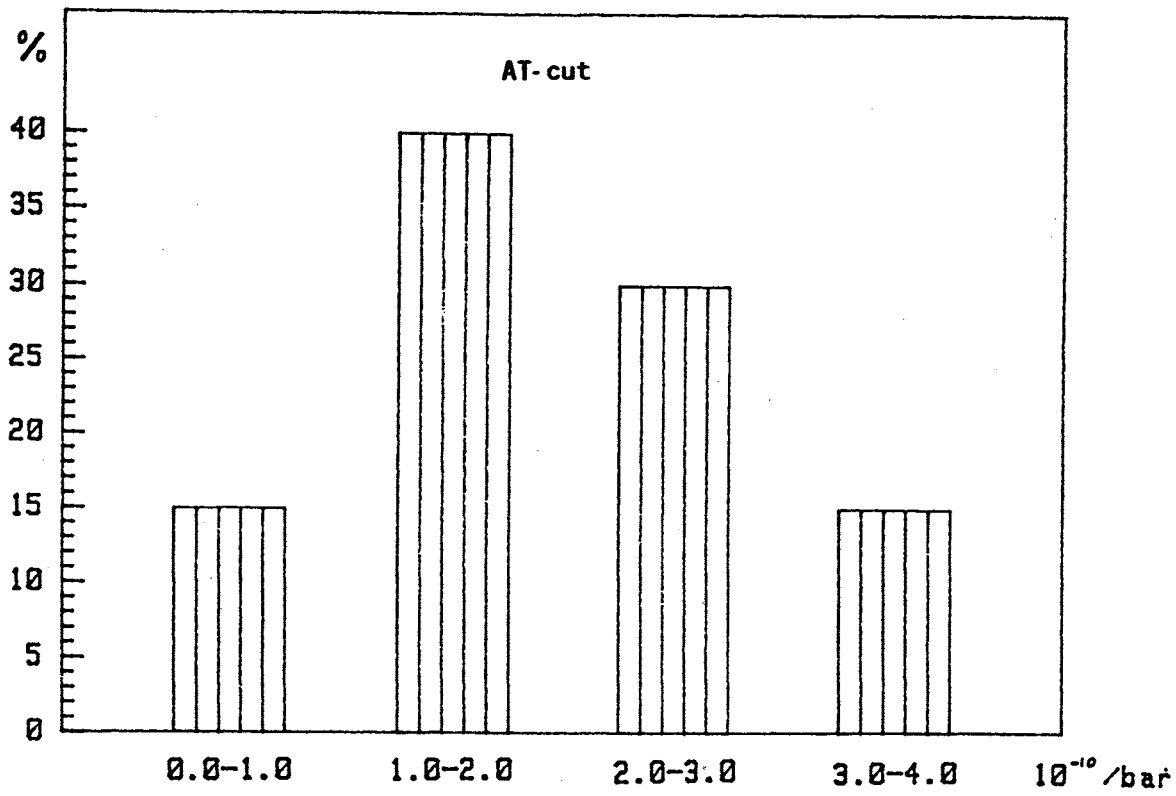


FIGURE 7 - Pressure sensitivity of the first seven resonators.

The vibration tests allow to determine the vibration spectrum and notably the critical frequency of the resonator[11][12]. According to the application type, the value of the frequency must be above the limit value determined with standards requirements. For example, the resonators taken on board aerospatial equipments must have a limit value of 2000 Hz for 20 g of acceleration. To determine the mechanical resonances of the resonator structure, the experimentaly method consists in sticking it on the vibrator support. The electric loads induced in the cristal-resonator (piezoelectic effect) are then amplified with a load amplifier ; the signal is analysed to obtain the low frequency mechanical resonance spectrum directly on a draw-table (Figure 8) . The first two resonators are tested and submitted to vibrations between 50 and 8000 Hz (20 g) in the y axis. On the recording signal, the peaks of amplitude give the mechanicals resonances of the resonator elastic structure.

Resonances close to 286 Hz, with a maximal amplitude around 5700 Hz , they are up to 2000 Hz (20 g).

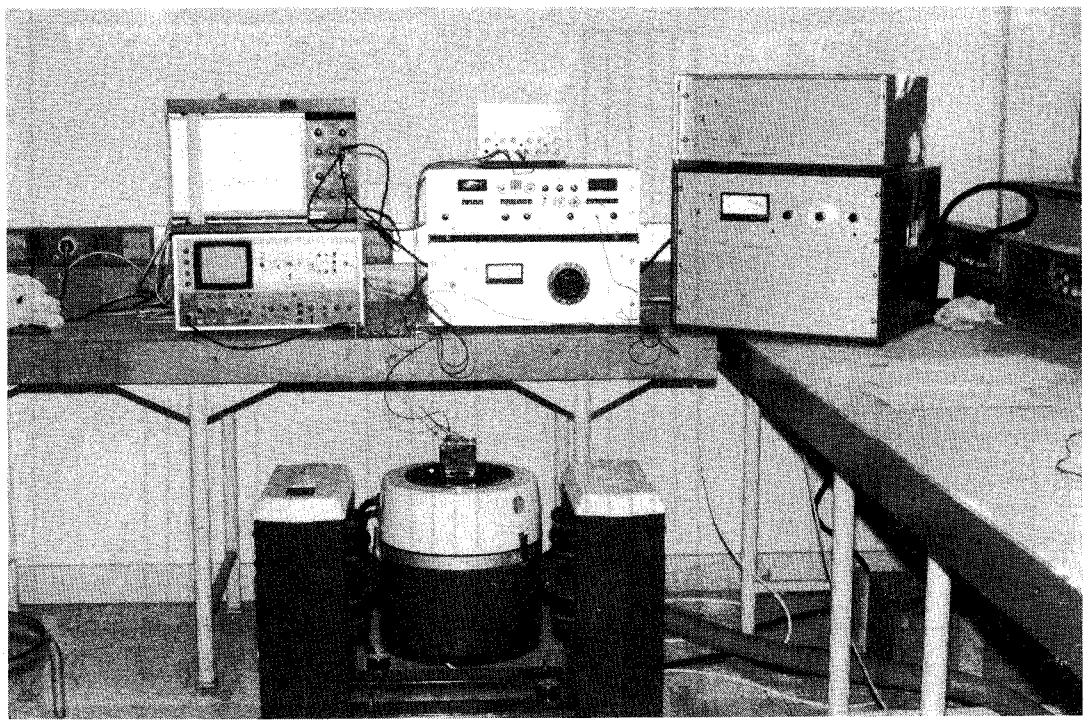


FIGURE 8 - Vibrator support and draw-table to determine the mechanical resonances.

5 - CONCLUSION

A new generation of miniaturized piezoelectric high stability resonators is studied with optimum reduction of environment effets. Their specifications are determined by a systematic characterization in pre-serie production ; the reproductibility of performances is obtained by statistical sensibilities.

The representative statistics curves of the first tests can be assimiled to gaussian curves in accordance with pre-serie production resonators.

It's with the complet results of the first thirty resonators that it will be possible to say that : the caracteristics and the sensibilities of this AT-cut miniaturized resonator make of it a qualified product for high-precision applications in severe and space environment.

Now, the same tests are performed on similar resonator with SC-cut quartz crystal in pre-serie production.

ACKNOWLEDGMENT

We wish to thank F. DEYZAC and D. JANIAUD of Aerospatial Agency O.N.E.R.A. - B.P. 72 - 92322 CHATILLON CEDEX (FRANCE) for tests and valuable discussions, and M. MOUREY of L.C.E.P. for statistics programs.

REFERENCES

(1) **DELAITE R.**, "Une nouvelle génération de résonateurs piézoélectriques miniaturisés de haute stabilité", Conférence, Actes du Premier Forum Européen Temps-Fréquence, Besançon, 18, 19, 20 mars 1987.

(2) **BESSON R.J.**, "Résonateur piézoélectrique à cristal auto-suspendu", brevet d'invention n° 78 022 61, 1978.

(3) **BESSON R.J.**, "Résonateur à quartz à électrodes non adhérentes au cristal", brevet d'invention n° 76 010 35, 9 janvier 1976.

(4) **DELAITE R.**, "Résonateur piézoélectrique à extremum de sensibilité vis-à-vis des contraintes extérieures de pression", brevet d'invention n° 85 090 58, 14 juin 1985. Extension à l'étranger en cours (U.S.A., Canada, CH, DE, GB, NL).

(5) **BESSON R.J., DELAITE R., GENESTIER G., MAITRE P., RENARD G., THIEBAUD D. et VALENTIN J-P,** "Structure mécanique symétrique pour résonateurs piézoélectriques", brevet d'invention n° 85 09 097, 14 juin 1985. Extension à l'étranger en cours (U.S.A., Suisse).

(6) **DELAITE R., MIGEOT T., VATERKOWSKI J-L et VIGNES T.**, "Etat d'avancement de la réalisation de 100 résonateurs miniatures", rapport de contrat Aide à l'Innovation ANVAR n° X.84.04.015.I, L.C.E.P., E.N.S.M.M. Besançon, Mars 1987.

(7) **BESSON R.J.**, "A new piezoelectric resonator design", Proceeding of 30th Annual Frequency Control Symposium, Fort Monmouth, 1976.

(8) **DELAITE R.**, "Influence de la structure de maintien sur la sensibilité accélérométrique et barométrique des résonateurs à quartz de haute stabilité", thèse de Docteur en Sciences pour l'Ingénieur, Université de Franche-Comté, n° 005, 17 juin 1985.

(9) **DELAITE R.**, "Résonateur à quartz pour environnement sévère", revue de Physique Appliquée, n° 20, pp 741-752, octobre 1985.

(10) **DELAITE R. and VALENTIN J-P**, "Analysis of force-frequency effect of piezoelectric circular resonators strained by edge angular sectorstresses", Il Nuovo Cimento (Europhysics Journal), Volume 6 D, n° 5, pp 574-584, Décembre 1985.

(11) **DELAITE R.**, "Résonateur piézoélectrique à sensibilité réduite pour environnement aérospatial", Conférence, Actes de la 2ème Journée Nationale Science et Défense, Ecole Polytechnique (Palaiseau), 5 décembre 1985.

(12) **MIGEOT T.**, "Rapport interne d'activité", L.C.E.P., E.N.S.M.M. Besançon, octobre 1987.

(13) **DELAITE R. et VALENTIN J-P**, "Compensation des effets mécaniques dans un cristal de coupe SC", revue de Physique Appliquée, février 1986.

2 ND EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 16 - 17 - 18 MARCH 1988

Session RO-03 Number 2

The Radiation Susceptibility of BVA-AT Resonators to Simulated Medium Earth Orbit Radiation

J. J. Suter, A. G. Bates, and J. M. Cloeren

The Johns Hopkins University
Applied Physics Laboratory
Johns Hopkins Road
Laurel, MD 20707 USA

Introduction

Piezoelectric quartz crystal resonators are widely used in satellite communication and guidance systems. The system performance of satellites requires quartz crystal resonators with superior stability, low system noise levels, and excellent phase noise characteristics. The apparent susceptibility of quartz crystal resonators to proton radiation manifests itself as frequency shifts. Previous experiments have indicated that this phenomenon is dependent on radiation dose, dose rate, particle flux or fluence, and kinetic energy levels [1]. Early studies indicated that the mechanisms within the quartz crystal responsible for frequency shifts depend on the total doses accumulated within the crystal blank [2]. Additionally, impurities introduced in quartz during the growth process cause resonators made from these quartz bars to be sensitive to ionizing radiation. In particular, aluminum defect centers have been demonstrated to greatly increase the radiation sensitivity of quartz to radiation at high dose levels (>100 rad). However, a quantitative study of the radiation sensitivity of quartz crystals to lower dose levels did not reveal any correlation between impurity content and radiation sensitivity [3].

In this paper, we will report on the radiation susceptibility of BVA-AT quartz crystal resonators and show that their sensitivity to low levels of gamma radiation (1.25 MeV) is less than those compared to AT resonators with adhered electrodes. This paper will also discuss the effectiveness of preconditioning techniques, together with a modest interpretation of the test results in terms of the materials aspects of the quartz electrode interface.

Previous work on the radiation susceptibility of quartz crystal resonators at our laboratory focused on AT quartz crystal resonators with adhered electrodes. Those studies concluded that the radiation susceptibility of AT resonators to low doses of radiation (<10 rad) are independent of the aluminum defect center concentration of the crystal blank. Furthermore, those studies concluded that the susceptibility of quartz crystals to low doses of radiation cannot be extrapolated from high dose radiation tests. It was also shown that the response of crystal oscillators to proton radiation encountered in space can be modeled with gamma radiation from cobalt 60 isotope sources. It was also concluded from the present studies that the radiation susceptibility of quartz crystal resonators in correlated to their manufacturing process and also to interactions between the quartz material and electrodes. This last observation motivated an investigation of the radiation response of electrodeless resonators of the BVA-AT design to possible radiation environment in medium earth orbits (1000–1400 km).

Gamma Radiation Studies

In our tests, BVA-AT resonators were mounted in test bed oscillators. The electronic components in the test fixtures were hardened to the appropriate radiation levels of the tests. Therefore, any changes in the frequency stability of the quartz crystals was due to resonator efforts only. Initial experiments reported the results on AT resonators using cobalt 60 and proton radiation; the effects of proton radiation on the frequency stability of quartz resonators can be modeled using gamma radiation from a cobalt 60 source [1]. Cobalt 60 irradiation of the BVA-AT quartz crystal resonators was performed with a cobalt 60 source, emitting photons with kinetic energies of both 1.33 MeV (50%) and 1.17 MeV (50%). The radioactivity of the source was 11.6 mCi. The accumulated doses and dose rates referred to in this study were all referenced to values at the quartz crystal blank, certifying consistent results that were independent of the activity of the cobalt 60 sources. The activity of the cobalt sources was calculated before each radiation test so that the decay of the source could be taken into account. The dose rate of the cobalt 60 radiation at the quartz crystal blank was also measured with thermoluminescent dosimeters that were read by a Victoreen 2810 dosimeter reader. These measurements confirmed the calculated dose rates to within 10%.

As shown in Fig. 1, the cobalt 60 radiation had to pass through 2.5 gm/cm^2 of equivalent aluminum shielding before entering the quartz crystal blank. The attenuation of the cobalt 60 radiation was, therefore, taken into account in each experiment so that a consistent dose rate at the quartz crystal was established. Furthermore, the plane of the quartz crystal resonators was placed perpendicular to the direction of the cobalt 60 photons, as shown in Fig. 1. In each experiment the dose rate, $D(r)$, of the cobalt 60 radiation at the quartz crystal resonator was kept at 0.1 rad/min by selecting an appropriate distance between the quartz blank and the radioactive source.

The sinusoidal output of the quartz crystal oscillator was applied to a mixer that was made part of a data acquisition system (see Fig. 2). This system is capable of measuring the quartz crystal's resonant frequency; frequency shift, df/f, as function of radiation dose; and Allan variance [3]. The operation of this system can be described as follows: the sinusoidal signal from the quartz crystal test fixture was mixed with a signal from a stable reference source (hydrogen maser or precision oscillator). The beat frequency (approximately 400 Hz) was applied to a Hewlett Packard HP 5335A counter that was interfaced to an HP 85 computer system. Concurrently, the beat frequency (frequency difference between the quartz resonator and reference oscillator) was converted by an HP 59303A digital-to-analog converter whose output was observed by an Esterline strip chart recorder. The strip chart recorder supplied an analog data file of the frequency stability of the quartz resonator, while the computer stored the information digitally for later data retrieval. The calculations of frequency shift and Allan variance of the quartz crystal resonator were performed by software developed for this work.

Low-level gamma radiation testing of quartz crystal resonators was initiated in order to measure df/f in a low-background radiation environment. Each of the BVA-AT resonators was incrementally exposed to 1 and 3 rad of cobalt 60 at a dose rate of 0.1 rad/min. The results of these tests are shown in Fig. 3, where the frequency shifts of the quartz resonators are presented as a function of accumulated radiation dose. These tests showed that in comparison with previ-

ously low-level radiation tests on AT resonators with adhered electrodes, the BVA-AT crystals have frequency shifts of a factor of 8 smaller. In the next series of tests, the frequency shifts in the electrodeless resonators as a function of accumulated dose and annealing time were measured.

Medium Earth Orbit Radiation Studies

A satellite orbiting the earth is periodically exposed to larger amounts of radiation over certain time intervals, while part of the orbit time is spent in a relatively radiation-free environment. This cycle—a period of exposure followed by a radiation-free interval—affects the quartz crystal resonator's frequency. The particular radiation environment selected for our experiments is the radiation spectrum of the TOPEX satellite orbit. This TOPEX satellite is tentatively scheduled for launch in 1991 with a mission duration of 3 years; its orbit has an altitude of approximately 1300 km, an inclination of 63°, and a period of about 121 min [4]. During orbit, instruments on board the spacecraft will be exposed to proton and electron radiation. The total radiation incident on the instrumentation will be the sum of this proton and electron radiation and also bremsstrahlung radiation. However, for radiation shielding thicknesses in excess of 2 gm/cm^2 the contribution of trapped proton radiation is dominant. Furthermore, for shielding in excess of of 4 gm/cm^2 the total accumulated dose is not significantly reduced. Therefore, the maximum dose at the quartz crystal resonator will be approximately 8000 rad over 3 years, since the quartz crystal oscillator test fixtures had a total of about 9 gm/cm^2 of equivalent aluminum shielding (refer to Fig. 1).

Figure 4 shows the magnitude of these radiation zones as a function of orbit time, with the shaded area representing the most intense radiation zone, the South Atlantic Anomaly. The peak value of the accumulated dose behind 4 gm/cm^2 of aluminum shielding in the South Atlantic Anomaly is about 10^{-3} rad/s; the TOPEX spacecraft spends approximately 42 min of its orbit in this radiation anomaly. Since the relative accumulated doses for the other zones are 3 orders of magnitude lower, it is safe to assume that most of the radiation accumulated in a quartz crystal resonator is contributed by the South Atlantic Anomaly. In other words, the spacecraft spends the remainder of the orbit (approximately 70 min) in a relatively radiation-free environment. The South Atlantic Anomaly represents an average dose rate of approximately 0.6 rad for 42 min (0.014 rad/min). That dose and dose rate were the basis for the radiation experiments on the electrodeless resonator; it also represents a value an order of magnitude lower than the dose rate (0.1 rad/min) in previous radiation tests.

The TOPEX radiation tests were conducted with a cobalt 60 source. It was shown that the gamma rays emitted by a cobalt source emulate proton energy (in terms of energy losses in quartz), and were therefore used in the radiation susceptibility studies of quartz crystal resonators. The test procedures for cobalt 60 irradiation of quartz crystal resonators were analogous to the ones described in the previous experiment, except that the distance between the cobalt source and resonator blank was increased so the dose rate at the crystal was the desired 1.4×10^{-2} rad/min.

The BVA-AT quartz crystal was subjected to three TOPEX orbit simulations. In each simulation the quartz crystal was exposed for 42 min to cobalt 60 radiation and allowed to recover

for 70 min in a radiation-free environment. This simulation technique deposited a total of 0.6 rad of cobalt radiation in the BVA-AT quartz crystal. Figure 5 illustrates the frequency shift induced in the quartz crystal resonator. During each orbit simulation the BVA-AT quartz crystal resonator, which was not preconditioned, demonstrated a radiation sensitivity of $5.0 \times 10\text{–}11$/rad. This frequency shift per accumulated dose was identical for all three orbits. It is apparent from Fig. 5 that during each radiation-free interval of 70 min, the resonator's frequency shift recovered by $1.0 \times 10\text{–}11$. Furthermore, the exposure of the quartz crystal to the simulated radiation environment manifested itself as step-like features.

Since preconditioning lowers the radiation sensitivity of quartz crystals, the BVA-AT resonator was irradiated with 20 krad of cobalt 60 radiation before the next series of orbit simulations. It appeared that df/f in the crystal resonator on a per rad basis was reduced by a factor of 2. However, the direction of the frequency shift was reversed, a phenomenon that has been observed in many quartz crystal resonators [1].

Discussion

It is apparent from the gamma radiation tests that the electrodeless BVA-AT quartz crystal resonator has a lower susceptibility to low doses of gamma radiation than the AT resonator with adhered electrodes. The sensitivity of AT resonators is apparently not correlated to the aluminum defect center concentration of the quartz crystal, since the BVA-AT resonator had an aluminum purity content about 20 times higher than the conventional AT resonator. It is, therefore, likely that the energy loss of the gamma radiation (1 and 3 rad exposures) is insufficient to overcome the binding energy of aluminum defect centers in the crystalline quartz. However, preconditioning to levels of 20 krad with gamma rays allowed some sodium ions to be removed form their original sites and, hence, improved the radiation susceptibility of the BVA-AT resonator.

References

[1] Norton, J. R., Cloeren, J. M., and Suter, J. J., "Results from Gamma Ray and Proton Beam Radiation Testing of Quartz Resonators," *IEEE Trans. Nucl. Sci.,* NS-31(5), 1230-1233 (1984).

[2] King, J. C., and Koehler, R., "Radiation Effects on Resonators," in *Precision Frequency Control,* Volume I, eds. E. A. Gerber and A. Ballato, Academic Press, New York, 147-158 (1985).

[3] Suter, J. J., and Maurer, R. H., "Low and Medium Dose Radiation Sensitivity of Quartz Crystal Resonators with Different Aluminum Impurity Content," *IEEE Trans. Ultrason., Ferroelec., and Freq. Contr.,* UFFC-34(6), 667-673 (1987).

[4] March, P. A., Horton, C. M., and Stassinopoulos, E. G., "TOPEX Radiation Environment," Jet Propulsion Laboratory Report D-2116 (1985).

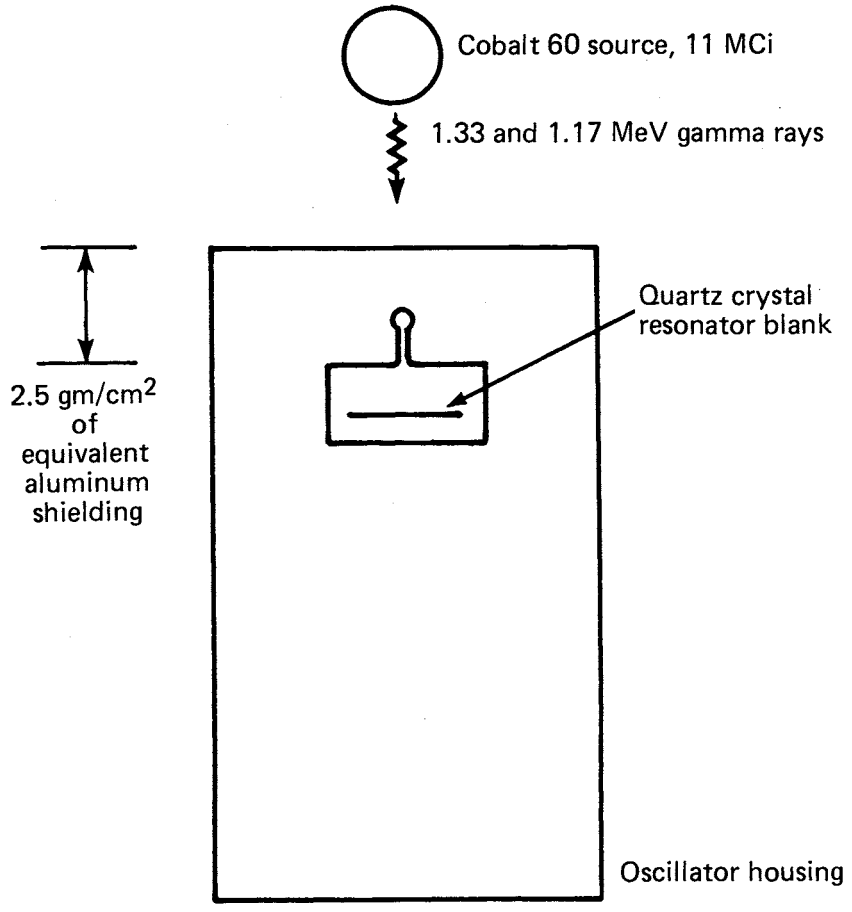


Fig. 1. Simplified diagram of radiation text configuration.

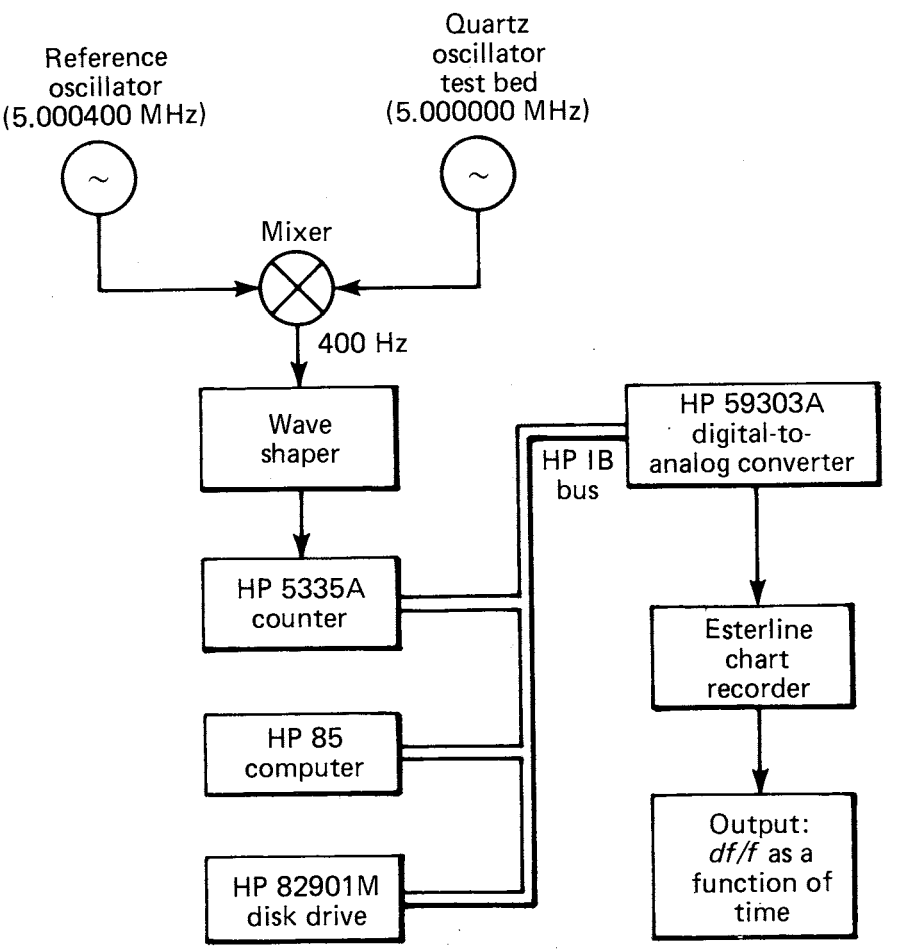


Fig. 2. Data acquisition system for frequency shift and Allan variance measurements of quartz crystal resonators.

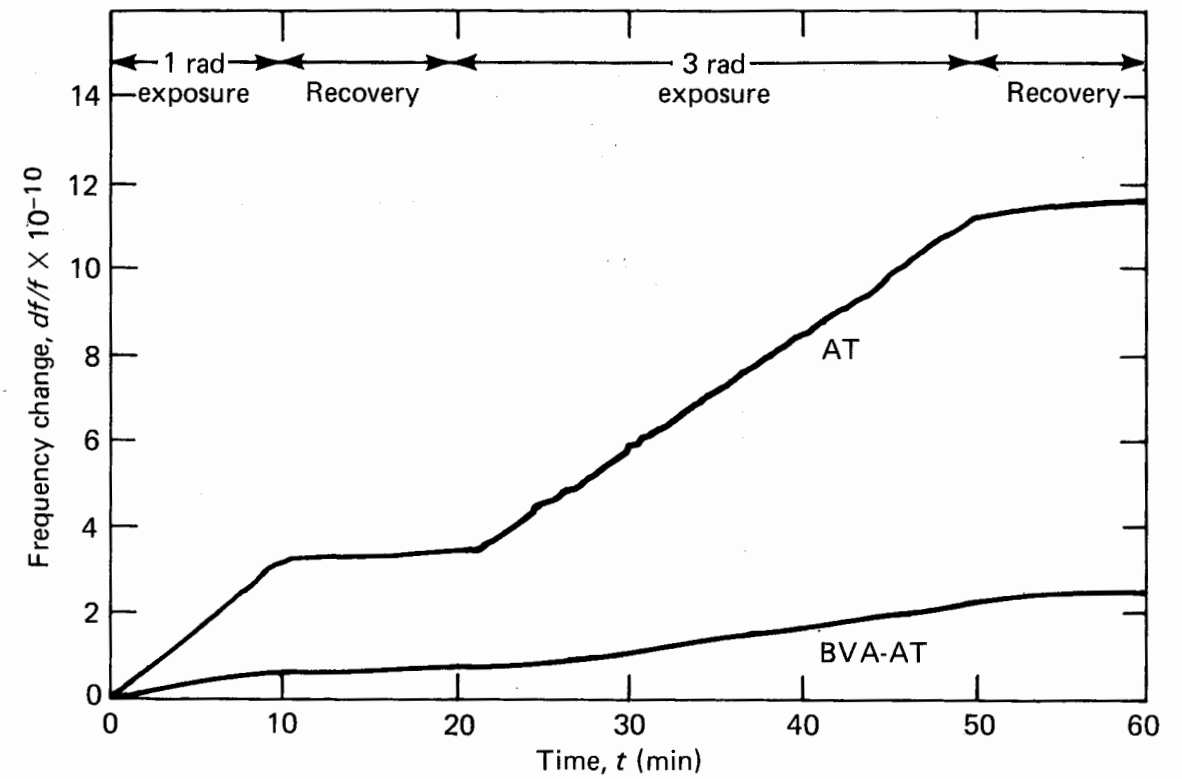


Fig. 3. Frequency shifts as a function of accumulated cobalt 60 dose for AT and BVA-AT quartz crystal resonators.

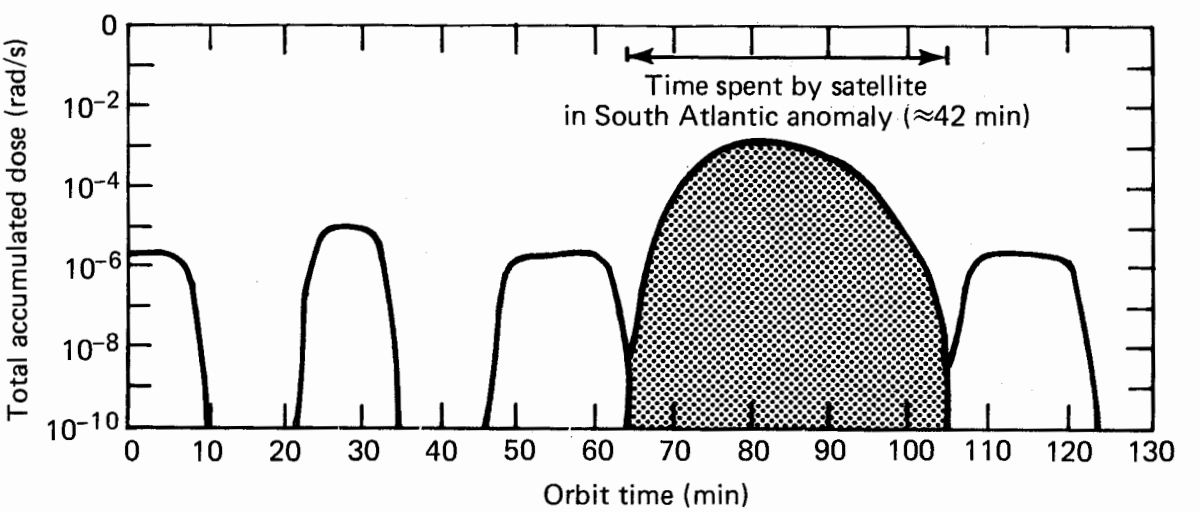


Fig. 4. Total accumulated dose behind 4 gm/cm^2 of aluminum shielding for a 1300-km earth orbit as a function of orbit time (after March et al., 1985).

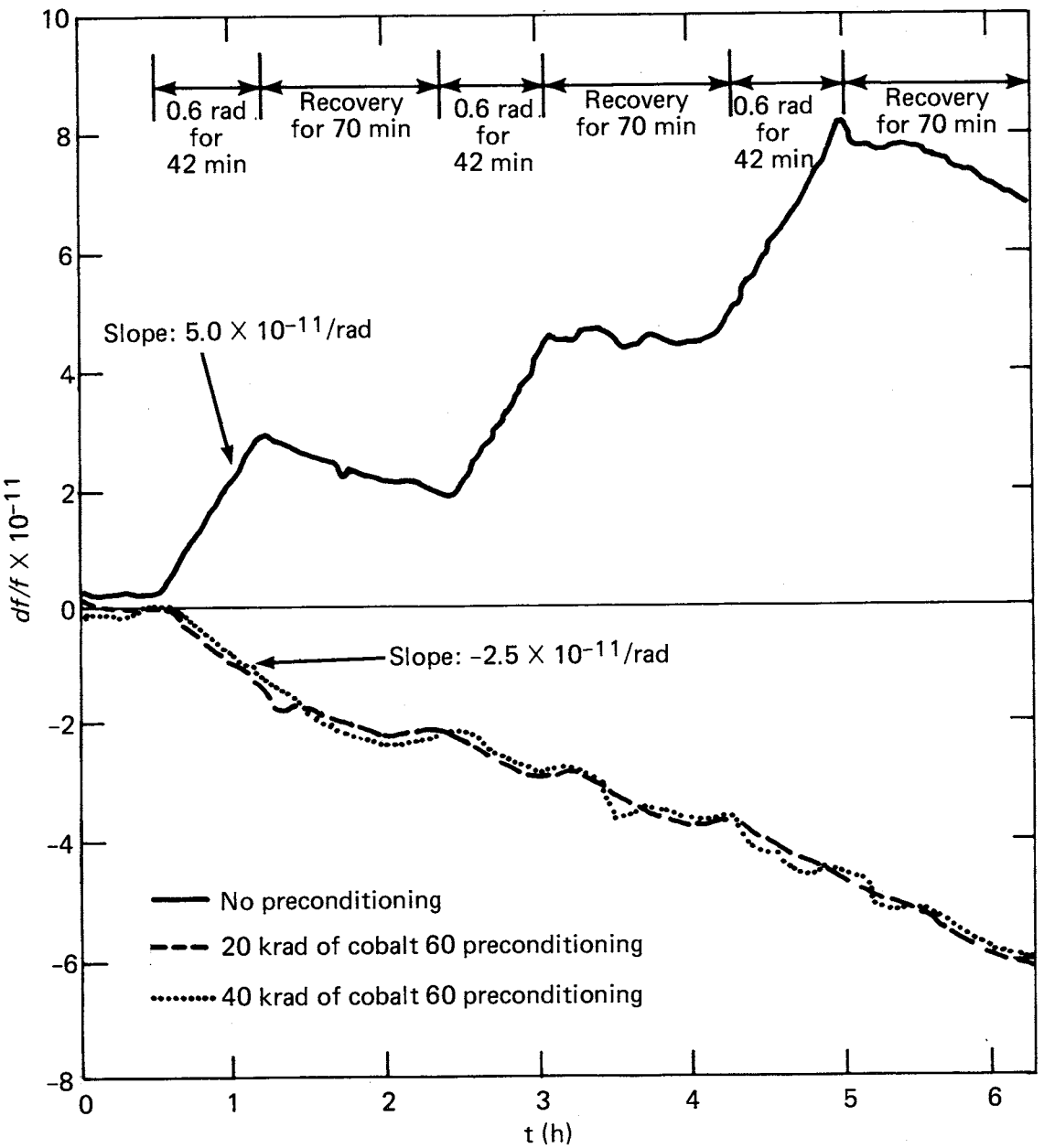


Fig. 5. Frequency shifts in a BVA-AT quartz crystal resonator for a 1.4×10^{-2} rad/min cobalt 60 simulated radiation environment.

2 ND EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 16 - 17 - 18 MARCH 1988

Session RO-03 Number 3

OPTIMISATION DE L'ORIENTATION DES RESONATEURS AUTO-SUSPENDUS

JP. AUBRY*, B. BONNOTTE*, JC. CRAVEUR**, F. DEYZAC** et D. JANIAUD**

* CEPE - 44, Avenue de la Glacière - BP 165 - 95105 ARGENTEUIL
** ONERA - BP 72 - 92322 CHATILLON Cedex

Introduction :

Les systèmes de télécommunications, de localisation ou de radars utilisent des oscillateurs ultra-stables. Les performances attendues de ces oscillateurs à quartz en terme de sensibilités accélérométrique et barométrique sont de plus en plus sévères et dépendent de celles des résonateurs qui les équipent.

Cet article a pour but de présenter les résultats des différentes phases de l'étude qui a permis d'optimiser l'orientation des résonateurs auto-suspendus dans le plan de la coupe SC.

L'ensemble de ces travaux a été soutenu par la Direction des Recherches Etudes et Techniques de la Délégation Générale pour l'Armement.

Le résonateur QAS (quartz auto-suspendu) "classique" a permis d'initialiser l'étude (figures 1 et 2). La lame QAS est constituée d'une couronne, de deux paires de ponts et d'une pastille vibrante. Les supports sont collés sur la couronne, à 90° de l'entre-axe des ponts.

Cette géométrie de lame assure le découplage de la pastille vibrante vis à vis du monde extérieur. Les valeurs de sensibilités accélérométrique et barométrique ainsi obtenues sont inférieures à celles mesurées pour un résonateur classique (pastille de quartz pleine).

Déroulement de l'étude :

Sous l'effet de sollicitations mécaniques extérieures au boîtier du résonateur, la fréquence de ce dernier varie. L'analyse de cette sensibilité permet de proposer une modélisation comportant trois phases indépendantes (figure 3).

1. Première phase : Optimisation de l'orientation du QAS "classique".

   La nature des efforts exercés sur la pastille vibrante est connue [1]. La compression diamétrale étant la cause principale de la sensibilité accélérométrique, un meilleur positionnement au zéro de la courbe de sensibilité à la compression (RATAJSKI) constitue la première phase de l'optimisation du QAS.

Cependant les valeurs de sensibilité accélérométrique, bien que satisfaisantes (de l'ordre de $8.10^{-10}/G$), sont insuffisamment reproductibles.

2. Deuxième phase : Filtre mécanique "extérieur" au cristal.

Une nouvelle géométrie de supports est définie (supports "Muguet") [1] permettant conjointement d'augmenter la reproductibilité des résultats et de minimiser la valeur moyenne de la sensibilité accélérométrique, inférieure au $10^{-10}/G$ pour les meilleurs résonateurs. De plus, le montage d'une platine intermédiaire entre la lame QAS et l'embase permet de diminuer d'un facteur 5 environ la sensibilité barométrique, indépendamment des supports utilisés.

Les performances atteintes sont satisfaisantes et répondent aux différents cahiers des charges proposés. Cependant, les spécifications requises pour ces équipements ne cessent d'évoluer : l'optimisation de la géométrie de la pastille vibrante devient alors indispensable.

3. Troisième phase : Filtre mécanique procuré par la géométrie QAS.

L'optimisation de la géométrie du résonateur QAS commence par la caractérisation des grandeurs force et contraintes. Ces grandeurs, ramenées par les ponts sur la pastille vibrante, dépendent de l'angle de compression diamétrale α (figure 4). Cette force a une composante de compression notée F_C et une composante de cisaillement notée F_Σ [1] (figure 5).

L'optimisation de la lame QAS consiste à rechercher la configuration transmettant le minimum d'efforts à la pastille vibrante, tout en respectant les spécifications imposées et en tenant compte des limites technologiques. Une géométrie "optimale", dont le dimensionnement est représenté sur la figure 6, a été choisie parmi les géométries étudiées. CEPE fabrique actuellement ce résonateur en remplacement du QAS "classique".

L'étude expérimentale de la compression diamétrale du QAS "classique" a permis non seulement de confirmer l'existence d'un cisaillement dans le plan (mis en évidence par éléments finis) mais aussi de montrer la grande sensibilité du QAS à ce type de sollicitation (figure 7).

Cependant, il n'est pas suffisant de disposer d'un filtre mécanique optimisé : Il faut déterminer l'orientation de ce filtre dans le repère cristallographique pour laquelle la variation de fréquence sous l'action d'un système de forces donné est minimale.

Cette variation de fréquence résulte des sensibilités aux efforts résiduels de compression et de cisaillement dont on connait les composantes (éléments finis). Il reste à estimer les deux coefficients de sensibilité notés k_c pour la compression et k_Σ pour le cisaillement (figure 8).

4. Quatrième phase : Optimisation de l'orientation du QAS (Nouvelle géométrie).

On appelle conventionnellement résonateur d'angle ψ un QAS dont les ponts font avec les axes cristallographiques de la coupe SC un angle ψ. Il convient de préciser que, si chaque QAS permet d'obtenir en totalité la courbe de compression, il ne permet de déterminer qu'un seul point de la courbe de cisaillement.

Un résonateur d'angle ψ ne fournit que la valeur du coefficient de sensibilité au cisaillement pour cet angle ψ.

Des lots de résonateurs d'angles divers ont été spécialement fabriqués par CEPE pour cette étude.

Des compressions diamétrales sont appliquées sur la pastille vibrante de chaque résonateur pour construire la courbe de sensibilité à la compression (RATAJSKI). Cette étude préliminaire permet de vérifier la cohérence des résultats avec les différentes références bibliographiques [2, 3].

Une compression est ensuite appliquée à 45° des ponts sur la couronne. La pastille vibrante n'est alors soumise qu'à un cisaillement dans le plan, entraînant la variation de fréquence :

$$\Delta f = k_\Sigma (\psi) \cdot F_\Sigma$$

La force de cisaillement étant connue, nous pouvons déterminer le coefficient de sensibilité au cisaillement. La courbe d'évolution de ce coefficient en fonction de l'angle ψ est représentée sur la figure 9.

Par application du principe de superposition, il est possible de modéliser la variation de fréquence d'un QAS d'angle ψ soumis à une compression d'angle α sur sa couronne.

$$1)\ \Delta f (\psi, \alpha) = k_c (\psi) \cdot F_c (\alpha) + k_\Sigma (\psi) \cdot F_\Sigma (\alpha)$$

soit sous une forme plus condensée :

$$2)\ \Delta f (\psi, \alpha) = \vec{k} (\psi) \cdot \vec{F} (\alpha)$$

où $k(\psi)$ est le vecteur sensibilité dans le plan d'un QAS d'orientation ψ et $\vec{F}(\alpha)$ le vecteur force transmis par les ponts sur la pastille vibrante. Cette relation vectorielle n'est autre que la généralisation de la relation scalaire de RATAJSKI appliquée au QAS.

Ce produit scalaire peut aussi s'écrire sous la forme :

$$\Delta f = B(\psi) \sin(2\alpha + \phi)$$

avec $B(\psi) = F\sqrt{k_c^2 + \beta k_\Sigma^2}$ et $\beta > 0$.

Les coefficients k_c et k_Σ ne s'annulent pas pour les mêmes valeurs de ψ. Il n'existe donc pas, dans le plan de coupe SC, d'orientation de "sensibilité nulle".

La coupe optimale est donc celle pour laquelle $B(\psi)$ est minimal, ce qui revient à déterminer les angles ψ vérifiant la condition

$$\frac{\partial B(\psi)}{\partial \psi} = 0$$

Quatre valeurs de ψ satisfont cette condition : 0°, 45°, 90°, 135°. De manière à affiner les résultats de cette modélisation, des essais complémentaires sont réalisés. Chaque résonateur est soumis à des compressions diamétrales sur sa couronne afin d'évaluer, pour l'orientation ψ qui lui est propre, la valeur maximale de la déviation de fréquence.

L'ensemble des résultats et le modèle analytique sont représentés sur la figure 10. Malgré sa simplicité, le modèle est en bon accord avec l'expérience. Il apparait de façon très nette que l'orientation optimale se situe entre $\psi = -10°$ et $\psi = 0°$. On admet que l'orientation $\psi = 0°$ est suffisamment proche de l'optimum pour être retenue dans l'immédiat.

Résultats expérimentaux :

Les résultats suivants ont été obtenus sur des lots représentatifs

- Quantité de pièces supérieure à vingt
- Cohérence des résultats au niveau du lot.

Pour une géométrie donnée, le tableau (I) donne le pourcentage de résonateurs dont la sensibilité accélérométrique est inférieure à une valeur donnée K.

Le tableau (II) donne la valeur de sensibilité accélérométrique minimale, la valeur moyenne ainsi que l'écart type obtenus sur le lot de résonateurs.

Géométrie de lame \ K	2.10^{-10}/g	3.10^{-10}/g	5.10^{-10}/g	7.10^{-10}/g	1.10^{-9}/g
QAS classique	0	15	55	80	90
QAS nouvelle géométrie, orientation non optimisée	44	75	100	100	100

(I)

Valeur de sensibilité accél. Minimale (10^{-10}/g)	Moyenne (10^{-10}/g)	Ecart type (10^{-10}/g)
2,5	5,7	2,6
1,1	2,6	1,1

(II)

Conclusion :

L'optimisation du résonateur QAS a été rendue possible grâce d'une part, à la mise en évidence de l'effet de cisaillement dans le plan (éléments finis) et d'autre part, à l'étude expérimentale du coefficient de sensibilité à ce cisaillement. Un modèle de comportement a été établi sous forme vectorielle, généralisant ainsi la relation de RATAJSKI.

Cette optimisation a conduit à réduire non seulement les efforts transmis à la pastille vibrante (amélioration du filtre mécanique) mais aussi l'influence des efforts résiduels sur la variation de fréquence (détermination d'une orientation de ce filtre dans le plan de coupe SC).

Actuellement, CEPE fabrique des résonateurs suivant la nouvelle géométrie et dont les ponts sont sur l'axe Z' ($\Psi = 90°$).

Des résonateurs d'orientation $\Psi = 0°$ sont en cours de réalisation pour confirmer les résultats de cette étude.

La progression régulière des résultats obtenus laisse entrevoir à moyen terme des valeurs de sensibilité accélérométrique de l'ordre du 10^{-10}/G à l'échelle industrielle.

BIBLIOGRAPHIE

1 JP. AUBRY, JC. CRAVEUR, F. DEYZAC, M. FRAGNEAU, D. JANIAUD.

Analyse des mécanismes de sensibilités accélérométrique et barométrique des résonateurs à quartz de type QAS.

1er Forum Européen du Temps Fréquence - Besançon - mars 87.

2 JM. RATAJSKI

The force sensitivity of AT cut quartz crystal
International Business - Machins Corporation
Federal System Division - Electronics Systems Center.

Force-frequency coefficient of singly rotated vibrating quartz crystals.

IBM J. Res Dev, vol 12, pp 92-99, 1968.

3 FL. NISSIM

Contribution à l'étude des propriétés des résonateurs à quartz de coupe à double rotation.

Thèse de Docteur - Ingénieur, Besançon - Juin 78.

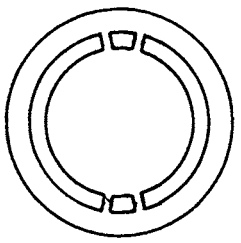


Fig. 1
Géométrie de la lame QAS "classique"

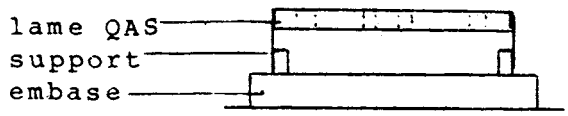


Fig. 2
Résonateur QAS "classique"

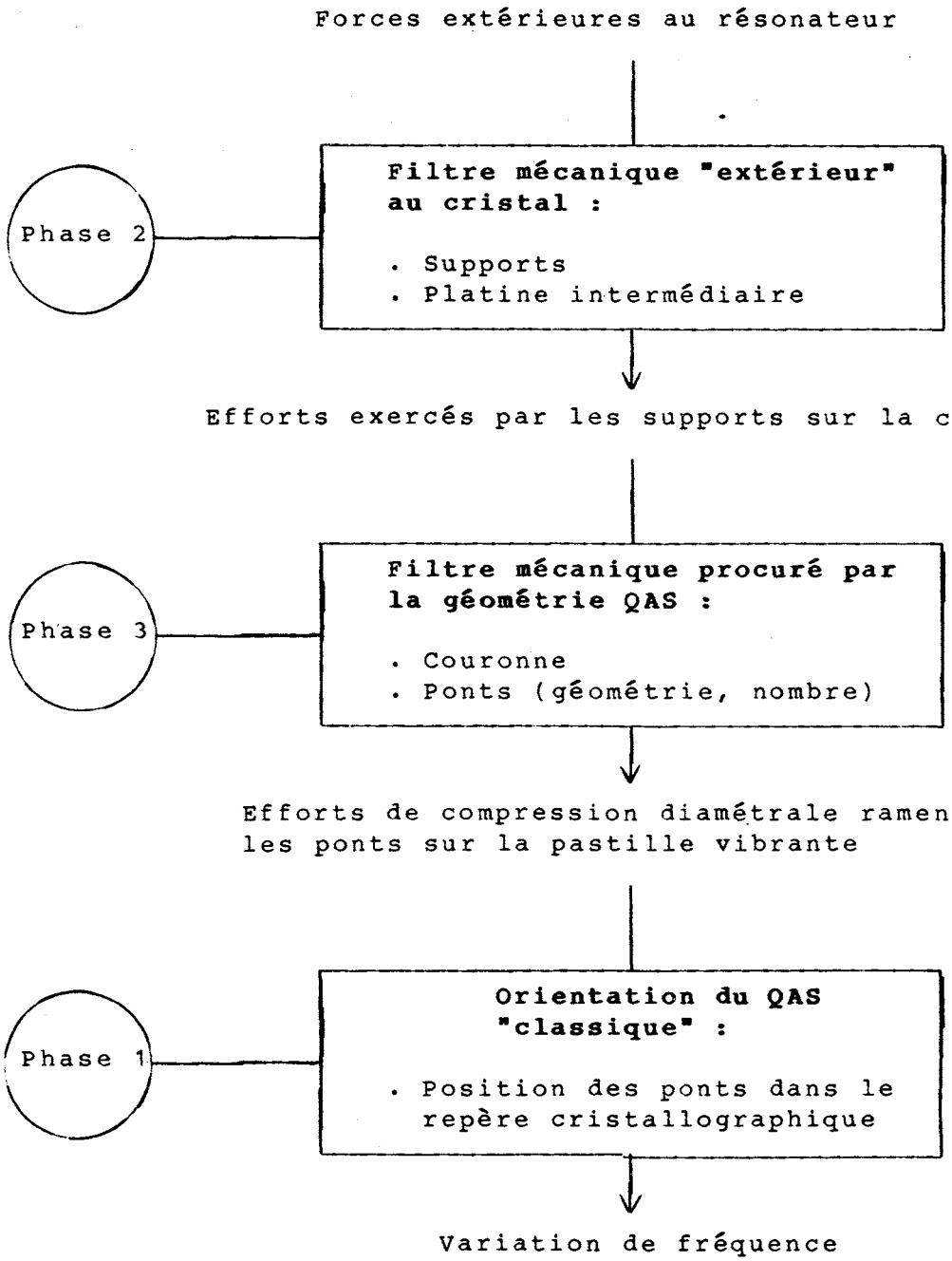


Fig. 3 : Modélisation de la sensibilité de la fréquence à une sollicitation extérieure.

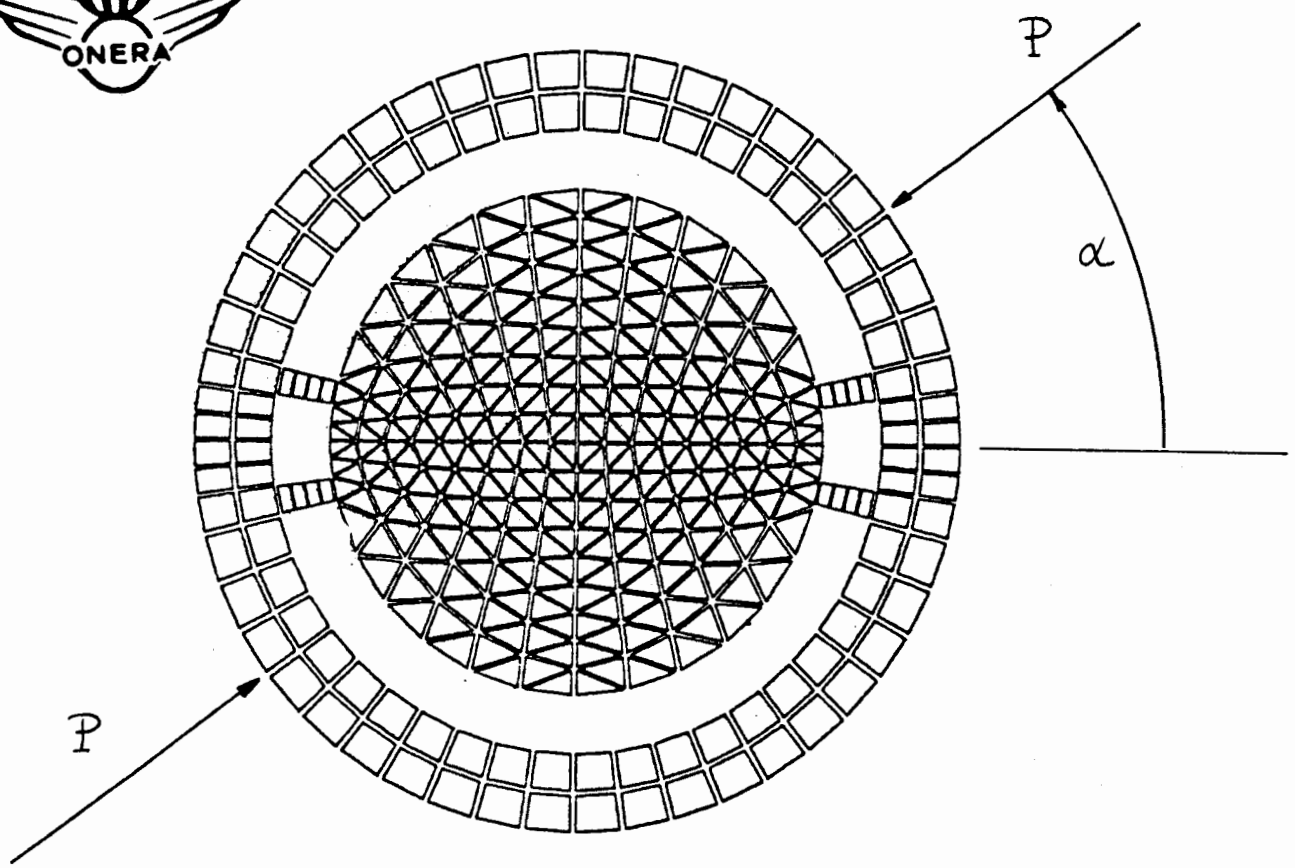


Figure 4 - Compression diamétrale d'un QAS "classique".

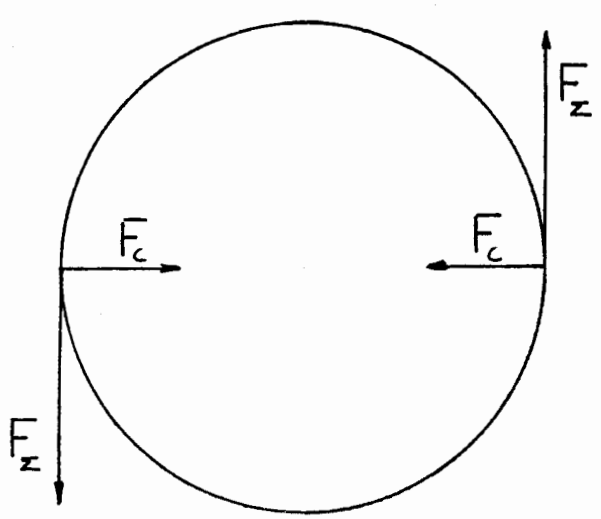


Figure 5 - Forces appliquées à la pastille vibrante.

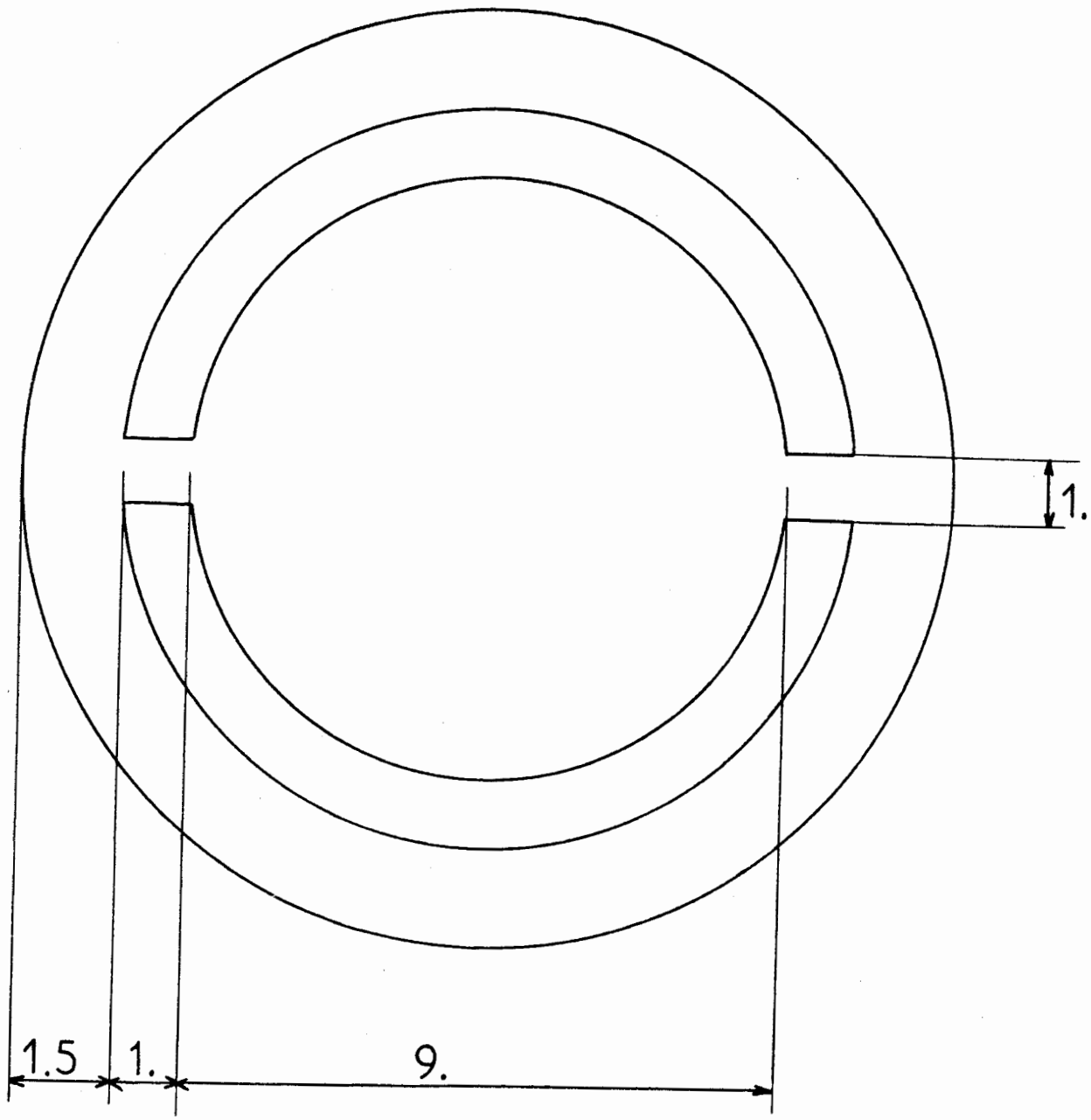


Figure 6 - QAS de nouvelle géométrie.

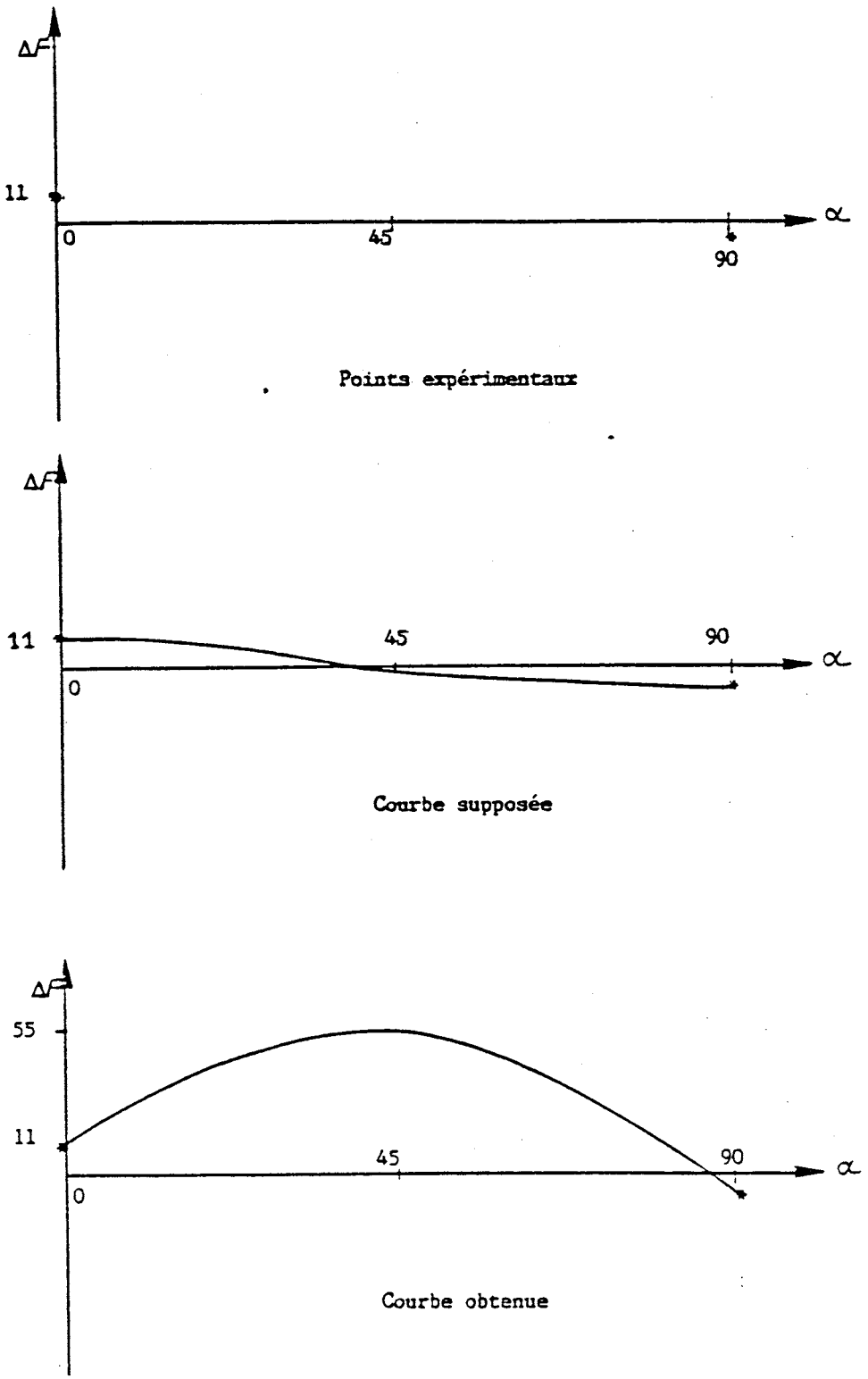


Figure 7 - Sensibilité au cisaillement d'un QAS.

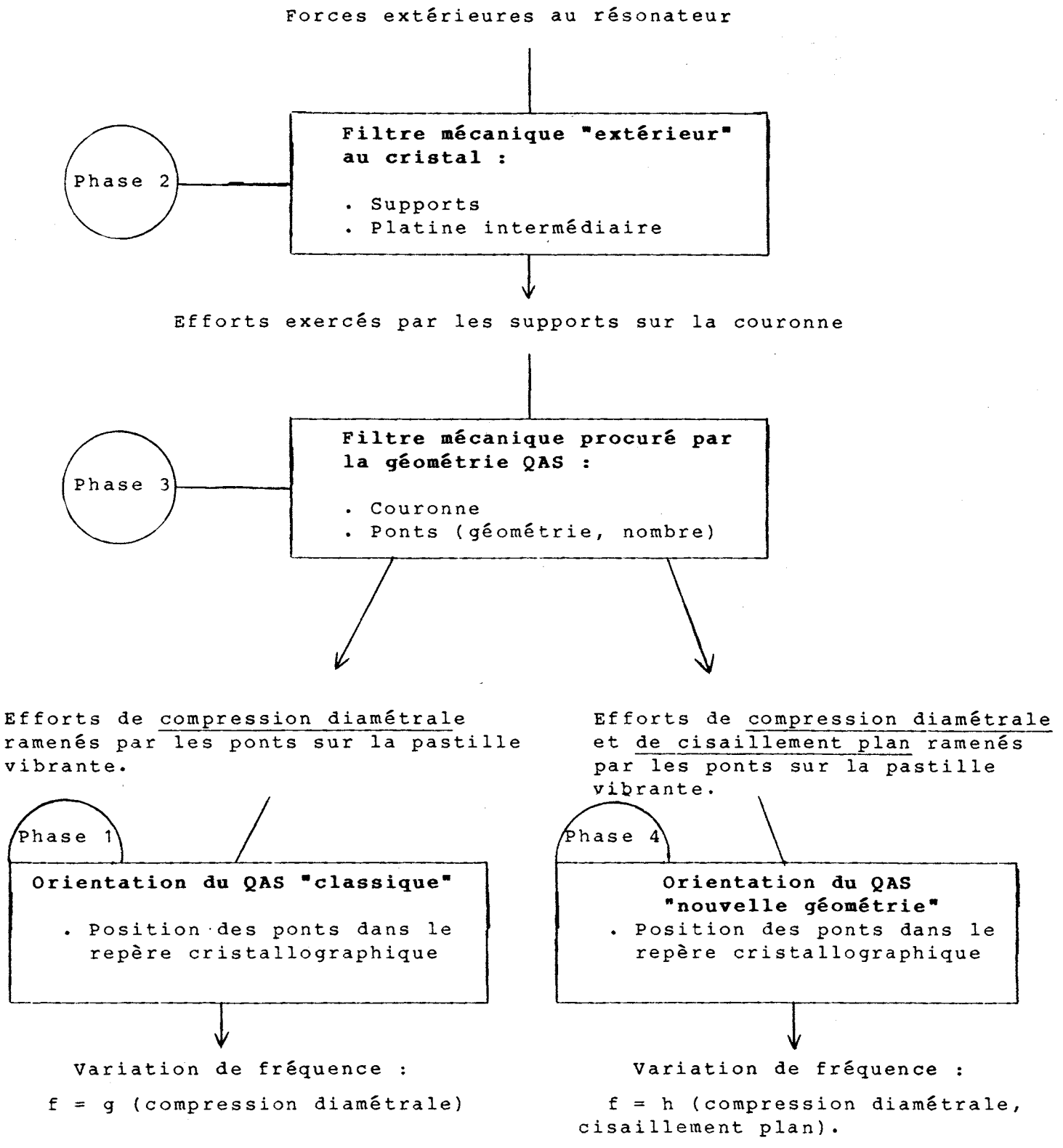


Fig. 8 : Optimisation globale du résonateur QAS.

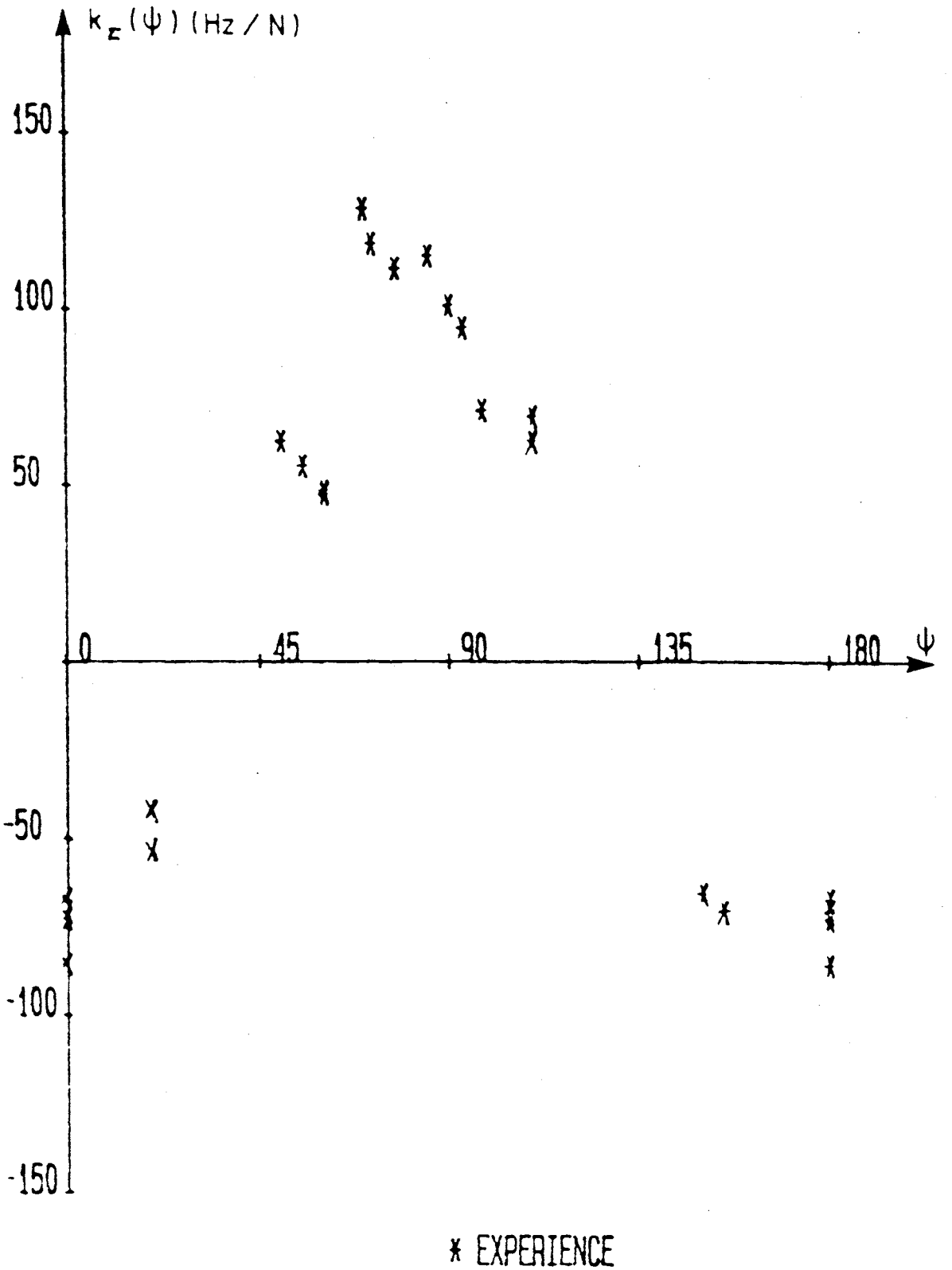


Figure 9 - Coefficient de sensibilité au cisaillement.

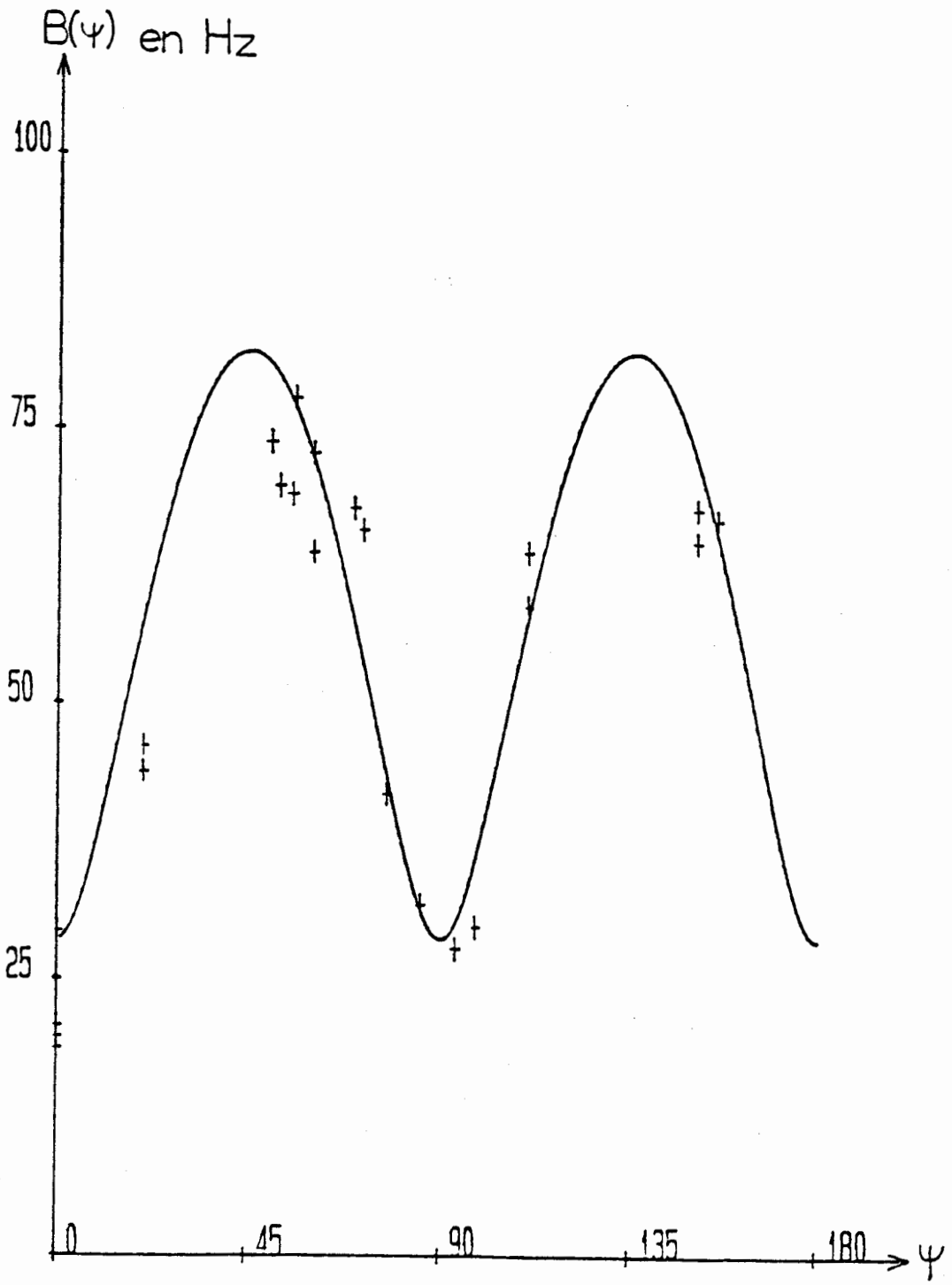


Figure 10 - Sensibilité maximale d'un QAS de coupe

2 ND EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 16 - 17 - 18 MARCH 1988

Session RO-03 Number 4

Space qualified 100 MHz BVA - VHF - Resonator

U.R. Peier, G. Jenni
Oscilloquartz SA, Neuchâtel, Switzerland

Summary

The developement of a 100 MHz fith overtone BVA - VHF - Resonator will be discribed. This resonator fits in a HC-40 type holder and has successfully passed the space qualification procedure:

- test of hermeticity
- X-ray irradiation up to 2 M rad
- thermal stress up to 400° C
- mechanical shocks up to 1290 g at 0 .9 msec
- random vibrations at 19.6 g RMS in a frequency range from 20 to 2000 Hz
- high temperature storage (125°C)
- burn-in test 2000 h and 5000 h
- ageing

All parameter measurement was done using IEC 444-3, the π-network with C_o-cancellation. This method turned out to be precise but not to be very handy for the user.

Results of all these tests and a summary of advantages and disadvantages of a 100 MHz resonators vs. the classical way of frequency-multiplication will be given in the paper.

The developement and the evaluation has been carried out under a contract of ESTEC (5190 / 82 / NL / GM).

Keywords

Quartz, Resonator, BVA, electrodeless, spacequalified, radiation hardened

1. Introduction

1.1. Reasons for using 100 MHz BVA resonators

The classical way for having a high stability source at 100 MHz is using a state of the art oscillator (Oscilloquartz 8600) at 5 MHz and multiplying by 20. In this case you will get the good long term stability (< $2 \cdot 10^{-11}$ / day) of the 8600 but you will get also the noise increased by 26 db. Fig. 1 gives a sketch of the situation. The BVA-VHF curve holds for a loaded Q of 65000. At Fourrier frequencies below 46 Hz the multiplied version is clearly better than the BVA-VHF but at frequencies higher than 46 Hz the noise of the new resonator becomes better: Compared with the multiplied version the noise floor is 18.5 db lower. For certain application this can be an advantage.

1.2. Restrictions

Many people thought that the "sandwich construction" of a BVA-resonator would never be able to survive shocks and random vibrations. People familiar with the construction of BVA-resonators found shock- and vibration-testing of BVA-resonators a very crazy idea - my person not excluded.

The tests made evidence of the contrary: The BVA-VHF design was very well able to survive shocks up to 1300 g's and random vibrations of 12.6 g/RMs (with peaks up to 80 g).

Testing "naked" crystals is a cumbersome task because such a crystal does not output a signal by itself. To get out a measurable quantity you have to put it in an oscillator or to use some passif techniques. If you try the first approach then you will not have problems with temperature but you certainly will have all the influences of the oscillator. On the other way with the passif method the measurement will be influenced by the circuit to a lesser degree and the influence will always be the same, but for each measurement you have to heat the crystal to its turn over temperature of about 75°C and wait long enough to ensure temperature stability. In the case of the BVA-VHF the thermal time constant is quite long. This means a waiting time of 90 minutes each time.

2. Construction of the BVA-VHF resonator

The BVA family of quartz is characterised by the fact that they do not have electrodes directly on the resonator. The electrodes are on a piece that we call "condensors" i.e. the electrodes are separated from the resonator by a gap which is for the BVA-VHF resonator in the range of 0.7 μ. The electrodes taken away from the resonator to avoid ion migration and therefore ageing. A sketch of the situation is shown in Fig. 2.

The two condensors and the plano-plano resonator are sandwiched together by three clips. The mounting structure has serveral functions:

1. a rigid cage decouples mechanically the sandwich from outside influances (stress due to temperature changes, atmospheric pressure, deformation of the enclosure etc)

2. protection of the very fragile sandwich (a 100 MHz resonator has a thickness of only 83 μ!)

On the surface of the condensor the electrodes are brought to the edge where they can be epoxy-bonded to a small nickel wire that is soldered to the feedthrough of the base.

The base is a standard HC-40/U on which we solder the base-plate. The cover is a modified version of the HC-40/U. To be consistant with our technology we need a pinch-off tube on the cover. Base and cover are coldwelded together.

To improve ageing the resonator undergoes a bake-out process at 180°C for 72 hours while pumping with a cryopump in the 10^{-8} Torr range. At the end of the bake-out cycle the enclosure is sealed by pinch-off. By this way we get at the same time a very severe test for the hermeticity of the cold-weld. The BVA-VHF enclosure is shown in Fig. 3.

3. Manufacturing of the BVA-VHF

The pieces were manufactured in the classical way. Sawing was done on a slurry-saw, angle correction was done by hand with a tolerancy of ± 30 sec of arc. To be able to align the three pieces of the sandwich, during roundgrinding a reference flat is produced. This alignment is necessary to avoid stresses due to the very different thermal expansion coefficients in the AT-blank (14.3 ppm/°C in X and 9.9 ppm/°C in Z') which might lead to frequency shifts.

Lapping is classical: We use planetray lapping and pin-lap machines. The slots are cut out by ultrasonic machining (after beeing polished) and the gap is created by chemical etch. The gold electrodes are deposited by evaporation under high vacuum.

The semi-terminated pieces were stored eather under vacuum or in a glovebox filled with dry nitrogen (< 40 ppm humidity). Mounting was done in the same glovebox.

4. Parameters

The frequency choosen was 99.55 MHz for the simple reason that our synthesizer was limited to 99 999 999.9 Hz. The results obtained in the test series is shown in Table. 1 with the extremas in Table 2. The measurements after the sealing test are taken as reference data.

As can be seen in Fig. 4 the VHF- BVA has a quite rich spectrum. Fig. 5. shows the same thing but in three dimensions.

4.1. Measurement method

Knowing that the conventional π-network method has some limitations in precision in the frequency range above 50 MHz, we have choosen IEC 444-3 as our measuring method. IEC 444-3 cancels out the influence of C_0 by using a resonating circuit. The schematic diagram can be found in Fig. 6. The measurement procedure is straight forward but really not very handy:

4.1.1. Initial set up

In a first step the parallel resistance R_{po} of the compensation circuit has to be determined. With the short circuit inserted the V_{BsP} and V_{AsP} respectivly readings are taken and K_p is calculated

$$K_p = \frac{V_{BsP}}{V_{AsP}}$$

Then the crystal is inserted and the parallel inductance L_p is ajusted to get a minimum voltage V_{BP} on channel B, then R_{po} can be computed

$$R_{po} = K_p \left(\frac{V_{AP}}{V_{BP}} - 1 \right) 2 R_T \qquad (R_T = 12.5\,\Omega)$$

Then the compensation must be fine tuned to get symmetrical phase readings (say ± 20°).

4.1.2. Frequency measurement (f_s)

The resonance frequency f_s is the frequency where the phase is zero.

4.1.3. Series resistance R_1

Since there is a resistance R_{po} (from the compensation circuit) in parallel to the quartz, R_1 cannot be measured directly. What is accessible is the total resistance R_0. It can be calculated from the values V_{A0} and V_{B0} and K with

$$K = \frac{V_{Bs}}{V_{As}}$$

V_{As} and V_{Bs} beeing the readings with the short circuit inserted.

$$R_0 = K \left(\frac{V_{A0}}{V_{B0}} - 1 \right) 2R_T$$

$$\frac{1}{R_1} = \frac{1}{R_0} - \frac{1}{R_{po}}$$

4.1.3. Motional inductance L_1

Two frequency readings $f_{+\phi}$ and $f_{-\phi}$ have to be taken at identical phase offsets $+\phi$ and $-\phi$. Fig. 4 shows also clearly that at 100 MHz the resonance circle becomes fairly small. A value of $\phi = 18°$ has been choosen. L_1 can be calculated from

$$L_1 = \frac{R_{po}\cot\phi - \sqrt{R_{po}^4 \cot^2\phi - 4B(R_{po} + 2R_T)}}{4\pi(R_{po} + 2R_T)(f_{+\phi} - f_{-\phi})}$$

with

$$B = R_{po}R_1^2 + R_{po}^2 R_1 + 2R_T R_1^2 + 4R_{po}R_1 R_T + 2R_{po}^2 R_T$$

With a handheld calculator these computations can easily be done but the measurement method itself is not handy to be implemented on an automatic measurement system. The resolution of the synthesizer becomes beeing the limiting factor for the accuracy. So the readings were taken manually and the parameter computation was afterwards done on a HP 85 computer.

4.2. Precision

With the equipment used the following precision can be obtained

Frequency $\frac{\Delta f}{f} : < \pm 10^{-7}$

Temperature : ± 0.5°C (turn over temperature)

resistance : < ± 5 %

5. Test procedure

The evaluation program was given by the ***European Space Agency*** **(ESA)**; the flowchart is given in Fig. 8. All 15 pieces were seal tested and after that test reference data were taken. Then the serie was splitted: 2 pieces were irradiated, 3 pieces were thermally cycled and 10 pieces were submitted to mechanical shocks. Then 2 pieces were selected for destructif testing (dry heat and shocks). At the end parameter measurement has been repeated before the pieces went to the 2000 h and 5000 h respectivly storage test.

5.1. Sealing test

ESA prescribed IEC 68.2.17, Test Qk with the following parameters:

Test 1

Severity 1000 hours ($q = 4 \cdot 10^6$ sec)

Test parameters:

internal volume : 1.7 cm³

immersion time : 30 minutes

max. leak rate : $R = 10^{-8}$ bar • cm³/sec

special parameters

$$L = \frac{P_0 V}{\theta} = \frac{1\ \text{bar} \times 1.7\ \text{cm}^3}{1000 \times 3600} = 4.7 \cdot 10^{-7}\ \text{bar} \cdot \text{cm}^3/\text{sec}$$

for air and for He

$$R = \frac{7.17 L^2 P t_1}{V} = 3.75 \cdot 10^{-7}$$

The test limit was therefore $4 \cdot 10^{-7}$ bar • cm³/sec. The test itself turned out to be not very significant because the outgassing of the glass seal was considerably high whereas the leakrate was

very small. In fact we had about the same readings for an open and a closed can! In the IEC procedure you can find at several places the remark "interpretation depends upon the experiance of the testing person" what says anything.

Fig. 7 gives the outgassing curves for some crystals, the readings are all well below the limit of $4 \cdot 10^{-7}$. Large leaks would anyway be detected during the bakeout procedure.

5.2. Irradiation test

Since Oscilloquartz is not equipped for doing such kind of tests we submitted the samples to ESA for the irradiation test. Two samples have been tested. #16 was submitted an X-ray dose of 1 MRad and #52 of 2 MRad. The dose was measured on surface of the enclosure.

The samples were returned to us and the parameter measurement was repeated (at the individual turn over point of the crystal). The resulting frequency shift was surprisingly small: -0.06 ppm and -0.216 ppm for the two samples. We are unable to tell wether these changes come from irradiation, early ageing or shocks and vibration during transportation from Switzerland to the Netherlands and back. But this result correlates well with the experiment[2] of Johns Hopkins University, Laurel, M. It seems to be that the BVA resonator design is a method to get radiation hardened crystals. We feel that the fact that the electrodes are not deposited on the resonator itself can avoid ion migration, an effect that could be increased by the X-ray energie.

5.3. Dry heat

The goal of this test was screening the enclosure. The test followed MIL-Std-883 D method 1010.2. It comprised 10 cycles from -65°C to +150°c and was executed by changing the samples from one oven to another. None of the samples failed.

5.4. Shock test (non destructing)

All of the samples were submitted to this test. The samples were fixed in a test fixture as in Fig. 9. The test followed IEC 68.2.27, Test Ea. Severity was half sine, 50 g ± 2%, 14 msec. We applied 3 shocks in the 3x2 axes, therefore 18 shocks per quartz.

An example is given in Fig. 10. As can be seen there is a large frequency offset during the test but after a recovery time the frequency returns to its initial value. Table. 3 shows a mean change of -0.063 ppm for the set of samples which is well below the measurement accuracy.

5.5. Dry heat (thermal steps)

By steps of 50°C the temperature was raised to 400°C. Exposure time at each level was 4 hours; the crystal was then cooled down to ambient temperature and brought up to its turn over temperature to get comparable parameter measurements.

Fig. 11. and Fig. 12. give the results. Up to 150°C there is nothing happening. At 200°C the solder on the feedtrough starts softening (fusion point 220°C). Up to 250°C the parameters do not show changes but you have to remember that at 250°C all solder connections are melting. Even the solder between the cage and the base will melt. The results show clearly that melting has not necessarly to lead to destruction. Though intended as a destructive test only one of the two samples has been destroid.

5.6. Shock testing (destructing)

IEC 68.2.27, Test Ea was applied, severity was to be increased by steps until destruction. The applied acceleration and duration can be seen from Table 4. In practice we had to stop at 1300 g due to limitation of the testing machine, though none of the samples was broken.

The frequency evolution is shown in Fig. 13. As in the case of the 50g shocks, there is a frequency shift of about 1 ppm after the first shock but the frequency comes back after a recovery time. The change in resistance is not significant since it is only half the value of the measuring precision.

5.7. Random vibration

The specs were: Random vibration at 19.6 g/RMS from 20 to 2000 Hz with a crossover at 50 Hz during 1.5 minutes per axis (3 axes tested). After the test the mean frequency change was found to be 0.012 ppm and the resistance change was about 0.5%. Both values are much smaller than the measurement precision so that can be said that there is no significant change.

A search for mechanical resonances in the range 20 to 2000 Hz was also negatif. The test shows the efficiency of the suspension: Vibration seems to be absorbed by the suspension structure and cannot reach the sandwich.

5.8. High temperature storage

This is a screening test that completes the program. The samples are stored at 125°C during 10 days. It seems to be a very efficient screening. After 10 days 2 of the 9 samples ran out of the specification. Fig. 14, 15 and 16 show the evolution. The dashed lines are the specification.

5.9. Ageing

The crystals have been mounted in in oscillator OSA 8657 and monitored for now 2 years. Ageing is found to be in the high 10^{-11}/day and low 10^{-10}/day range what is not really bad for a 100 MHz crystal but is not as good as we were hoping. That there seem to be seasonal influences is a more discourageing thing, though we are unable to tell wether they come from the crystal or (what is more likely) from the oscillator. Fig. 17. gives an ageing curve for the last two years. Measurement were taken every day at midnight, what is shown in Fig. 17 is a compressed plot.

6. Conclusions

A BVA-VHF structure can not only resist to mechanical environmental testing but has also passed the space qualification procedure tabled by ESA. The samples survived shocks up to 1300 g (0.9 msec) and random vibration of 19.6 g/RMS (20 ... 2000 Hz).

The BVA-VHF structure turned out have a good radiation hardedness. This fact has also been reported by other authors on this forum.

The BVA at 100 MHz is difficult to manufacture and frequency ajustment is a big but not insurmountable problem.

Oscillators equipped with BVA-VHF resonators exhibit a lower noise floor, but near the carrier they are not as good as a multiplied version of the OSA 8600.

Acknowledgement

The authors wish to thank Mr. P. Monier of ESTEC for the many encouraging discussions during the contract time. We wish to thank also Mr. P. Girardet of OSA for providing noise data.

References

1. B. Dulmet, ENSMM Besançon, Interim Report

2. J.J. Suter, A.G. Bates and J.M. Cloeren from Johns Hopkins University, Laurel, MD, elsewhere in these prodeecings

99V002/1

#	f [Hz]	R Ω	L1 [H]	C1 [F]	Q	Pi [°C]	C0 [pF]	Date dd.mmyy
2	99,549,918.00	53.26	8.23E-03	3.11E-16	96,665	77.0	3.16	12.02.85
3	99,550,106.00	61.85	8.98E-03	2.85E-16	90,791	77.4	3.20	12.02.85
5	99,548,535.00	125.46	1.48E-02	1.73E-16	73,764	76.6	3.10	12.02.85
6	99,550,466.00	154.14	1.03E-02	2.48E-16	41,757	76.6	3.10	12.02.85
7	99,550,388.00	125.46	1.10E-02	2.33E-16	54,665	76.7	3.10	12.02.85
9	99,550,008.00	129.16	1.33E-02	1.93E-16	64,172	72.0	3.10	12.02.85
14	99,550,142.00	94.84	1.05E-02	2.43E-16	69,254	76.1	3.10	12.02.85
16	99,549,783.00	71.58	9.21E-03	2.78E-16	80,487	76.7	3.14	12.02.85
17	99,549,943.00	90.44	9.72E-03	2.63E-16	67,191	76.5	3.06	12.02.85
19	99,549,893.00	73.59	9.46E-03	2.70E-16	80,442	77.1	3.13	12.02.85
20	99,549,728.00	63.88	1.03E-02	2.49E-16	100,495	74.5	3.13	12.02.85
23	99,549,977.00	77.88	8.79E-03	2.91E-16	70,631	76.1	3.17	12.02.85
32	99,550,240.00	73.08	8.44E-03	3.03E-16	72,211	75.9	3.15	12.02.85
33	99,550,070.00	71.58	8.31E-03	3.08E-16	72,624	76.6	3.16	12.02.85
36	99,550,631.00	52.31	7.57E-03	3.38E-16	90,509	74.4	3.10	12.02.85
42	99,550,482.00	59.15	7.32E-03	3.49E-16	77,432	77.8	3.10	12.02.85
45	99,550,041.00	66.89	1.04E-02	2.45E-16	97,444	76.5	3.16	12.02.85
46	99,550,193.00	82.56	1.00E-02	2.56E-16	75,759	76.3	3.19	12.02.85
47	99,549,836.00	56.97	9.62E-03	2.66E-16	105,592	75.2	3.16	12.02.85
51	99,549,590.00	84.42	1.07E-02	2.38E-16	79,476	75.9	3.17	12.02.85
52	99,550,352.00	66.45	1.02E-02	2.51E-16	95,721	74.2	3.19	12.02.85
53	99,550,283.00	61.85	1.01E-02	2.53E-16	102,275	75.6	3.16	12.02.85
54	99,550,142.00	79.58	1.16E-02	2.21E-16	91,033	79.4	3.20	12.02.85
55	99,550,108.00	69.65	1.06E-02	2.41E-16	95,385	77.3	3.15	12.02.85
57	99,550,121.00	117.49	1.04E-02	2.45E-16	55,590	76.0	3.18	12.02.85
58	99,549,901.00	96.38	9.76E-03	2.62E-16	63,309	78.5	3.16	12.02.85
mean=	99,550,033.73	83.07	9.98E-03	2.62E-16	79,411	76.27	3.14	
sigma=	392.22	26.75	1.60E-03	3.94E-17	16,435	1.45	0.04	

Table 1. Parameter measurement - *data taken after seal test (reference data)*

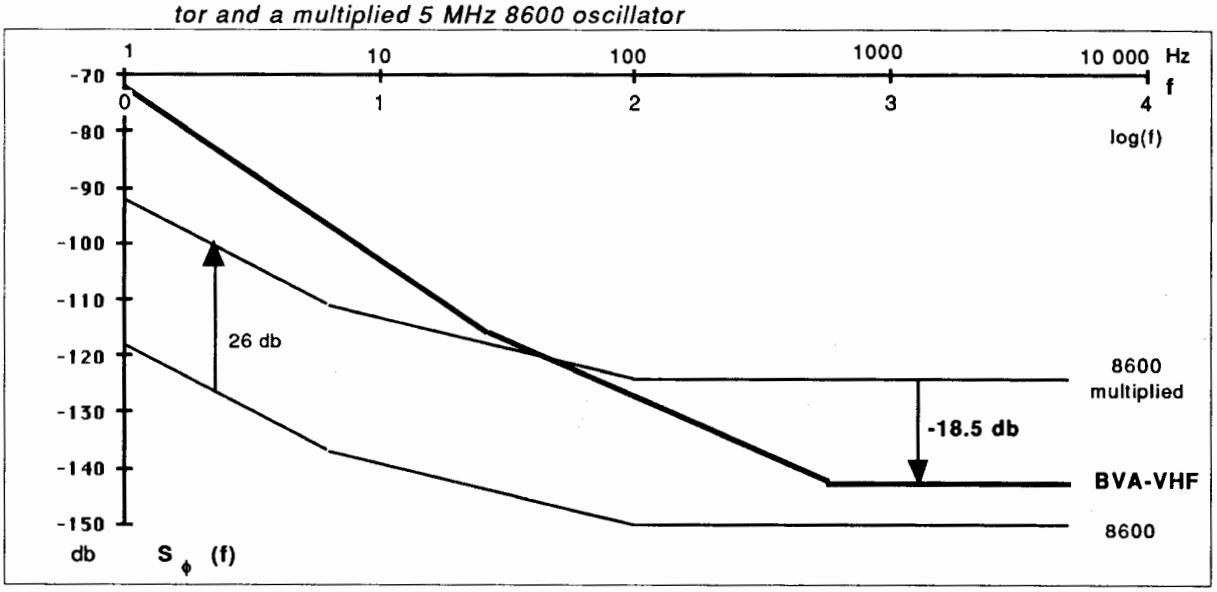
#	Characteristics	Symbol	Value min.	Value max	Unit
1	nominal frequency	f_0	99.550000		MHz
2	Reference Temperature	T_0	25		°C
3	overtone		5		
4	Load capacitance	C_L	∞		pF
5	Drive level	P	100		µW
6	Frequency ajustement	$\frac{\Delta f}{f}$	-14.7	6	ppm
7	resonance restistance	R_1	52.3	125.5	Ω
8	Frequency change in operating temperature range	$\frac{\Delta f}{f}$	N/A		ppm
9	Resistance change in operating temperature range	$\frac{\Delta R_1}{R_1}$	N/A		%
10	Turn over temperature	P_t	74.2	78.5	°C
11	Frequency change vs. drive level	$\frac{\Delta f}{f}$	N/A		ppm
12	Resistance change vs. drive level	$\frac{\Delta R_1}{R_1}$	N/A		%
13	Motional inductance	L_1	7.3	14.8	mH
14	Motional capacitance	C_1	0.173	0.349	fF
15	Statical capacitance	C_0	3.1	3.2	pF
16	Quality factor	Q	54000	106000	

Table 2. Parameter deviation in the batch

Table 3. Parameter changes during *nondestructif shock testing* (50 g)

#	f [ppm]	R [%]	L1 [%]	C1 [%]	Q
2	-0.3	-6.8	-7.6	8.2	-0.8
3	0.1	-3.4	-4.3	4.4	-0.8
7	-0.4	3.5	3.1	-3.0	-0.4
19	1.1	0.5	-0.9	0.9	-1.4
23	-0.2	-5.4	-5.7	6.1	-0.3
32	0.1	1.2	-1.1	1.1	-2.2
33	-0.5	5.7	2.9	-2.9	-2.6
47	-0.1	0.3	0.7	-0.6	0.3
53	-0.2	-0.6	0.5	-0.5	1.1
54	-0.2	-5.9	-1.2	1.2	5.0
55	0.0	-27.6	-21.5	27.3	8.5
mean =	-0.063	-3.51	-3.19	3.84	0.57
sigma =	0.42	8.92	6.92	8.54	3.31

Fig. 1. Short term stability vs. Fourrier frequency - *A comparison between an VHF-Resonator and a multiplied 5 MHz 8600 oscillator*



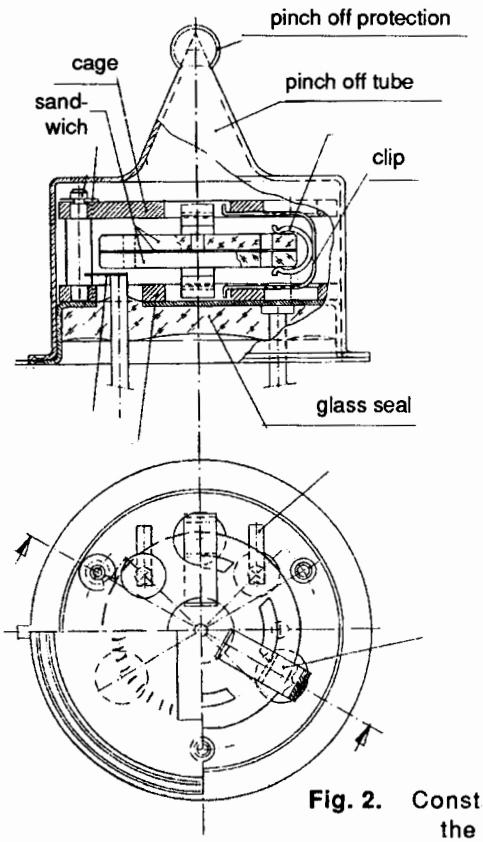


Fig. 2. Construction of the BVA-VHF

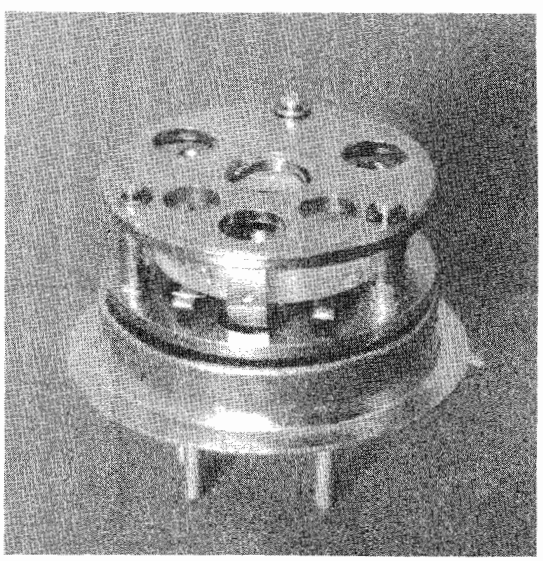


Fig. 3. BVA-VHF Resonator

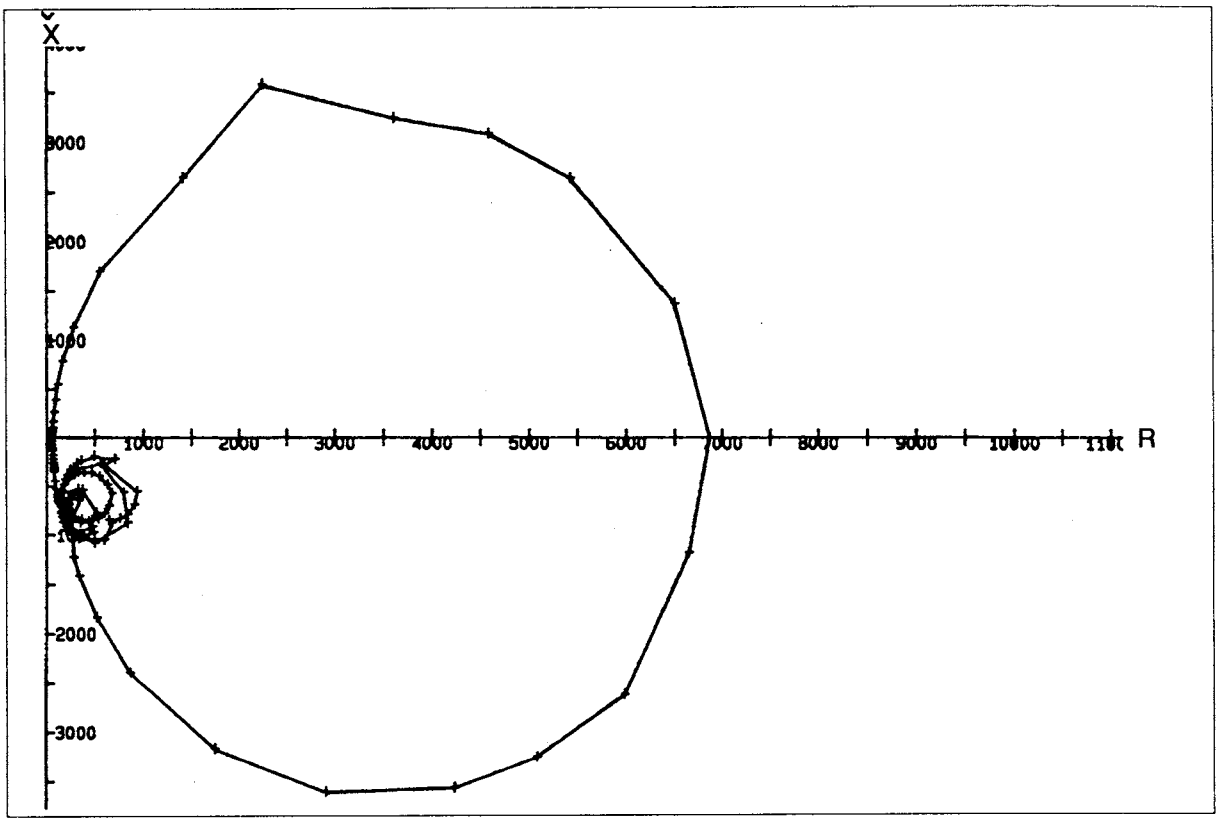


Fig. 4. Resonance curve of the BVA-VHF - *Impedance plane*

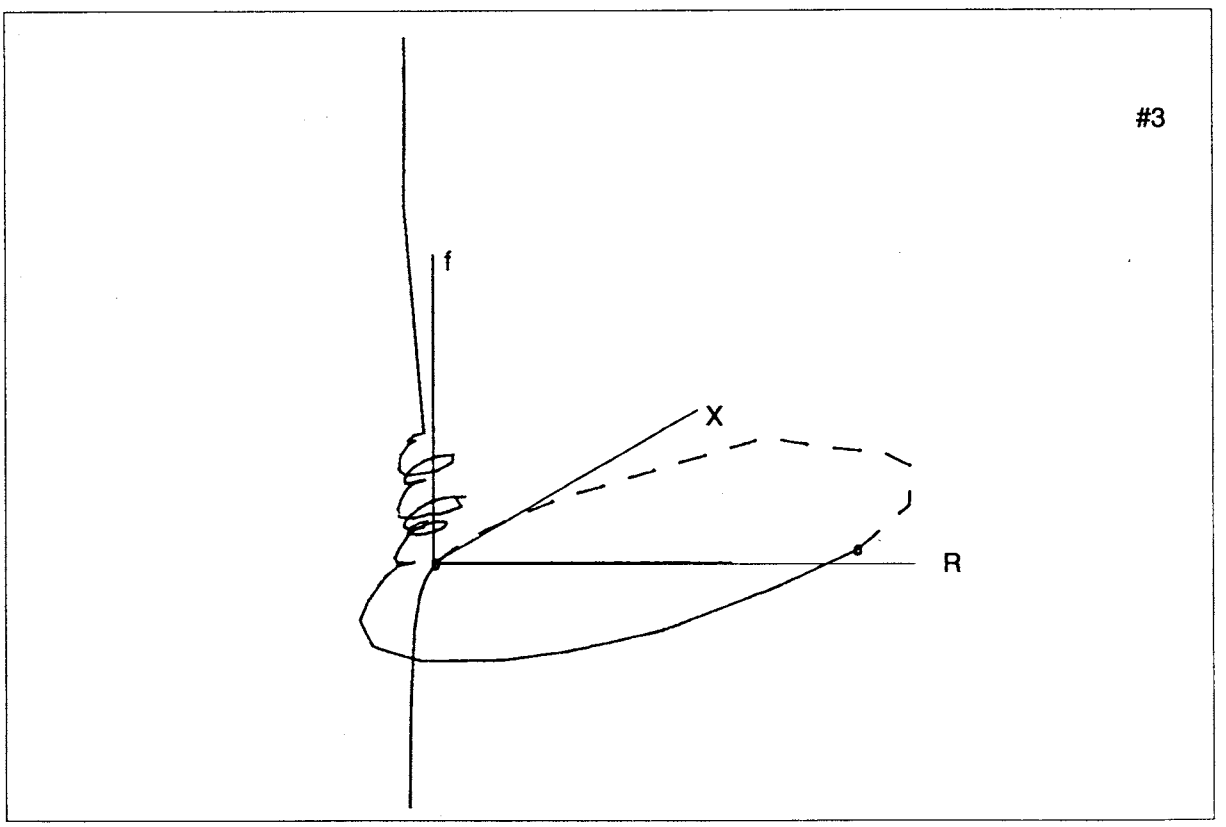


Fig. 5. Impedance curve in 3 dimensions (f, R and X) - *same quartz as in Fig. 4*

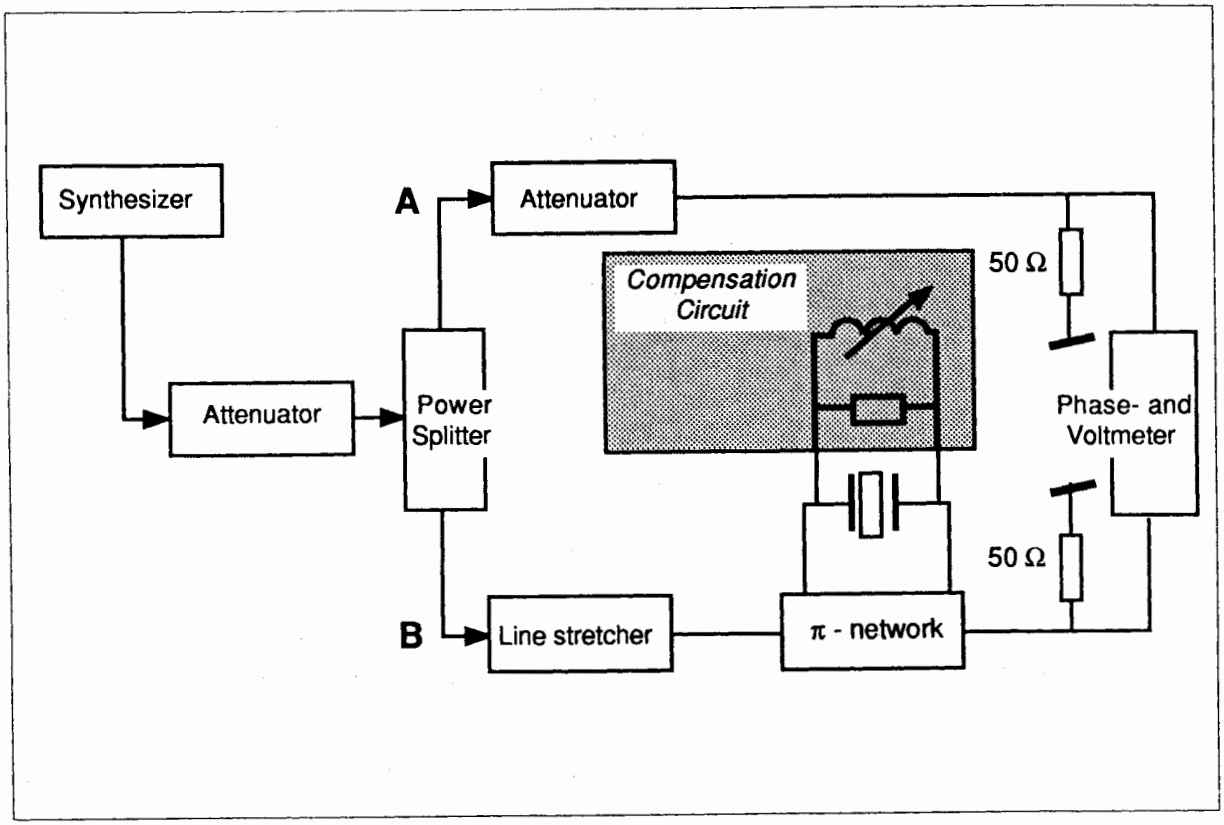


Fig. 6. Compensation circuit for C_0 cancellation - *method following IEC 444-3*

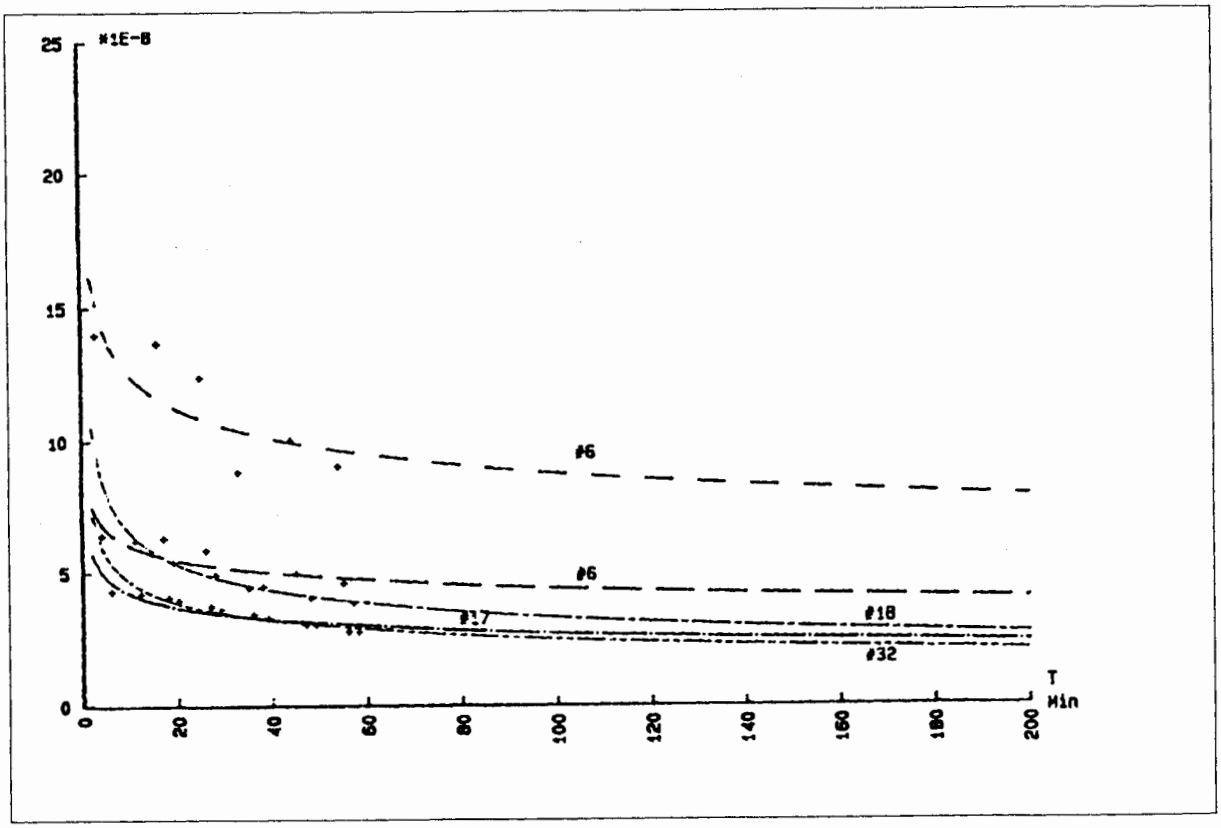


Fig. 8. Outgassing curve - *seal test for #6, 17, 18, 32*

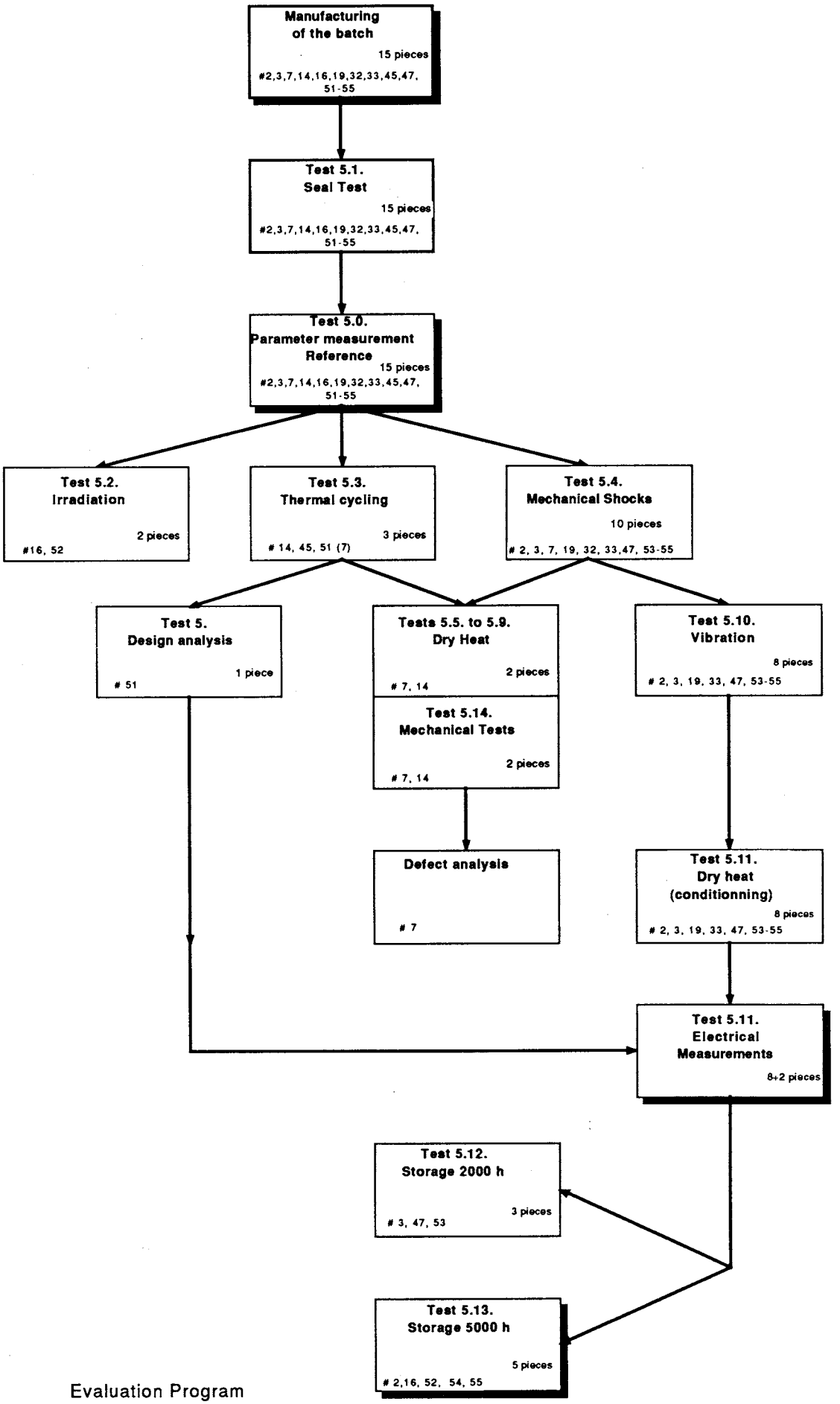


Fig. 7. Evaluation Program

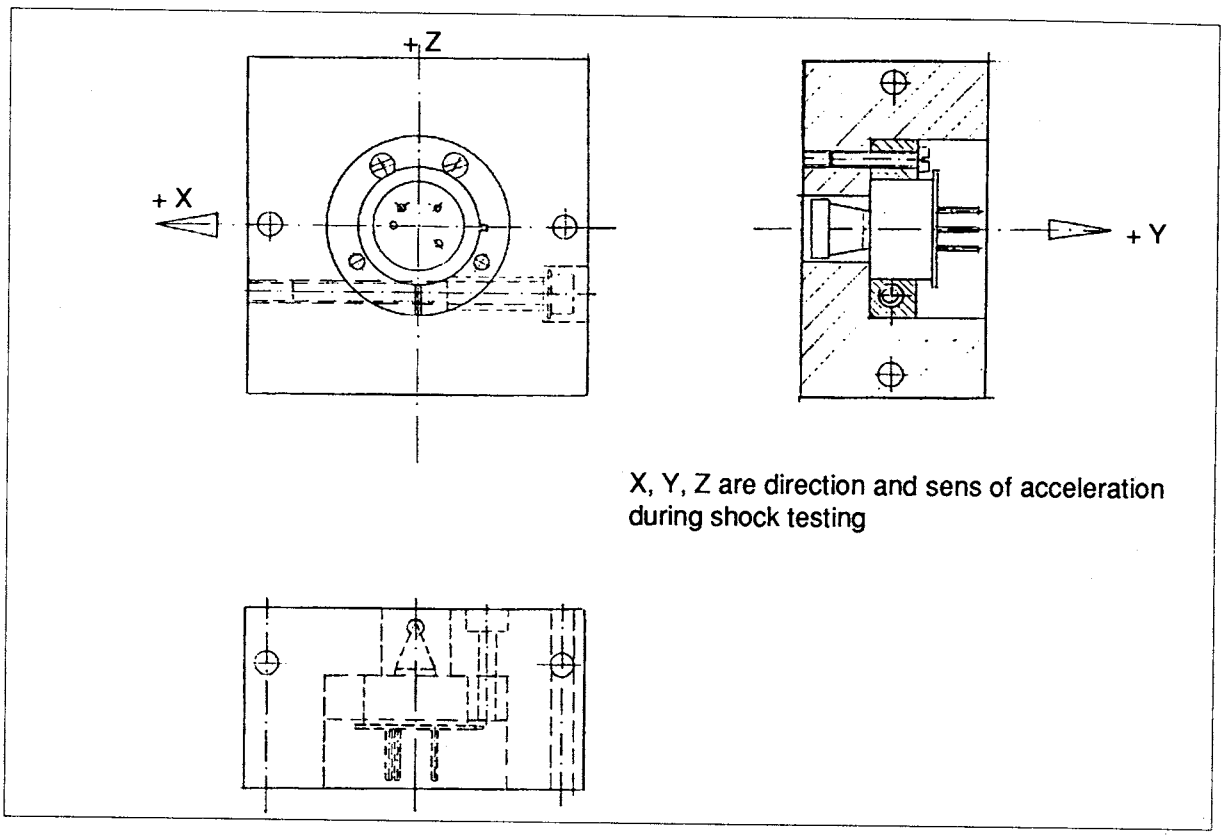


Fig. 9. Fixture for shock testing

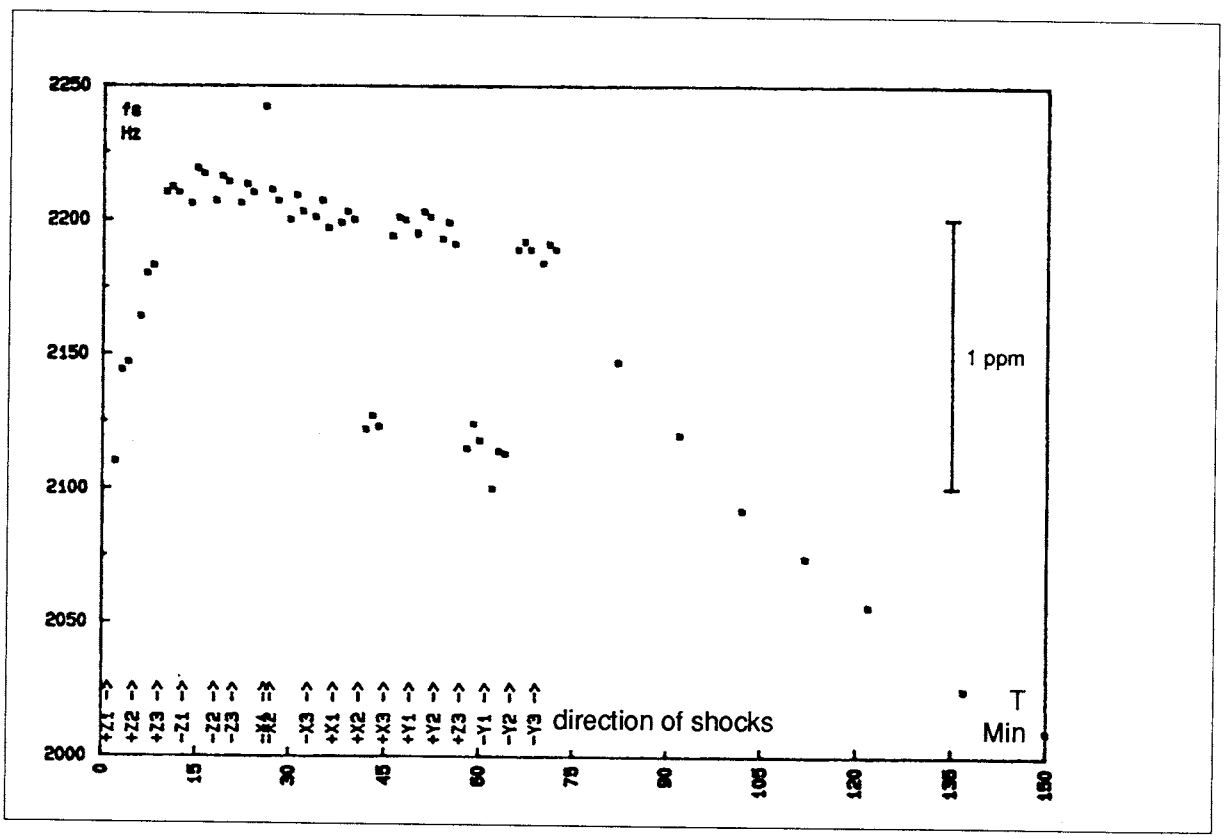


Fig. 10. Frequency shifts during *shock testing* (50 g) and *recovery time*

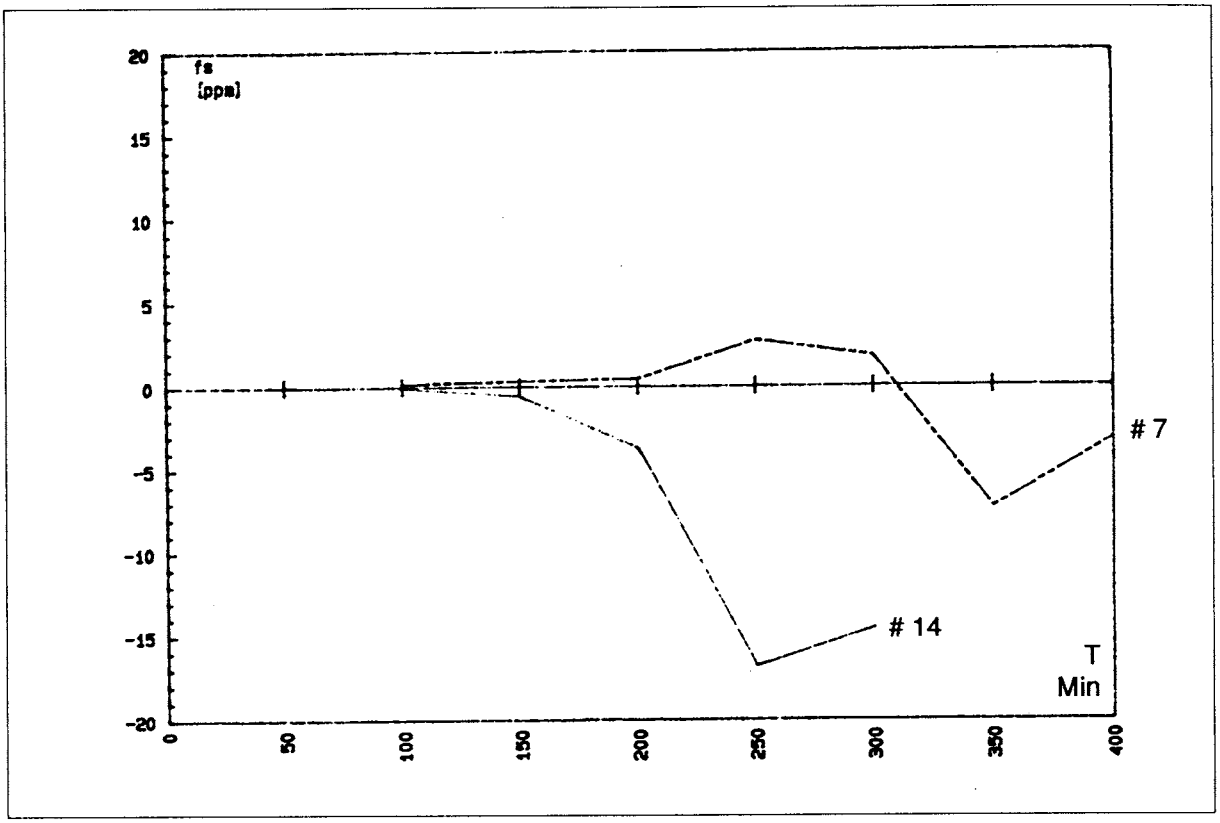


Fig. 11. Frequency shifts during *dry heat*

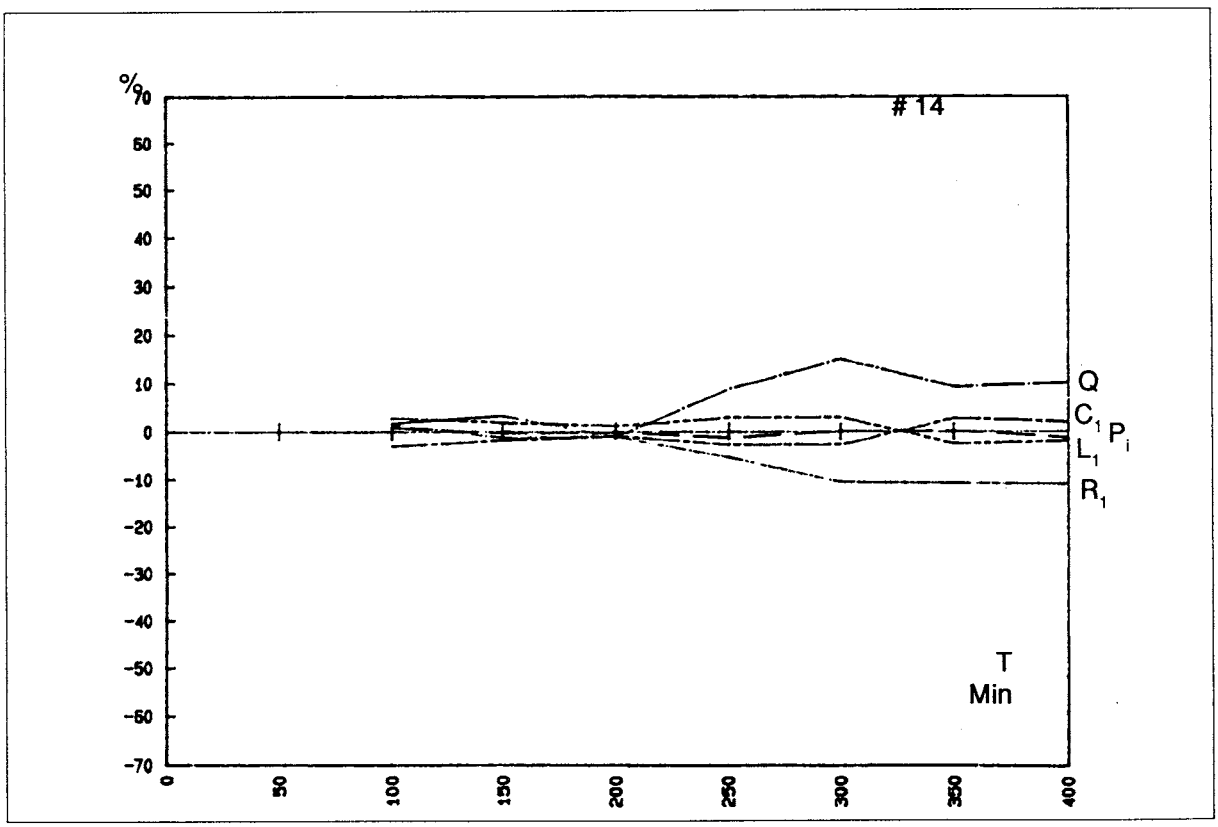
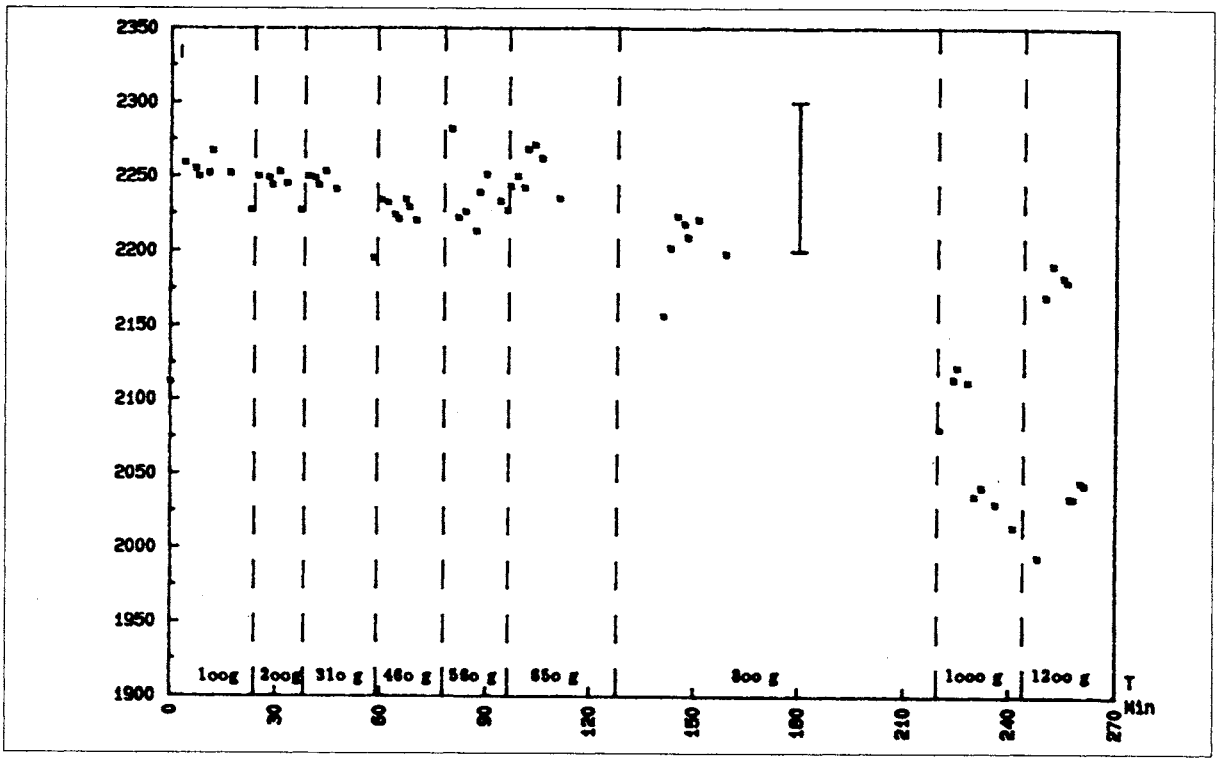


Fig. 12. Parameter changes during *dry heat*



g	100	200	270	300	450	540	640	700	780	900	1190	1290
msec	9	3.8	3	2.5	1.4	1.1	1	1.2	1.1	1	0.9	0.9

Fig. 13. Frequency shifts during *destructif shock testing*

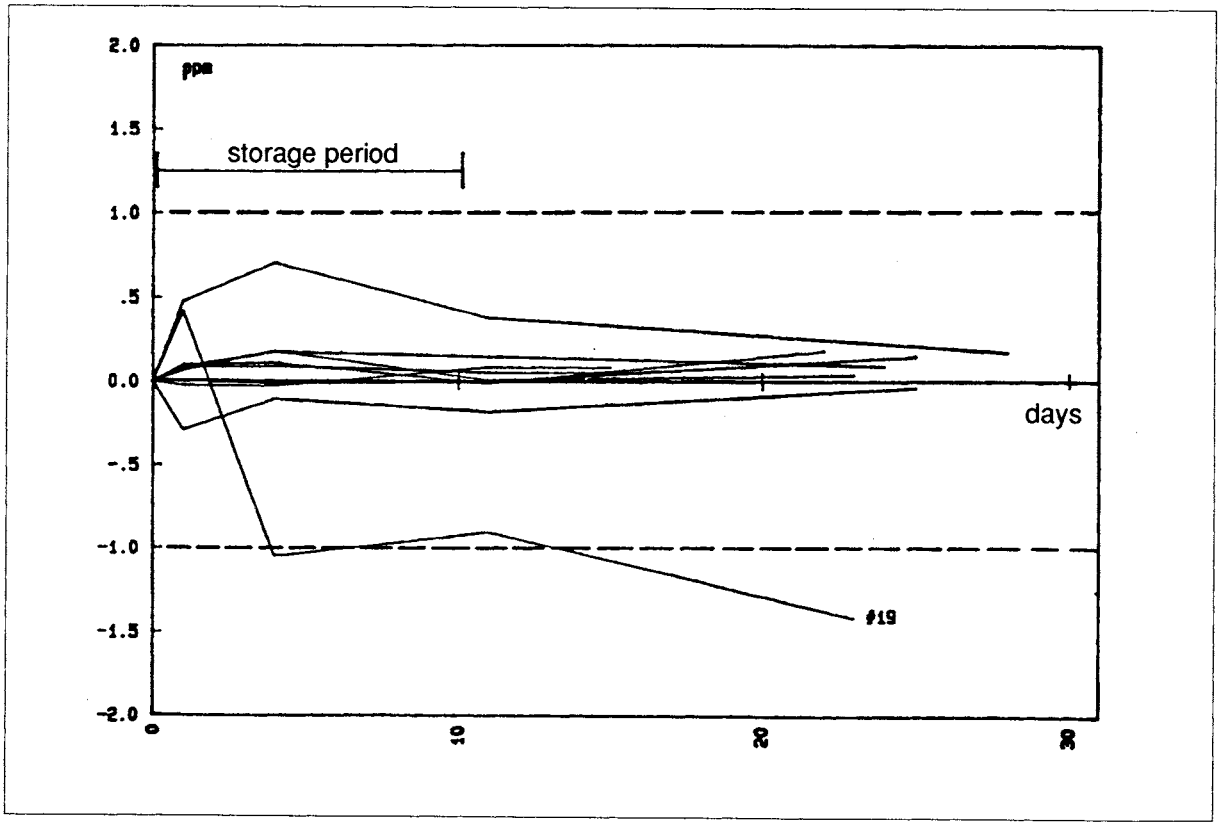


Fig. 14. Frequency evolution during *high temperature storage* (125°C)

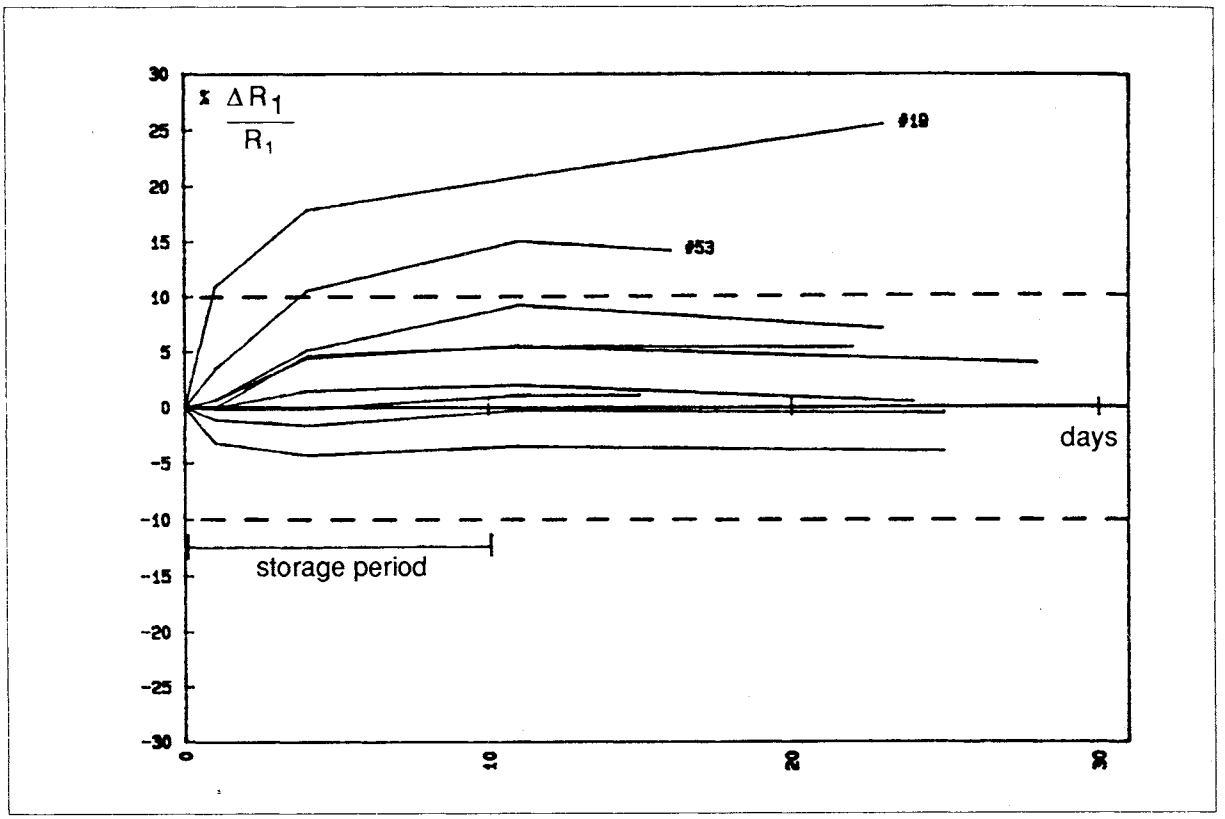


Fig. 15. Resistance evolution during *high temperature storage* (125°C)

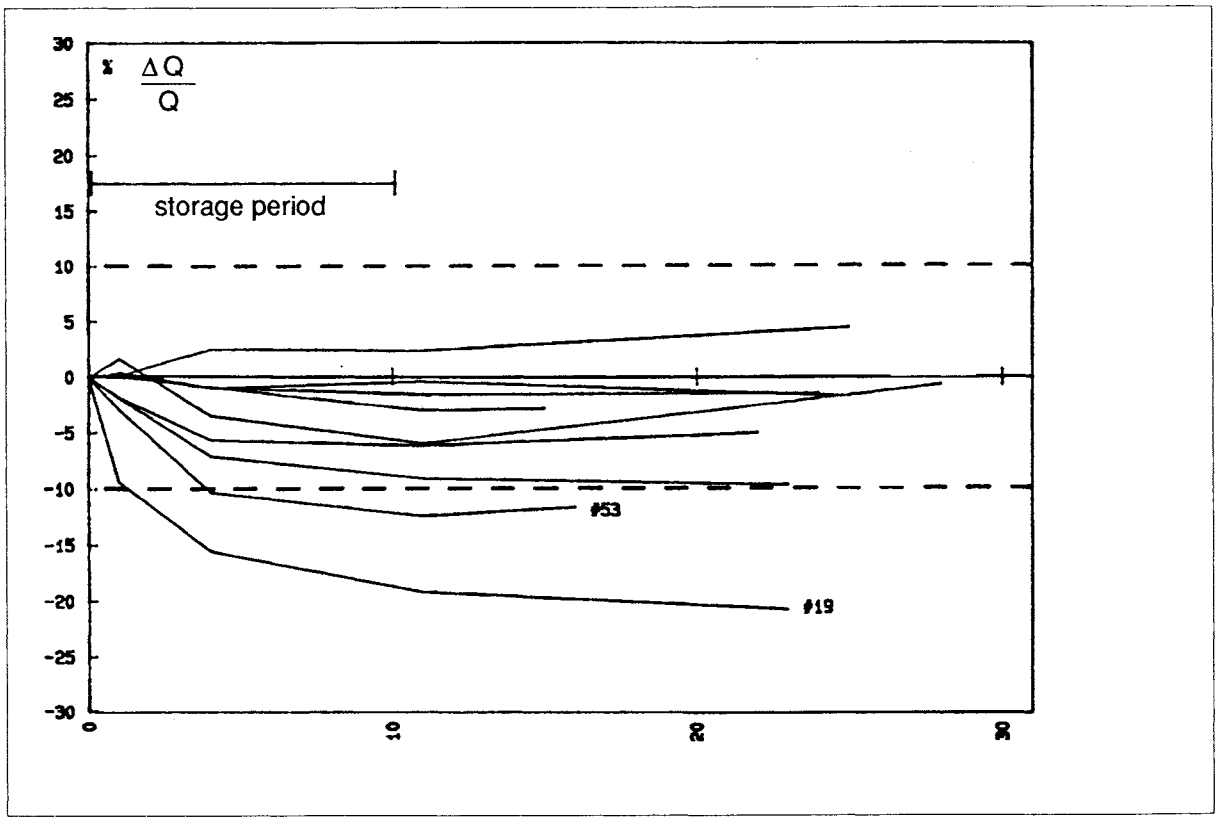


Fig. 16. Changes in quality factor Q during *high temperature storage* (125°C)

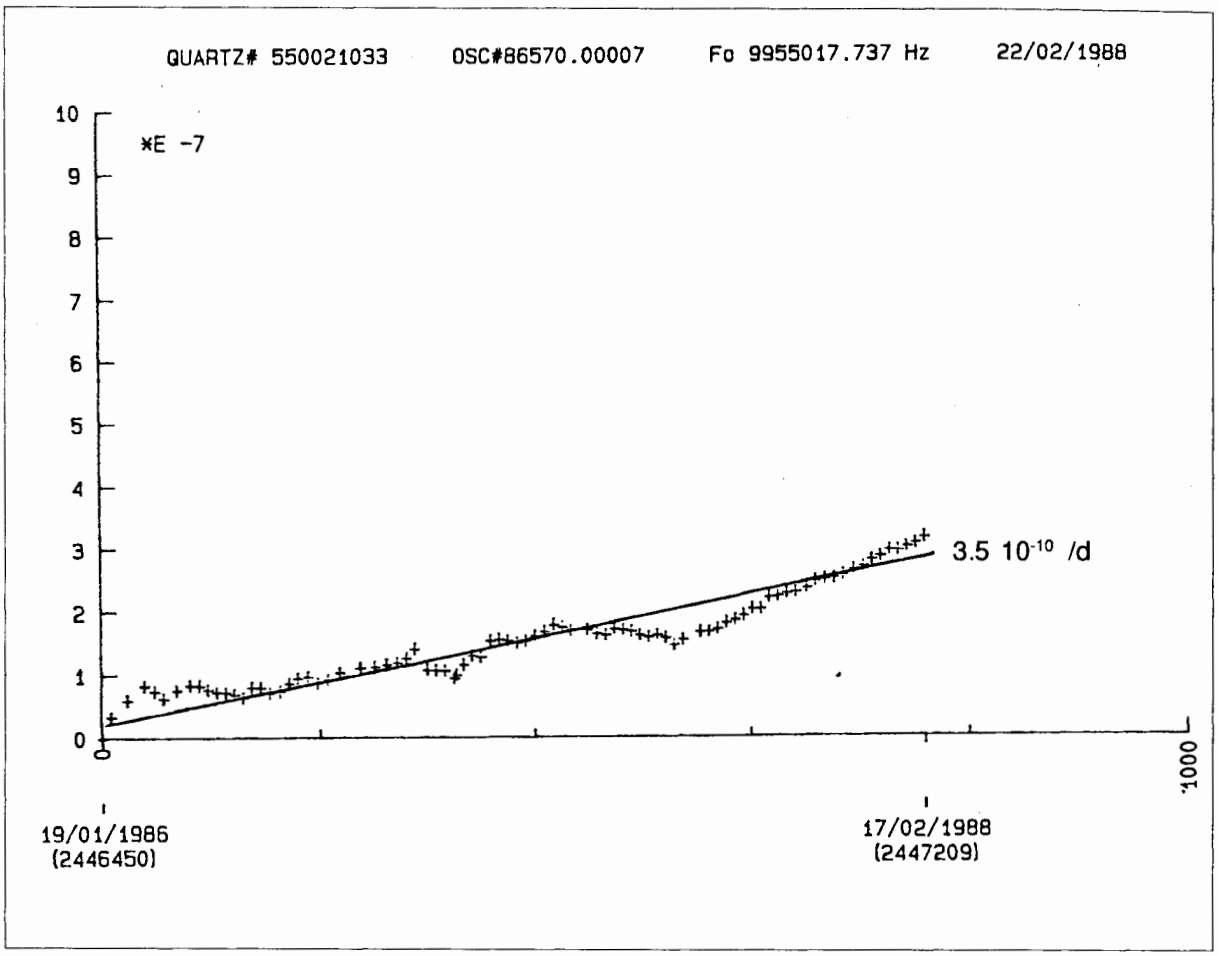


Fig. 17. Ageing curve - *for 2 years*

2 ND EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 16 - 17 - 18 MARCH 1988

Session RO 703 Number 5

Results of the Qualification Programme of a BVA-Type Crystal Oscillator for Satellite Use.

W. Schäfer
Institute of Navigation
University of Stuttgart
Breitscheidstr. 3, D-7000 Stuttgart 1

1. Summary

A commercial crystal oscillator using a BVA-type quartz-resonator has been modified for space use on-board an ERS-1 experiment. The mechanical design has been changed to be compatible with vibration loads and the electrical power consumption has been reduced. After recalibration, the oscillator was integrated into the experiment and successfully subjected to a full qualification programme. The results of initial mechanical investigations on the resonators itself, the electrical characteristics after modification and the tests performed on the integrated unit are described.

2. Application

The oscillator described in this paper is the frequency and time reference for an experiment designed to provide very Precice Range- And Range-rate information (PRARE) of a fast moving spacecraft. It will be launched as an experimental package on the first European Remote Sensing Satellite (ERS-1) by mid 1990.

The measurement principle of PRARE is based on round trip group delay and carrier phase measurements of microwave signals transmitted by the spacecraft using regenerative and coherent ground transponders. Two frequency bands (2.2 GHz and 8.5 GHz) are used to determine the total electron content of the propagation path to allow for ionospheric correction. The data provided will be one range measurement and one range-rate measurement per second to a maximum of four ground stations simultaneously. Range accuracy is below ±10 cm and the range-rate uncertainty is 1.8 mm with an integration time of one second each.

3. Oscillator Requirements

The ERS-1 orbit will be about 777km with an inclination of 98°. The relative velocity to ground is in the order of 1 cm/µs. This results in a datation requirement of about 10 µs per day combined with low phase noise characteristics to have a doppler resolution of 0.1 Hz at 8.5 GHz. The absolute frequency itself is not critical due to the two-way system design.

Besides these electrical characteristics, which are well within capabilities of commercially available units, requirements specific to space applications have to be observed, like low power consumption, low mass, the use of non-magnetic materials and tolerance against launch stresses. The unit must operate in a hard vacuum environment and use an appropriate thermal design. The requirements are summarized in Table I below.

- TIME: 10 μSEC/DAY CORRESPONDING TO 6.5 CM ALONG TRACK
- AGING: 10^{-11} /DAY
- FREQUENCY SETTING: 10^{-8} (TWO-WAY SYSTEM)
- EXTREMELY LOW 1/f NOISE
- LOW POWER CONSUMPTION
- RUGGED MECHANICAL DESIGN
- SURVIVE LAUNCH STRESSES
- OPERATE IN HARD VACUUM
- APPROPRIATE THERMAL DESIGN
- APPROPRIATE MATERIALS SELECTION (LOW OUTGASSING)
- LOW MAGNETIC SUSCEPTIBILITY
- TOLERANCE AGAINST COSMIC RADIATION

Table I: Oscillator Requirements

4. Oscillator Procurement

Review of commercially manufactured units at the beginning of the project (1984) showed no readily available designs. The use of rubidium clocks was ruled out because of their relatively high power consumption, which is in the same order as the total power assigned to the instrument in stand-by mode. Quartz resonators showing the most promising results with respect to aging, resistance to cosmic rays and very high Q-factor are manufactured by Oscilloquartz, Switzerland.

The oscillator design using these resonators was well proven. In an agreement with the manufacturer it has been decided to use the resonator and sustaining circuit (the part inside the dewar) as built from the production line and modify the mechanical and electrical hardware at the institute.

As a first step, sample resonators alone have been vibration tested with the crystals hard-mounted on the shaker and in all axis according to values given in Table II.

- SINE:	PEAK VALUES	RANDOM:	13.4 g RMS
	10 - 30 Hz : 10 g		duration: 2 min
	30 - 100 Hz : 15 g		

Table II: Vibration Levels for ERS-1

No mechanical failures have been observed and the frequency changes were below 1·10E-9.

5. Manufacturing

A new buffer design has been developed at the institute, which reduced power and voltage requirements. Due to the better thermal isolation in vacuum, the supply voltage has been reduced to 15V, which is compatible with voltages present in the experiment. The total power consumption is now 1.2 to 1.8W in vacuum, depending on temperature. The mechanical construction is based on a high strength glass dewar (DURAN glass by Schott) which is embedded in a silicone foam mounted in a milled aluminum enclosure. The foam is of an open cell type, which retains its elastic properties in vacuum and under prolonged mechanical stress. Power transistors are mounted directly to the case for good thermal contact. Total mass is now 800 gr.

The electrical circuits have been carefully tested for magnetic susceptibility. It was found, that the main cause of frequency shifts were ferrite materials, changing their inductance due to external fields. All inductors are now wound on toroidal cores made of a low permeability iron powder material. These modifications lowered the susceptibility by an order of magnitude, it is now in the range of 1·10E-12 for 250 Gauss (0.025 T).

6. Oscillator Performance

After tuning and recalibration by the manufacturer, the six units produced have typical characteristics as shown in Table III below:

- SIGMA TAU:	0.1 S	: 1.55·10 E-12
	10 S	: 3.9 ·10 E-13
- SINGLE SIDEBAND NOISE:	1 Hz	: -120 dB
	10 Hz	: -146 dB
	100 Hz	: -151 dB
	1000 Hz	: -153 dB
- AGING:		1.1·10E-11/d
- THERMAL STABILITY IN VACUUM:		3.55·10E-11/K

Table III: Oscillator Characteristics

7. Test Programme

A full qualification programme required by the ERS-1 project has been successfully performed with the PRARE engineering model, consisting of the following tests:

- VERIFICATION OF PHYSICAL PROPERTIES
- THERMAL TEST
- SINE VIBRATION TEST, all axis
- RANDOM VIBRATION, all axis
- MAGNETIC CLEANLINESS TEST
- THERMAL VACUUM TEST
- EMC TEST

During thermal vacuum test, the S-band carrier frequency (2248 MHz) was monitored to determine temperature dependency. After an initial stabilization phase of about 2.5 days, the temperature coefficient was about 3.55·10E-11/K (see Figure 1).

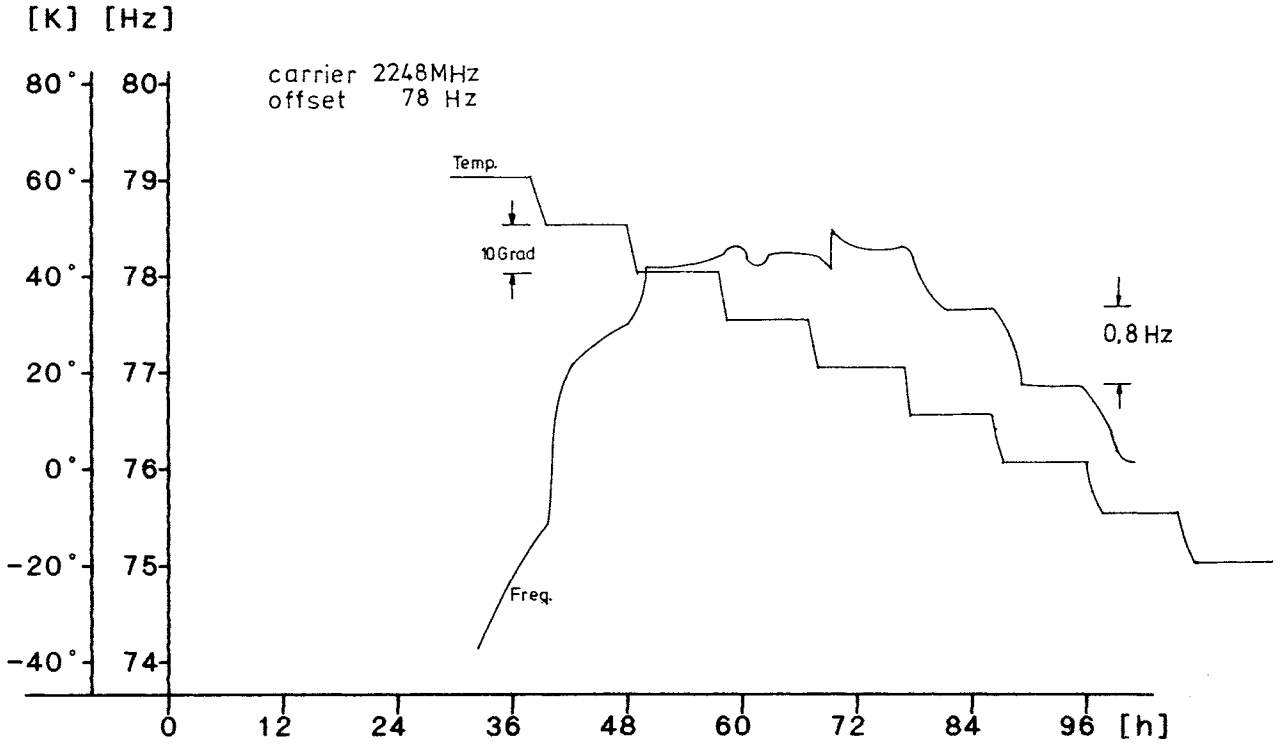


Figure 1: Temperature Stability in Vacuum (1·10E-8 mbar)

8. Acknowledgements

The PRARE project is financed by the German "Bundesministerium für Forschung und Technologie" under grant 01 QS 154.
Satellite Integration is provided by ESA.

2 ND EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 16 - 17 - 18 MARCH 1988

Session RO-03 Number 6

ENVIRONMENTAL EFFECTS ON THE MEDIUM AND LONG TERM FREQUENCY STABILITY OF QUARTZ OSCILLATORS

Fred L. Walls
Time and Frequency Division
National Bureau of Standards
Boulder, Colorado 80303

ABSTRACT

The medium and long term frequency stability of most quartz oscillators is degraded by various environmental effects, the most important of which appear to be acceleration, temperature, load change, pressure, and possibly humidity. In this paper we show preliminary data which indicate that the medium and long term frequency stability of some oscillators can be improved by controlling the pressure and humidity around the oscillator. On one 5 MHz oscillator controlled by a fifth-overtone, AT-cut resonator of the BVA style, we measured a fractional frequency stability, $\sigma_y(\tau) = 3\ (\pm 1) \times 10^{-13}$, in a bandwidth of 100 Hz for measurement times of 0.03 s to 460,000 s after removing a drift of -7.7×10^{-13}/day. Initial measurements on a different oscillator controlled by a traditional 5 MHz, fifth-overtone, AT-cut resonator indicated up to a factor of 2 improvement in the medium term stability when the pressure and humidity were controlled. These data were obtained on only one oscillator of each type and may or may not be representative. If these improvements can be obtained in other precision crystal oscillators, then crystal oscillators may become usable in some applications generally thought to require atomic standards.

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

INTRODUCTION

Most precision, quartz-crystal-controlled oscillators exhibit a flicker-of-frequency floor which is approximately given by the phenomenological equation $\sigma_y(\tau) = 2.8 \times 10^{-7}/Q$ for measurement times of order Q/ν. Here, $\sigma_y(\tau)$ is the fractional frequency stability, ν is the oscillation frequency, and Q is the unloaded quality factor of the resonator [1-7]. For times much longer than Q/ν, the frequency stability generally is degraded by various environmental effects. The most important effects appear to be acceleration, temperature, load change, pressure and possibly humidity. Some oscillators may also show a sensitivity to magnetic field probably caused by the electronics. The level of flicker frequency appears to be very closely tied to the inverse of the forth power of the unloaded quality factor, Q^{-4}. Gagnepain has shown that the predominant contribution to the flicker level is acoustic scattering losses [6]. In this paper we show experimentally that when the obvious environmental effects such as acceleration, temperature, pressure, humidity, and load changes are controlled, the frequency stability (of at least one crystal controlled oscillator) is constant within a factor of 2 for measurement times from 0.03 s to at least 460,000 s. This indicates that the observed flicker-of-frequency spectrum is not just a superposition of multiple noise processes over a narrow range of times, but rather represents a fundamental noise process within the oscillator and quite likely is due to the quartz resonator, which extends over 7 orders of magnitude in Fourier frequency or averaging times. If such performance can be duplicated in other oscillators, then the areas in which quartz controlled oscillators can be used may potentially be expanded.

MEASUREMENTS OF MEDIUM AND LONG TERM FREQUENCY STABILITY

We have recently begun measurements on the medium and long term frequency stability of several bulk-wave, quartz-resonator-controlled oscillators. One of the oscillators uses a traditional fifth-overtone, 5 MHz, AT-cut resonator while the other uses a fifth-overtone, 5 MHz, AT-cut resonator of the BVA design [8]. They are both precision oscillators exhibiting very good

frequency stability for measurement times from 0.03 s to 100 s. For these tests the changes in acceleration, vibration, and load impedance were minimized. The initial tests were performed in a room with temperature changes of about ± 1 K and the naturally occurring changes in atmospheric pressure and humidity.

Figure 1 shows the frequency offset from the NBS time scale for the oscillator that used the traditional AT-cut resonator for averaging times of two hours. Figure 2 shows the fractional frequency stability, $\sigma_y(\tau)$, as a function of measurement time with the drift removed. These data are consistent with random-walk-frequency noise and are similar to those observed by others with AT resonators subject to temperature fluctuations [3,5]. Figure 3 shows the initial frequency stability results when this oscillator was installed inside a simple chamber which stabilized the pressure and humidity. At longer times there is about a factor of 2 improvement in the frequency stability over that obtained while it was open to the atmosphere.

Figure 4 shows the frequency offset from the NBS time scale for the 5 MHz oscillator which used the fifth-overtone, AT-cut, BVA resonator. The temperature stability was of order ± 1 K. Figure 5 shows the fractional frequency stability with the drift removed. Figure 6 shows the frequency of the BVA controlled oscillator after installing it inside a very simple enclosure which largely isolated it from changes in both the external pressure and humidity and controlled the temperature to about ± 0.1 K. The average frequency shift of several in 10^{11} from the value shown in figure 4 appears to be largely due to changes in the load impedance which occurred when the length of the cable was changed. Initially, the frequency drift rate was nearly zero but returned to near the original value after one week of operation. Figure 7 shows the fractional frequency stability, $\sigma_y(\tau)$, as a function of measurement time (exclusive of the first 8 days) with the drift removed. Degrading the temperature stability to ± 1 K degrades the frequency stability approximately 50%. We have also included the short-term-stability performance for a measurement bandwidth of 100 Hz. The 90% error bars on the medium and long term data have been calculated for flicker noise. Figure 8 shows the time residuals after removing the second difference.

These data indicate that, at least for the one BVA oscillator, the noise can be modeled as white phase noise in the short term (which is most likely due to the electronics [1]), flicker frequency in the medium term, and frequency drift in the long term. The random walk frequency noise observed in the traditional AT-cut, resonator-controlled oscillator and in the BVA-AT-cut oscillator open to the atmosphere is insignificant when the pressure and humidity are controlled (under benign laboratory conditions). The mechanism by which the atmospheric fluctuations perturb the frequency are open to speculation. Perhaps the pressure and humidity fluctuations cause small changes in the temperature and/or temperature gradients within the oven.

CONCLUSION

We have shown that the medium term frequency stability of two different 5 MHz oscillators controlled by two different types of AT-cut resonators are somewhat sensitive to changes in atmospheric pressure and possibly humidity. Stabilizing the pressure and humidity, in addition to the normal parameters in an otherwise controlled laboratory setting, improves the frequency stability for measurement times in the region of hours to days. In particular the frequency stability of the BVA controlled oscillator was spectacular, $\sigma_y(\tau) = 3\ (\pm\ 1) \times 10^{-13}$ for measurement times from 0.03 to 460,000 s (measurement bandwidth 100 Hz). Further tests on oscillators controlled by both AT- and SC-cut resonators of several different designs are necessary in order to verify that these effects are generic to unsealed precision oscillators. If this indeed proves to be true, then quartz crystal controlled oscillators engineered to minimize these effects may well be used in areas which have traditionally been dominated by atomic frequency standards.

ACKNOWLEDGEMENTS

I am very grateful to many colleagues over the years for enlightening discussions on the topic. I would especially like to mention Dr. R. Besson, Dr. J. J. Gagnepain, Dr. J. Vig, C. Stone, and J. M. Cloeren, who very kindly shared some of his unpublished data on this topic.

REFERENCES

1. 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, IM-24, 15-20 (1975).

2. J.J. Gagnepain, J. Ubersfeld, G. Goujon and P. Handel, "Relation Between 1/f Noise and Q Factor in Quartz Resonators at Room and Low Temperatures, First Theoretical Interpretation," In Proc. of the 35th Annual Symposium on Frequency Control, 476-483 (1981).

3. J.J. Gagnepain, "Fundamental Noise Studies of Quartz Crystal Resonators, Proc. of the 30th Annual Symposium on Frequency Control, 84-92 (1976).

4. J.J. Gagnepain, M. Oliver and F.L. Walls, "Excess Noise in Quartz Crystal Resonators," in Proc. of the 37th Annual Symposium on Frequency Control, 218-225 (1983).

5. Y. Noguchi, Y. Teramachi and T. Musha, "1/f Frequency Fluctuation of a Quartz Crystal Oscillator, Proc. of the 35th Annual Symposium on Frequency Control, 484-491 (1981).

6. M. Planat and J.J. Gagnepain, "Temperature Dependence of 1/f Noise in Quartz Resonators in Relation with Acoustic Attenuation," Noise in Physical Systems and 1/f Noise, (Edited by A. D'Amico and P. Mazzetti, Elsevier Science Publishers B.V., 1986), pp. 323-326.

7. T.E. Parker, "Characteristics and Sources of Phase Noise in Stable Oscillators," Proc. of the 41st Annual Symposium on Frequency Control, 99-110 (1987).

8. R. Besson, "A New Piezoelectric Resonator Design," Proc. 30th Symposium on Frequency Control, 78-83 (1976).

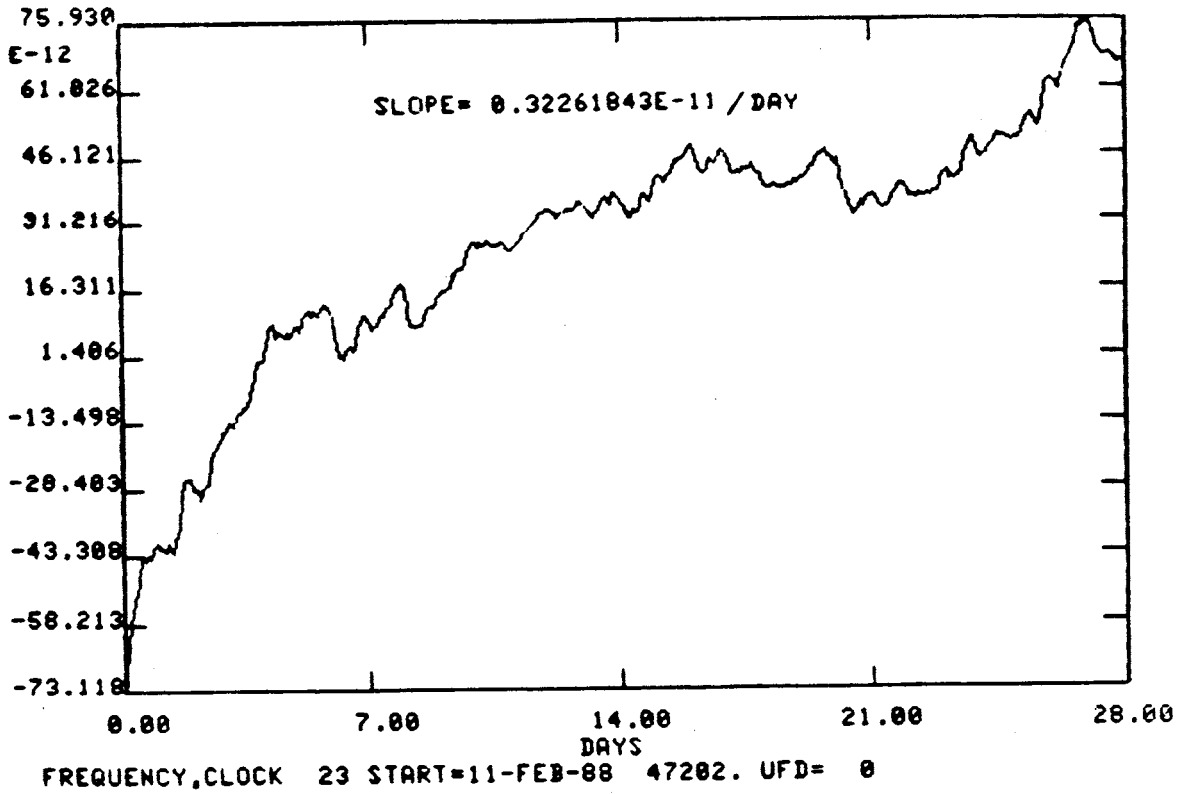


Figure 1. Frequency offset from the NBS time scale for a 5 MHz oscillator (clock 23) controlled by a fifth-overtone, AT-cut resonator of traditional design.

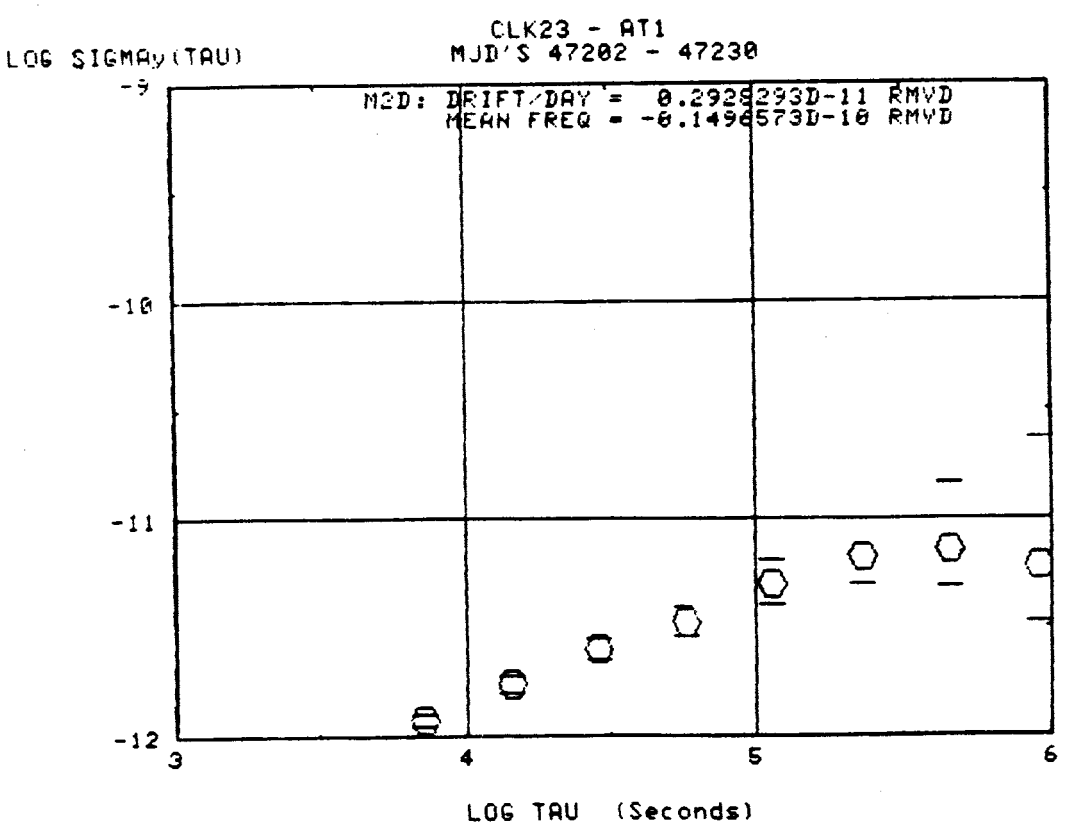


Figure 2. Fractional frequency stability, $\sigma_y(\tau)$, as a function of measurement time for clock 23 with the drift removed. The noise is nearly pure random walk frequency.

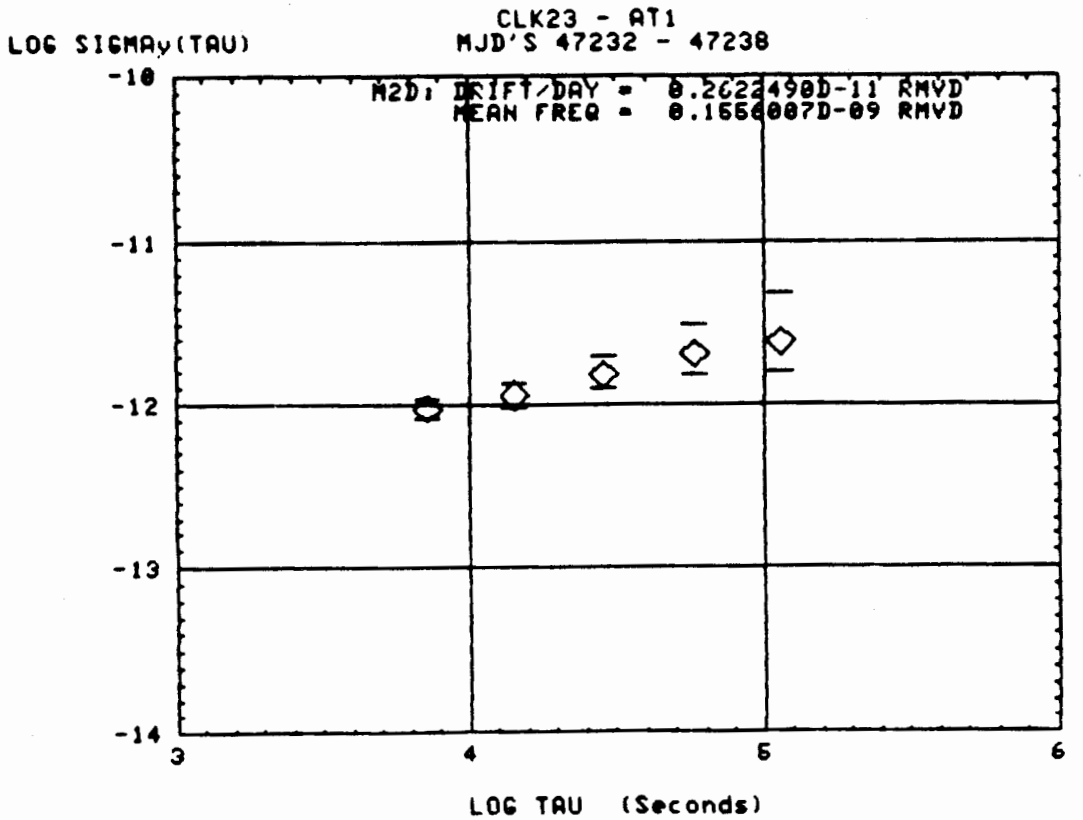


Figure 3. Fractional frequency stability, $\sigma_y(\tau)$, of clock 23 after it was installed inside a chamber to control the atmospheric pressure and humidity surrounding it.

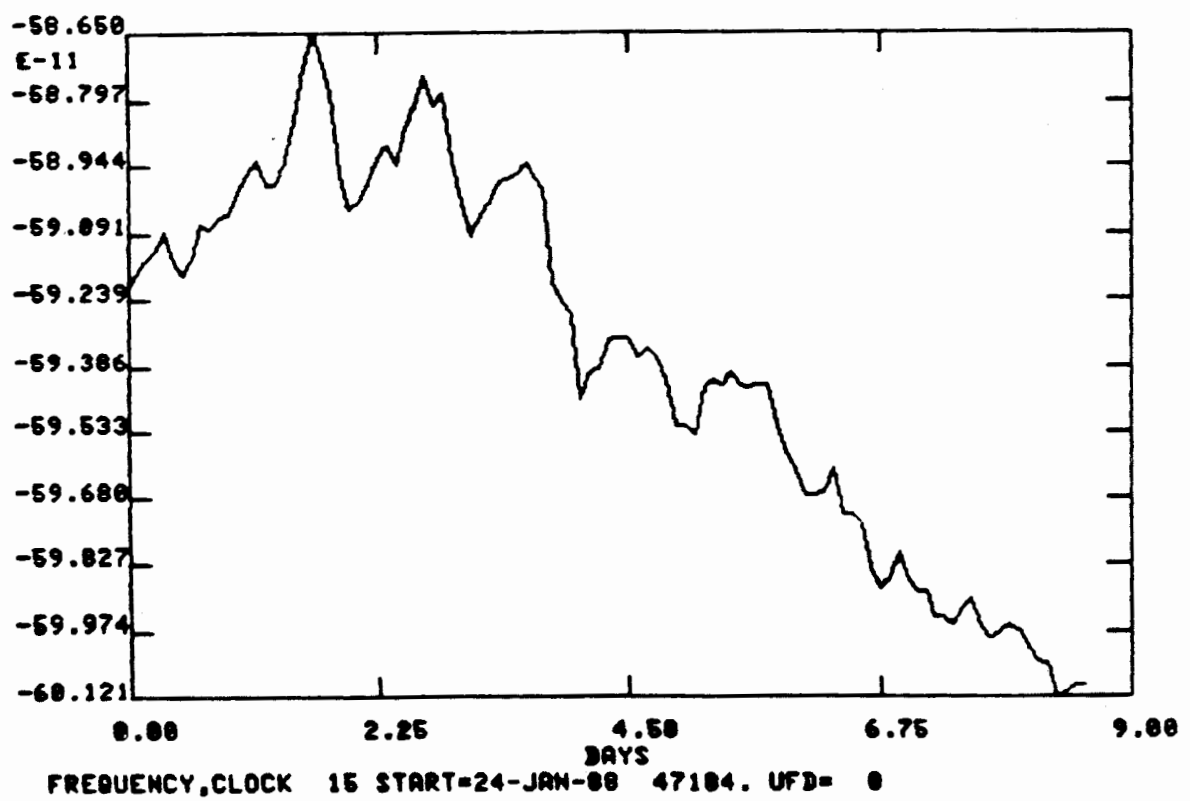


Figure 4. Frequency offset from the NBS time scale for a 5 MHz oscillator (clock 15) controlled by a fifth-overtone, AT-cut resonator of the BVA style [8].

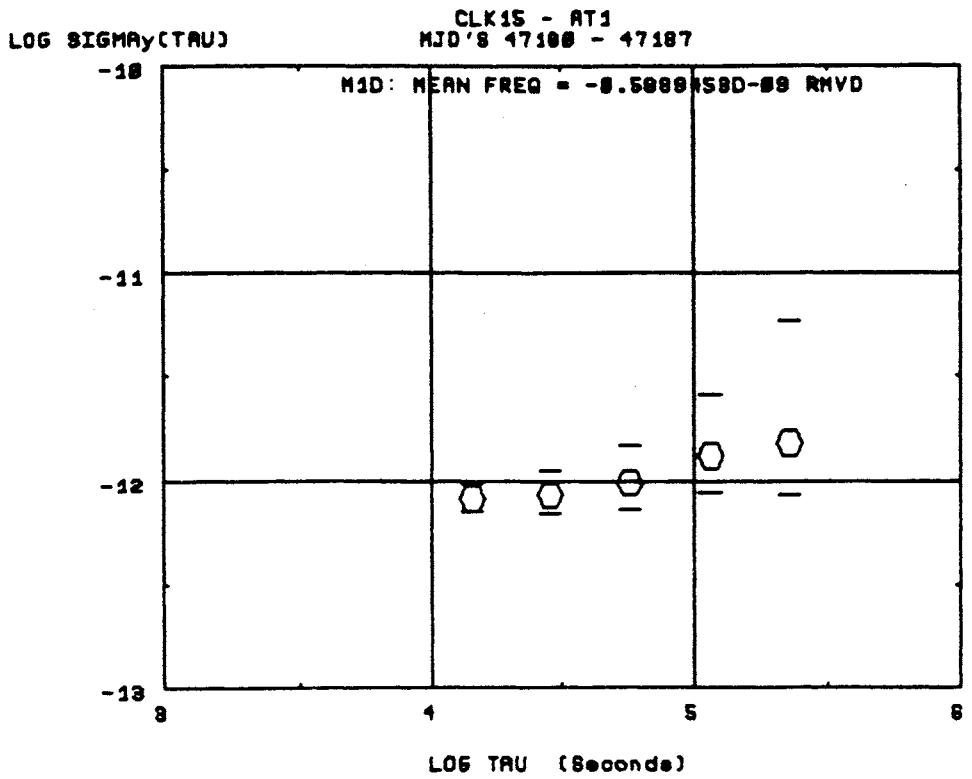


Figure 5. Fractional frequency stability, $\sigma_y(\tau)$, of clock 15 as a function of measurement time with the drift removed. The short term frequency stability is the same as that shown in figure 7.

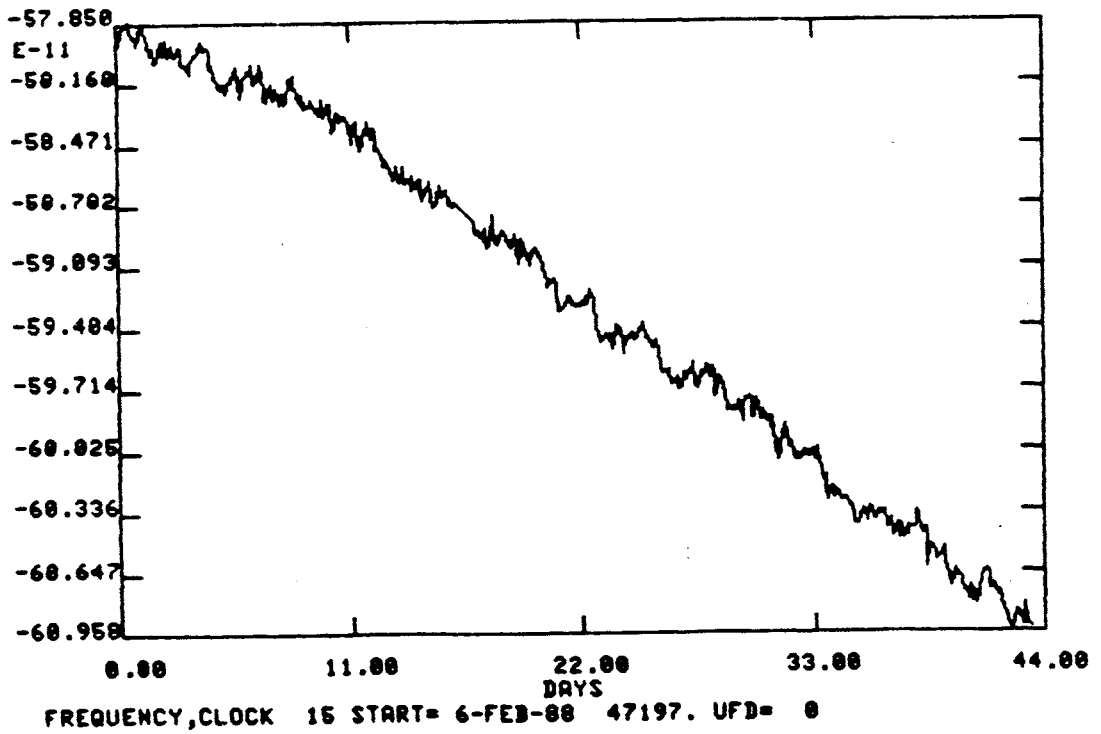


Figure 6. Frequency offset of clock 15 after installing it inside a simple chamber to control the atmospheric pressure and humidity surrounding it, and adding temperature control to approximately ± 0.1 K.

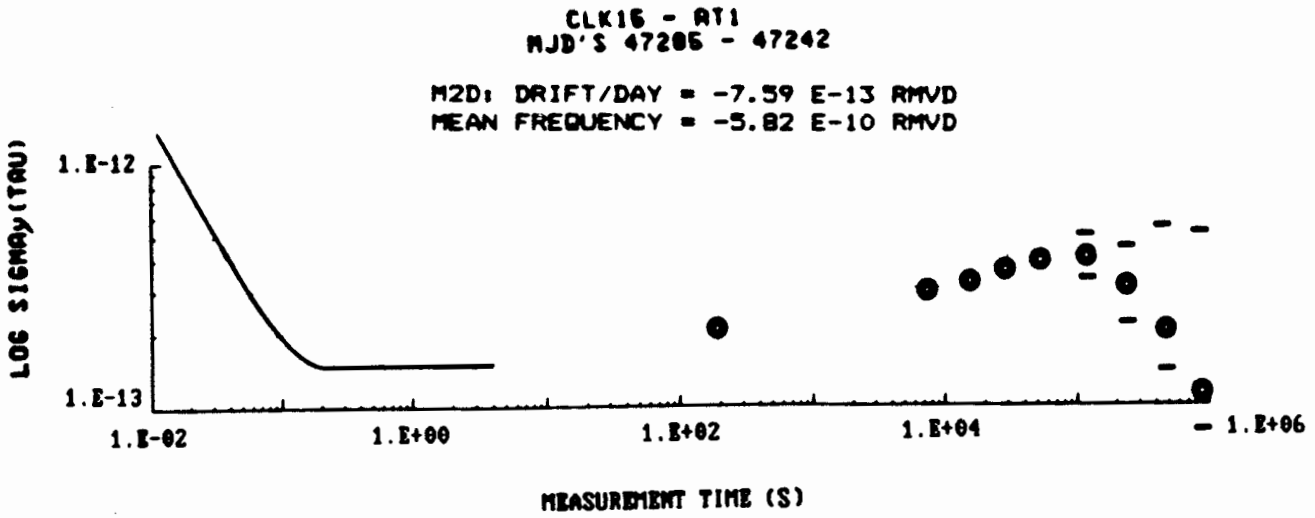


Figure 7. Fractional frequency stability, $\sigma_y(\tau)$, of clock 15 inside the chamber and additional temperature control to approximately ± 0.1 K with the drift removed. The short term frequency stability in a measurement bandwidth of 100 Hz is also indicated.

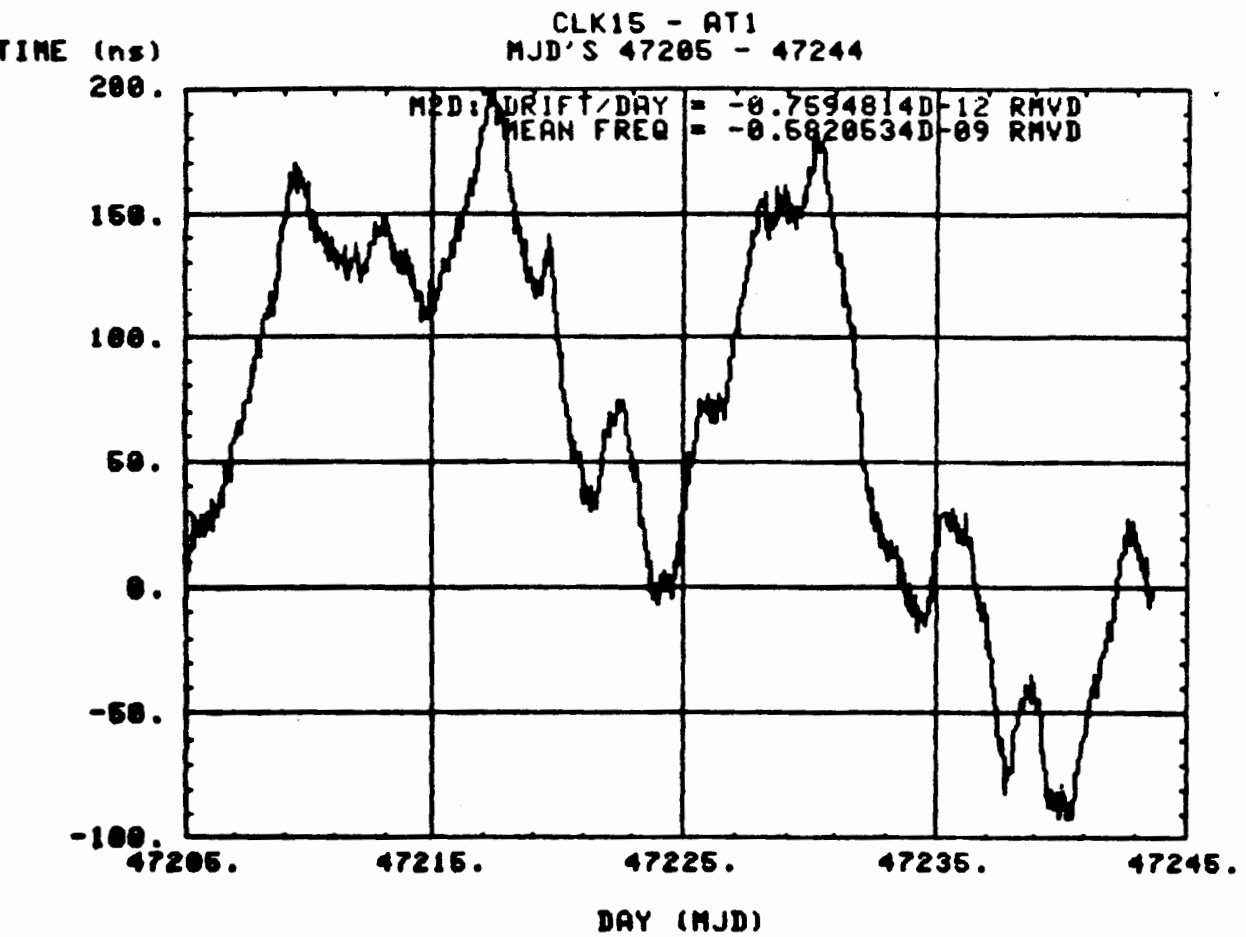


Figure 8. Time residuals of clock 15 inside the chamber after removing a 2nd difference as an approximation to the frequency drift.

RO-04

S A W

1. Novel Quartz Cuts for SAW Devices
   D.E. Bower, T.N. Oliver
   STC Components, Harlow - England

2. Aging in SAW Oscillators
   A. D'Almeida, P. Leblois, D. Hauhen
   CNRS, Besançon - France
   S. Calisti, J.Y. Duquesnoy, L. Penavaire
   Thomson-Sintra ASMDTAS, Valbonne - France

3. An Analysis of the Normal Acceleration Sensitivity of ST-Cut Quartz Surface Wave Resonators with the Substrate Extending Beyond the Supports
   H.F. Tiersten, D.V. Shick
   Rensselaer Polytechnic Institute Troy, New-York - USA

4. Bulk Waves Excited into an Anisotropic Half Space by a Metallic Interdigital Transducer
   Y. Zhang, M. Planat
   CNRS, Besançon - France

5. Accoustic Surface Transverse Wave Filters at 2 to 2.8 GHz for Oscillators and Timing Recovery
   B. Fleischmann, G. Riha, H.W. Woerz
   Simens AG, Munich - FRG

2 ND EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 16 - 17 - 18 MARCH 1988

Session RO-04 Number 1

NOVEL QUARTZ CUTS FOR SAW DEVICES

D. E. BOWER and T. N. OLIVER

STC Components
Quartz Crystal Division
Harlow, England

Abstract

We report on the electrical performance over temperature of SAW two-port resonators fabricated on a range of Quartz cuts. The LST cuts described by Shimizu, although frequency-stable over temperature, show significant changes in loss and device Q which limits their practical application. The strain-sensitive FST cut reported by Hauden et al. is well suited to sensor applications such as a SAW Accelerometer.

Introduction

The great majority of practical SAW devices are fabricated on singly-rotated cuts of Quartz oriented between ST and AT. These cuts all have a zero first-order temperature coefficient (TCF), a reasonable coupling factor (k^2) and a zero power-flow angle.

Many workers have searched for Quartz orientations with improved SAW characteristics [1 - 10]. They have found no cuts which combine a zero first-order TCF and a coupling factor significantly improved over ST. There are many cuts with improved second-order TCF, but these typically have a poor power-flow angle which degrades SAW performance. Moreover, most of these cuts are doubly-rotated, which imposes severe difficulties during the manufacture and X-ray alignment of the Quartz substrates.

The theoretical results are sensitive to the precise values of the Quartz physical constants inserted in the calculations, as well as the film overlay characteristics. A new set of self-consistant physical constants has been reported recently [11] which may alter some of the conclusions derived from older work.

More recently Shimizu [10] has reported on leaky-surface-wave (LST) cuts with excellent theoretical temperature performance.

Objectives

We aimed to evaluate a range of Quartz cuts for suitability for SAW resonator fabrication. These narrow-band devices are commonly employed as the stabilising element in oscillators and for receiver front-end filtering at both VHF and UHF frequencies.

Figure 1 lists the orientations we employed for the work and also shows the angle conventions employed. The FST cut described by Hauden [5] is

predicted to be very sensitive to strain and would be useful for a sensor application such as the SAW Accelerometer described in [12].

Device Fabrication and Testing

We used a test mask containing two-port SAW-resonator designs with either single or split-finger transducers. The main design parameters :

Design wavelength (lamda)	:	12 um.
Transducer/Mirror Aperture	:	50 lamda
Transducer Lengths	:	30 lamda
Mirror type/Length	:	518 shorted metal strips
Metallisation	:	1500 A Aluminium

A number of devices were fabricated on each of the cuts listed in Figure 1. We tested the electrical responses in an un-tuned 50 ohm jig and several parts were also tested over the temperature band -20 to +60 deg C.

Results

Typical electrical responses for each of the cuts tested are shown in Figures 2,3 and 4. As expected, the ST and AT results show a clean, single resonance peak with Q (centre frequency/3dB bandwidth) around 12000. Both single and split-finger transducer structures show very similar results.

The LST results (74.33 and 75.00 deg cuts) shown in Figure 3 are more problematic. The split-finger transducer results (lower plots) show tripled resonances, and the dominant central peak has a Q around 4000 to 5000. The single-finger transducer results (upper plots) show a very distorted structure, probably due to internal reflections within the transducers, and the Q values degrade to 2000 and below.

The FST results shown in Figure 4 are broadly similar to those of the LST cuts, though the Q values of around 8000 to 10000 seen on the split-finger version are only slightly inferior to ST or AT. However, the SST electrical results, showing a depressed central resonance flanked by two dominant secondary resonances for the split-finger case and a very complex structure for the single-finger case, are not easily explained.

We also investigated the X33 cut [8] and a 75.33 degree LST variant but the electrical results were too poor to justify further work.

The frequency v. temperature performance of each of the cuts is shown in Figures 5,6 and 7. The results show the expected parabolic behaviour and the well known turnover shift between the ST and AT cuts is clearly observed. The FST cut shows a second-order TCF slightly better than ST cut, though the SST cut is significantly worse on this parameter. The parabolic behaviour of the LST cuts is in an inverted sense compared with ST, and the second order TCF values are very small for the 75.00 cut version (split-finger).

However, this very stable frequency v. temperature behaviour for LST is not the complete story : Figure 8 and 9 show plots of the resonance peak for

the 75.00 cut LST versions at temperatures around -25,25 and 75 deg C. As can be seen, the device loss and Q degrade significantly over this temperature range which would rule out many potential applications.

The FST cut also demonstrated a frequency v. strain sensitivity around 6 to 8 times that of ST cut as measured in a mass-loaded cantilever-beam system. Under similar conditions, the LST 75.00 cut showed a strain sensitivity estimated at < 0.2 times that of ST.

Conclusions

The single-rotation cuts between ST and AT orientations remain the appropriate choices for most SAW resonator requirements. The LST 75.00 cut may have limited applicability if the device Q and loss variability with temperature are not significant. The FST cut appears well matched to the SAW Accelerometer application and further work is in progress.

Acknowledgement

T. N. Oliver would like to acknowledge the support of the UK Science & Engineering Research Council and Dr. R. Brewster (Aston University).

References

1) Williams & Cho. 'Numerical Analysis of Doubly Rotated Cut SAW Devices' Proc. 1979 Ultrasonics Symp. pp 627-631

2) Newton. 'A Study of the Propagation Characteristics of the Complete Set of SAW Paths on Quartz with Zero Temperature Coefficient of Delay' Proc. 1979 Ultrasonics Symp. pp. 632-636

3) Oates & Williams. 'Effects of Temperature-Dependent Anisotropy in RAC Devices and a Cut of Quartz for a Temperature-Compensated RAC' Proc. 1979 Ultrasonics Symp. pp. 691-695

4) Lukaszek & Ballato. 'What SAW can learn from BAW' Proc. 1980 Ultrasonics Symp. pp. 173-183

5) Hauden & Theobald. 'Dynamic Thermal Sensitivity of SAW Quartz Oscillators'. Proc. 1980 Ultrasonics Symp. pp. 264-267

6) Shimizu & Yamamoto. 'SAW Propagation Characteristics of Complete Cut of Quartz and New Cuts with Zero Temperature Coefficient of Delay' Proc. 1980 Ultrasonics Symp. pp. 420-423

7) Williams et al. 'Temperature Stable SAW Devices using Doubly Rotated Cuts of Quartz' Proc. 1980 Ultrasonics Symp. pp. 429-433

8) Webster 'X-cut Quartz for Improved Surface Acoustic Wave Stability' J. Appl. Phys. Vol. 56, pp. 1540-1542 (1984)

9) Sinha. 'Propagation Characteristics of Surface Waves in Quartz and their influence on Device Performance' Proc. 1985 Ultrasonics Symp. pp. 216-227

10) Shimizu et al. 'A New Cut of Quartz with Extremely Small Temperature Coefficient by Leaky Surface Wave'
Proc. 1985 Ultrasonics Symp. pp. 233-236

11) James 'A New Measurement of the Basic Elastic and Dielectric Constants of Quartz' IEE Colloquium : Physics and Device Applications of Acoustic Waves (London 22 May 1987)

12) Bower et al. 'A High-Linearity SAW Accelerometer'
Proc. 1987 Frequency Control Symp. pp. 544-547

CUT	EULER ANGLES		
	θ	ϕ	ψ
S T	42.75	0	0
A T	35.25	0	0
S S T	-49.2	0	23
L S T	75.0	0	0
F S T	-41.5	6.34	26

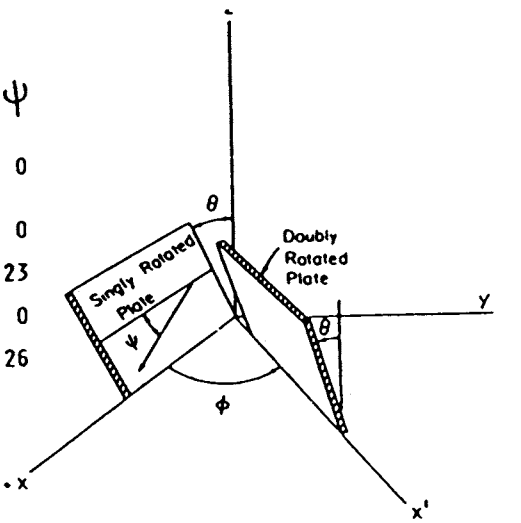


Fig.1 Quartz Cuts and Definitions

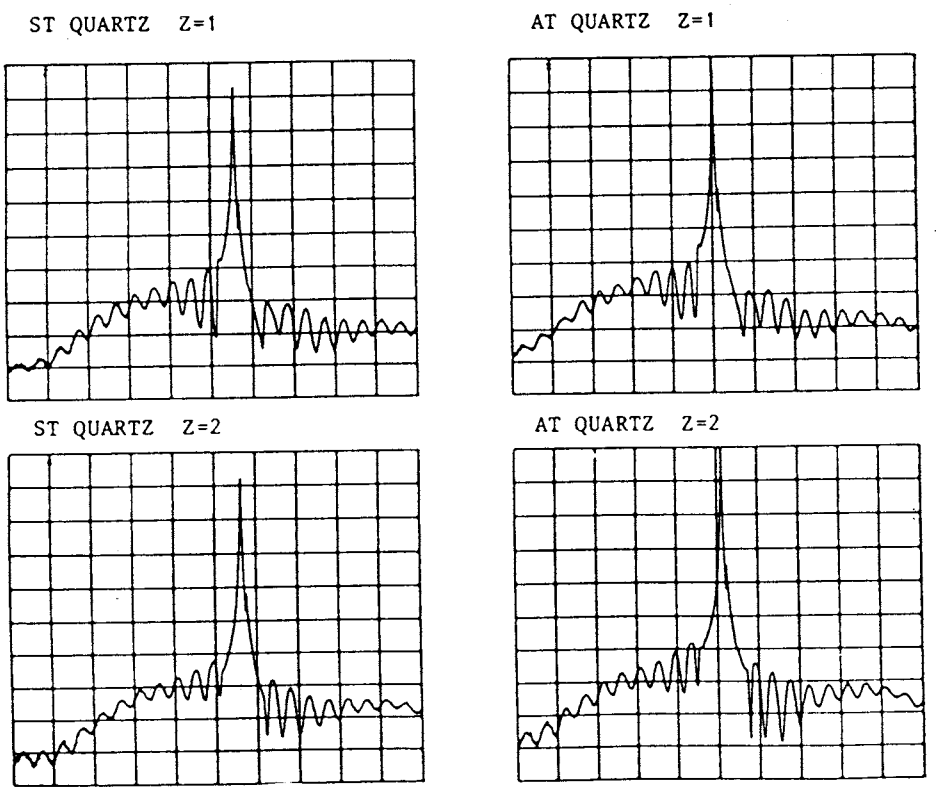


Fig.2 ST and AT Attenuation Characteristics

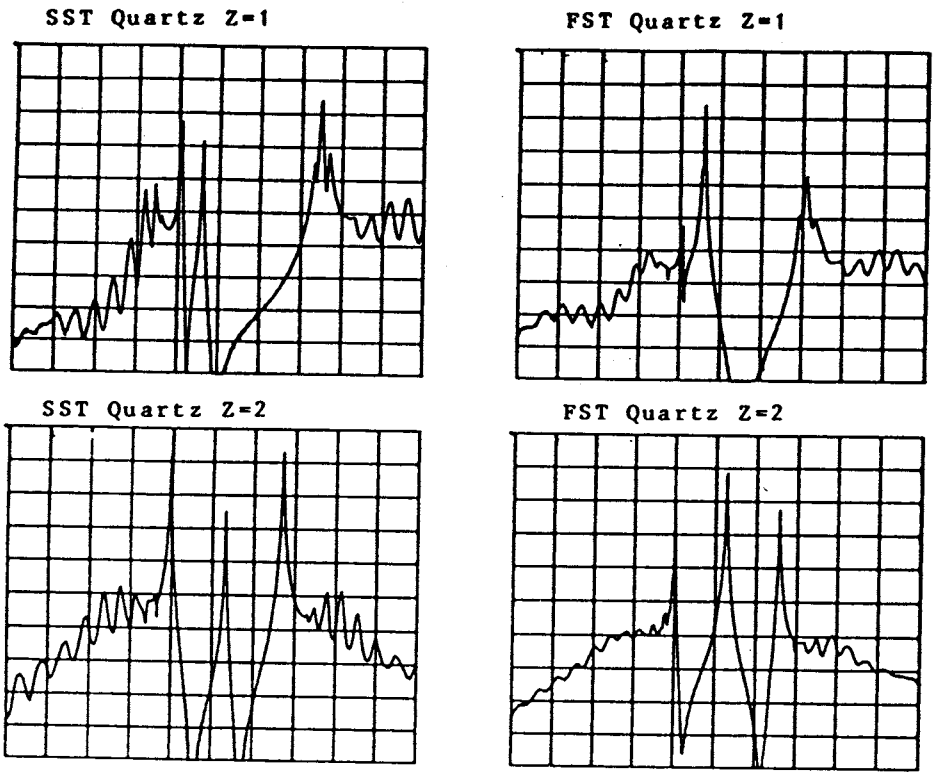


Fig. 3 SST and FST Attenuation Characteristics

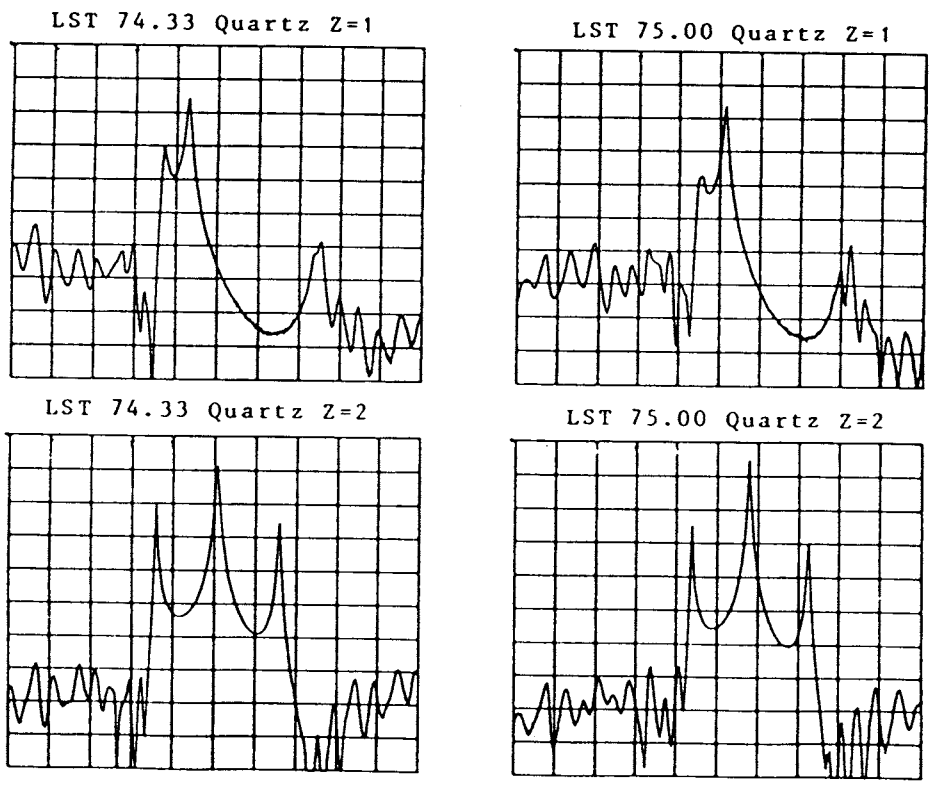


Fig. 4 LST Attenuation Characteristics

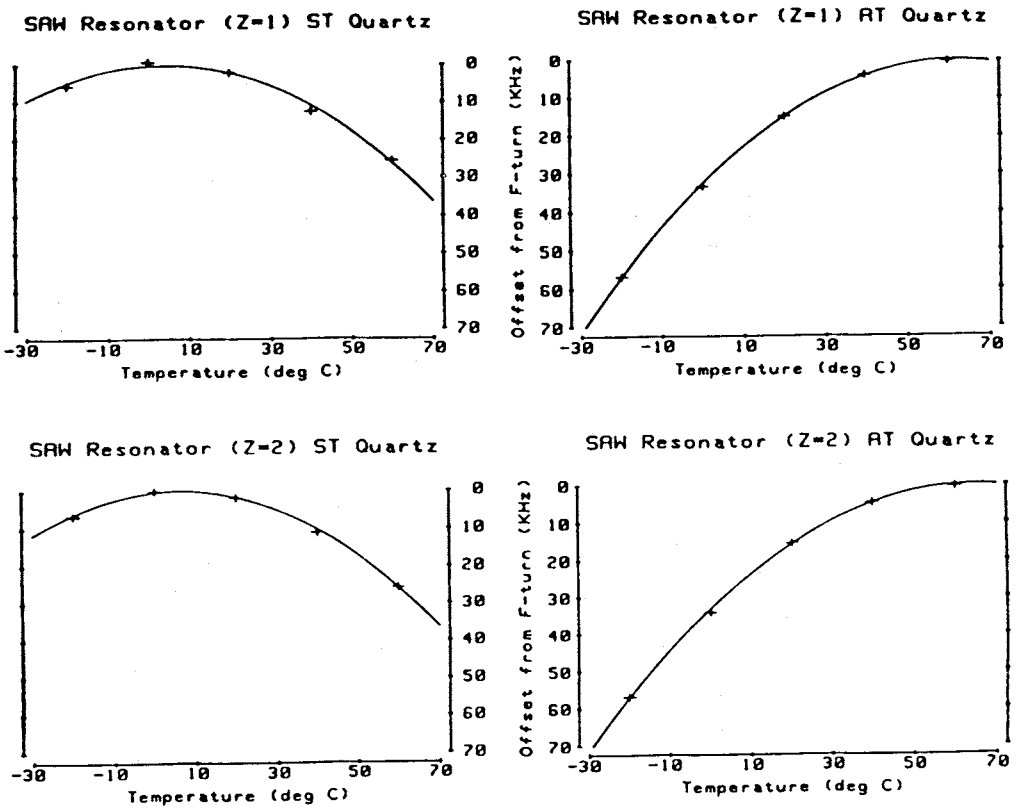


Fig. 5 ST and AT Temperature Characteristics

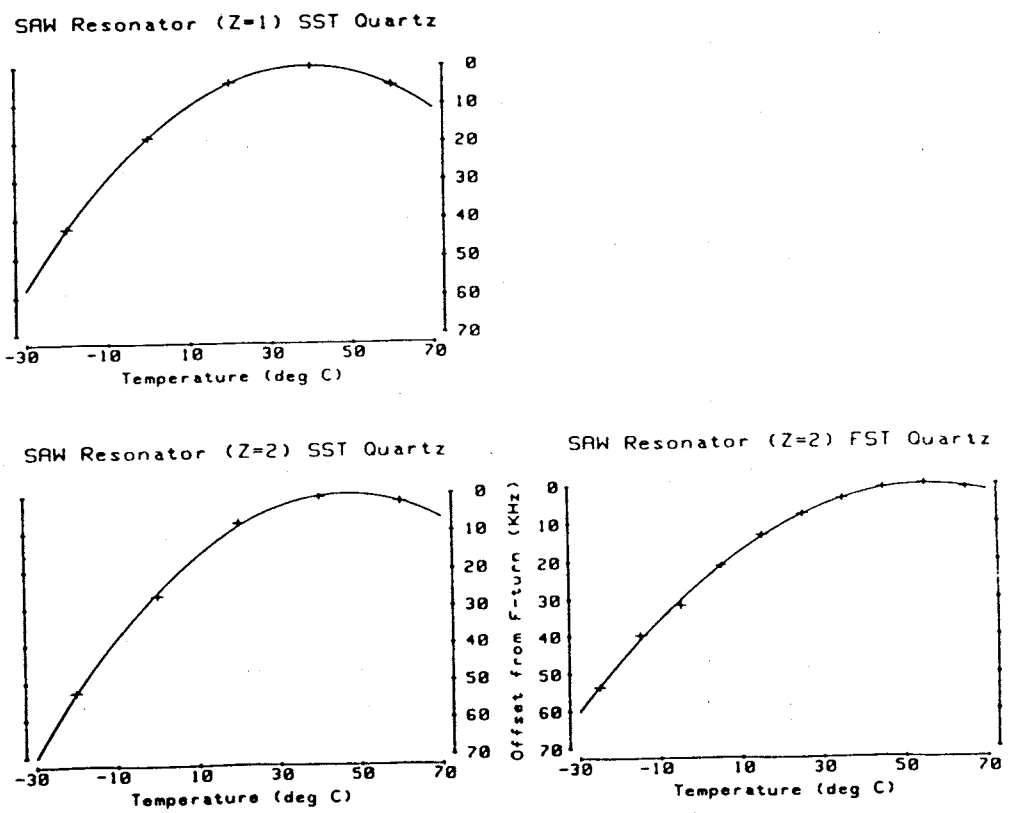


Fig. 6 SST and FST Temperature Characteristics

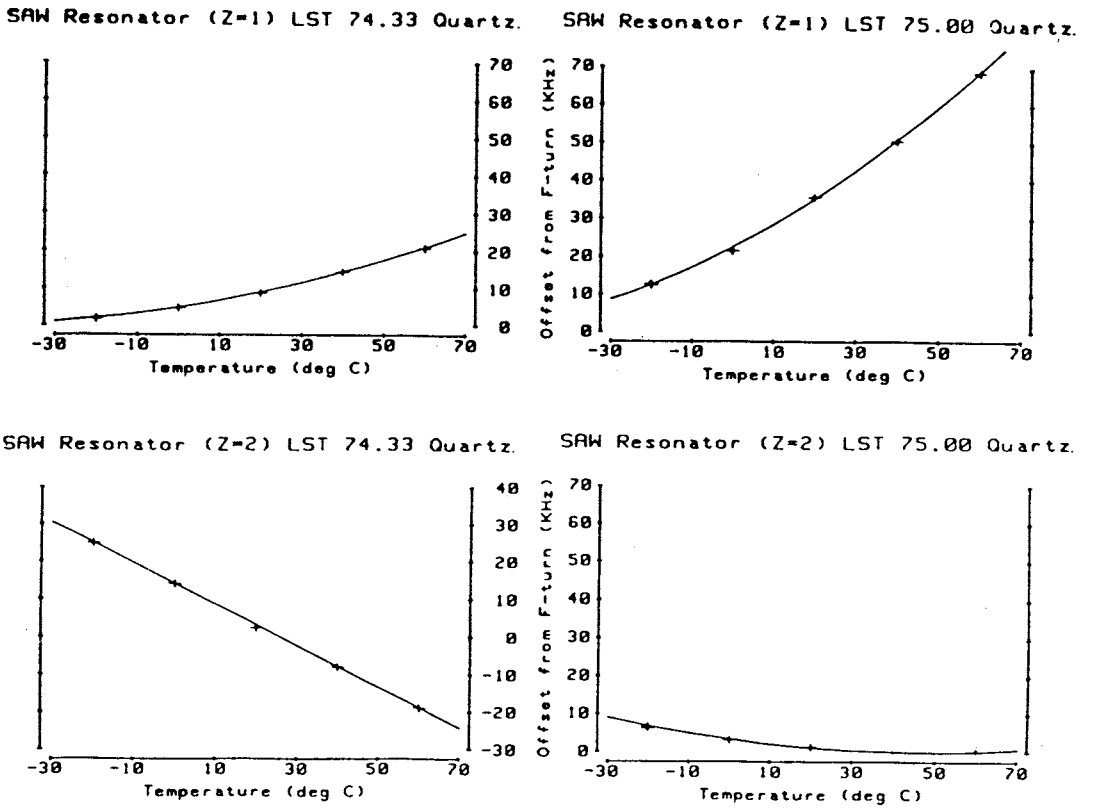


Fig. 7 LST Temperature Characteristics

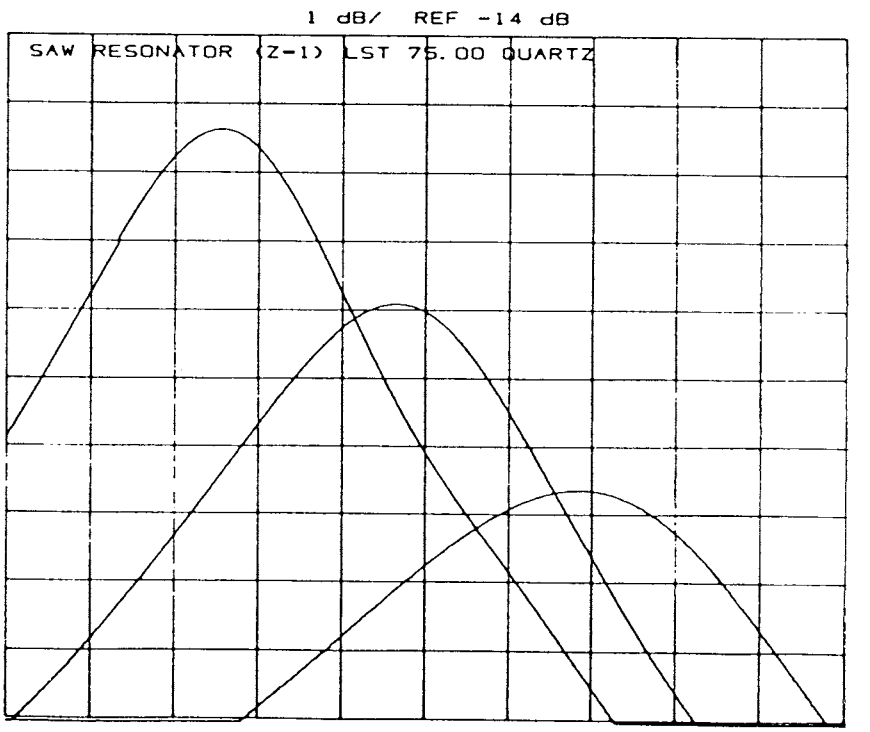


Fig. 8 LST Temperature/Loss Variation (Single Finger)

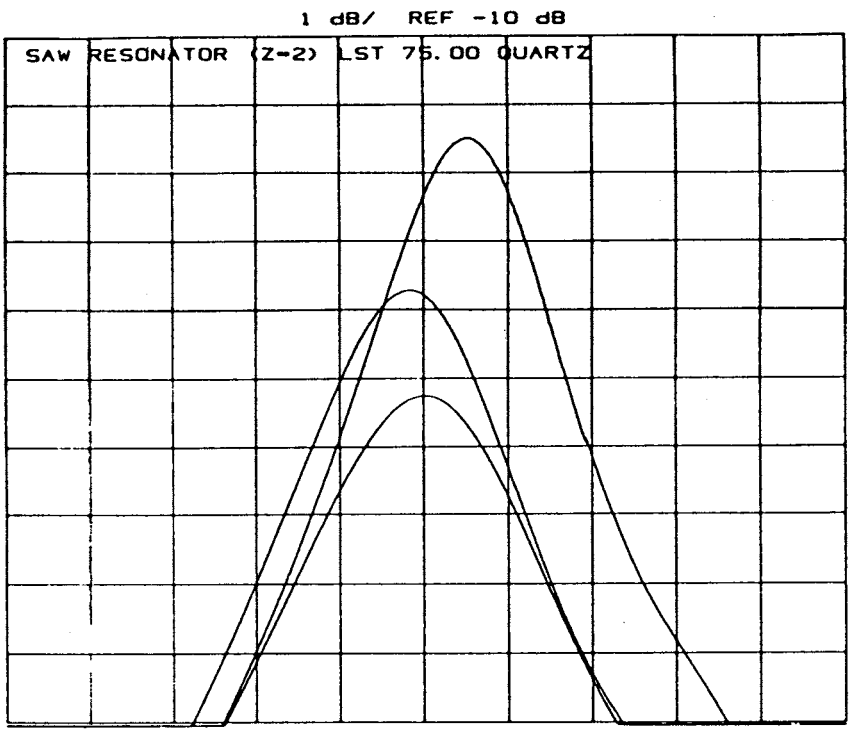


Fig. 9 LST Temperature/Loss Variation (Split Finger)

2 ND EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 16 - 17 - 18 MARCH 1988

Session RO-04 Number 2

AGING IN SAW OSCILLATORS

A. D'ALMEIDA*, P. LEBLOIS*, D. HAUDEN*,
S. CALISTI**, J.Y. DUQUESNOY**, L. PENAVAIRE**

*Laboratoire de Physique et Métrologie des Oscillateurs du CNRS
associé à l'Université de Franche-Comté-Besançon
32, avenue de l'Observatoire - 25000 Besançon - France

**Thomson-Sintra ASMDTAS,
399 route des Crêtes, B.P. 38, 06561 Valbonne Cedex - France

The SAW oscillators have a low phase noise close to their high frequency carrier [2,7]. The paper presented at the last forum [8] described different technological parameters acting on the SAW oscillator frequency stability and especially long term drift : material quality, machining residual stress, annealing, electrode metallization, resonator fastening and scaling conditions.

Short and long term frequency stability measurements have been carried out for several months on series of SAW oscillators built following different technological ways and the results obtained are presented and expected values for aging are about 1 ppm/year. So, an analysis comparing the influence of the different technological parameters on the oscillator stability is made. Finally, it will have to allow to choose the best adapted process.

This work is supported by Direction des Recherches et Etudes Techniques (France)

1. INTRODUCTION

S.A.W. quartz resonators are very interesting to UHF-oscillators with very low phase noise because excitation power applied on them can be much greater than this one applied on B.A.W. resonators before nonlinear effects appearing (15 to 20 dB more) [1]. Nevertheless, two problems are still to solve : the aging and the accelerometric sensitivity. Many studies have been done in several laboratories [2-6] and frequency stabilities better than 1 ppm/year have been reached and in the best cases measured accelerometric sensitivities are few parts of 10^{-9}/G.

In the last paper (EFTF 87) pamameters influencing aging were analyzed and technological points were underlined like stress relaxation of the substrate, metallization influence. Since, SAW resonators built following different technological processes have been concerted in oscillators to measure the frequency shifts versus time.

Fabrication and electrical characterization of SAW resonators and accelerometric sensitivities are performed at Thomson Sintra DTAS (Valbonne, France) ; oscillators and aging measurements are performed at LPMO (Besançon, France).

2. SAW RESONATORS

The SAW resonators used in the study are two ports grooved reflectors resonators (2 transducers in the cavity) at 320 MHz. Each reflector has 700 grooves 2.5 µm wide with a periodicity of 5 µm. The space between the reflectors is 100 λ_o and the transducer length is 45 λ_o and the acoustic beam aperture is 100 λ_o (λ_o is the acoustic wavelength).

They are realized on quartz crystal substrate with orientation $\theta = 36^\circ$.

Transducers are metallized in Aℓ-Ca 2 % with thickness of 700 Å obtained by lift-off technics. The grooves are etched by means of reactive ionic etching. The resonators used in a case closed under N_2-He 10 % gas.

Different technological parameters

Previous technological investigations were a stress relaxation of polishing and metallization by means of bakings and the processes used involves to one baking before metallization (1 H. at 400 °C) and one baking after metallization (1 H. at 350 °C) under $N_2 + H_2$ 10 % gas. Moreover, the Aℓ-Ca 2 % metal layer is deposited on the substrate without Cr thin layer. The influence on aging of that layer was not proved but chromium diffusion in aluminium increases electrical resistivity. Finally, authors have underlined the influence of that layer on the flicker noise level [7].

In order to look at the process influence on the aging values, 5 different ways was used to built SAW resonators.

L_1 : Cr layer, no baking (3236)

L_1' : no Cr, no baking (3336)

L_2 : no Cr, first baking (3436)

L_3 : L_2 + second baking (3636)

L_4 : L_3 + Ti-Pt-Au layer on connection pads (1104).

3. ELECTRICAL CHARACTERISTICS

For each process, several SAW resonators were electrically characterized : insertion losses, loaded-Q factor, unloaded-Q factor, turnover point.

Results are summarized in Table I.

One can see that unloaded-Q factor of L_4 is 10 % greater than for L_3 and 20 % greater than for L_1. The improvement of the Q-factor is easily explained by the decreasing of the resistive losses on the connection pads. SAW resonators are built from standard quality quartz-crystal (about 2.10^6) and corresponding theoretical calculation for unloaded Q factor at 320 MHz gives 17 200.

Frequency-temperature characteristics of SAW resonators were plotted and any difference is notable between the processes. All turn over points are arround 62 $^{\circ}$C and the measured second order temperature coefficient is equal to $-29.10^{-9}/K^2$.

In brief, if bakings and the Cr layer cancellation do not change the electrical characteristics, the gold overlayer of pads improves the SAW resonator Q-factors.

4. VIBRATION SENSITIVITIES

In consideration of the resonator weight, the mechanical vibration filtering must be performed by a special mounting. Figure 1 presents the two-dimensional mounting where the SAW resonator is sticked directly on the case bottom. If the sensitivities along X-axis and Z-axis are generally satisfactory ($\sim 10^{-9}/G$), G-sensitivity along Y-axis is 10 times longer according to the cases and the glue. Moreover, the direct mounting on the case is disadvantageous versus aging and temperature control (temperature gradients and thermoelastic effects).

A mounting has been calculated to improve the Y-axis G-sensitivity. Figure 2 shows accelerometric sensitivities along the three axes as a function of the vibration frequency.

5. S.A.W. RESONATOR-OSCILLATOR

Temperature controlled oscillators have been built with S.A.W. resonators under atmospheric pressure at 320 MHz. The oscillating loop consists in the resonator connected with an integrated amplifier and a directional coupler. The output signal is amplified by an output amplifier (40 dB) given an isolation coefficient better than 60 dB [8]. Frequency stabilities are measured through a multiplexing system.

6. ACQUISITION AND SIGNAL PROCESSING SET-UP

The scheme of the acquisition and signal processing built in is represented on Figure 3. Every S.A.W. oscillators are connected to a multiplexer 72 to 4 lines having a frequency bandwidth of 500 MHz with a rejection better than 50 dB. Moreover, a temperature sensor, a pressure sensor and an hygrometric sensor enable to measure simultaneously the surrounding conditions of oscillators. Physical quantities and oscillator output signals parameters (amplitude and frequency) are measured.

Measured values are stored in the computer to be processed before restitued as graphs or printed results.

7. S.A.W. OSCILLATOR AGING AND SHORT TERM STABILITY

Amplitude level of oscillators is controlled permanently and their frequency is accounted twice a day.

Figure 4 represents the relative frequency shifts (in ppm) versus time up to 35 weeks, where the time origin for every oscillators is translated to the same point.

Table II summarizes the characteristics and the dates when oscillators were starting to run.

Looking more in details to each technological series, figures 5, 6 and 7 show the aging of oscillators of L_1', L_3, $L_1(d)$ and $L_4(f,k)$.

Short term stabilities are measured using the phase-bridge method [9, 10] where a very good synthesizer is used as frequency reference. The relative frequency stability $\sigma(\tau)$ is plotted in the temporal domain for τ values from 0.1 sec to 100 sec.

Table III summarizes the frequency shifts obtained on the two last weeks (22/2/88 to 7/3/88), the power in the oscillating loops and the short term stability $\sigma(\tau)$ measured over 1 sec.

Aging analysis elements

1. The aging behavior of S.A.W. resonator oscillators seems related to the baking conditions and the electronic circuit. According to the baking conditions, series L_2 (b) and L_1' (a and e) have a lower aging than other oscillators. But resonator h (L_2) and g (L_1') have a larger aging. In these cases the electronic circuit thickness was changed (.8 mm against 1.6 mm before). On figure 4, we can also notice that aging is generally larger when oscillators are starting to run for every oscillators built with circuit at .8 mm of thickness. But after 12 weeks, aging becomes on the same order for every oscillators (Table III).
2. Moreover, the first results for sticked substrates and connections show a very large aging may be due to the glue degasing.
3. It is very difficult to correlate clearly oscillation powers and aging.
4. The mean value of aging is separated from the slow frequency fluctuations [11], then correlation can be made with the environmental parameter fluctuations. Figure 8 shows clearly that oscillators a (L_1'-type) and c (L_3-type) have frequency fluctuations following strong temperature variations (for instance at weeks 6, 11, 22, 24, ...). At the opposite, oscillator (d) is not correlated because it is more temperature insulated by a regular thermal oven. Except this one, other oscillators have a

temperature control included within the oscillator case (integrated circuit temperature controller) and its time constant seems a little too small.

Short term stability

Table III gives also short term stabilities which are generally better than 1.10^{-10} over 1 sec. Particularly, short term stability of the last type L_4 is very good ($\sim 2.10^{-11}$/1 sec).

In conclusion, after 35 weeks, aging analysis shows that baking conditions must be studied deeply to conclude definitively. Actual oscillator is around 2.10^{-6}/year. Any studied resonator was closed under vacuum, yet.

Two steps will be explored : resonator in regular case and all quartz mounting and expected aging will be better than 1.10^{-6}/year.

REFERENCES

[1] M. Planat, J.J. Gagnepain, C. Lardat and L. Penavaire, "Nonlinear characteristics of SAW grooved resonators", Proc. of the 1980 IEEE Ultrasonics Symposium.

[2] W.J. Tanski, R.A. St Cyr, P.G. Dragonetti and E.G. Kosco, "A radar system application of a 840 MHz SAW resonator stabilized oscillator", IEEE Trans. Sonics Ultrason., vol. SU-28, n° 3, may 1981.

[3] G.K. Montress, T.E. Parker and M.J. Loboda, "Extremely low phase noise SAW resonator oscillator design and performance", Proc. of the 1987 Ultrasonics Symposium.

[4] T.E. Parker, "Analysis of aging data on SAW oscillators", Proc. of Annual Frequency Control Symposium 1980.

[5] T.E. Parker, "Random and systematic contributions to long term frequency stability in SAW oscillators", Proc. of the 1983 Ultrasonics Symposium.

[6] W.R. Shreves, "Active aging of SAW resonators", Proc. of the 1980 Ultrasonics Symposium.

[7] W.J. Tanski, "The influence of a chrome film bonding layer on SAW resonator performance", Proc. of the 1985 Ultrasonics Symposium.

[8] S. Calisti, Ph. Defranould, D. Hauden, "Influence des paramètres technologiques sur la stabilité en fréquence des résonateurs à ondes de surface", 1er Forum Européen Temps-Fréquence, Besançon, mars 1987.

[9] J.J. Gagnepain, "Stabilité et pureté spectrale : concepts", Besançon, mars 1985.

[10] J. Groslambert, "Méthodes expérimentales de mesures", Besançon, mars 1985.

[11] T.E. Parker, "Very long period random frequency fluctuations in SAW oscillators", Proc. of the 37th Annual Frequency Control Symposium, 1983.

Type	N°	Insertion losses	Loaded Q	Unloaded Q	Number of resonators
L_1	3236	5,54 dB	7384	14988	3
L_1'	3336	6,13 dB	7649	15307	4
L_2	3436	6,13 dB	7936	15718	6
L_3	3636	6,18 dB	8180	16156	5
L_4	1104	5,49 dB	8384	17916	3

Table I

Mean values of electrical characteristics for each resonator type

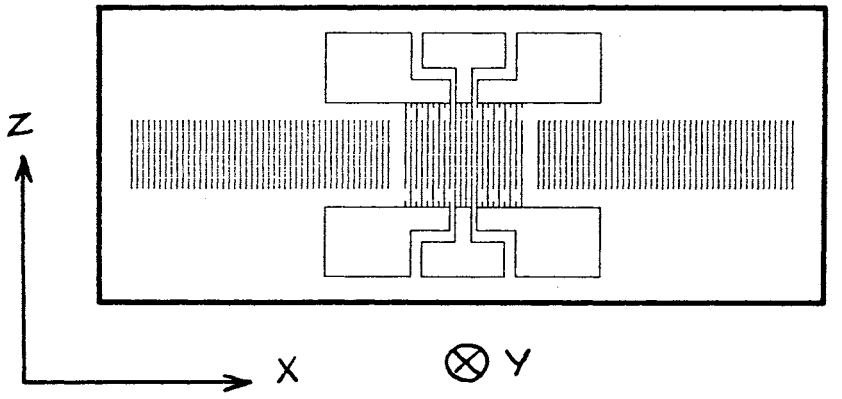


Fig. 1

Axis position for accelerometric sensitivity measurements

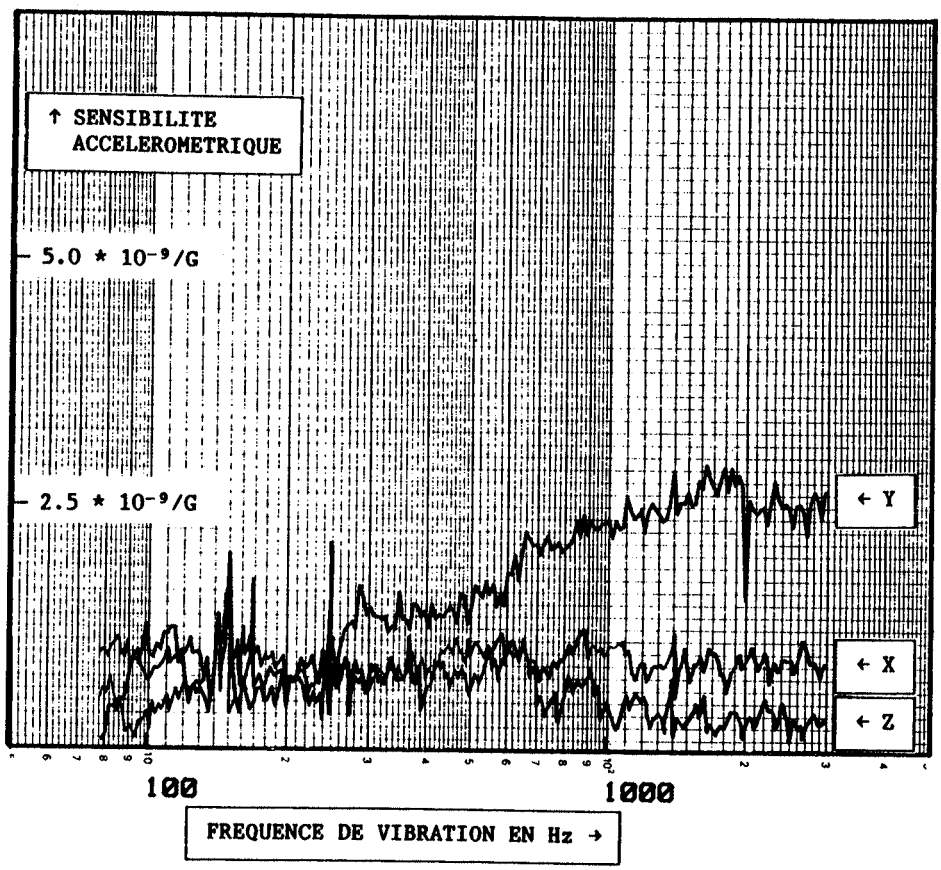


Fig. 2 : Accelerometric sensitivity of SAW resonator at 320 MHz on the 3-axis versus vibration frequency

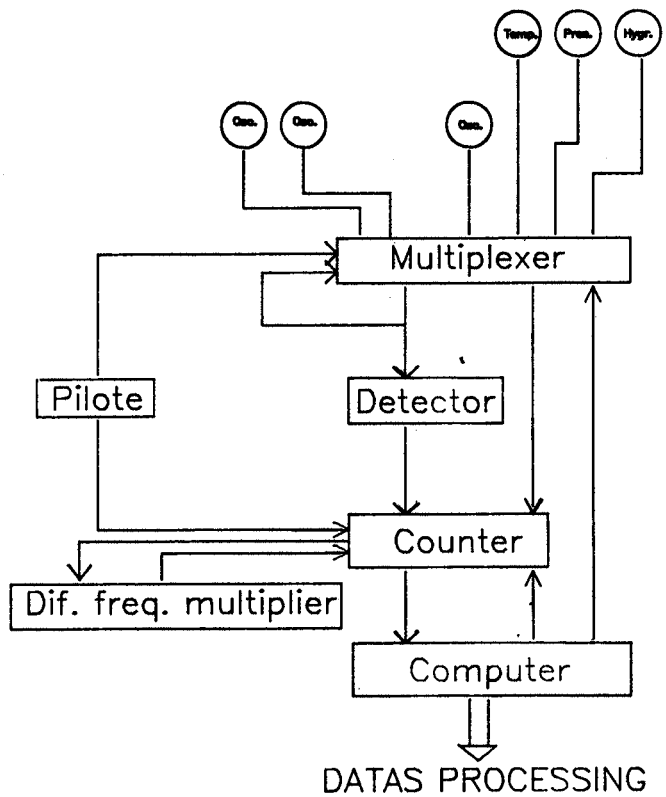


Fig. 3- Synoptic of the measuring bench

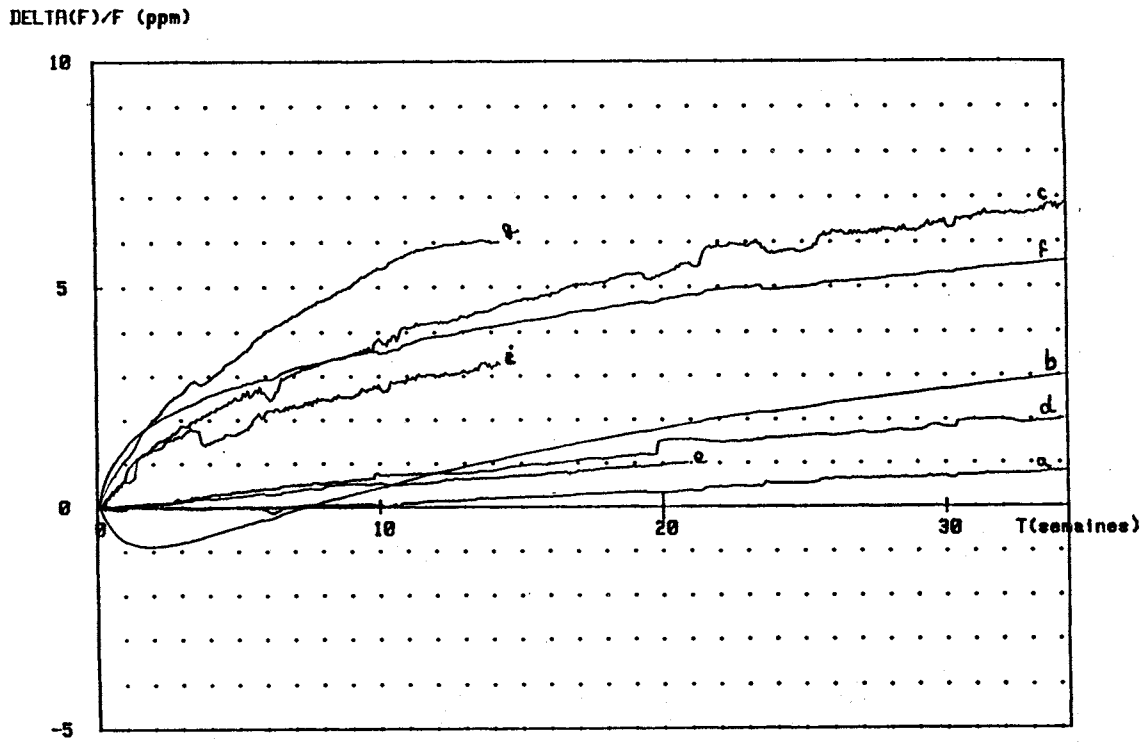


Fig. 4
Aging of SAW oscillators at 320 MHz during 35 weeks

RESONATOR REFERENCE	Type	RESONATORS Mounting	Encaps.	Connections	ELECTRONICS Circuit thickness (mm)	Transfo (y/n)	AGING STARTING DATE	LETTER ON FIGURES
3236/04	L1	free	N2+He	brazed			06/02/87	d
3436/07	L2	free	N2+He	brazed	1.6	yes	12/07/87	b
3636/03	L3	free	N2+He	brazed	1.6	yes	12/07/87	c
3336/06	L'1	free	N2+He	brazed	1.6	yes	12/07/87	a
1104/04	L4	free	N2+He	brazed	1.6	yes	12/07/87	k
3336/08	L'1	free	N2+He	brazed	1.6	yes	12/10/87	e
1104/05	L4	free	N2+He	brazed	1.6	yes	12/07/87	f
3336/07	L'1	free	N2+He	brazed preaging (1 week at 65 °C)	0.8	yes	27/11/87	g
3436/06	L2	free	N2+He	brazed	0.8	yes	27/11/87	h
3636/01	L3	free	N2+He	brazed	0.8	yes	27/11/87	j
3636/07	L3	free	N2+He	brazed	0.8	yes	27/11/87	i
1104/17	L4	sticked	N2+He	sticked + 10 cycles thermiques	1.6	yes	14/12/87	
1104/11	L4	sticked	N2+He	sticked + 10 cycles thermiques	1.6	yes	21/01/88	

Table II

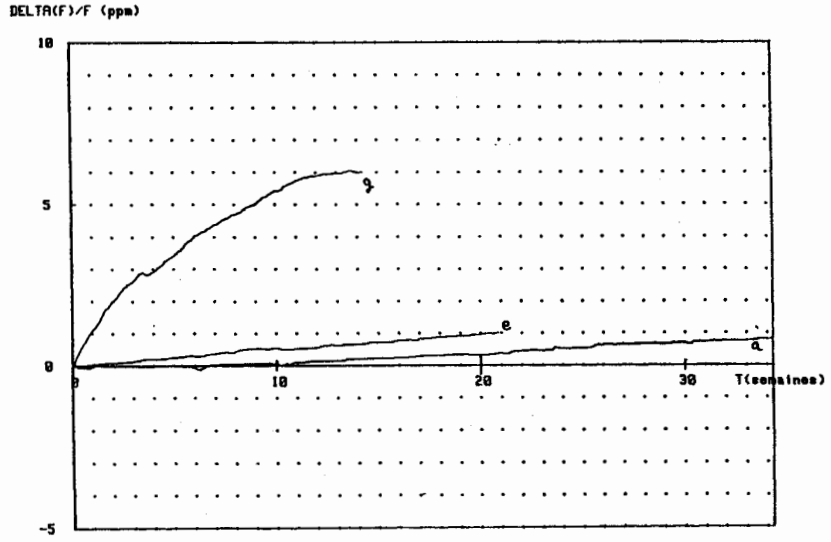


Fig. 5 : Aging of SAW oscillators type L_1'

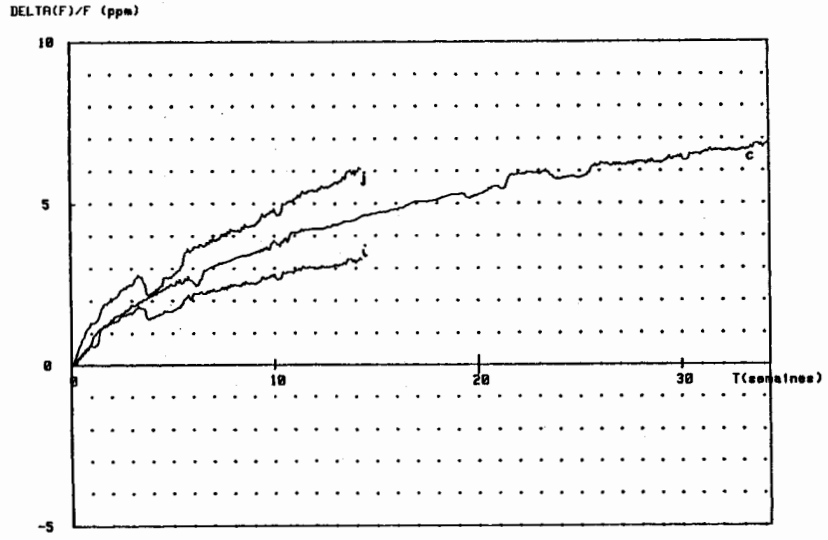


Fig. 6 : Aging of SAW oscillators type L_3

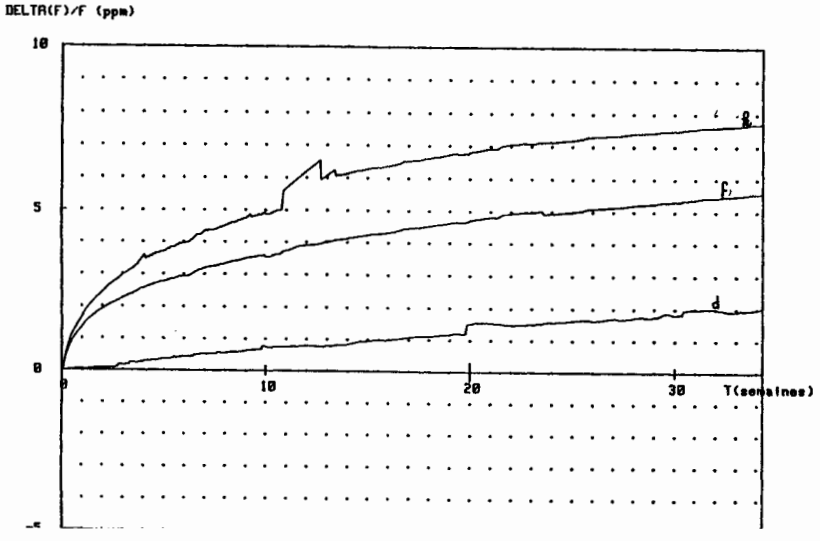


Fig. 7 : Aging of SAW oscillators type L_4 (f and k) and type L_1 (d)

RESONATOR REFERENCE	AGING STARTING DATE	DAY AGING (2 LAST WEEKS) /day	SHORT TERM STABILITY /1.s	OSCILLATOR power (dBm)	LETTERS on FIGURES
3236/04	06/02/87	12.10^{-9}	3.10^{-10}	- 4.95	d
3336/06	12/07/87	$2.6.10^{-9}$	3.10^{-11}	- 6.99	a
3336/08	12/10/87	7.10^{-9}	7.10^{-11}	-17.69	e
3336/07	27/11/87	49.10^{-9}	9.10^{-11}	- 4.31	g
3436/07	12/07/87	11.10^{-9}	6.10^{-11}	- 7.45	b
3436/06	27/11/87	12.10^{-9}	9.10^{-11}	- 3.87	
3636/03	12/07/87	12.10^{-9}	6.10^{-11}	- 5.69	c
3636/01	27/11/87	43.10^{-9}	8.10^{-11}	-10.22	j
3636/07	27/11/87	21.10^{-9}	8.10^{-11}	-23.01	i
1104/04	12/07/87	$8.7.10^{-9}$	3.10^{-11}	- 7.21	f
1104/05	12/07/87	$8.1.10^{-9}$	2.10^{-11}	- 2.60	k
1104/17	14/12/87	-110.10^{-9}	$1.5.10^{-10}$	-25.23	sticked

Table III

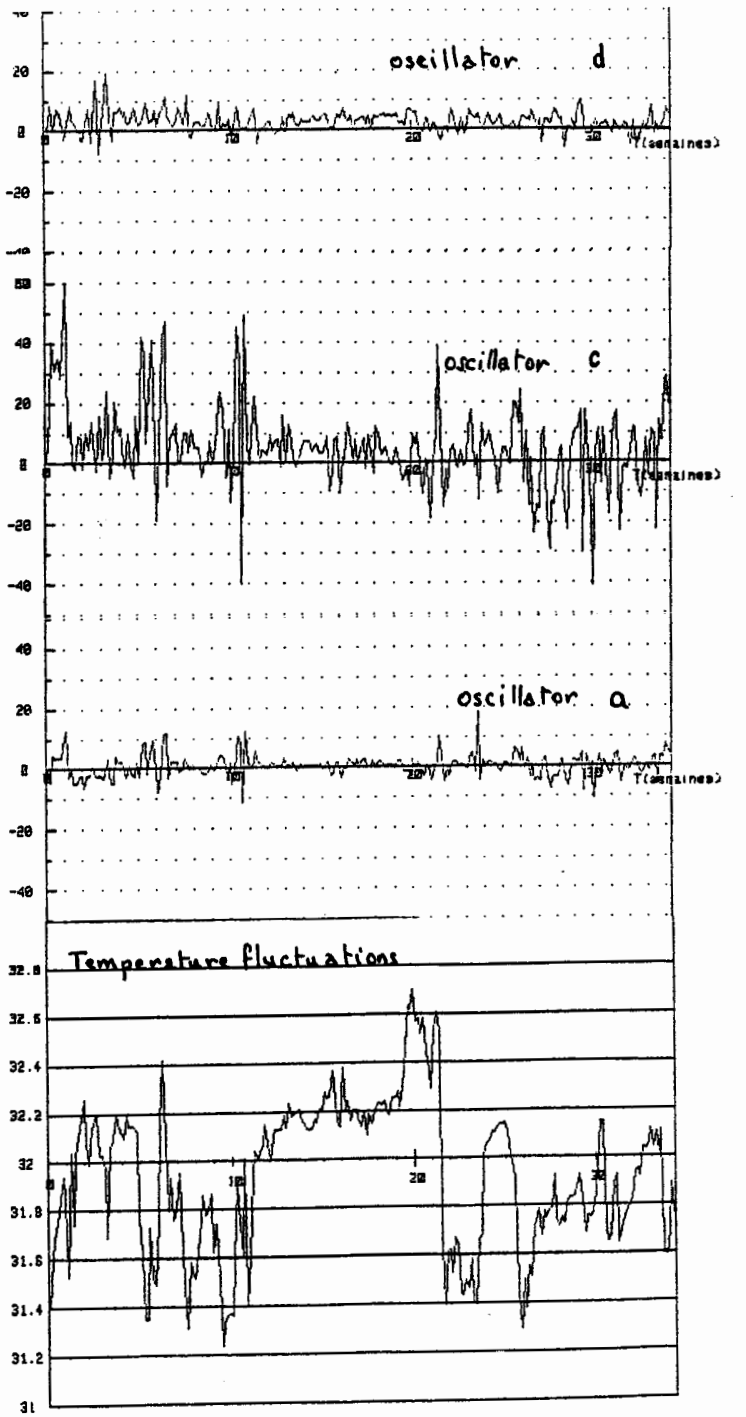


Fig. 8 : Correlation between temperature fluctuations and slow fluctuations around mean aging values

2 ND EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 16 - 17 - 18 MARCH 1988

Session RO-04 Number 3

AN ANALYSIS OF THE NORMAL ACCELERATION SENSITIVITY OF ST-CUT QUARTZ SURFACE WAVE RESONATORS WITH THE SUBSTRATE EXTENDING BEYOND THE SUPPORTS

H.F. Tiersten and D.V. Shick
Rensselaer Polytechnic Institute
Troy, New York 12180-3590, USA

ABSTRACT

An analysis of the normal acceleration sensitivity of ST-cut quartz surface wave resonators with the substrate extending beyond rectangular supports is performed. The calculated results show that for certain planar aspect ratios the sensitivity vanishes and that, as a consequence of the overhangs, the slope of the sensitivity versus aspect ratio curve is much shallower than that obtained in any earlier work. Hence, very low normal acceleration sensitivies can be obtained with this support configuration relatively easily.

1. INTRODUCTION

The sensitivity of a surface wave resonator to an acceleration field that causes cylindrical flexure about the thin dimension is quite large because the flexural deformation is greatest near the surfaces of the substrate, which is exactly where the wave is located. Nevertheless, in previous work[1,2] on the normal acceleration sensitivity of ST-cut quartz surface wave resonators supported along rectangular edges it has been shown that the sensitivity vanishes for certain planar aspect ratios. This is a result of the fact that for ST-cut quartz cylindrical flexure in the propagation direction and normal to it causes changes in frequency of opposite sign. Both the cases of simple[1] and rigid[2] supports were considered and it was shown that the slope of the sensitivity versus aspect ratio curve for the rigidly-supported case was about 1/4 of that for the simply-supported case.

In this work the configuration in which the substrate extends beyond rectangular supports is treated. This is the support configuration employed in the work at Raytheon[3,4] and it has additional parameters to control the normal acceleration sensitivity, which can be used to advantage. In the treatment the biasing deformation state is obtained by solving the equations of anisotropic static flexure, which is possible because the relatively unimportant corner regions of the configuration are ignored in the treatment. The flexural biasing deformation state is employed in an existing perturbation equation[5] along with the recently determined[6] proper continuous representation of the acoustic surface wave mode shape for arrays of reflecting grooves to determine the normal acceleration sensitivity. It is shown that in the vicinity of the zero-

crossings the slope of the normal acceleration sensitivity versus aspect ratio curve is about 1/10 of that for the case of rigid supports. This means that, for the support configuration considered in this work, very low normal acceleration sensitivities can be achieved over a reasonably wide range of aspect ratios. Recently, normal acceleration sensitivities of a number of parts in 10^{11} per g have been observed[7] with this configuration at Raytheon.

2. PERTURBATION EQUATIONS

For purely elastic nonlinearities the equation for the first perturbation of the eigenvalue obtained from the perturbation analysis[5] mentioned in the Introduction may be written in the form

$$\Delta_\mu = H_\mu / 2\omega_\mu , \quad \omega = \omega_\mu - \Delta_\mu , \tag{2.1}$$

where ω_μ and ω are the unperturbed and perturbed eigenfrequencies, respectively, and

$$H_\mu = - \int_V \hat{K}^n_{L\gamma} g^\mu_{\gamma,L} \, dV , \tag{2.2}$$

where V is the undeformed volume of the piezoelectric plate. In (2.2) g^μ_γ denotes the normalized mechanical displacement vector, and $\hat{K}^n_{L\gamma}$ denotes the portion of the Piola-Kirchhoff stress tensor resulting from the biasing state in the presence of the g^μ_γ, and is given by

$$\hat{K}^n_{L\gamma} = \hat{c}_{L\gamma M\alpha} g^\mu_{\alpha,M} , \tag{2.3}$$

where

$$\hat{c}_{L\gamma M\alpha} = T^1_{LM} \delta_{\gamma\alpha} + c_{3L\gamma M\alpha KN} E^1_{KN} + c_{2L\gamma KM} w_{\alpha,K} + c_{2LKM\alpha} w_{\gamma,K} \tag{2.4}$$

and

$$T^1_{LM} = c_{2LMKN} E^1_{KN} , \quad E^1_{KN} = \frac{1}{2} (w_{K,N} + w_{N,K}) . \tag{2.5}$$

The quantities T^1_{LM}, E^1_{KN} and w_K denote the static biasing stress, strain and displacement field, respectively. Thus, in this description the present position $\underset{\sim}{y}$ is related to the reference position $\underset{\sim}{X}$ by

$$\underset{\sim}{y}(X_L, t) = \underset{\sim}{X} + \underset{\sim}{w}(X_L) + \underset{\sim}{u}(X_L, t) . \tag{2.6}$$

The coefficients c_{2LMKN} and $c_{3L\gamma M\alpha KN}$ denote the second and third order elastic constants, respectively.

The normalized eigensolution g^μ_γ and $\hat{f}^\mu$ is defined by

$$g^\mu_\gamma = \frac{u^\mu_\gamma}{N_\mu} , \quad \hat{f}^\mu = \frac{\tilde{\varphi}^\mu}{N_\mu} , \quad N^2_\mu = \int \rho u^\mu_\gamma u^\mu_\gamma \, dV , \tag{2.7}$$

where u_γ^μ and $\tilde{\varphi}^\mu$ are the mechanical displacement and electric potential, respectively, which satisfy the equations of linear piezoelectricity

$$\tilde{K}_{L\gamma}^{\ell} = c_{L\gamma M\alpha} u_{\alpha,M} + e_{ML\gamma} \tilde{\varphi}_{,M} ,$$

$$\tilde{\mathcal{D}}_L^{\ell} = e_{LM\gamma} u_{\gamma,M} - \epsilon_{LM} \tilde{\varphi}_{,M} , \tag{2.8}$$

$$\tilde{K}_{L\gamma,L}^{\ell} = \rho \ddot{u}_\gamma , \quad \tilde{\mathcal{D}}_{L,L}^{\ell} = 0 , \tag{2.9}$$

subject to the appropriate boundary conditions, and ρ is the mass density. Equations (2.8) are the linear piezoelectric constitutive relations and (2.9) are the stress equations of motion and charge equation of electrostatics, respectively. The upper cycle notation for many dynamic variables and the capital Latin and lower case Greek index notation is being employed for consistency with Ref.5, as is the remainder of the notation in this section.

The substitution of (2.3) in (2.2) yields

$$H_\mu = - \int_V \hat{c}_{L\gamma M\alpha} g_{\alpha,M}^{\mu} g_{\gamma,L}^{\mu} \, dV . \tag{2.10}$$

Since g_α^μ denotes the normalized surface wave mode shape and from (2.4) $\hat{c}_{L\gamma M\alpha}$ depends on the biasing state, H_μ can be evaluated when the surface wave mode shape and biasing state are known.

3. EQUATIONS FOR ANISOTROPIC STATIC FLEXURE

With the aid of Mindlin's plate equations[8-10] it has been shown that the equation of anisotropic static flexure with X_2 normal to the major surfaces can be written in the form[11]

$$K_{AB,AB}^{(1)} + F_{B,B}^{(1)} + F_2^{(0)} = 0 , \tag{3.1}$$

where we have introduced the convention that A, B, C, D take the values 1 and 3 and skip 2 and

$$K_{AB}^{(1)} = \int_{-h}^{h} X_2 K_{AB} \, dX_2 , \quad F_B^{(1)} = h[K_{2B}(h) + K_{2B}(-h)] , \tag{3.2}$$

and in this instance

$$F_B^{(1)} = 0 , \quad F_2^{(0)} = K_{22}(h) - K_{22}(-h) - 2h\rho a_2 = - 2h\rho a_2 , \tag{3.3}$$

since K_{2K} vanishes on the major surfaces and where a_2 is the acceleration in the X_2-direction. From Eqs.(4.25) of Ref.11 the constitutive equations for the stress-resultants take the form

$$K_{AB}^{(1)} = \frac{2}{3} h^3 \gamma_{ABCD} E_{CD}^{(1)} , \tag{3.4}$$

where Voigt's anisotropic plate elastic constants are given by

$$\gamma_{RS} = c_{RS} - c_{RW}c^{-1}_{WV}c_{VS}, \quad R,S = 1,3,5; \; W,V = 2,4,6, \tag{3.5}$$

in the compressed notation, and where we have introduced the scheme shown. The plate strains $E^{(1)}_{CD}$ in (3.4) are given by

$$E^{(1)}_{CD} = \frac{1}{2}\left(w^{(1)}_{C,D} + w^{(1)}_{D,C}\right), \tag{3.6}$$

and from the relaxation of the stress resultants $K^{(1)}_{2L}$, the vanishing of the plate shear strains $E^{(0)}_{2A}$ and (3.6) we have the respective relations

$$E^{(1)}_{W} = -c^{-1}_{WV}c_{VS}E^{(1)}_{S}, \quad w^{(1)}_{A} = -w^{(0)}_{2,A}, \quad E^{(1)}_{CD} = -w^{(0)}_{2,CD}, \tag{3.7}$$

which, respectively, are given in Eqs.(4.24), (3.34) and (3.37) of Ref.11. Since from Eqs.(3.31) of Ref.12 we have $E^{(0)}_{KL} = 0$, when the plate deflection $w^{(0)}_{2}$ has been found, we know the three-dimensional strain field from the relation

$$E_{KL} = \frac{1}{2}\left(w_{L,K} + w_{K,L}\right) = X_2 E^{(1)}_{KL}, \tag{3.8}$$

as in Eqs.(3.32) of Ref.12.

Although we now have the plate strains from (3.8), we cannot yet determine the $\tilde{c}_{L\gamma M\alpha}$ from (2.4) because, as noted in Ref.12, we need the displacement gradients $w_{L,K}$ or equivalently the rotations Ω_{KL}, which are defined by

$$\Omega_{KL} = \frac{1}{2}\left(w_{L,K} - w_{K,L}\right). \tag{3.9}$$

As in Ref.12, we require that the associated plate rotations that accompany the plate strains that arise from the relaxation of the plate stress resultants $K^{(1)}_{2L}$ satisfy the appropriate three-dimensional rotation gradient-strain gradient relations, i.e.,

$$\Omega_{KL,M} = E_{ML,K} - E_{MK,L}. \tag{3.10}$$

From (3.22) and (3.23) of Ref.12 we have the expressions for the plate rotations

$$\Omega^{(n)}_{AB} = \frac{1}{2}\left(w^{(n)}_{B,A} - w^{(n)}_{A,B}\right),$$

$$\Omega^{(n)}_{2A} = \frac{1}{2}\left[(n+1)w^{(n+1)}_{A} - w^{(n)}_{2,A}\right]. \tag{3.11}$$

From (3.33) of Ref.12, we have

$$\Omega_{13}^{(0)} = \text{arbitrary constant} = 0, \quad \Omega_{13}^{(1)} = 0, \tag{3.12}$$

and from (3.35) and (3.34), respectively, of Ref.12, we have

$$\Omega_{2A}^{(0)} = -w_{2,A}^{(0)}, \quad \Omega_{2A}^{(1)} = E_{2A}^{(1)}, \tag{3.13}$$

which with (3.24) of Ref.12 enables us to write

$$\Omega_{2A} = -w_{2,A}^{(0)} + X_2 E_{2A}^{(1)}, \quad \Omega_{13} = 0, \tag{3.14}$$

which is the same as (3.36) of Ref.12. Now, from (3.8), (3.14) and

$$w_{K,L} = E_{KL} + \Omega_{LK}, \tag{3.15}$$

we have the desired three-dimensional displacement gradients when the plate deflection $w_2^{(0)}$ has been found.

4. ANALYSIS FOR FLEXURAL BIASING STATE

A plan view and cross-section of the substrate extending beyond the supports is shown in Fig.1 along with the coordinate system. Since the active surface wave region is in the interior of the plate away from the corner regions, certain simplifying approximations may be made in the treatment while still retaining a high degree of accuracy in the interior region of the plate. First the corner regions are ignored and the first plan view in Fig.2 is replaced by the second. Then we assume that the four overhangs transfer all load as shear and moment in the short cantilevered spanning direction to the interior of the now simply-supported plate, as shown in the cross-section in Fig.2. Clearly, equivalent shears and moments exist for the cross-section in the other direction, which is not shown. However, the expressions for these other shears and moments are indicated at the bottom of Fig.2. These assumptions are most accurate the greater the distance from the corner region, which for each direction corresponds to the active region of the substrate. In addition, these assumptions simplify the problem considerably in that for these conditions an exact solution of the equations of anisotropic static flexure may be obtained by decomposing the problem into three parts, each of which may be solved exactly. The three parts consist of a vertical loading in the interior of the simply-supported plate, edge moments per unit length m_3 at $X_3 = \pm b$ with zero moments at $X_1 = \pm a$ and edge moments per unit length m_1 at $X_1 = \pm a$ with zero moments at $X_3 = \pm b$. Since each of these problems may readily be solved exactly, the combined solution may readily be obtained by superposition by virtue of linearity.

For the case of the interior vertical loading the substitution of (3.3), (3.4) and $(3.7)_3$ in (3.1) yields the equilibrium equation for flexure of the thin plate in the form

$$\frac{2}{3} h^3 \gamma_{ABCD} w_{2,CDAB}^{(0)} + 2h\rho a_2 = 0. \tag{4.1}$$

Since the plate is simply-supported along the edges, the boundary conditions take the form

$$w_2^{(0)} = 0, \quad K_{11}^{(1)} = 0 \quad \text{at} \quad X_1 = \pm a,$$
$$w_2^{(0)} = 0, \quad K_{33}^{(1)} = 0 \quad \text{at} \quad X_3 = \pm b. \tag{4.2}$$

Since for singly-rotated Y-cut quartz plates (such as the ST-cut) $\gamma_{15} = \gamma_{35} = 0$, we can obtain a series solution for the simply-supported plate in the form

$$w_2^{(0)} = \sum_m \sum_n A_{mn} \cos \alpha_m X_1 \cos \varkappa_n X_3, \tag{4.3}$$

which satisfies (4.2) provided

$$\alpha_m = m\pi/2a, \quad \varkappa_n = n\pi/2b, \quad m \text{ and } n \text{ odd}. \tag{4.4}$$

Substituting from (4.3) into (4.1) and employing the orthogonality of the trigonometric functions, we obtain

$$A_{mn} = \frac{48\, \rho a_2}{h^2 mn\pi^2 R_{mn}}, \tag{4.5}$$

where

$$R_{mn} = -[\gamma_{11}\alpha_m^4 + \gamma_{33}\varkappa_n^4 + (2\gamma_{13} + 4\gamma_{55})\alpha_m^2 \varkappa_n^2]. \tag{4.6}$$

For the case of the edge moment from the X_3-overhang loading, the differential equation (3.1) and constitutive equations (3.4) take the form

$$K_{AB,AB}^{(1)} = 0, \quad K_{AB}^{(1)} = -\frac{2}{3} h^3 \gamma_{ABCD} w_{2,CD}^{(0)}, \tag{4.7}$$

since there is no interior loading for this portion of the problem, for which the edge conditions take the form

$$w_2^{(0)} = 0, \quad K_{11}^{(1)} = 0 \quad \text{at} \quad X_1 = \pm a,$$
$$w_2^{(0)} = 0, \quad K_{33}^{(1)} = \bar{m}_3 = \rho g h \ell_b^2 \quad \text{at} \quad X_3 = \pm b. \tag{4.8}$$

Again, since $\gamma_{15} = \gamma_{35} = 0$, we can obtain a series solution in the form

$$w_2^{(0)} = \sum_{m=1,3,5}^{\infty} f_m(X_3) \cos \frac{m\pi X_1}{2a}, \tag{4.9}$$

the substitution of which in (4.7) yields

$$\gamma_{11}\left(\frac{m\pi}{2a}\right)^4 f_m - 2(\gamma_{13}+2\gamma_{55})\left(\frac{m\pi}{2a}\right)^2 f_m'' + \gamma_{33} f_m'''' = 0 , \tag{4.10}$$

for each m. Since the problem is symmetric in X_3, we take

$$f_m = \cosh \zeta_m X_3 , \tag{4.11}$$

the substitution of which in (4.10) yields

$$\zeta_m^{(1)}{}_{(2)} = \frac{m\pi}{2a}\left[\frac{\gamma_{13}+2\gamma_{55}}{\gamma_{33}} \pm \sqrt{\frac{(\gamma_{13}+2\gamma_{55})^2 - \gamma_{11}\gamma_{32}}{\gamma_{33}^2}}\right]^{1/2} , \tag{4.12}$$

along with the negatives, which do not yield independent symmetric solutions. Thus, for each m, we take a sum of the two solutions in the form

$$f_m = A_m \cosh \zeta_m^{(1)} X_3 + B_m \cosh \zeta_m^{(2)} X_3 . \tag{4.13}$$

The satisfaction of the edge conditions $(4.8)_3$ for each n yields

$$B_m = S_m A_m , \tag{4.14}$$

where

$$S_m = -\cosh \zeta_m^{(1)} b / \cosh \zeta_m^{(2)} b . \tag{4.15}$$

In order to satisfy the remaining edge condition $(4.8)_4$ we take the form

$$w_2^{(0)} = \sum_{m=1,3,5}^{\infty} A_m(\cosh \zeta_m^{(1)} X_3 + S_m \cosh \zeta_m^{(2)} X_3)\cos \frac{m\pi X_1}{2a} , \tag{4.16}$$

and from the orthogonality of the $\cos m\pi X_1/2a$ in the interval $-a < X_1 < a$, we obtain

$$A_m = (-1)^{\frac{m-1}{2}} \frac{4\bar{m}_3}{m\pi}\left[-\frac{2}{3} h^3 \gamma_{33}(\zeta_m^{(1)2} - \zeta_m^{(2)2})\cosh \zeta_m^{(1)} b\right]^{-1} . \tag{4.17}$$

For the case of the edge moment from the X_1-overhang loading the solution is obtained in the same way, but with the treatment of the X_1-

and X_3-coordinates interchanged. Consequently, we simply write the resulting solution by analogy with the previous one, thus

$$w_2^{(0)} = \sum_{m=1,3,5}^{\infty} \bar{A}_m(\cosh \bar{\zeta}_m^{(1)} X_1 + \bar{S}_m \cosh \bar{\zeta}_m^{(2)} X_1)\cos \frac{m\pi X_3}{2b}, \qquad (4.18)$$

where

$$\bar{\zeta}_m^{(1)}_{(2)} = \frac{m\pi}{2b}\left[\frac{\gamma_{13}+2\gamma_{55}}{\gamma_{11}} \pm \sqrt{\frac{(\gamma_{13}+2\gamma_{55})^2 - \gamma_{11}\gamma_{33}}{\gamma_{11}^2}}\right]^{1/2}, \qquad (4.19)$$

and

$$\bar{A}_m = (-1)^{\frac{m-1}{2}} \frac{4\bar{m}_1}{4\pi}\left[-\frac{2}{3} h^3 \gamma_{11}(\bar{\zeta}_m^{(1)^2} - \bar{\zeta}_m^{(2)^2})\cosh \bar{\zeta}_m^{(1)} a\right], \qquad (4.20)$$

$$\bar{S}_m = -\cosh \bar{\zeta}_m^{(1)} a / \cosh \bar{\zeta}_m^{(2)} a, \quad \bar{m}_1 = \rho g h \ell_a^2. \qquad (4.21)$$

Now, from the combined $w_2^{(0)}$, which consists of the sum of (4.3), (4.16) and (4.18), we can obtain the biasing $w_{K,L}$ from Eqs.$(3.7)_3$, $(3.7)_1$, (3.8), (3.14) and (3.15).

5. RESONANT SURFACE WAVE MODE SHAPE

In this section we present the proper continuous representation of the acoustic surface wave mode shape in resonators with grooved reflectors, which was obtained in recent work[6]. The straight-crested surface wave displacement field may be written in the known form[13,14],

$$u_j = \alpha_j(X_2) e^{i\xi(X_1 - Vt)}, \quad \alpha_j = \sum_{m=1}^{4} C^{(m)} A_j^{(m)} e^{i\beta_m X_2}. \qquad (5.1)$$

A plan view of the resonator showing the reflecting arrays of grooves, the coordinate system and the associated planar geometry is shown in Fig.1, and we note that the plane $X_2 = 0$ denotes the ungrooved surface of the substrate and the axis points down. It has been shown[6] that the variable-crested resonant surface wave mode shape with variable amplitude along the transmission path is very accurately approximated by

$$u_j = \cos \frac{\pi X_3}{2w} \mathrm{Re}\left[\alpha_j(X_2)\hat{C}^R(X_1) e^{i\xi(X_1 - s)} + \alpha_j^*(X_2)\hat{C}^L(X_1) e^{-i\xi(X_1 - s)}\right] e^{-i\omega t}, \qquad (5.2)$$

where the variations along the transmission path are given by

$$\hat{C}^R(X_1) = \left[r_1 e^{-\alpha_1 N} e^{-\beta_2 X_1} - r_2 e^{-\alpha_2 N} e^{-\beta_1 X_1} \right] \bar{C}/d ,$$

$$\hat{C}^L(X_1) = r_1 r_2 \left[e^{-\alpha_1 N} e^{-\beta_2 X_1} - e^{-\alpha_2 N} e^{-\beta_1 X_1} \right] \bar{C}/d . \qquad (5.3)$$

In (5.3) $\bar{C}$ denotes the amplitude of the input wave, N denotes the number of grooves,

$$d = r_1 e^{-\alpha_1 N} - r_2 e^{-\alpha_2 N} , \quad \beta_{1,2} = \alpha_{1,2} N/(\ell - s) , \qquad (5.4)$$

and $\alpha_{1,2}$ and $r_{1,2}$ are given in Eqs.(32) and (34) of Ref.6. The amplitude of the standing surface wave mode, i.e., the part multiplying cos ωt, along the transmission path is plotted in Fig.3.

6. ACCELERATION SENSITIVITY

From Sec.5 we now know g_γ^μ and from Secs.3 and 4 we know $\hat{c}_{L\gamma M\alpha}$ for normal acceleration, i.e., flexure with the substrate extending beyond rectangular supports. Hence, we can now evaluate H_μ in Eq.(2.10). Such calculations have been performed using the known values of the second order[15] and third order[16] elastic constants of quartz. The calculated acceleration sensitivity is plotted as the solid curve in Fig.4 as a function of the planar aspect ratio a/b for b = 5.0 mm and given overhangs ℓ_a and ℓ_b, which are denoted in the figure. In addition, the normal acceleration sensitivities that were calculated in earlier work for the cases of simple[1] and rigid[2] edge supports are plotted in Fig.4 as a function of the planar aspect ratio a/b for the same value of b. It can be seen from the figure that all the sensitivity curves go through zero near the same value of a/b. The values of ℓ_a and ℓ_b for the overhang configuration treated in this work were chosen so that this would occur in order to facilitate a comparison of the configurations. It is quite clear from the figure that in the vicinity of the zero-crossings the slope of the sensitivity curve is much shallower for the configuration with the overhangs treated in this work than for the other cases treated in earlier works. This means that for the configuration treated here very low normal acceleration sensitivities can be obtained (parts in 10^{11} per g) over a reasonably large and, indeed, practical range of aspect ratio. As noted in the Introduction, such sensitivities have recently been observed at Raytheon with this support configuration. Figure 5 shows these same curves in the vicinity of the zero-crossings plotted to an expanded scale.

ACKNOWLEDGEMENTS

This work was supported in part by the Army Research Office under Contract No. DAAG-29-85-K-0075.

REFERENCES

1. D.V. Shick and H.F. Tiersten, "An Analysis of the Acceleration Sensitivity of ST-Cut Quartz Surface Wave Resonators Supported Along the Edges," Proceedings of the 40th Annual Symposium on Frequency Control, U.S. Army Electronics Technology and Devices Laboratory, Fort Monmouth, New Jersey and Institute of Electrical and Electronics Engineers, New York, IEEE Cat. No. 86CH2330-9, 262 (1986).

2. 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," Proceedings of the 41st Annual Symposium on Frequency Control, U.S. Army Electronics Technology and Devices Laboratory, Fort Monmouth, New Jersey and Institute of Electrical and Electronics Engineers, New York, IEEE Cat. No. 87CH2427-3, 282 (1987).

3. G.K. Montress, T.E. Parker and J. Callerame, "A Miniature Hybrid Circuit SAW Oscillator Using an All Quartz Packaged Resonator," 1985 Ultrasonics Symposium Proceedings, IEEE Cat. No. 85CH2209-5, Institute of Electrical and Electronics Engineers, New York, 277 (1985).

4. J.A. Greer, T.E. Parker, M. Rothschild and D.J. Ehrlich, "A Novel Technique for Trimming the Frequency of a Sealed Surface Acoustic Wave Resonator," Proceedings of the 41st Annual Symposium on Frequency Control, U.S. Army Electronics Technology and Devices Laboratory, Fort Monmouth, New Jersey and Institute of Electrical and Electronics Engineers, New York, IEEE Cat. No. 87CH2427-3, 351 (1987).

5. H.F. Tiersten, "Perturbation Theory for Linear Electroelastic Equations for Small Fields Superposed on a Bias," J. Acoust. Soc. Am., 64, 832 (1978).

6. H.F. Tiersten, J.T. Song and D.V. Shick, "A Continuous Representation of the Acoustic Surface Wave Mode Shape in Arrays of Reflecting Grooves," J. Appl. Phys., 62, 1154 (1987).

7. J.A. Greer, private communication.

8. R.D. Mindlin, "An Introduction to the Mathematical Theory of the Vibration of Elastic Plates," U.S. Army Signal Corps Eng. Lab., Fort Monmouth, New Jersey (1955). Signal Corps Contract DA-36-03956-56772.

9. R.D. Mindlin, "High Frequency Vibrations of Crystal Plates," Quart. Appl. Math., 19, 51 (1961).

10. H.F. Tiersten, Linear Piezoelectric Plate Vibrations (Plenum, New York, 1969), Chap.13.

11. H.F. Tiersten and B.K. Sinha, "Temperature Dependence of the Resonant Frequency of Electroded Doubly-Rotated Quartz Thickness-Mode Resonators," J. Appl. Phys., 50, 8038 (1979).

12. D.S. Stevens, H.F. Tiersten and B.K. Sinha, "Temperature Dependence of the Resonant Frequency of Electroded Contoured AT-Cut Quartz Crystal Resonators," J. Appl. Phys., 54, 1709 (1983).

13. J.J. Campbell and W.R. Jones, "A Method for Estimating Optimal Crystal Cuts and Propagation Directions for Excitation of Piezoelectric Surface Waves," IEEE Trans. Sonics Ultrason., SU-15, 209 (1968).

14. B.K. Sinha and H.F. Tiersten, "Elastic and Piezoelectric Surface Waves Guided by Thin Films," J. Appl. Phys., 44, 4830 (1973). Sec.II.

15. R. Bechmann, "Elastic and Piezoelectric Constants of Alpha-Quartz," Phys. Rev., 110, 1060 (1958).

16. R.N. Thurston, H.J. McSkimin and P. Andreatch, Jr., "Third Order Elastic Constants of Quartz," J. Appl. Phys., 37, 267 (1966).

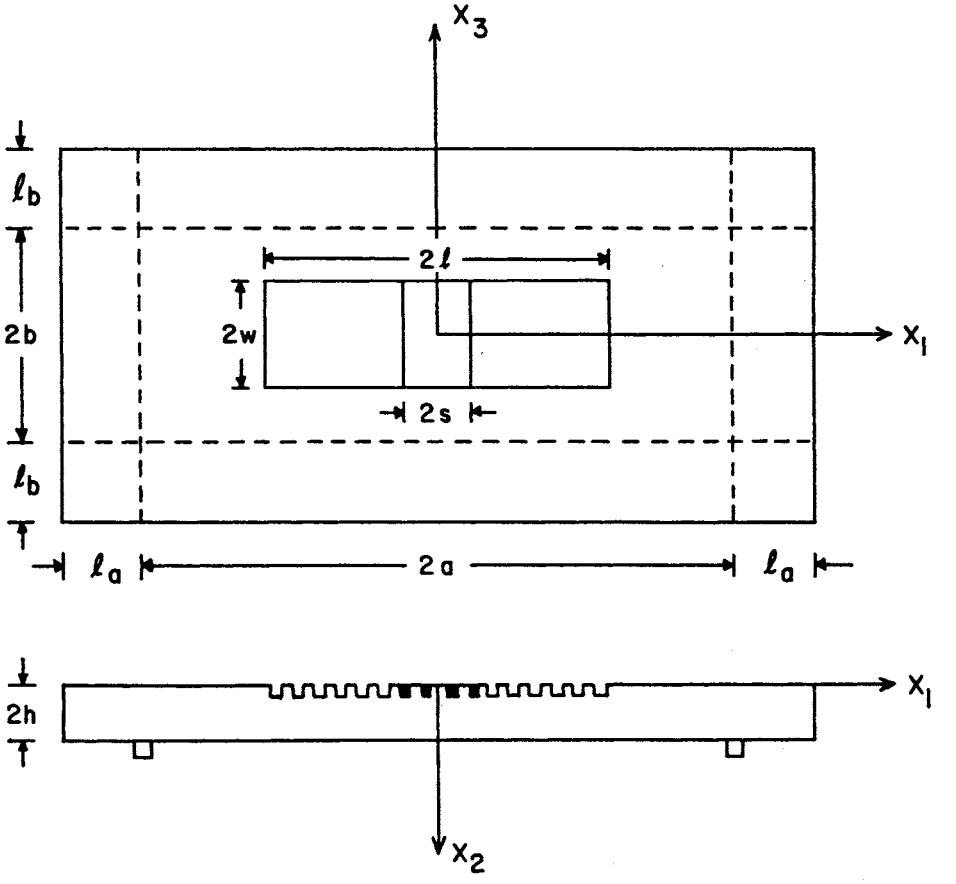


Figure 1 Plan View and Cross Section of Rectangular ST-Cut Quartz Plate with the Substrate Extending Beyond Rectangular Supports

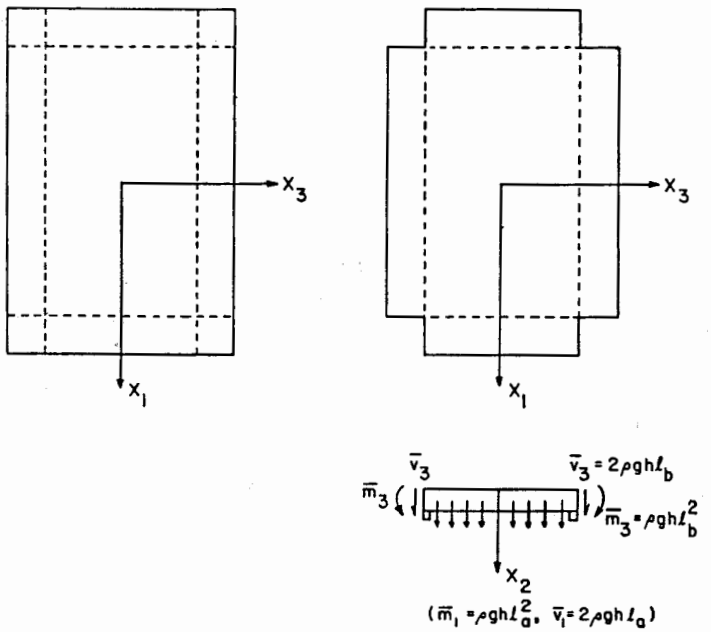


Figure 2 Plan View and Cross-Section of Plate for Simplifying Approximation Made in the Analysis for the Biasing State

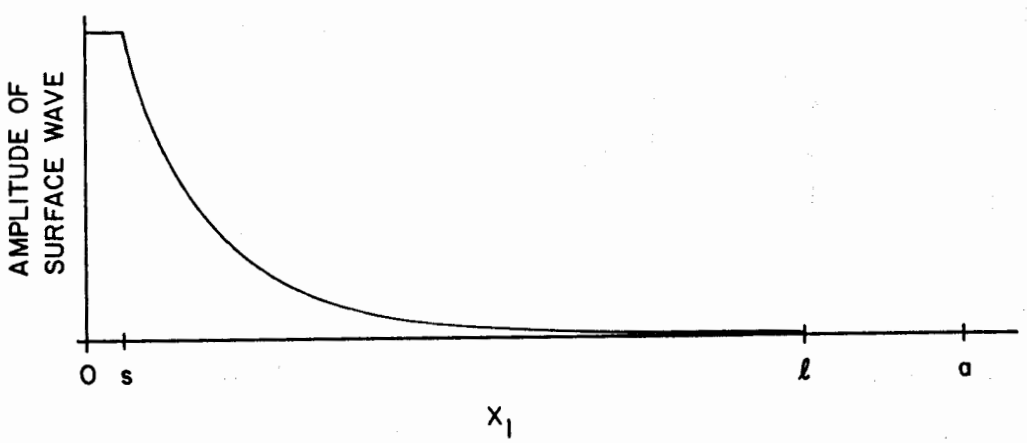


Figure 3 Amplitude of Resonant Surface Wave Mode Along Transmission Path

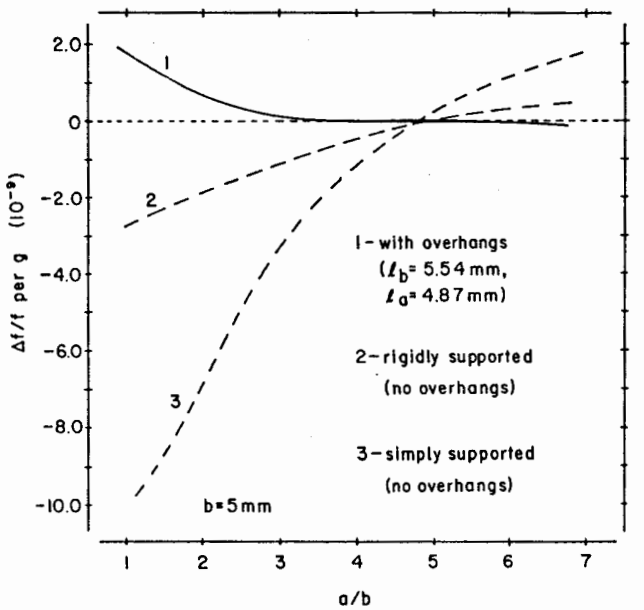


Figure 4 Calculated Normal Acceleration Sensitivity Versus Planar Aspect Ratio a/b for the Support Configuration Treated in this Work (Solid Line) and the Configurations Treated in Two Earlier Works (Dotted Lines)

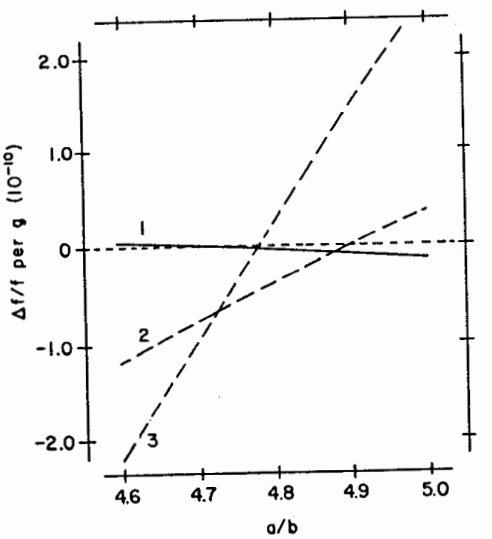


Figure 5 Portion of Figure 4 Near the Zero-Crossings Plotted to an Expanded Scale

2 ND EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 16 - 17 - 18 MARCH 1988

Session RO-04 Number 4

BULK WAVES EXCITED INTO AN ANISOTROPIC HALF SPACE BY A METALLIC INTERDIGITAL TRANSDUCER

Y. ZHANG and M. PLANAT

Laboratoire de Physique et Métrologie des Oscillateurs du CNRS
associé à l'Université de Franche-Comté-Besançon
32, avenue de l'Observatoire - 25000 Besançon - France

Abstract

The response of an elastic half space to any excitation consists of two parts. The first one corresponds to the propagation of the well known non dispersive Rayleigh wave. The other one is made of the simultaneous emission of dispersive bulk waves.

In the Fourier integral approach, the Rayleigh wave comes from an isolated pole of the integrant. On the other hand, bulk wave contributions arise from all values of the wavenumber in the reciprocal space. In most cases the bulk wave field can be approximated by the ray approximation with the result that the energy propagates in a definite direction which is normal to the slowness curve.

We here study the case where the excitation of waves is by an interdigital transducer deposited on a quartz substrate. It is shown that orientation and width of the main lobe is very sensitive both to the applied frequency and crystal orientation. Radiation pattern is computed for singly rotated quartz cuts with the transducer respectively parallel to X-axis (shear mode excitation) and perpendicular to X-axis.

1. INTRODUCTION

An elastic half-space subject to surface disturbances may generate Rayleigh waves ; these waves are used in a wide domain of applications including signal processing, oscillators, sensors, material characteri-

zation, ... The simultaneous excitation of bulk waves has historically been considered as a spurious effect in devices using an interdigital transducer as a source for surface wave excitation. However they can present advantages as smaller temperature effects, less sensitivity to surface contamination and higher frequency operation [1], but the main interest lies in the fact that the radiated field is angularly focused and the direction of emission can be scanned by the applied excitation frequency [2].

In the so-called surface skimming bulk wave (SSBW) devices, the IDT couples only to the horizontal shear displacement ; however in the general case, it can also excite longitudinal and vertical shear vibrations, depending on the substrate.

There are two main ways of solving the fundamental problem of the wave excitation. In the first one an effective permittivity is defined from the relation between the Fourier transform of charge density and the electrical potential at the surface [3]. This approach is mainly numerical and it is difficult to distinguish one wave from the others. The second approach is based on the classical paper by Lamb [4]. The electroelastic potential inside the medium is computed and the mechanical displacement obtained as a forced problem [5].

Previously [5, 6] the mechanical displacement was evaluated asymptotically at or near the surface. In this paper a definite integral representation of the displacement is obtained at any point of the semi-infinite anisotropic solid and asymptotic formulae are obtained for the far field solution [7]. The analysis is applied to singly rotated quartz cuts with the transducer respectively parallel to X-axis (shear mode excitation) and perpendicular to X-axis.

2 - GENERAL FORMULATION OF THE PROBLEM

We consider the excitation by a very wide interdigital transducer located on a low piezoelectric crystal. We solve the problem by taking the Fourier transform of fields along the direction Ox_1, and choosing exponential dependence along the depth Ox_2 (Fig. 1).

Three components (r = 1-3) are obtained of wave numbers k_1 and $k_2^{(r)}$ along respectively Ox_1 and Ox_2 axis, corresponding to wave motions carrying energy away from the surface. In polar coordinates (R, θ_o), the displacement takes the following form

$$u_j(R,\theta_o) = \int_{-\infty}^{+\infty} \sum_{r=1}^{3} \left[\hat{\phi}_o \frac{B_r\, a_j^{(r)}}{\Delta}\right](k_1)\, e^{Rf^{(r)}(k_1)}\, dk_1 \qquad (1)$$

with

$$f^{(r)}(k_1) = -\,i\,[k_1 \cos\theta_o + k_2^{(r)}(k_1) \sin\theta_o]$$

In these relations $k_2^{(r)}$ (r = 1-3) are the roots of the Christoffel equation, $a_j^{(r)}$ are the corresponding eigenvectors, Δ is the determinant of the boundary condition and B_r the weight of each mode r. The symbol $\hat{\phi}_o$ denotes the Fourier transform (expressed at the surface) of the applied voltage ϕ. Dispersion relations $k_2^{(r)}(k_1)$ are represented on Fig. 2 for the case of a Y-cut quartz crystal.

Inserting the integral inside the summation, we evaluate the integral for each mode r by using the steepest descent path method.

The integration is performed in the complex k_1 plane. We deform the path to get the major part of the integral around the so-called saddle point $k_1 = k_{1c}^{(r)}$ defined by the zero first order derivative of the phase.

The location of the saddle point is crucial in the continuous deformation of the original path to the steepest descent one. A first kind of path is represented in full lines on Fig. 3 for a wide observation angle $\theta_o = \theta_1$. In this case, the contribution from the saddle point is the only one representing the far field solution because no singularity has been enclosed in the integration loop.

A second kind of path is represented in dotted lines for a lower observation angle $\theta_o = \theta_2 < \theta_1$. The path includes both a pole k_{1R} and a branch point k_{1B}. The pole is defined by the relation $\Delta(k_{1R}) = 0$ and corresponds to the Rayleigh wave propagation. The corresponding displacement can be obtained from the residue theorem [5]. Due to the exponential attenuation along the depth, its contribution is negligeable except in the closed vicinity of the surface. The branch point may be excluded from the path by

using a loop beginning and ending at infinity and encircling the branch point. Its contribution can be computed by integrating along the steepest descent paths Γ^+ and Γ^-.

3 - CONTRIBUTION OF THE SADDLE POINT. APPLICATION TO SURFACE SKIMMING BULK WAVES IN SINGLY ROTATED QUARTZ CRYSTAL CUTS WITH PROPAGATION NORMAL TO X-AXIS

For sufficiently deep observations angles θ_o, the contribution of the saddle point is dominant and takes the following form

$$u_j(R,\theta_o) = \sum_{r=1}^{3} \left\{ \frac{\exp\left[R\, f^{(r)} + i(\pi/4)\ \mathrm{Sgn}\, f_2^{(r)}\right]}{\left[2\pi R \left| f_2^{(r)} \right|\right]^{1/2}} \cdot \frac{B_r\, a_j^{(r)}}{\Delta}\, \hat{\phi}_o \right\}(k_{1c}^{(r)}) + O\left(R^{-3/2}\right) \qquad (2)$$

where $f_2^{(r)}$ is the second order derivative of the function $f^{(r)}$ and the expression inside the brackets has to be evaluated at the saddle point $k_{1c}^{(r)}$.

In equation (2) the saddle point $k_{1c}^{(r)}$ happens when the first order derivative of phase vanishes, i.e. when

$$f'\,(k_{1c}^{(r)}) = -\ \mathrm{cotg}\ \theta_o \qquad (3)$$

On the other hand, since the wave length along direction Ox_1 is equal to twice the transducer periodicity, it is apparent from Fig. 1 that waves interfer constructively for an excitation frequency f given by

$$f = V_c \ / \ (\lambda_o \cos\ \theta_c) \qquad (4)$$

where θ_c is the propagation angle for the main excited wave and V_c is the corresponding phase velocity. As shown on Fig. 2, the angle θ_o is defined from the normal to the dispersion curve and represents the power flow direction.

From eq. (3) we see that the saddle point location depends on the observation angle θ_o. For different values of θ_o the propagation velocity of the wave that contributes is different and the main direction of emission is scanned by the applied frequency (eq. (4)). Moreover the total displacement at angle θ_o is the summation of all modes r.

Therefore the validity of eq. (2) depends on both crystalline configuration and the value of θ_o. In particular when $\theta_o \to 0$ or π, the path of integration may include a pole, corresponding to Rayleigh wave excitation or a branch point. In the last case, the saddle point approachs a branch point and expression (2) gives an incorrect result : this case will be studied in § 4.

Here we restrict ourselves to the case of singly rotated cuts of quartz with propagation normal to X-axis. This simple class of cuts is convenient for supporting the so-called surface skimming bulk waves (SSBW), because only horizontal shear motion is coupled to piezoelectric excitation.

Main characteristics of SSBW for this class of cuts were studied recently [7]. A complete analytic treatment is possible and yields an expression of power flow density in radial direction given by

$$P_r (R, \theta_o) = \alpha_f \alpha \phi_o^2 / R \qquad (5)$$

where α_f and α are respectively transducer and material factors.

For operation at the surface, the cut with a rotation angle $\theta = 31.5^\circ$ exhibits the highest velocity corresponding to the highest possible frequency. This cut is near the well known AT-cut $\theta = 35^\circ$ which has been used for oscillators with frequencies up to 3.4 GHz [8]. The lowest velocity is obtained for $\theta = 121.5^\circ$. Both cuts have zero power flow angle.

Efficiency was studied as a function of θ. At shallow angles, the best efficiency (180 W/m^2) is obtained near the ST-cut ($\theta = 42^\circ$) and remains large (100 W/m^2) for the $\theta = 31.5^\circ$ cut (Fig. 4). On the contrary, material efficiency α remains almost constant at shallow angles for BT-cut ($\theta = 131^\circ$) and $\theta = 121.5^\circ$ cut, and exhibits a peak for observation angles near 90°. These latter cuts may be suitable for vertical radiation application.

Plots of power density P_r versus observation angle θ_o are given on Fig. 5 for the two extreme cases of the highest velocity cut $\theta = 31.5^\circ$ and the lowest velocity cut $\theta = 121.5^\circ$. It is clearly observed that the width and height of the main lobe for $f = f_o$ are sensitive to the crystalline configuration. Moreover, it is apparent that the main lobe direction depends on the operating frequency more strongly for the lowest velocity cut.

4 - CONTRIBUTION OF BRANCH POINTS : APPLICATION TO SINGLY ROTATED CUTS OF QUARTZ WITH PROPAGATION ALONG X-AXIS

By looking at Fig. 2, we see that there exist a double root for the function $k_2(k_1)$ at points k_{1b} where the tangent to the dispersion curve is vertical. These so-called branch points, where the integrant in (1) becomes double valued, have to be excluded from the closed contour by using an extra loop in the complex k_1-plane (cf. Fig. 3).

Branch points are found in the deformation of contour only at shallow angles. Consequently the steepest descent paths encircling branch points are vertical [6]. After lengthy calculations it can be shown that a branch point contribution takes the form

$$u_j = i\, k_o\, \hat{\phi}_o(k_{1b})\, \exp(-ik_b x_1)(c_j/2\pi) \int_0^\infty \frac{\eta^{1/2}}{a + b\eta} e^{-R\eta}\, d\eta \qquad (6)$$

where $a = \Delta^2/2$, b and c_j are material coefficients and Δ is the determinant of boundary condition and has been computed at the branch point.

Three distinct cases can be considered :

a) <u>when $a = 0$, i.e. $\Delta = 0$</u>

The field exactly satisfies boundary condition. In this case

$$u_j \propto R^{-1/2} \qquad (7)$$

This corresponds to the case of SSBW waves considered in § 3.

b) <u>when $a \neq 0$ and $|a| \gg |b|\eta$</u>

The integral I in (6) decays like $R^{-3/2}$

$$I = \frac{\sqrt{\pi}}{2a} R^{-3/2} \qquad (8)$$

c) <u>in the general case</u>

$$I = R^{-3/2}\, J(Z) / a \qquad (9)$$

where

$$J(Z) = \Gamma(3/2)\, e^{Z/2}\, Z^{1/4}\, W_{-3/4,\ 1/4}(Z) \quad ; \quad Z = (a/b)R$$

and Γ and W are respectively gamma and Withaker functions.
In the limit $R \to \infty$, expression (9) tends to (8).

The analysis was applied to singly rotated cuts of quartz with propagation along X-axis.

Mechanical displacements corresponding respectively to fast shear, slow shear and longitudinal branches are represented on Fig. 6, as a function of rotation angle θ. The results are for a propagation distance $R = 100\ \lambda_0$ and an excitation voltage $\Phi_0 = 1$ V.

For the fast shear mode and cuts located near the 32 ° cut (Fig. 6a), we have $\Delta = 0$. From (7) the decay is proportional to $R^{-1/2}$. The displacement is approximately shear horizontal and consequently the wave is a quasi-SSBW. For other cuts the decay is proportional to $R^{-3/2}$ and the contribution is lower.

For the slow shear mode (Fig. 6b), the main contribution again comes from the $R^{-1/2}$ term although the determinant Δ nearly vanishes only at $\theta = 120$ °. Main contributions happen at $\theta = 0$, $\theta = 80$ ° and $\theta = 150$ °. The SSBW for this mode exists in the vicinity of $\theta = 120$ °, since only the u_3 is dominant.

Finally for the longitudinal mode (Fig. 6c), the decay is proportional to $R^{-3/2}$. The wave is strong only around the ST-cut ($\theta = 41$ °) and its polarization is longitudinal.

5 - CONCLUSION

As it has been shown in this paper the piezoelectric excitation of waves in an anisotropic half space is yet an open problem and a rich variety of propagation modes can be obtained. Besides the well-known non dispersive Rayleigh wave, a number of dispersive waves can be excited and studied asymptotically at large distances. In this paper we only were interested for the case where the ray approximation is valid and new results were given for quartz crystals.

For rotated Y-cuts of quartz the geometrical approximation can break down for the shear mode, since the dispersion curve exhibits an inflexion point (Fig. 2). At this point the field is considerably enhanced and higher order terms in the development (2) have to be taken into account.

On the other hand the present analysis can be readily extended for the case of crystals, as $LiNbO_3$, where the piezoelectric coupling is high. By using the effective permittivity concept introduced recently [3], it is possible to obtain a complete physical interpretation of the radiation pattern.

REFERENCES

[1] M.F. LEWIS, "Surface skimming bulk waves, SSBW", Proceedings of the Ultrasonics Symposium, Phoenix, Arizona, 1977, pp. 744-752.

[2] M.A. GOODBERLET and D.L. LEE, "The excitation and detection of surface-generated bulk waves", IEEE Trans. Sonics Ultrason., 1984, SU-31, pp. 67-76.

[3] R.F. MILSON, N.H.C. REILLY and M. REDWOOD, "Analysis of generation and detection of surface and bulk acoustic waves by interdigital transducers", IEEE Trans. Sonics Ultrason., 1977, SU-24, pp. 147-166.

[4] H. LAMB, "On the propagation of tremors over the surface of an elastic solid", Philosophical Transaction of the Royal Society (London), A 203 (1904) 1, pp. 1-42.

[5] M. PLANAT and P. SCHIAVONE, "Efficiency of surface interdigital transducers on anisotropic media", J. Appl. Phys., 1985, 57(1), pp. 49-52.

[6] K. YASHIRO and N. GOTO, "Analysis of generation of acoustic waves on the surface of a semi-infinite piezoelectric solid", IEEE Trans. Sonics Ultrason., 1978, SU-25, pp. 146-153.

[7] Y.W. ZHANG and M. PLANAT, "Anisotropy dependence of the focused field radiated into a piezoelectric half space by an interdigital transducer", Appl. Phys. Lett., 1988, 52(6), pp. 456-458.

[8] K.F. LAU, K.H. YEN, R.S. KAGIWADA and A.M. KONG, "High frequency temperature stable SBAW oscillators", IEEE Ultrasonics Symp. 1980, p. 240-244.

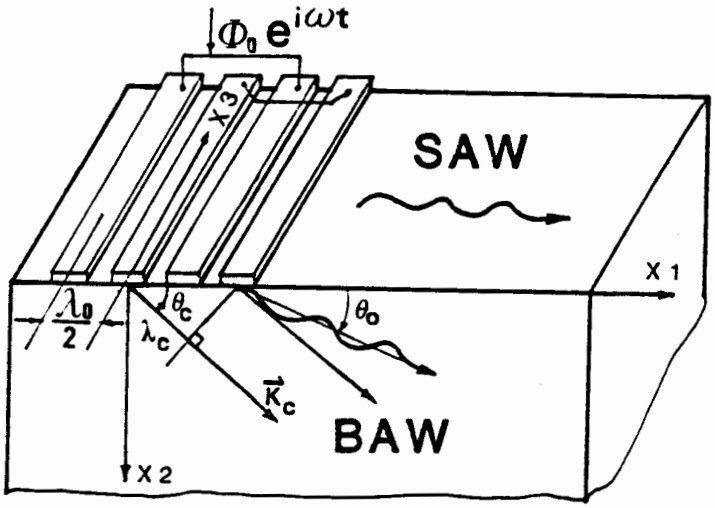


Fig. 1 : Semi-infinite geometry of surface and bulk wave excitation by an interdigital transducer

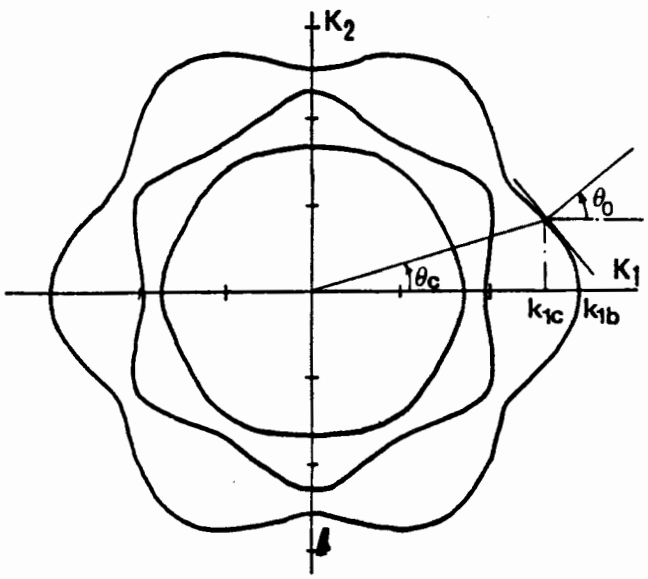


Fig. 2 :
Wave vector diagram (or dispersion relation) of a Y-cut quartz crystal

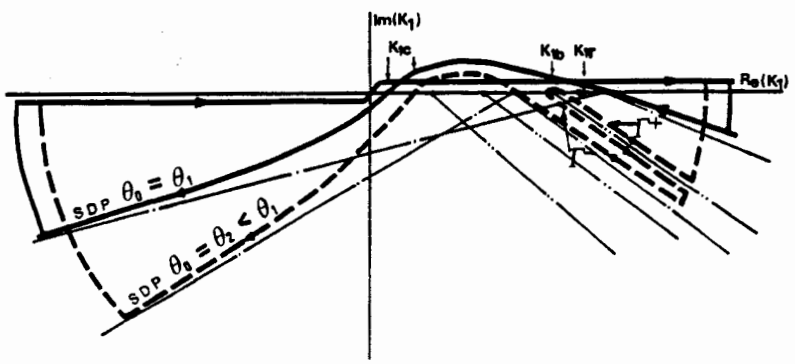


Fig. 3 :
Schematic of the steepest descent path and the singularities in the complex wave number plane

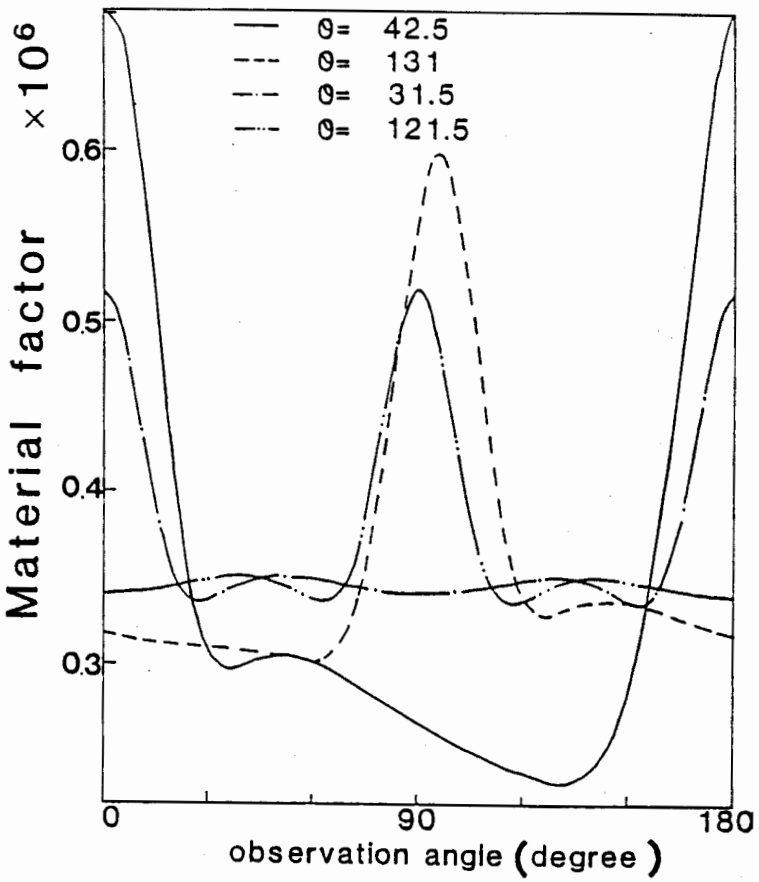


Fig. 4 :
Material efficiency of bulk radiation as a function of observation angle for surface shear horizontal waves in quartz

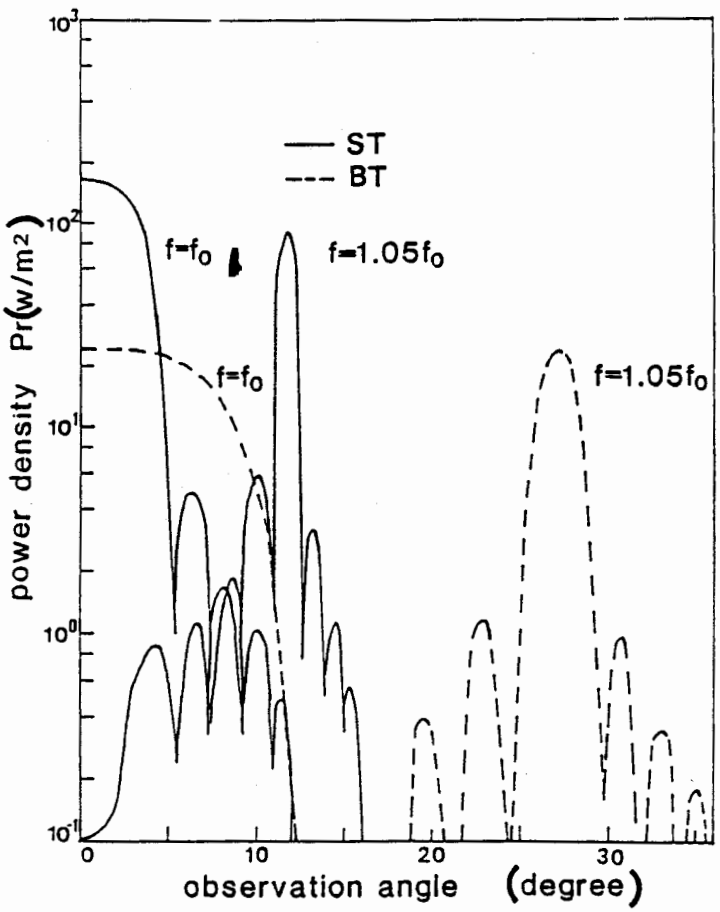


Fig. 5 :
Radiation diagram of bulk waves from a 50 pairs of fingers transducer (f_0 is the cut-off frequency).

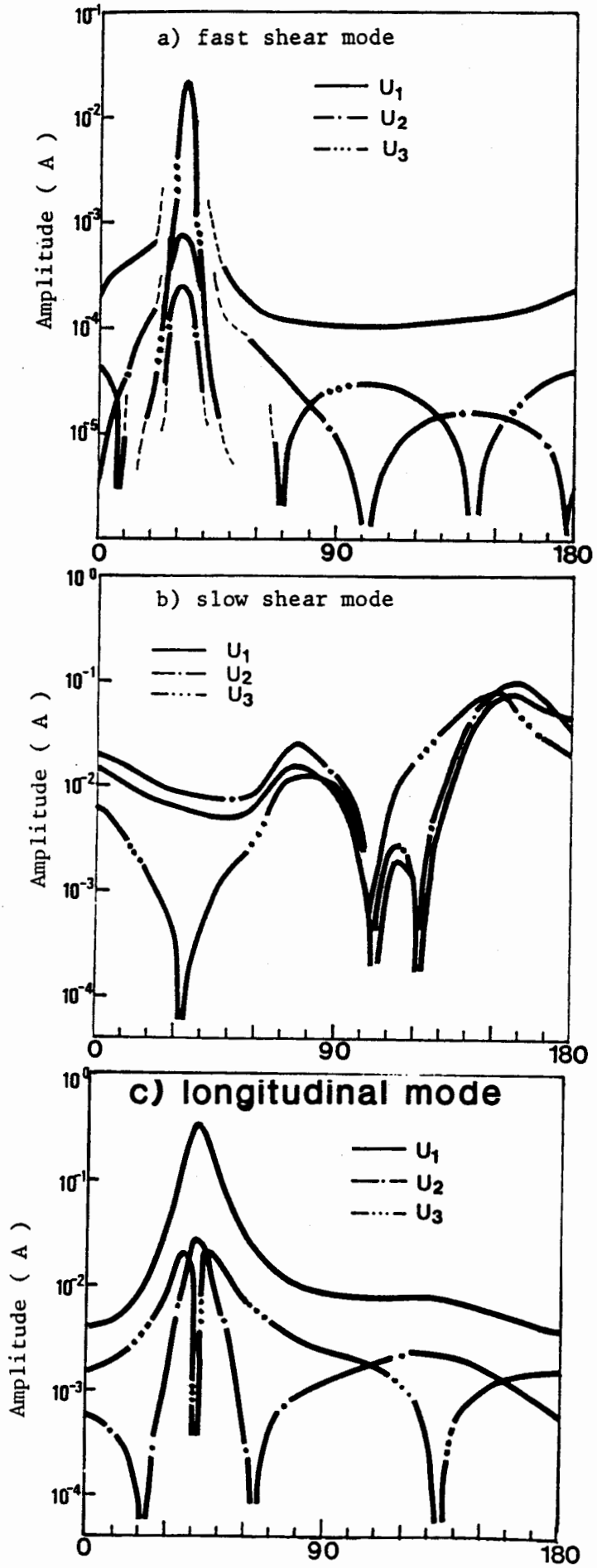


Fig. 6 :
Branch point contribution for a rotated Y-X cut of quartz

a) fast shear mode

b) slow shear mode

c) longitudinal mode

2 ND EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 16 - 17 - 18 MARCH 1988

Session RO-04 Number 5

ACOUSTIC SURFACE TRANSVERSE WAVE FILTERS AT 2 TO 2.8 GHZ FOR OSCILLATORS AND TIMING RECOVERY

B. Fleischmann, G. Riha and H.-W. Woerz,

Siemens AG,
Corporate Research and Development,
Munich, Federal Republic of Germany

SUMMARY

Surface Transverse Wave (STW) filters for the frequency range from 2 to 2.8 GHz were designed and produced to study the applicability of Surface Acoustic Waves (SAW) for timing recovery filters in optical links and for high quality factor filters in oscillators of this frequency range. The filters were fabricated with a standard photolithographic process on temperature stable rotated Y-cut quartz substrates. They operate at higher harmonics.

Two different retiming filters were designed for 2.49 GHz, which recently became a standardized frequency for data transmission at 2.4 Gbit/s, and for 2.78 GHz. The 2.49 GHz-filters have an insertion loss of 12.5 dB and a quality factor of 1100. The insertion loss of the 2.78 GHz-filters is 12 dB higher; the Q factor is 1350.

Oscillator filters were designed for 1.98 and 2.65 GHz. Both have Q factors of 3400 with an untuned insertion loss of 21 and 25 dB, respectively. Oscillators built with the 1.98 GHz-filters have a single sideband phase noise at 1 kHz offset from the carrier of -100 dBc/Hz.

The reproducibility of the electrical parameters is an important criterion for the applicability of the filters. Our measured standard deviations of mounted filters are quite low. E. g. the value for the centre frequencies is typically 40 ppm for each type.

I. INTRODUCTION

Digital communication systems of the next generation with a data transmission rate of 2.4 Gbit/s require narrowband filters for frequency synthesis and timing recovery. Many of these systems will probably operate at the recently standar-

dized SONET (**S**ynchronous **O**ptical **NET**work) bit frequency of 2488.32 MHz. To demonstrate the usefulness of SAW filters at this high frequency we investigated the production and the performance of SAW delay lines in the frequency range from 2 to 2.8 GHz with medium and high Q factors.

To fulfill the demands for high reproducibility and mass production a standard photolithographic process was used. The linewidth resolution of this method gives an upper frequency limit of just above 1 GHz for fundamental mode SAW filters using the Rayleigh Wave (RW). By operation at a higher harmonic the usable frequency range can be extended up to 2 GHz, but conversion of the RW to bulk waves increases the insertion loss and reduces the maximum Q value [1].

Therefore we used the STW on AT-cut quartz, which has a 60 % higher propagation velocity than the RW [2] and smaller mode conversion losses. The resulting frequency limit is approximately 3 GHz.

Four different filter types, two with medium and two with high Q values, were designed and fabricated . A compact hybrid oscillator was developped for one type and results thereof will also be discussed.

II. FABRICATION PROCESS

Our standard photolithographic process with 10:1 reduction projection printing and lift-off technique was used to produce the filters. It has a linewidth resolution of about 0.7 microns. Substantially smaller linewidths can be achieved with electron beam exposure [3], but decreased reproducibility and large exposure times make this process prohibitive for mass production.

Nevertheless, at the resolution limit of the projection printing deviations of the fingerwidths have to be considered. They are smallest under the condition of a purely periodic geometry, i.e. constant finger and gap widths, and a metallization ratio of about 0.5. These aspects mean restrictions to the filter design.

It should be mentioned here that operation at a higher harmonic implies general benefits on the reproducibility. That is because equal relative rather than absolute linewidth deviations lead to equal frequency and loss scattering, presuming an optimal filter design. It means that using a fundamental frequency design, which has smaller linewidths, requires a higher production accuracy to achieve the same scattering of electrical parameters.

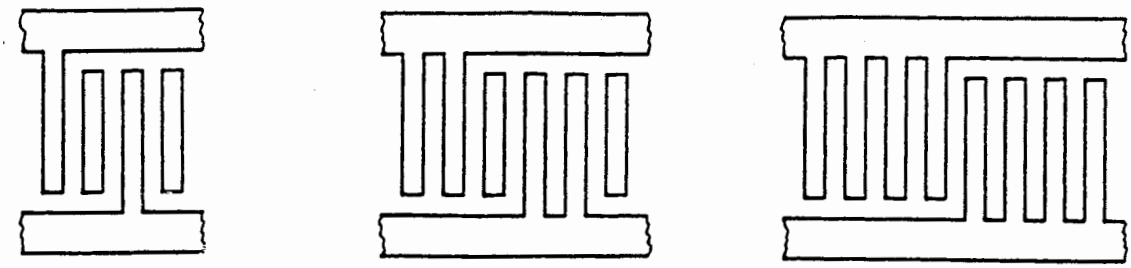
The metallization thickness, too, has a large influence on the frequency stability. In the investigated frequency range

the dependency of the wave velocity from the thickness of the metal layer is about 10 ppm/Å. Therefore it is indispensable to maintain the metallization thickness within only a few Ångstrøms over the whole wafer.

III. ANALYSIS METHOD AND FILTER DESIGN

A detailed prescription of our analysis method is given in reference [4] and the references cited therein. It contains a Green's function method to calculate the electrostatic charge distribution on the transducer and a modified scattering matrix model to calculate transfer functions. This method gives good results for surface waves, including the STW. Conversion of the STW to bulk shear waves, which occurs at higher harmonics, has been considered as propagation loss.

Studies of the excitation of higher harmonics have led to the definition of the relative coupling coefficient k_r^2, which is a useful measure for the excitation strength of a specific transducer type at a specific harmonic frequency [5]. The higher k_r^2, the better the coupling of the SAW, normalized to one wavelength. For our filters three different interdigital transducer (IDT) types with a high insensitivity of k_r^2 to variations of the metallization ratio were used. Figure 1 shows schematic descriptions of one period of each type.



a) b) c)

Figure 1: Three types of SAW transducers

Common features of all four designs are the use of uniform, i.e. unweighted, transducers and of energy trapping structures between these [6]. The basic shape of the transfer function is therefore sinx/x squared with a nominal sidelobe suppression of 26 dB. It can be degraded by secondary effects like propagation loss and multiple transit signals to values between 25 and 20 dB. This is in oscillator and timing recovery applications normally sufficient.

III.A 2 GHz filter with high Q value

The first filter type to be presented, type A, consists of two split-isolated transducers [7], shown in figure 1a. It is similar to the well-known splitfinger type, which is also designated as split-connected. Split-isolated transducers have the advantage that transducer capacity and radiation conductance are reduced by half [8]. This allows a reduction of diffraction effects, which could decrease the maximally achievable Q factor, by doubling the transducer's aperture without significant change of its impedance.

Both types excite the fundamental frequency, i.e. one period is one wavelength, and the third harmonic, i.e. one period is three wavelengths, with the necessary insensitivity and strength. Other, higher harmonics are less useful because of weaker coupling and higher sensitivity to linewidth deviations.

The filter was designed for a centre frequency of 1.98 GHz. It consists of two long transducers to achieve a high Q value.

III.B 2.49 GHz filter with medium Q value

The transducer type shown in figure 1b may be used for fundamental and fifth harmonic operation. Here, every third finger is floating. This has the advantage of completely suppressing the third harmonic, which would be excited in a three fingers up - three down configuration [8].

Filter type B was designed for passive timing recovery with a Q value of 1100. A three transducer arrangement to suppress electromagnetic feedthrough was chosen [5].

III.C 2.65 GHz filter with high Q value

Type C uses two long transducers of the same kind as the 2.49 GHz filter, but with smaller linewidths. It was designed for use in oscillators.

III.D 2.78 GHz filter

The last type (D) operates with the transducer design shown in figure 1c. It excites the fundamental frequency and the third, fifth and seventh harmonic with reasonable strength. Operating at the seventh harmonic, the achievable frequencies are 17 % higher than with a splitfinger type at the third harmonic. In applications where a broadband suppression of spurious signals is necessary one transducer of this kind may be combined with one of another type to form a filter in which only the desired harmonics coincide. Other harmonics would be suppressed. For this type, too, the three

transducer configuration was chosen.

The most important design parameters are summarized in table 1. Type A was fabricated on two different quartz cuts to achieve different turnover temperatures, which are mainly dependent on the cut angle. Pure aluminium was used for the transducer structures. For high-Q filters thin metallization thicknesses are indispensible to avoid excessive mode conversion losses. For example type A filters have about 10 dB higher insertion loss with 400 instead of 250 Å metal thickness.

type	cut angle	geometric period	number of IDTs	used harmonic	metal thickness
A	37.5, 37	1.926 μm	2	3	25, 30 nm
B	37.5	1.700 μm	3	5	25, 30 nm
C	36	1.600 μm	2	5	25 nm
D	36	1.600 μm	3	7	20 nm

Table 1: Important design parameters

IV. RESULTS

Five wafers of type A, two of type B and one of each of the others were fabricated. Between twenty and fifty filters of each type were mounted into standard TO-8 packages with an organic adhesive and bonded with gold wires. Twenty filters of type D were additionally mounted into TO-39 packages.

It was already mentioned that the filters' passbands are shaped like sinx/x squared. This can be seen in figures 2 to 4.

IV.A Type A filters

Figure 2 is a plot of the measured frequency response of a filter of type A, fabricated on 37.5° rotated Y-cut with 250 Å of metallization. The untuned insertion loss is about 21 dB. The Q value, defined by $Q = \pi f \tau$, where f is the centre frequency and τ the group delay time, is 3400.

On 37° rot. Y-cut the wave velocity is about 0.11 % higher. The centre frequency increases by the same amount. The piezoelectric coupling on both cuts is almost identical. Deviations of the insertion loss lie within the scattering of this parameter.

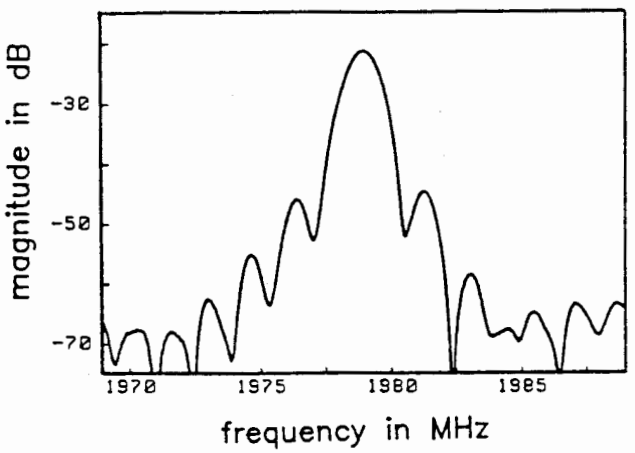
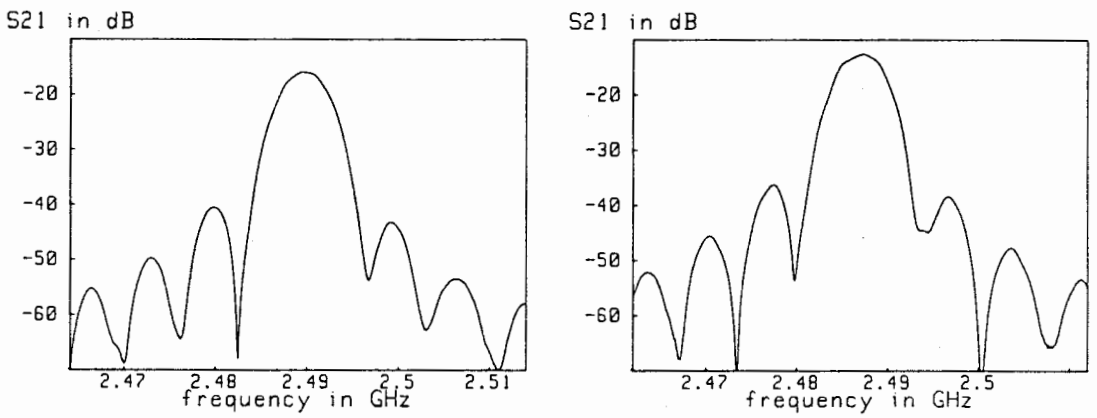


Figure 2: Frequency response of a filter of type A

IV.B Type B filters

Filters of types B and D suffer less from mode conversion to bulk waves because of their shorter transducers. This can be seen in figure 3, where the transfer functions of filters of type B with 250 and 300 Å of metallization are plotted. It can be seen that the insertion loss of the filter with 300 Å metallization thickness is only 12.5 dB and thereby even 2.5 dB lower than of the filter with 250 Å.

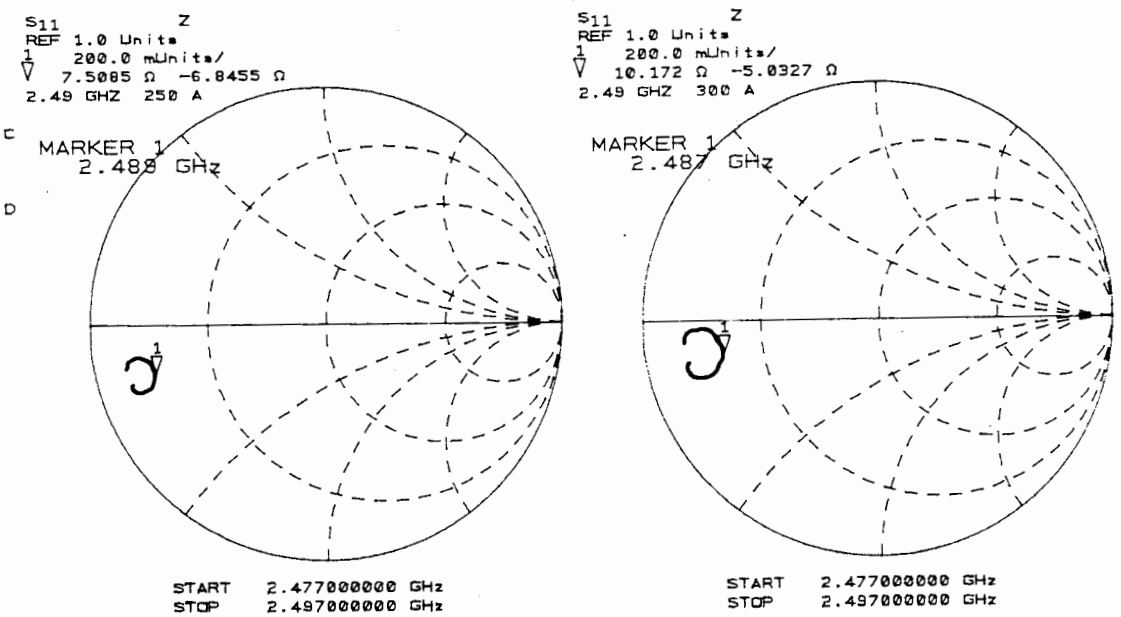


a) 250 Å b) 300 Å

Figure 3: Frequency responses of type B filters with 250 and 300 Å metallization thickness

The reduced insertion loss is mainly not a consequence of smaller pad and finger resistances. Figures 4a and 4b show that with 300 Å the "circle" is wider, i.e. the radiation resistance has increased. That is the cause for the decrease in insertion loss.

This result defeats predictions that the shear horizontal wave is no longer bound to the surface at frequencies above



a) 250 Å b) 300 Å

Figure 4: Input impedances of type B filters with 250 and 300 Å metallization thickness

the stopband of the grating (in our case the grating is formed by the transducers; the stopband frequency is close to 1.5 GHz). It implies that the piezoelectric coupling of the STW increases with increasing layer thickness [9], even at frequencies above the stopband. It should be repeated here that conversion to shear horizontal bulk waves at the finger edges cannot be prevented. So the character of the STW becomes similar to a leaky surface wave. The similarity of the wave to the bulk shear wave directly above the stopband [10,11] can in our opinion be attributed to the peak in the mode conversion ratio at the so-called enter frequency [12]. The enter frequency f_e is the harmonic mean of the surface wave velocity v_{saw} and the velocity of the slowest bulk wave v_{bulk} which contributes to the surface wave divided by twice the grating period p_{gr}, i.e.:

$$f_e = \frac{v_{saw} \cdot v_{bulk}}{v_{saw} + v_{bulk}} \cdot \frac{1}{p_{gr}}$$

In the case of the STW, which is only slightly slower than the bulk shear wave, the enter frequency is close to the bragg frequency, i.e. the stopband frequency of the grating.

The lower insertion loss of the filters with 300 Å metallization thickness results also in reduced suppression of double and triple transit signals and enlarges the ripple

in the passband. Compared to these effects, distortions by electromagnetic feedthrough and diffraction are very small.

IV.C Type C filters

The second high-Q filter to be presented is the 2.65 GHz filter. The measured loaded Q value of these filters is as high as for the 2 GHz filters, i.e. 3400 (the group delay time is 410 ns). The untuned insertion loss is 25 dB. The Smith chart in figure 5 shows that the loss can be reduced by matching or by choosing lower line impedances.

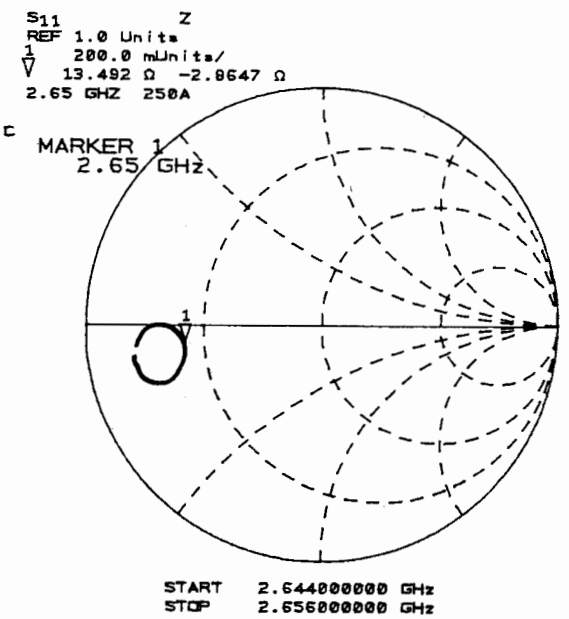


Figure 5: Impedance of a filter of type C

Electromagnetic feedthrough is a general problem in high frequency SAW filters. Capacitive components from the

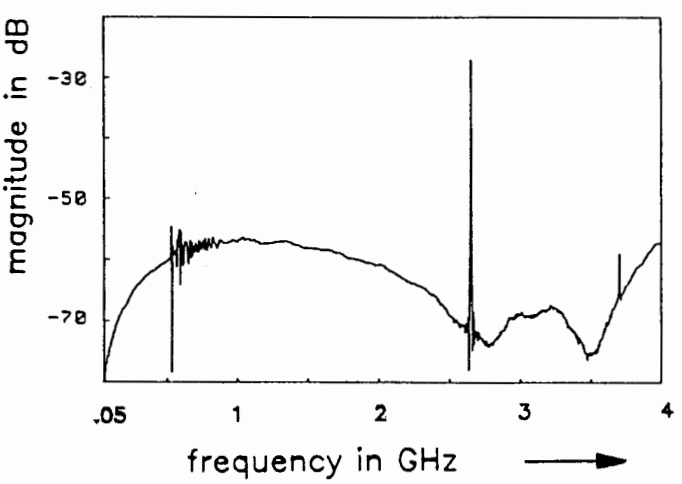


Figure 6: Broadband frequency response of a filter of type C

bonding pads and inductive components from the transducer fingers and the bonding wires add and may cause a high spurious level and even distortions of the passband. But because the phase of the inductive crosstalk depends on the geometry of the layout it can be designed to be phase opposite to the capacitive one. This technique was used in all our designs. Figure 6 shows as an example the measured broadband frequency response of a filter of type C. Below 2.5 GHz the capacitive crosstalk is dominant; above 3.5 GHz inductive coupling dominates. In between there is a broad region where both components interfere destructively.

IV.D Type D filters

The frequency response of a type D filter is plotted in figure 7. The insertion loss of filters mounted in TO-8 packages is 25 dB. It is 10 dB higher than of type B with 250 Å, due to the almost 50 % lower relative coupling coefficient. As the double and triple transit signal suppression is also higher, the frequency response is very smooth. For the investigation of the reproducibility filters of this type were mounted into TO-39 packages to reduce the influence of the bonding wires. Due to an increase of the electrical mismatch the insertion loss increased by about 5 dB. But this measure led to a very high reproducibility. The standard deviation of the centre frequency on one wafer is only 26 ppm. This value is excellent for filters at 2.8 GHz.

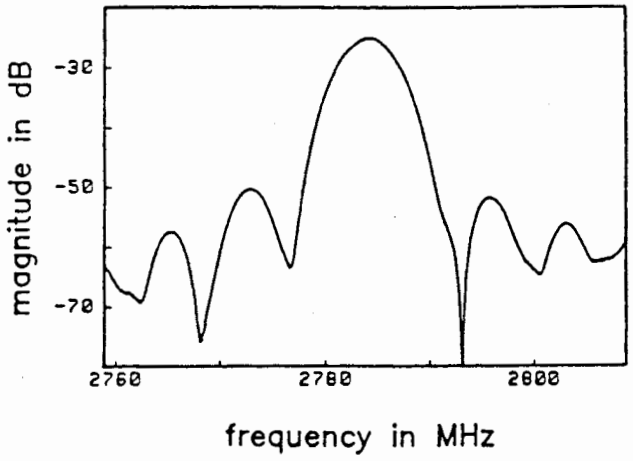


Figure 7: Measured frequency response of a filter of type D

IV.E Statistical results

For applications in either oscillator or timing recovery circuits low scattering of the electrical parameters is of outstanding importance. Table 2 is a comprehensive description of the measured mean values and standard deviations of the most important parameters of all nine wafers. It can be seen that variations, especially of the centre frequency and even from wafer to wafer, are very small. The table demon-

strates the high degree of reproducibility which is achievable using standard photolithography. It can still be improved if the inductivity of the bonding wires is maintained more accurately and by improving the photolithography.

type	centre freq. mean [MHz]	σ [ppm]	insert. loss mean [dB]	σ [dB]	3 dB-bandw. [MHz]	group delay [ns]	cut angle [°]	Al h. [Å]
A	1979.190	45	21.27	0.17	1.25	545	37.5	250
A	1979.093	31	21.21	0.26	1.24	547	37.5	250
A	1979.122	54	21.44	0.45	1.19	550	37.5	250
A	1981.331	23	21.61	0.38	1.24	539	37.0	250
A	1980.712	53	21.95	0.43	1.18	536	37.0	300
B	2489.582	38	15.06	0.29	4.55	148	36.0	250
B	2487.330	48	12.52	0.37	4.48	141	36.0	300
C	2650.705	46	25.34	0.40	1.75	410	36.0	250
D	2784.730	26	30.32	0.37	5.37	155	36.0	200

Table 2: Measured filter parameters (σ = standard deviation)

V. OSCILLATOR DESIGN AND PERFORMANCE

SAW oscillators permit direct generation of GHz frequencies without the need for multiplication circuits [13]. Thus, the phase noise increase inherent in the multiplication process is avoided. To study the performance achievable with high-Q STW filters the Institut für Hochfrequenztechnik of the Technische Universität München developped a 2 GHz STW oscillator.

The hybrid oscillator was developped for type A filters. Its block diagram is shown in figure 8. The oscillator was made in microstrip technology on an RT-Duroid substrate measuring

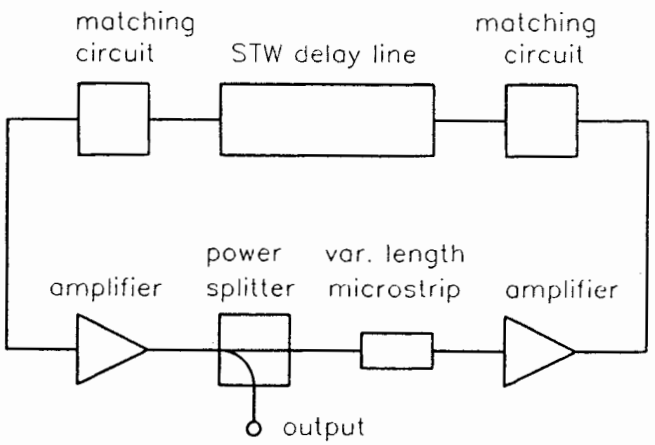


Figure 8: Block diagram of 1.98 GHz STW oscillator

1x2 inches. Two identical silicon bipolar low noise MMIC amplifiers (Avantek MSA-0835) with a gain of almost 18 dB were used. The output coupler is a 3 dB Wilkinson divider. A more detailed description of the oscillator is given in [14].

The oscillator's phase noise characteristic was studied theoretically [15] and experimentally. Figure 9 shows the measured single side band phase noise between 1 Hz and 300 kHz offset from the carrier. The measurement was realized with an HP3048A phase noise analyzer using the frequency discriminator method. In our case it is accurate for modulation frequencies up to 300 kHz. At 1 kHz offset from the carrier the single sideband phase noise is -100 dBc/Hz. The output power is 6.5 dBm, the calculated noise floor is at -158 dBc/Hz. These are good results considering the high oscillation frequency of almost 2 GHz.

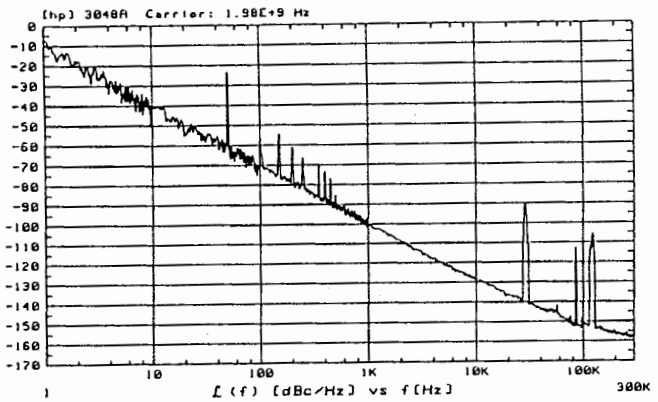


Figure 9: Measured single sideband phase noise of 1.98 GHz STW oscillator

VI. CONCLUSION

The advantageous characteristics of acoustic Surface Transverse Waves on rotated Y-cut quartz substrates are high propagation velocity and purely horizontal polarization. They can be used for low loss, high-Q SAW filters at frequencies up to 3 GHz. With standard photolithography the scattering of the electrical filter parameters can be made sufficiently small for oscillator and timing recovery applications. The achievable high Q factors make direct frequency synthesis with low phase noise at GHz frequencies possible. Further improvements on the filter performance, especially concerning propagation loss and Q factor, are to be expected when operation at fundamental frequencies is feasible on a production scale. This requires photolithographic techniques with high reproducibility and a linewidth resolution of less than 0.5 microns.

ACKNOWLEDGEMENTS

The authors would like to thank L. Eichinger, Professor P. Russer and R. Weigel for development of the oscillators, H. Angerer, B. Kamp and U. Knauer for fabrication of the filters and W. Pahl for help with the measurements.

REFERENCES

[1] B. Fleischmann, W. Ruile and G. Riha, "Rayleigh-Mode SAW-Filters on Quartz for Timing Recovery at Frequencies above 1 GHz", 1986 IEEE Ultrasonics Symposium Proceedings, pp. 163-167

[2] M. Lewis, "Surface Skimming Bulk Waves", 1977 IEEE Ultrason. Symp. Proc., pp. 744-752

[3] K. F. Lau, K. H. Yen, A. M. Kong and K. V. Rousseau, "SBAW Oscillators at 3 to 5 GHz Range", 1983 IEEE Ultrason. Symp. Proc., pp. 263-266

[4] W. Ruile, F. Müller and G. Riha, "Metal RACs with Track Interference Weighting", 1986 IEEE Ultrason. Symp. Proc., pp. 147-151

[5] B. Fleischmann, W. Ruile, G. Riha and A. R. Baghai-Wadji, "Reproducible SAW-Filters at 2.5 GHz", 1987 IERE International Conference on Frequency Control and Synthesis, pp. 29-35, 1987

[6] B. A. Auld and B.-H. Yeh, "Theory of Surface Skimming SH Wave Guidance by a Corrugated Surface", 1979 IEEE Ultrason. Symp. Proc., pp. 786-790

[7] A. J. De Vries, R. L. Miller and T. J. Wojcik, "Reflection of a Surface Wave from three Types of Interdigital Transducers", 1972 IEEE Ultrason. Symp. Proc., pp. 353-358

[8] H. Engan, "Surface Acoustic Wave Multielectrode Transducers", IEEE Transactions on Sonics and Ultrasonics, Nov. 1975, pp. 395-401

[9] T. Nishikawa, A. Tani, C. Takeuchi and J. Minowa, "1.6 GHz SH-Type Surface Acoustic Wave Filter", Japanese Journal of Applied Physics, 1981, Supplement 20-4, pp. 29-32

[10] D. F. Thompson and B.A. Auld, " Surface Transverse Wave Propagation under Metal Strip Gratings", 1986 IEEE Ultrason. Symp. Proc., pp. 261-266

[11] C. A. Flory and R. L. Baer, "Surface Transverse Wave Mode Analysis and Coupling to Interdigital Transducers" to be published in 1987 IEEE Ultrason. Symp. Proc.

[12] M. N. Islam, H. A. Haus and J. Melngailis, "Bulk Radiation by Surface Acoustic Waves Propagating under a Grating", IEEE Trans. on Son. and Ultras., March 1984, pp. 123-135

[13] K. F. Lau, K. H. Yen, R. S. Kagiwada and A. M. Kong, "High Frequency Temperature stable SBAW Oscillators", 1980 IEEE Ultrason. Symp. Proc., pp. 240-244

[14] L. Eichinger, B. Fleischmann, P. Russer and R. Weigel, "A 2 GHz Surface Transverse Wave Oscillator with low Phase Noise", to be published in 1988 IEEE MTT-S International Microwave Symposium Proceedings

[15] T. E. Parker, "1/f Phase Noise in Quartz Delay Lines and Resonators", 1979 IEEE Ultrasonics Symposium Proceedings, pp. 878-881

RO-05

Oscillators

1. Microwave Oscillator Using Partly Metallized Dielectric Resonators
   M. Valentin, R. El Cheikh
   CNRS, Besançon - France

2. Thermal Hysteresis in Quartz Resonators and Oscillators
   J. Beaussier
   ONERA, Châtillon - France

3. Miniature Oscillators
   K. Schwaller, K. Schulthess, W. Zingg
   Micro Crystal, Grenchen - Switzerland

2 ND EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 16 - 17 - 18 MARCH 1988

Session RO-05 Number 1

MICROWAVE OSCILLATOR USING PARTLY METALLIZED DIELECTRIC RESONATOR

Michelle VALENTIN et Rida E. EL CHEIKH

Laboratoire de Physique et Métrologie des Oscillateurs du CNRS
associé à l'Université de Franche-Comté-Besançon
32, avenue de l'Observatoire - 25000 Besançon - France

Abstract

New stabilized UHF integrated oscillators using a partly metallized dielectric resonator in the feedback circuit, are presented in frequency band (800 / 1 000 MHz) for radio-telephone and future land mobile satellite communication systems. With these techniques a frequency stability $\sigma_y(\tau)$ of less than 4.10^{-9} for τ = 1 s without oven, an output power of more 16 dBm, a mechanical tuning over 51 MHz and an electrical tuning over 1 MHz are obtained.

1. INTRODUCTION

In modern communication systems, Microwave Integrated Circuits (MIC's) with small dimensions, low weight, low cost, are required. The performance requirements of these systems can only be satisfied through the development of stable, very low phase noise microwave sources. Until recently, these requirements were met by the use of large Invar cavities in the oscillator circuits. Now, MIC technology uses bipolar transistors, field effect transistors, Gunn diodes, Impatt diodes interconnected by microstrip lines. The size of metallic cavity is not compatible with the structure of MIC. With the recent availability of new microwave dielectric with large relative dielectric constant ε_r, large Q factor, high temperature stability, large cavities are replaced by dielectric resonators (DR's) : these functional ceramic devices trap the microwave energy : greater the relative dielectric constant, larger becomes the trapped energy.

The composition used (Zr, Sn)TiO_4 developped by Thomson-CSF Corbeville Laboratories [1] offers an adjustable temperature coefficient of resonance frequency within 1 ppm/ °C between -20 and +20 ppm/ °C, low microwave loss and high ε_r (typically 37).

Stabilized microwave oscillators using dielectric resonator have been reported by many authors [2][3][4] in the optimum frequency range 4 - 25 GHz. The size of the final oscillator is determined by the size of resonator which is not compatible with MIC below 2 GHz.

In the range of UHF frequencies, from 300 MHz to 3 GHz, in order to avoid this disadvantage, some size reduction is achieved by using quarter wave dielectric rectangular resonators, with three metallized faces ; we report herein the performance of new dielectric resonator oscillators (DRO's) constructed using these partly metallized dielectric resonators.

2 - DIELECTRIC RESONATOR FEEDBACK CIRCUIT

Microwave oscillators are classified according to the oscillation mechanism into two typical types, the negative resistance type and the feedback type [5]. For the simplicity we adopt a feedback type ; the circuit is composed of a feedback element, the dielectric resonator (DR) between the output port and the input port of an amplifier. The amplifier stage G is a broad band monolithic amplifier (21 dB gain, noise figure : 5.5 dB, 50 Ω input/output impedance) easily cascadable.

Fig. 1 shows two basic configurations proposed in this paper : a two-port dielectric resonator oscillator and a four-port dielectric resonator oscillator. The auxiliary output of four-port dielectric resonator oscillator can be used for electronic varactor tuning.

Fig. 1 : Block-diagram of dielectric resonator oscillator
1a - two-port dielectric resonator 1b - four-port dielectric resonator

The circuit is fabricated on a dielectric substrate (RT Duroid 6010) of 1.27 mm thickness and a $ZrSnTiO_4$ resonator is placed in the vicinity of two 50 Ω microstrip lines. The microstrip lines are magnetically coupled through the dielectric resonator. The resonator frequency can easily be changed by adjusting an air-gap spacing e between the resonator and a frequency tuning plate (Fig. 2).

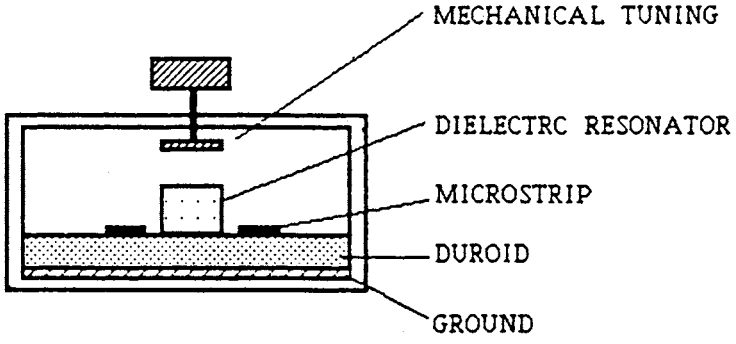


Fig. 2 : A representative configuration for mechanical tuning of a dielectric resonator

A low relative dielectric constant material (polyfoam) can be placed between resonator and Duroid to raise the resonator above the ground plane and increase the loaded Q of resonator. The housing (conductive wall) of the circuit forms a waveguide-below-cutoff which behaves like a pure inductance and prevent radiation losses and spurious coupling.

The resonator scattering parameters of several dielectric resonators are measured using an automatic network analyser. The insertion losses are adjusted easily (5 to 20 dB) with the coupling constant between resonator and microstrip by changing the resonator position.

The $\lambda/4$ resonators used have an unloaded Q (Q_0) of 600 measured with the additional conductor losses of three metallized faces (two opposed faces and one extremity) : the size reduction is realized but metallic losses increase considerably, since the surface area in which the current flows are reduced by ε_r.

Resonant frequencies of dielectric rectangular resonator of dimensions 10 mm × 10 mm × 18 mm and 7 mm × 8 mm × 16,5 mm are 0.88 GHz and 1,01 GHz respectively.

3. CHARACTERIZATION OF DIELECTRIC RESONATOR OSCILLATOR AND PRELIMINARY RESULTS

Bias dependence

The dc bias circuit plays an important role in the performance of the oscillator.

Fig. 3 presents output power P, oscillation frequency f as a function of bias voltage V_B at ambient temperature. The output power increases with bias voltage. A pushing figure of -575 kHz/V is obtained for voltage range from 8 V to 11 V.

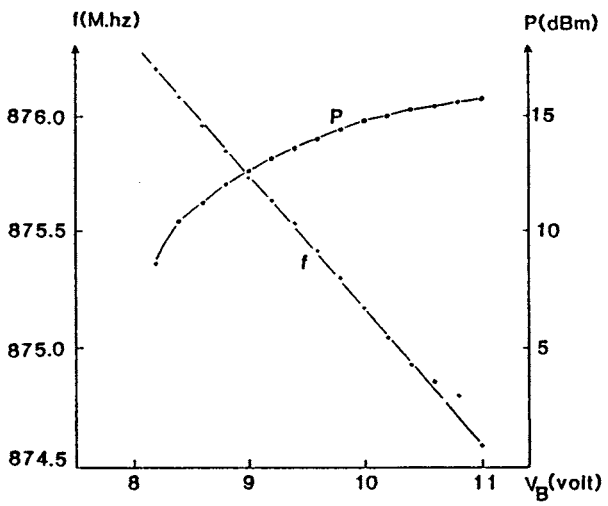


Fig. 3 :
Oscillation characteristics of DRO as a function of bias voltage V_B

Mechanical tuning performance

Fig. 4 shows mechanical tuning characteristics as a function of the air-gap spacing "e". The tuning of range over 51 MHz is obtained. The output variation is less 1 dB.

Electrical tuning performance

A voltage controlled dielectric resonator oscillator (VCO) is built by coupling a varactor diode to the dielectric resonator. The oscillator is tunable over 1.146 MHz with a 0 to -30 V varactor voltage ; variations in output power are negligeable. For voltage range from -10 V to -30 V a frequency drift of-30 kHz/V is found. Fig. 5 shows electrical tuning characteristics as a function of varactor voltage.

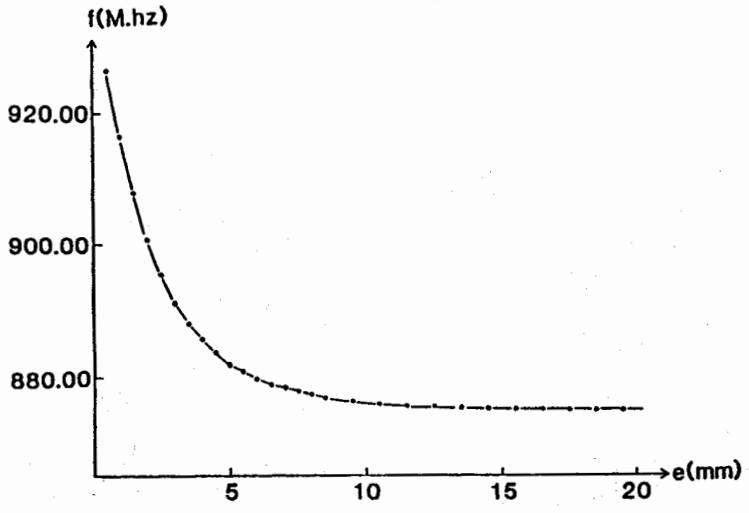


Fig. 4 : Mechanical tuning characteristics as a function of air-gap thickness (e)

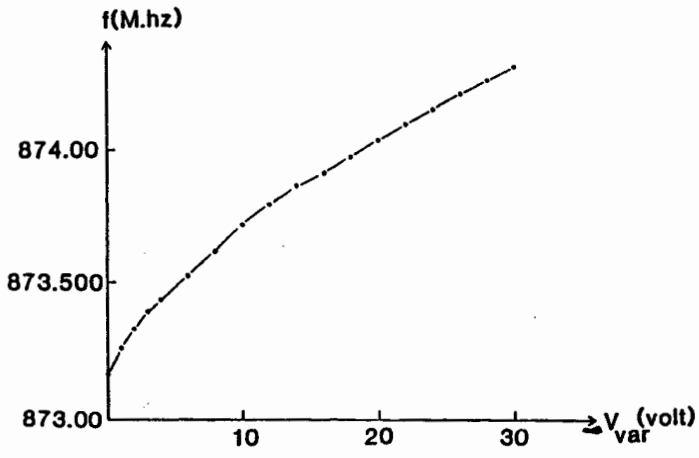


Fig. 5 : Electrical tuning of DRO as a function of varactor voltage

Temperature sensitivity

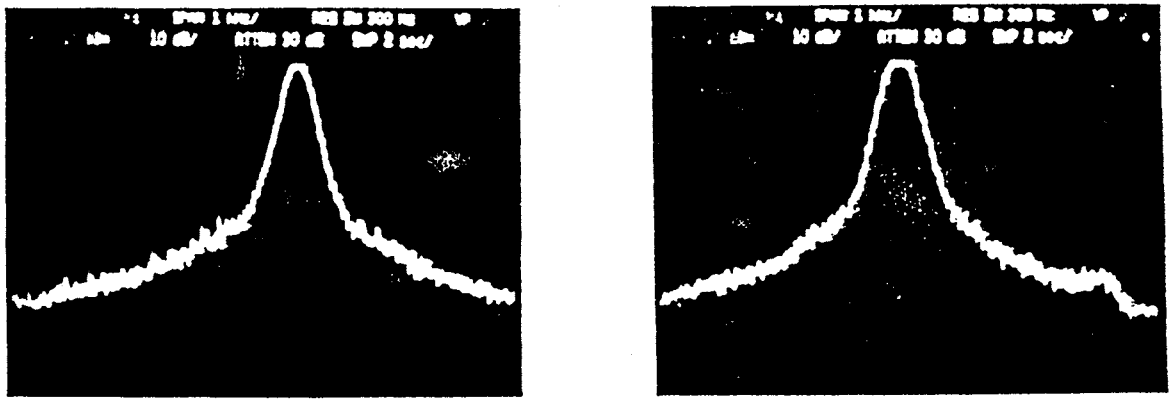
The characterization of DRO's temperature sensitivity is important for the design of temperature stable oscillators.

The frequency drift is found to be approximately -4 ppm / °C for temperature range from 20 °C to 50 °C.

Spectrum and noise

The influence of resonator configuration and coupling coefficient on the spectral shape are important. Fig. 6 presents spectra of a DRO with and without electrical tuning, measured with a commercial spectrum analyser.

The bandwidth was set at 300 Hz, the horizontal scale is 1 kHz/division and the vertical scale is 10 dB/division. At a frequency 1 kHz away from the carrier, the carrier to noise ratio is 40 dBc.



a b

Fig. 6 : UHF DRO output spectrum (SPAN : 1 kHz/ RES BW 300 Hz REF : +16 dBm 10 dB/ ATT : 30 dB SWP 2 sec/)

a) without electrical tuning

b) with electrical tuning

The performance of frequency sources is usually described in terms of accuracy, reproductibility and stability. The characterization of the stability is usually the most important information to the user.

When the oscillator is carefully designed, the dominant source of oscillator noise appears to be the oscillator electronic, the contribution due to dielectric resonator being negligeable [6][7].

Oscillator stability has been measured in the time domain by evaluating the fractional frequency fluctuations $\langle\sigma_y(\tau)\rangle$ versus the averaging time τ and expressed as Allan's variance [8].

Fig. 7 presents the short term stability for a four-port dielectric resonator oscillator without oven. Frequency stability is 3.5×10^{-9} in the range of 0.01 sec to 10 seconds.

The other device (two-port resonator in the feedback circuit) presents a frequency stability of 5.5×10^{-9} in the same range.

A compensation for temperature is studied now for frequency stability improvement.

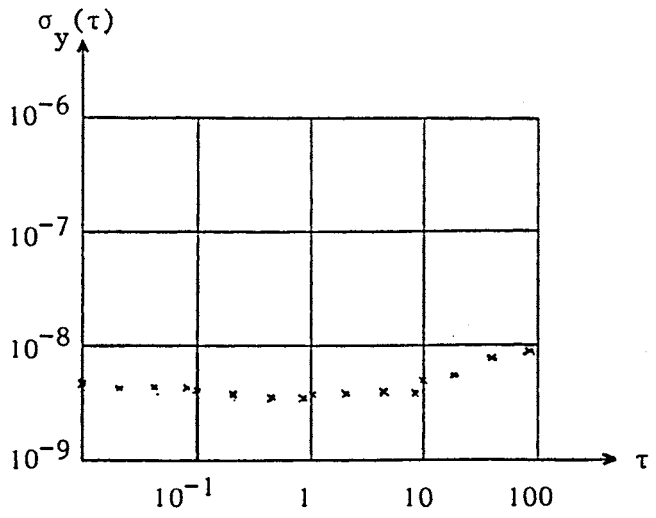


Fig. 7 : Short term frequency stability of DRO

4. CONCLUSION

A new type of UHF oscillator with a partly metallized dielectric resonator in the feedback circuit has been proposed and performances of the oscillator have been described. This compact stabilized oscillator exhibited 50 mW output power in the band 800 - 1 000 MHz, a frequency stability $\sigma_y(\tau)$ of 3.5×10^{-9} in the range of 0.01 second to 100 seconds, an easy mechanical or electrical tuning. This new type of oscillator is simple and compact, therefore, has wide applications for various UHF systems.

REFERENCES

[1] J.C. MAGE, B. MARCILHAC, "Amélioration de la stabilité en température des matériaux pour résonateurs diélectriques", Actes des 5èmes Journées Nationales Microondes, Nice, p. 41-43, 1987.

[2] H. ABE, Y. TAKAYAMA, A. HIGASHISAKA, H. TAKAMIZAWA, "A highly stabilized low-noise GaAs FET integrated oscillator with a dielectric resonator in the C-band", IEEE Trans. Microwave Theory Tech., vol. MTT-26, pp. 156-162, 1978.

[3] C. TSIRONIS, D. HENNINGS, "Highly stable FET DROs using new linear dielectric resonator material", Electronics Letters, vol. 19, n° 18, pp. 741-743, 1983.

[4] E.C. NIEHENKE, R.D. HESS, "A microstrip low-noise X-band voltage-controlled oscillator", IEEE Trans. Microwave Theory Tech., vol. MTT-27, pp. 1075-79, 1979.

[5] M.J. LOBODA, T.E. PARKER, G.K. MONTRESS, "Frequency stability of L-band, two-port dielectric resonator oscillators", IEEE Trans. Microwave Theory Tech., vol. MTT-35, pp. 1334-38, 1987.

[6] M. VALENTIN, "Bruit de phase et de fréquence de résonateurs diélectriques", In "Interaction of Radiation with Matter", a volume in honour of Adriano Gozzini - Scuolo Normale Superiore - PISE, pp. 174-185, 1987.

[7] M. OLIVIER, J. GROSLAMBERT, M. VALENTIN, "Mesure du bruit de phase, par des méthodes actives et passives de sources microondes de haute pureté spectrale, Actes des 5èmes Journées Nationales Microondes, Nice, pp. 56-58, 1987.

[8] L.S. CUTLER, C.L. SEARL, "Some aspects of the theory and measurement of frequency fluctuations in frequency standards", Proc. IEEE, vol. 34, n° 2, pp. 136-154, 1966.

2 ND EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 16 - 17 - 18 MARCH 1988

Session RO-05 Number 2

HYSTERESIS THERMIQUE DANS LES RESONATEURS ET LES OSCILLATEURS A QUARTZ

J. Beaussier
Office National d'Etudes et de Recherches Aérospatiales
BP 72, 92322 Châtillon Cedex, France

ABSTRACT

First part of the paper is devoted to the definition of static thermal hysteresis, to the description of encountered difficulties for an accurate measurement of that sort of disturbance and to the presentation of experimental results. In the second part, main physical phenomena able to induce thermal hysteresis are investigated for the quartz itself. It is shown that the defect appears on elastic coefficients. Mechanical and thermal hysteresis have many similar aspects which can be used to formulate some assumptions about physical phenomena. Non linear elasticity and moving dislocations are involved. Third part deal with possibility to reduce thermal hysteresis. In addition to a previous quartz selection, interest of thermal processing is discussed.

RESUME

La première partie de l'exposé définit l'hystérésis thermique statique, expose les difficultés rencontrées pour évaluer ce type de perturbation et présente quelques résultats expérimentaux. La seconde partie analyse les phénomènes susceptibles de provoquer l'hystérésis thermique sur le quartz lui-même, ce défaut devant être attribué au coefficient élastique du matériau. Il existe une étroite similitude entre l'hystérésis mécanique et l'hystérésis thermique et c'est à partir de cette similitude que des hypothèses ont pu être formulées sur l'origine de l'hystérésis thermique du résonateur. Des phénomènes d'élasticité non linéaire au niveau des défauts cristallins et particulièrement des dislocations doivent être mis en cause. La troisième partie traite des possibilités de réduction de l'hystérésis thermique et de l'intérêt d'une sélection préalable du quartz ainsi que de traitements thermiques pour obtenir ce résultat.

1 - INTRODUCTION

Les critères de réversibilité et d'irréversibilité des dérives des oscillateurs à quartz présentent un grand intérêt et sont très exploités à l'ONERA. Les influences des températures, des accélérations, des pressions, des champs magnétiques, peuvent le plus souvent être considérées comme réversibles : la fréquence y reprend sa valeur initiale lorsque cesse l'action extérieure. La notion d'irréversibilité, typiquement définie par les dérives temporelles, doit être étendue soit à l'action de ces mêmes paramètres agissant par leurs niveaux élevés, soit à des paramètres d'une autre nature, par exemple les rayonnements ionisants.

Mais, en tout état de cause, la réversibilité n'est qu'une approximation. La thermodynamique montre depuis longtemps que tous les processus réels sont irréversibles à une échelle suffisamment petite. La résolution très élevée des mesures de fréquence permet actuellement de confirmer cette déduction.

L'hystérésis thermique caractérise la part d'irréversibilité des dérives du même nom. Les modèles thermiques couramment exploités pour les résonateurs à quartz n'introduisent pas ce phénomène qui, selon plusieurs auteurs, échappe à toute modélisation. Mais l'hystérésis s'est avérée importante et limitative soit pour les oscillateurs de précision à compensation numérique directe [1], soit pour des capteurs à fréquence variable utilisant les propriétés du quartz [2]. La présente communication résume des expérimentations et quelques réflexions faites à l'ONERA sur ce thème. Les difficultés qu'implique la mesure de l'hystérésis font l'objet d'une première partie. La seconde partie traite des phénomènes physiques susceptibles, au niveau du résonateur, de provoquer le phénomène. Les moyens de réduire l'hystérésis sont évoqués en guise de conclusion.

2 - DEFINITION ET CARACTERISATION EXPERIMENTALE DE L'HYSTERESIS THERMIQUE

Les phénomènes d'hystérésis thermique décelables dès qu'un résonateur à quartz est soumis à des cycles thermiques fermés dont l'amplitude dépasse quelques degrés sont néanmoins complexes et difficiles à caractériser avec précision. Un grand nombre de causes font que la fréquence ne retrouve pas la même valeur après un cycle qui fait revenir une variation de température à sa valeur initiale. La figure 1 montre en 1b la réponse d'un résonateur à quartz à un créneau de température obtenu par variation du point de consigne du thermostat et appliqué conformément à 1a. La mesure de la température est faite au plus près du résonateur et si possible du quartz en utilisant, par exemple, le mode thermométrique de la coupe SC. La droite A est la dérive temporelle supposée linéaire et indépendante de la température dans le domaine exploré. En O, dès les premières variations, apparaît la sensibilité thermique dynamique proportionnelle à la dérivée dT/dt. La mesure de la dérive thermique statique correspondant à la température T_M du palier doit attendre la fin de ce phénomène. Lorsque la température revient à T_0, il faut agir de même pour la décroissance $T_M - T_0$, avant la mesure de l'hystérésis proprement dite, en D sur la figure. Les temps de mesure peuvent, dans ces conditions, atteindre plusieurs heures et une correction de dérive temporelle s'avère nécessaire.

D'autres causes d'erreurs sont liées à la précision des mesures de température et, en particulier, à celle du retour aux conditions initiales. Ces erreurs peuvent être statiques ou dynamiques. En dynamique, si la mesure de la température est faite sur le boîtier du résonateur, il faut tenir compte de la constante de temps boîtier-quartz dont une valeur courante est de quelques dizaines de seconde mais qui peut dépasser 1000 s pour certains résonateurs. En statique, les capteurs de température apportent leurs erreurs propres et les gradients thermiques empechent la définition d'une température exacte du quartz. Dans ces conditions, il apparaît évident que la mesure de l'hystérésis thermique nécessite des équipements de précision élevée associés à un traitement numérique complexe des résultats. Ces mesures ne peuvent être effectuées que sur des résonateurs de haute qualité à dérive temporelle réduite (10^{-10} ou 10^{-11}/jour).

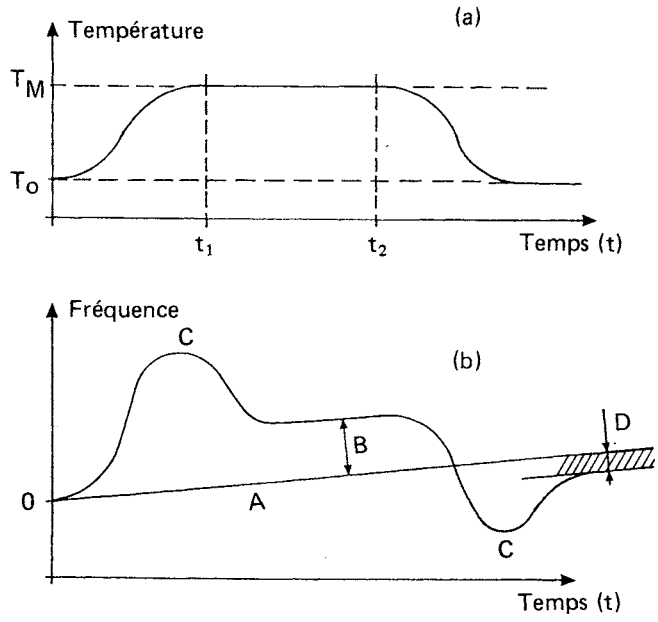


Fig. 1 - Dérive temporelle et dérives thermiques avec hystérésis.
A - Dérive temporelle ; B - Dérive thermique pour le palier T' ;
C - Sensibilités thermiques dynamiques ; D - Hystérésis thermique.

L'utilisation de diagrammes fréquence température donne une représentation plus précise des phénomènes. On rappelle que tous les oscillateurs performants fonctionnent à la température de leur point d'inversion (Fig. 2) et que leur diagramme fréquence température présente une forme parabolique au voisinage de ce point. La concavité de la parabole est tournée vers le bas pour une coupe SC et vers le haut pour une coupe AT. Des cycles thermiques d'amplitude totale ± 2,5°C, ± 5°C, ± 10°C sont effectués de part et d'autre du point d'inversion. Deux modes opératoires peuvent être utilisés pour ces balayages ; le premier avec un balayage sinusoïdal à fréquence très basse, par exemple 10^{-4} Hz. L'hystérésis thermique, atteinte lorsque les cycles ne se rapprochent plus de la courbe statique (Fig. 2a), est indiscernable si la résolution de la mesure n'est pas suffisante. Le second mode opératoire met en oeuvre un balayage incrémental de la température [1]. Il y a 5 incréments par trajet pour chacune des gammes précitées dont les amplitudes sont, par conséquent, de 0,5°C, 1°C et 2°C. Pour chaque incrément, la durée de la mesure peut atteindre de 2 à 4 h. Ce mode opératoire qui s'est avéré supérieur au précédent est représenté sur la figure 2b. Le passage de l'un à l'autre des 5 points se fait par un trajet en festons imposé par les caractéristiques dynamiques, particulièrement en coupe AT. Le cycle limite atteint de part et d'autre de la courbe statique caractérise l'hystérésis thermique qui peut alors être définie par l'écart ΔF_M entre les deux courbes à la température T_1 d'inversion (Fig. 3). Une constante de temps boîtier-quartz excessive par rapport à la sensibilité dynamique peut masquer celle-ci. Cette constante de temps produit aussi des festons mais une différence importante peut être constatée avec le cas précédent : les festons dus à la constante de temps ont au départ une tangente horizontale et passent de l'autre côté de la courbe au point d'inversion.

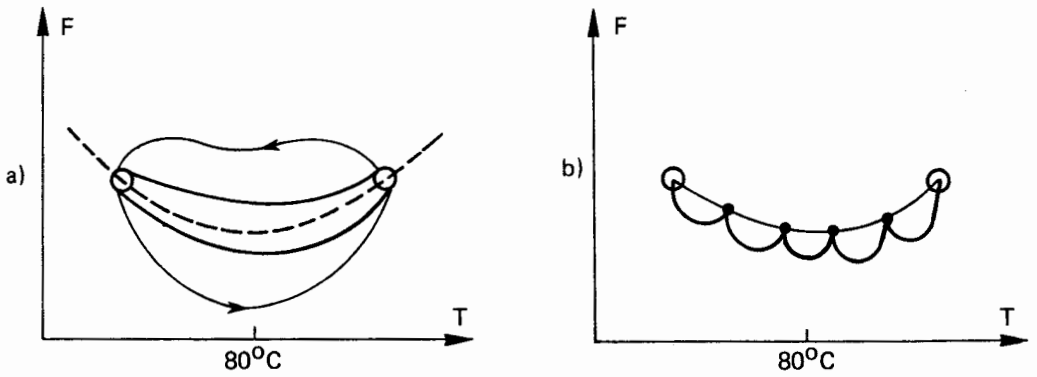


Fig. 2 - Diagramme fréquence/température et hystérésis sur résonateur de coupe AT. (a) - en variation thermique sinusoïdale ; (b) - en incrément de température.

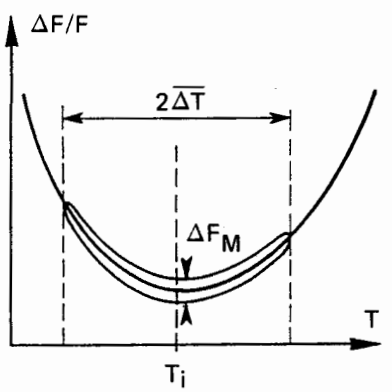


Fig. 3 - Cycle limite d'hystérésis.

Sans entrer dans le détail des calculs nécessaires à l'acquisition et au traitement des mesures, on peut signaler que des approches polynomiales de degré 2 ou 3, par moindres carrés ou par Tchebycheff ont été simultanément utilisées. On obtient les valeurs moyennes ΔT des balayages thermiques montants et descendants ainsi que les écarts correspondants ΔF_M des fréquences d'inversion F_i et des coefficients thermiques du second ordre b_i liés à la concavité des courbes représentées sur la figure 3.

3 - MODELISATION DE L'HYSTERESIS THERMIQUE

3.1 - Modélisation globale

L'écart maximal $\Delta F_M/F$ en fonction de ΔT peut être modélisé par une relation quadratique :

$$\frac{\Delta F_M}{F_i} = K_H \, \Delta T^2$$

Le coefficient K_H qui caractérise par conséquent l'hystérésis thermique est par ailleurs égal à la variation du coefficient thermique du second ordre. En effet, la caractéristique fréquence-température du quartz s'exprime autour du point d'inversions par :

$$\frac{\Delta F_M}{F_i} = b_i (T - T_i)^2$$

On passe de la caractéristique statique théorique aux courbes d'hystérésis montante et descendante en introduisant, dans cette relation, une variation $\pm \delta F_i$ de F_i et $\pm \delta b_i$ de b_i. On obtient dans ces conditions :

$$\left(\frac{\Delta F}{F_i}\right)_1 = + \frac{\delta F_i}{F_i} + (b_i - \delta b_i)(T_M - T_i)^2$$

et

$$\left(\frac{\Delta F}{F_i}\right)_2 = - \frac{\delta F_i}{F_i} + (b_i + \delta b_i)(T_M - T_i)^2$$

L'intersection des deux courbes donne la relation :

$$\frac{\delta F_i}{F_i} = \delta b_i \ (T_M - T_i)^2 = K_H \ \Delta T^2 \qquad \text{avec } K_H = \delta b_i$$

Pour tous les types de résonateurs ayant fait l'objet de mesures à l'ONERA, les coefficients K_H qui s'expriment en $\Delta F/F/°C^2$ se situent entre 0,5 et 1.10^{-10} pour une coupe SC et entre 1 et $1,5.10^{-10}$ pour une coupe AT. La courbe de la figure 4 a été tracée avec $K_H = 1,5.10^{-10}$ et une excursion maximale $\Delta T = 10°C$ [1]. Avec une excursion de $\pm$ 50°C, les écarts de fréquence dus à l'hystérésis thermique atteindraient la valeur de $3,75.10^{-7}$ et ne seraient que trois fois inférieurs avec la meilleure valeur obtenue pour K_H. L'hystérésis thermique est par conséquent importante et rapidement prohibitive pour les oscillateurs non thermostatés. Les valeurs mesurées à l'ONERA sont en bon accord avec celles publiées par ailleurs [3,4].

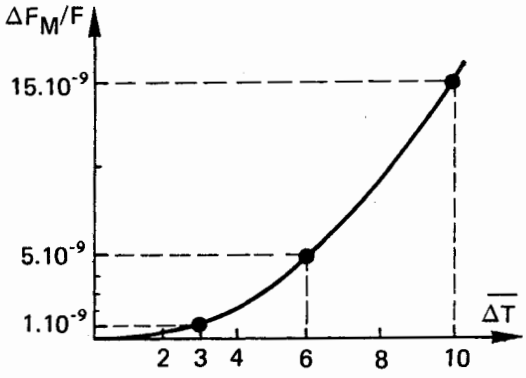


Fig. 4 - Hysteresis en fonction de l'excursion thermique.

3.2 - Paramètres susceptibles de provoquer l'hystérésis thermique

Les paramètres à considérer en priorité sont ceux qui se retrouvent au premier ordre dans l'expression de la fréquence :

$$F = \frac{n}{2e_1}\sqrt{\frac{E_{ij}}{\rho}}$$

- L'épaisseur e_i de la pastille de quartz et la masse volumique ρ ne peuvent être mis en cause. Les coefficients de dilatation sont bien connus, leurs non linéarités très faibles et aucun phénomène d'hystérésis n'est décelable sur les dilatations propres du quartz au cours d'un cycle thermique fermé.

- Il n'en est pas de même des coefficients élastiques E_{ij} dont les variations thermiques au premier et surtout au second ordre sont publiées avec une très large dispersion et pour lesquels des phénomènes d'hystérésis thermique sont constatés. Il apparaît par conséquent judicieux d'introduire seulement les variations des coefficients élastiques dans une modélisation d'hystérésis thermique en précisant toutefois davantage l'expression de ces coefficients [5].

Le coefficient E_{ij} de l'expression générale de la fréquence définit la relation entre les contraintes $\tilde{T}_i$ et les déformations vibratoires $\tilde{S}_j$: $\tilde{T}_i = E_{ij}\ \tilde{S}_j$.

Pour les effets de différentes perturbations agissant sur la fréquence, en particulier les perturbations mécaniques telles les forces, accélérations, pressions, il est bien connu qu'il faut utiliser les coefficients élastiques dits du second ordre, E_{ijk} qui introduisent les déformations statiques $\overline{S_K}$. La relation précédente s'écrit alors :

$$\tilde{T}_i = \left(E_{ij} + E_{ijk}\ \overline{S_K}\right) \tilde{S}_j$$

A partir de cette expression, deux types d'hystérésis thermique peuvent être définis.

- L'hystérésis sur E_{ij} qui sera désignée par <u>hystérésis intrinsèque</u> du quartz. Un cycle thermique fermé provoque un décalage ΔE entre la valeur initiale E_0 et la valeur finale E_i du coefficient élastique :

$$E_0 - E_i = \Delta E$$

  Ce décalage est provoqué par des phénomènes internes au quartz pour lesquels il est possible de formuler des hypothèses qui feront l'objet du paragraphe suivant.

- L'hystérésis sur la déformation statique $\overline{S_K}$ qualifiée d'<u>hystérésis de structure</u>. Bien que les mesures effectuées concernent des résonateurs de haute qualité fabriqués avec soin, il ne faut pas exclure des phénomènes d'élasticité non linéaire au niveau des liaisons supports-quartz. Ces phénomènes sont susceptibles, lors d'un cycle fermé, de provoquer des contraintes et, par conséquent, des déformations $\overline{S_K}$ non parfaitement réversibles. Il s'agit par conséquent d'un effet thermique introduisant de manière indirecte l'hystérésis dans le mécanisme bien connu de la sensibilité du résonateurs aux forces et aux accélérations. Cet effet qui présente quelques similitudes avec l'hystérésis intrinsèque sera rappelé en conclusion.

4 - ETUDE PHYSIQUE DE L'HYSTERESIS INTRINSEQUE

4.1 - Approches énergétiques et mécaniques de l'hystérésis intrinsèque

L'étude physique de l'hystérésis thermique des coefficients élastiques doit être abordée par deux approches distinctes. La première utilise un bilan energétique, la seconde, associée à une transposition d'échelle, utilise des similitudes avec des phénomènes d'élasticité non linéaire.

Pour le bilan energétique, il faut rappeler que si le coefficient élastique ne revient pas à sa valeur initiale après un cycle fermé mécanique ou thermique, il y a variation d'énergie interne du résonateur et cette variation correspond à une énergie dissipée. Les phénomènes physiques qui provoquent l'hystérésis sont par conséquent des phénomènes dissipatifs. Dans un résonateur à quartz, les formes suivantes d'énergie sont à considérer.

- L'énergie mécanique dynamique fournie par l'excitation piézoélectrique. C'est une énergie interne à la mise en route et ensuite une énergie d'entretien qui compense les dissipations vibratoires.

- L'énergie mécanique statique est une énergie interne élastique liée aux contraintes mécaniques du quartz. Une partie de cette énergie est transmise par les supports.

- L'énergie thermique est fournie soit sous forme statique par chauffage extérieur, soit sous forme dynamique par dissipation vibratoire. Elle se répartit en énergie interne ou en énergie dissipée selon les liaisons thermiques de la pastille de quartz, le système pouvant être soit isotherme, soit adiabatique.

Il résulte de ce premier bilan que si un phénomène dissipatif est nécessaire pour provoquer l'hystérésis, l'inverse n'est pas vrai et que seuls ceux de ces phénomènes provoquant une diminution non compensée d'énergie interne statique sont à mettre en cause. Cette propriété sera utilisée pour déterminer la nature des phénomènes physiques susceptibles de provoquer l'hystérésis.

Pour les similitudes thermomécaniques, il faut rappeler que les coefficients élastiques responsables de l'hystérésis thermique apparaissent comme la pente ou la dérivée du diagramme contrainte-déformation. Il est par conséquent judicieux de rechercher des similitudes entre l'hystérésis thermique d'une part et, d'autre part, les très nombreux résultats acquis dans le domaine de l'élasticité non linéaire des solides. Bien que ces résultats correspondent à des niveaux élevés de contraintes et de déformations, on peut montrer que cette similitude est possible.

Deux types de phénomènes principaux présentés respectivement en figures 5a et 5b caractérisent l'élasticité non linéaire. La plasticité se définit par une courbure plus ou moins rapide en AP du diagramme contrainte/déformation et par un retour à contrainte nulle selon la droite PP'. La déformation résiduelle OP' qui en résulte est la forme la plus classique de l'hystérésis mécanique H. La viscosité introduit le temps dans le comportement mécanique. La déformation élastovisqueuse de la figure 5b croît d'abord rapidement, puis de plus en plus lentement en tendant asymptotiquement vers une valeur stable. Si la charge constante est

supprimée le phénomène inverse se produit mais une déformation résiduelle permanente H peut apparaître en fonction de la charge.

Ces notions peuvent être transposées au niveaux très faibles de contraintes mécaniques rencontrés dans les résonateurs à quartz. En effet et d'une manière générale, un comportement parfaitement élastique tel que l'exprime la loi de Hooke n'est qu'une approximation qui idéalise les structures réelles. L'hystérésis mécanique apparaît dès les plus faibles contraintes et dès que la précision et la résolution des dispositifs de mesure s'avèrent suffisantes. Le matériau garde toujours une mémoire de la dernière déformation subie et ce point a déjà été signalé en introduction. La figure 6 traduit cette propriété schématisée "en loupe" autour de O et souvent qualifiée de "microplasticité".

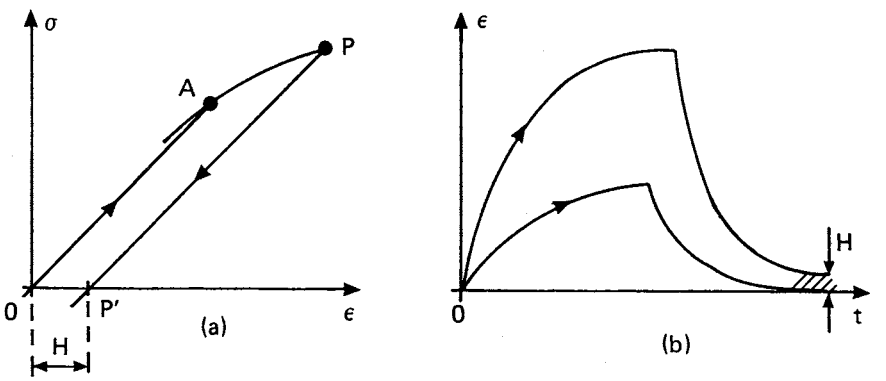


Fig. 5 - Elasticité non linéaire et hystérésis mécanique
(a) - Plasticité ; (b) - Viscosité

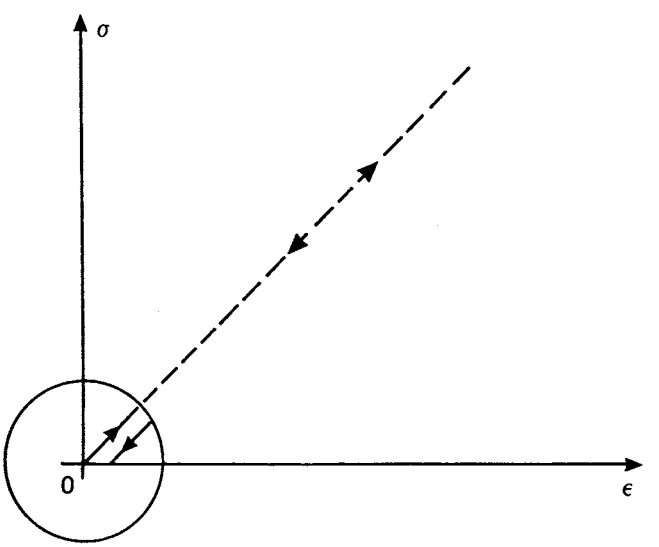


Fig. 6 - Hystérésis mécanique à faible niveau (microplasticité).

Il est évidemment très tentant d'attribuer à cet hystérésis résiduel les mêmes causes et les mêmes effets que ceux constatés aux niveaux élevés avec les phénomènes de plasticité et de viscosité. Il suffit pour cela de supposer que tout matériau élastique, en particulier le quartz, présente un comportement partiellement plastique ou partiellement viscoélastique lié, non au matériau lui-même, mais à ses défauts et à ses inhomogénéités internes. Même si ces imperfections sont en très petit nombre, leur influence sur les propriétés physiques peut être considérable. Ceci n'est qu'une hypothèse, déjà faite pour d'autres phénomènes irréversibles, par exemple la sensibilité aux rayonnements ionisants. Elle peut être appuyée par des considérations énergétiques sur l'état vibratoire du quartz.

4.2 - Phénomènes dissipatifs vibratoires et hystérésis mécanique

La mise en vibration d'un solide tel que le quartz est toujours dissipative et pour qu'il en soit ainsi il faut faire appel aux phénomènes de viscosité et de plasticité précédemment décrits. La figure 7a schématise dans les deux cas la dissipation sur une période qui se représente graphiquement par la surface hachurée comprise entre le diagramme aller et le diagramme retour, ces deux courbes étant forcément différentes en vibration dissipative. Le diagramme dissipatif en vibration viscoélastique de la figure 7a1 présente la forme elliptique classique qui ne justifie en rien le terme d'hystérésis. On rappelle que la dissipation est dans ce cas provoquée par un frottement proportionnel à la vitesse de variation de la contrainte. La figure 7a2 représente le même type de diagramme en vibration dite plastique. L'énergie dissipée par période est toujours égale à la surface hachurée qui montre toutefois une différence essentielle avec la précédente. On retrouve en effet la discontinuité avec points singuliers en P et P' où le coefficient élastique n'a pas la même valeur en trajet aller et en trajet retour. Cette particularité s'interprête en régime vibratoire par un frottement sec ne dépendant que du signe de la vitesse. Une comparaison entre les énergies correspondantes aux deux types de dissipation est possible en introduisant l'énergie par période et par unité de volume à σ constant, sous la forme générale $D = J\,\sigma^n$. J est alors une constante caractéristique du matériau. La dissipation est viscoélastique pour n = 2. Elle est plastique pour des valeurs de n fractionnaires et comprises en général entre 2 et 3.

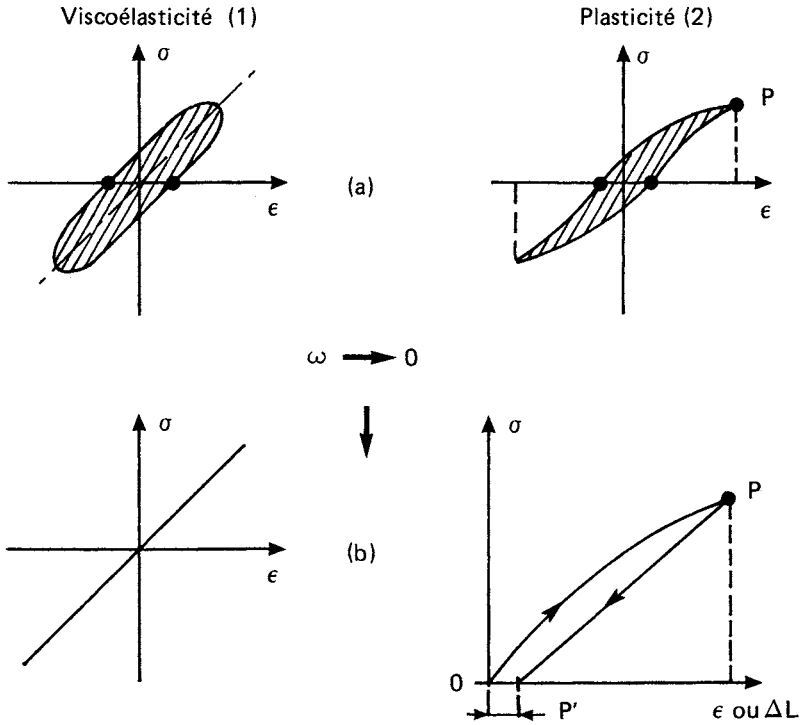


Fig. 7 - Dissipation vibratoire et hystérésis.

Si, maintenant, on effectue un transfert dynamique statique en supposant que la fréquence diminue et tende vers 0, l'ellipse de la figure 7a1 va progressivement se confondre avec la droite 7b1. On tend vers la loi de Hooke sans hystérésis mécanique décelable. Par contre, la même opération sur le diagramme 7a2 introduit en 7b2 le phénomène caractéristique d'hystérésis mécanique statique qui apparaît en déformation résiduelle sur OP' après le cycle retour. Le diagramme est alors analogue à celui de la déformation plastique de la figure 5a. En traçant la courbe dérivée $E = f(\epsilon)$ de ce dernier diagramme, il est possible d'obtenir (Fig. 8) une meilleure représentation de l'hystérésis mécanique du coefficient élastique, la différence $\Delta E = E_0 - E_I$ apparaissant directement en ordonnée. Deux éléments conditionnent cette différence. D'une part, une discontinuité $E_1' - E_0'$ sur la valeur du coefficient élastique au point de rebroussement P, d'autre part, une différence de pente $dE/d\epsilon$ entre les trajets OP et PP'. En supposant le retour PP' linéaire, soit $(dE/d\epsilon) = cte$, l'hystérésis mécanique du coefficient élastique s'exprime par :

$$(\Delta E)_{\epsilon = 0} = \left(E_1' - E_0'\right)_{\epsilon = \epsilon_M} + \left(\frac{dE}{d\epsilon}\right)_{OP} \epsilon_M$$

Cet hystérésis mécanique correspond comme on l'a montré à une dissipation d'énergie interne statique provoquée, notamment, par un phénomène de frottement sec. Cette dissipation diminue le travail des forces mécaniques $\mathcal{F}\Delta L$ sur le trajet PP' de la figure 7.

4.3 - Hystérésis thermique

Il reste maintenant à établir dans quelles conditions les hypothèses physiques établies au paragraphe précédent pour l'hystérésis mécanique du coefficient élastique sont transposables à l'hystérésis thermique. Une élévation de température appliquée au quartz provoque, comme pour les actions mécaniques, deux sortes d'effets.

Une dilatation thermique statique dont la similitude avec la dilatation mécanique est mise en évidence par la forme différentielle de l'équation d'état :

$$d\epsilon = \frac{1}{E} d\sigma + \alpha_L \, dT$$

Lorsque les actions mécaniques et thermiques sont séparées, cette relation s'intègre, pour le thermique, en :

$$\epsilon = \frac{\Delta L}{L} = \alpha_L \, \Delta T$$

Sous l'aspect dilatation, une transposition d'échelle de coefficient α_L permet d'obtenir, en abscisse, la similitude.

Une variation des coefficients élastiques. Cette variation est alors liée à l'accélération de phénomènes dissipatifs internes caractérisés par une énergie d'activation et régie de ce fait par la loi exponentielle classique $P = P_0 e^{-A/KT}$. Une similitude mécanique résulte du fait que les mêmes phénomènes dissipatifs peuvent aussi être activés par la mise en vibration du quartz.

A température décroissante, et comme en mécanique, il faut, pour justifier l'hystérésis thermique après retour aux conditions initiales, faire appel à des non réversibilités créées par des dissipations du type frottement sec. Les phénomènes physiques dissipatifs internes au quartz étant limités et bien connus, une brève énumération permet de mettre en évidence ceux susceptibles de provoquer l'hystérésis.

L'effet thermoélastique a été considéré dès qu'il s'est avéré nécessaire de définir une limite au coefficient de surtension. Cette limite a été désignée par le terme "surtension thermodynamique". Elle est due à la chaleur latente de dilatation qui provoque des échanges thermiques inversés à chaque demi période et non entièrement réversibles. Cet effet, surtout important en vibration de flexion, est assimilable à une viscoélasticité et ne provoque pas d'hystérésis.

Les diffusions d'impuretés sont considérées comme très importantes pour la stabilité temporelle et la sensibilité du résonateur aux rayonnements ionisants et, de ce fait, très étudiées [6]. Les impuretés atomiques diffusent sous l'action des gradients de concentration (loi de Fick) et ces transports sont eux aussi activés par la température et par la mise en vibration du quartz. Comme pour l'effet thermoélastique, il y a déphasage entre la diffusion et les actions qui la provoque. Ce déphasage crée une dissipation de type viscoélastique, donc non susceptible d'hystérésis.

Les migrations de défauts structuraux ou macroscopiques, en particulier des dislocations cristallines du quartz, apparaissent dans ces conditions comme la cause principale de l'hystérésis thermique pouvant seule justifier les ordres de grandeur constatés. Les déplacements des dislocations, eux aussi amplifiés par la température et la vibration, s'effectuent le long de plans appelés plans de glissement et sont contrariés par une résistance de type frottement sec. Pour le quartz, sauf autour de la dislocation et sauf sur la surface de glissement, le matériau reste parfaitement élastique. Ce schéma correspond bien à l'hypothèse de "microplasticité partielle" faite précédemment. La mobilité des dislocations du quartz a été longtemps controversée et n'a été confirmée que très récemment. L'existence de l'hystérésis thermique du quartz semble appuyer cette thèse.

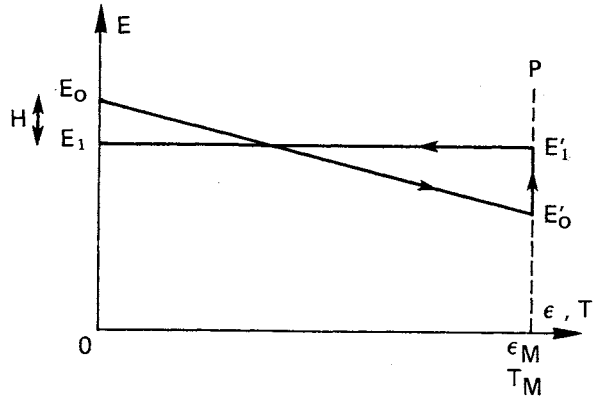


Fig. 8 - Hystérésis du coefficient élastique au cours d'un cycle mécanique ou thermique.

Ainsi une similitude des effets permet de mettre en évidence de nombreux points communs à l'hystérésis thermique et à l'hystérésis mécanique. La courbe de la figure 8 peut, en particulier, être tracée avec en abscisse soit une échelle de déformation, soit une échelle de température. Dans les deux cas, elle correspond bien aux résultats expérimentaux sur l'hystérésis thermique du coefficient élastique. Une relation analogue à celle de l'hystérésis mécanique peut être établie sous la forme :

$$(\Delta E)_{T = T_1} = \left(E'_1 - E'_0\right)_{T_1 + \Delta T} + \left(\frac{dE}{dT}\right)_{OP} \Delta T$$

5 - REDUCTION DE L'HYSTERESIS ET CONCLUSION

Les procédures expérimentales décrites au § 2 ont mises en évidence le fait que l'hystérésis thermique se stabilisait et tendait vers une limite après un certain nombre de cycles de température. Une observation identique a été faite sur les capteurs à quartz à fréquence variable. Tout se passe comme s'il était nécessaire d'éliminer d'abord des effets de mémoire liés à l'antériorité thermique et cette situation évoque avec l'hystérésis magnétique une similitude évidente bien qu'à une échelle beaucoup plus réduite. Il apparaît par conséquent judicieux de rechercher une réduction de l'hystérésis par des cycles thermiques à des températures très supérieures à celles des points d'inversion, par exemple 150°C. Ce travail a été fait. Il s'est avéré parfois positif, souvent sans effet et il a été difficile d'en obtenir des résultats concordants. Toutefois, dans la plupart des cas où, par traitement thermique, une réduction de l'hystérésis a pu être obtenue, celle-ci présentait au départ une valeur plus importante.

Ce fait doit être interprêté compte tenu de l'existence de deux sources possibles d'hystérésis, l'hystérésis intrinsèque du quartz qui fait l'objet de l'essentiel de cette communication, mais aussi l'hystérésis de structure évoquée au § 3.2. Dans ce dernier cas, les traitements thermiques sont susceptibles de provoquer des stabilisations et même des changements d'état aux points défectueux des assemblages, particulièrement dans les liaisons supports-quartz. Les phénomènes d'hystérésis de structure peuvent s'en trouver réduits.

Par contre, l'hystérésis intrinsèque du quartz est due, comme on l'a montré, aux coefficients élastiques et l'hypothèse a été faite d'une influence prépondérante sur ce phénomène des glissements des dislocations internes au quartz. Si cette hypothèse est justifiée, il n'apparaît pas que des traitements à température élevée, qui accélèrent évidemment les phénomènes en jeu, puissent néanmoins, après retour à la température initiale, réduire l'hystérésis de manière notable. Il devrait par contre y avoir une relation directe entre l'hystérésis thermique et le coefficient de surtension du matériau, ce dernier étant en général d'autant plus élevé que le taux de dislocation est faible. Un autre critère de sélection du quartz est le taux d'impuretés atomiques qui, toujours d'après les hypothèses précédentes, ne devrait pas avoir une influence notable sur l'hystérésis thermique. Mais, l'étude du phénomène avec divers échantillons de quartz, artificiel, naturel, balayé, à surtension élevée (premium Q) etc. n'a pas encore été faite à l'ONERA. On a pu toutefois constater une légère diminution de l'hystérésis, d'un facteur 1,5 environ avec des résonateurs de coupe SC comparés à des résonateurs de coupe AT.

En conclusion, l'hystérésis thermique des résonateurs à quartz est un phénomène important bien que actuellement très masqué par l'utilisation généralisée du thermostat dans les oscillateurs à haute stabilité. Il se manifeste néanmoins dans ce dernier cas pour l'erreur dite de "rétracabilité". Mais pour les oscillateurs à compensation directe, numérique ou non, l'hystérésis thermique constitue un obstacle majeur à l'amélioration des performances. Il en est de même pour les capteurs à fréquence variable en général non thermostatés. Les études et les réflexions menées à l'ONERA sur ce sujet se poursuivent.

BIBLIOGRAPHIE

[1] - F. Deyzac, "Oscillateurs à quartz thermostaté compensé", 1er Forum Européen Temps Fréquence, Besançon, 18-20 mars 1987, T.P. ONERA n° 1987-29.

[2] - J. Beaussier, "Application aux capteurs à fréquence variable des études effectuées à l'ONERA sur la stabilité des oscillateurs à quartz", La Recherche Aérospatiale, mai-juin 1982, pp. 151-163.

[3] - M.J. Forster, "Thermal hysteresis of AT + SC cut quartz crystal resonators", 36th AFCS, 1982, pp. 140-158.

[4] - J. Vig, R. Filler, J. Kosinski, "SC cut resonators for temperature compensated oscillators", 36th AFCS, 1982, pp. 181-186.

[5] - J. Beaussier, "Quelques aspects de physique des défaillances et de fiabilité prévisionnelle dans les oscillateurs à quartz de haute stabilité", 5ème Colloque International de Fiabilité et de Maintenabilité, Biarritz, 6-10 octobre 1986, T.P. ONERA n° 1986-151.

[6] - R. Brendel, "Effet électroélastique et diffusion des impuretés dans le quartz", Thèse de Doctorat, Besançon, 1983.

2 ND EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 16 - 17 - 18 MARCH 1988

Session RO-05 Number 3

M I N I A T U R E O S C I L L A T O R S

K. SCHWALLER, K. SCHULTHESS AND W. ZINGG

MICRO CRYSTAL, DIV. OF SMH, CH-2540 GRENCHEN

SUMMARY

Two novel oscillators are presented. The first one is a fully surface mount designed (SMD) crystal clock oscillator sealed under vacuum in a ceramic package. Main applications of this cost effective XO are in computers and digital equipments. The second one is a miniature OCXO hermetically sealed in a DIL-compatible metal package. Its performances are a trade-off between TCXO's and more cumbersome OCXO's and it is assigned to the mobile telephone market.

1. INTRODUCTION

New driving forces in the microelectronics industry like miniaturization, surface mount technology (SMT) and the boom in the area of mobile communication and digital processing equipment push the quartz crystal manufacturers to develop more performant, customer friendly and competitive products.

2. SURFACE MOUNT OSCILLATOR

At present there are only few crystal clock oscillators fully suitable for surface mounting available on the market. In the last year MICRO CRYSTAL developped a new product dubbed MCSO and shown in photo 1.

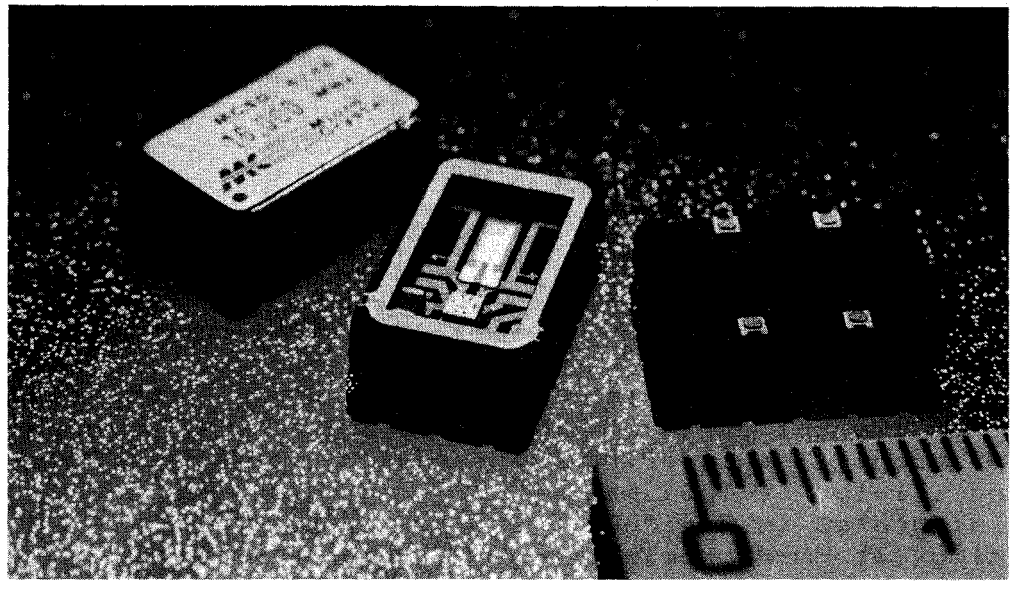


Photo 1 MICRO CRYSTAL's surface mount oscillator (MCSO)

The MCSO is a cost effective and performant product intended primarily for the microcomputer industry. The MCSO consists of an **HCMOS oscillator** and an advanced **miniature** chemically milled rectangular **AT quartz** resonator (1). The layout of the connecting pathes is configured on the bottom of the custom designed ceramic chip carrier. The tracks are gold plated and the IC is connected with gold wire bonding. The final frequency adjustment is performed with the quartz chip mounted in the ceramic carrier. After degassing and a functional control, the oscillator is ready for sealing. A conventional covar lid (gold plated) with attached eutectic 80/20 preform is used for hermetic reflow sealing under vacuum. The four contacting pads on the bottom of the package (fig. 1) are provided with bumps in order to improve solder wetting and cleaning operation after surface mounting. This SMD-oscillator withstands temperatures up to 260°C for 20 sec and it is adequate for vapor phase and wave soldering processes.

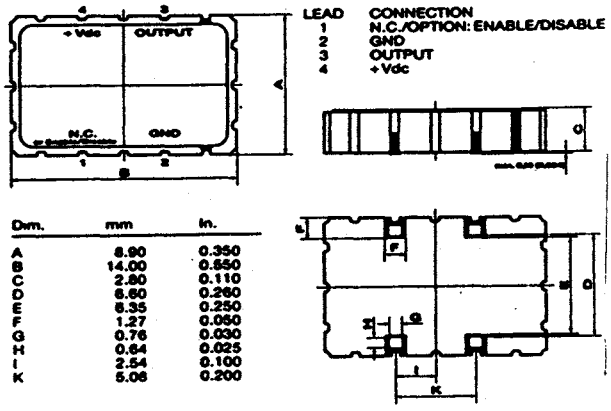


Fig. 1 Dimensions of MCSO*

The MCSO is available in a broad frequency range extending from 100 kHz up to 35 MHz (50 MHz soon). The inverter type oscillator (V3503) has been integrated in 2 micron technology at EM-Microelectronic (Marin, Switzerland). The HCMOS chip (1,6x1,6 mm2) includes seven divider stages which are accessible by three internal select terminals and give provision for division up to 128 by powers of 2. Supply voltage is 5 V dc and the dependence of current on nominal frequency is shown in figure 2.

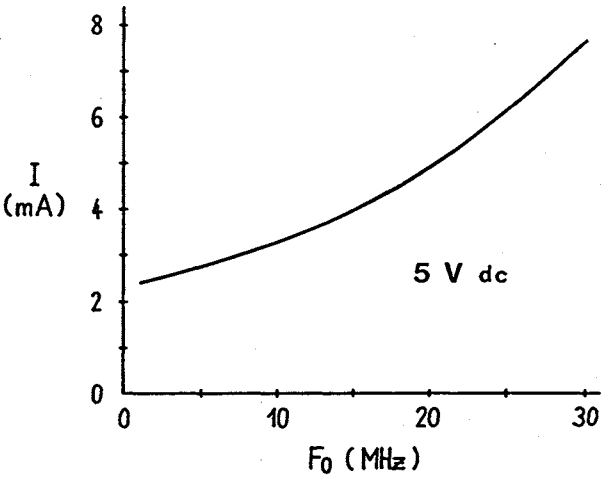


Figure 2 Current vs frequency of MCSO with 12 pF load capacitance.

The output signal is TTL and CMOS compatible and the transition time amounts typically to 6 ns (fig. 3). The operating temperature range of the MCSO is -40 to 85°C.

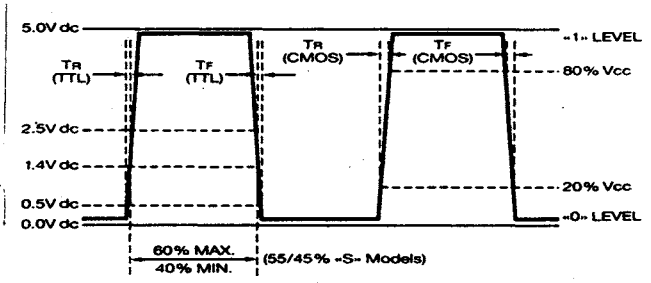


Fig. 3 MCSO output signal*

* The MCSO is mechanically and electrically compatible with the MSO from Motorola

3. MINIATURIZED OCXO

The demand for miniaturized crystal oscillators with improved stability (typ. 1 to 5 ppm) over a large temperature range (e.g. -30 to 70°C) is growing steadily. Such oscillators are needed today in mobile communication and transmission equipment (e.g. cordless telephone and cellular mobile radio) for civilian as well as for military applications. Crystal oscillators can be classified according to their performance into three main groups:

a) XO:
Simple oscillators without any compensation exhibit a frequency stability which is fully reliant on the performance of the quartz itself. An overall stability of +/- 10 ppm can be achieved as a best case (with AT crystals) in the temperature range from -30 to + 70°C.

b) TCXO:
Temperature compensated crystal oscillators which generally use temperature sensitive components and analog compensation networks are very common. Although TCXO's can be small sized and thrifty in power consumption, the improvement in stability as compared to XO's is typically of one order of manitude (e.g. +/- 2 ppm over -30 to +70°C).

c) OCXO:
Oven controlled crystal oscillators are used in applications requiring higher accuracy and stability. Standard performance is 10^{-7} in a temperature range from -10 to 60°C, but top level devices (e.g. BVA controlled oscillators) reach a stability of 10^{-9} between -30 and +60°C. Generally speaking, OCXO's are cumbersome, greedy for power consumption and expensive.

3.1. GOAL AND DESIGN OF NEW OCXO

During the last years miniaturized quartz crystals appeared on the market (2,3). Photochemically milled AT crystals in the frequency range 8 to 30 MHz and ZT crystals at 3.6 MHz are now produced by MICRO CRYSTAL. These tiny resonators are packaged in a ceramic crystal carrier which is favorable for a compact oscillator construction. Further improvement in compactness is possible with the use of an integrated oscillator chip comprising heater and oven regulation functions which has been implemented in the SARSAT localisation buoy (4). Taking advantage of these potentialities, MICRO CRYSTAL developped a **miniature OCXO** (5) mounted in a 4-pin DIL-compatible metal package (Photo 2).

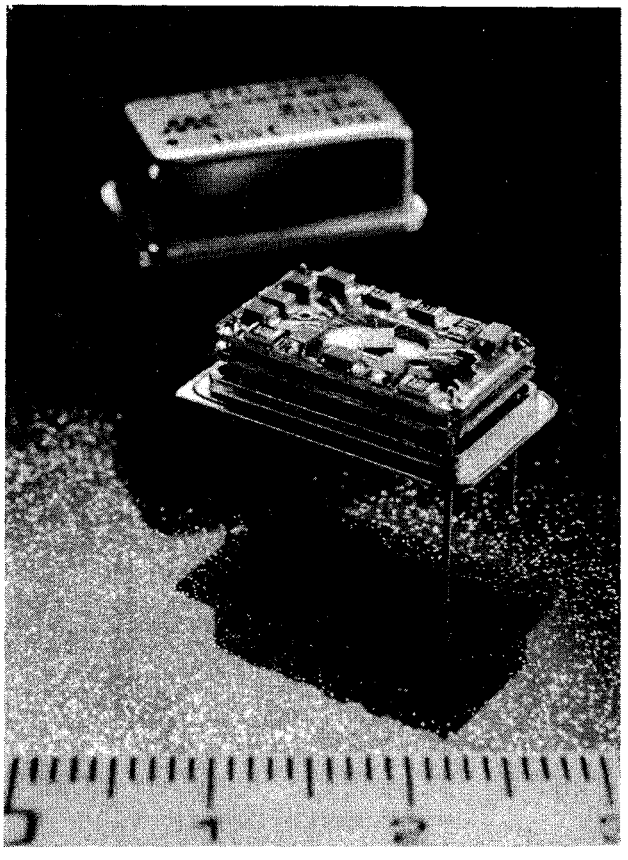


Photo 2 OCXO in DIL-package

Nominal frequency	(MHz)	3.6 and 10 to 24
Temperature range	(°C)	-30 to 70
Frequency stability		10^{-6} to 10^{-7}
Short term stability	($\sigma_y(\tau)$)	10^{-9} ($0,1 \leq \tau \leq 10s$)
Aging		+/- 1 ppm/y
Supply voltage	(V)	12
Power consumption max.	(mW)	500
Warm-up time	(min)	10
Output	(V)	.3 (pp-sinusoidal)
Dimensions	(mm)	20.8 x 13.1 x 7.0

Table I Technical characteristics of miniature OCXO

The goal of this product is to fill a gap **between TCXO's and high stability OCXO's.** Main applications are in the mobile radio and cordless telephone business as well as in the instrumentation field.

The targeted performances of this miniature OCXO are listed in table I. Reduction in power consumption is achieved through appropriate mounting of the gate-array circuit chip and the quartz package (Fig. 4).The OCXO consists of two superposed PC boards. The electronic circuitry of the output stage is laid on the lower level whereas the auxilliary components of the gate-array (oscillator and heating loop adjustment) are located on the upper board.

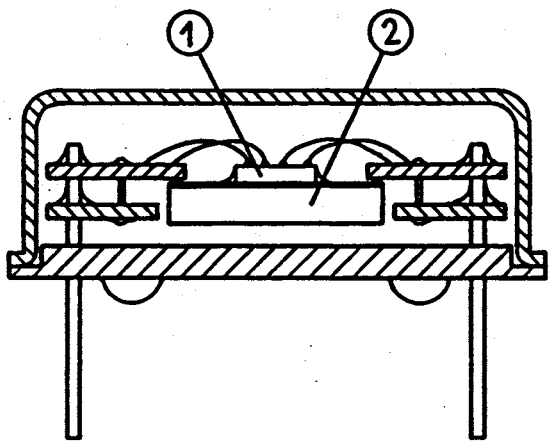


Fig. 4 Schematic View of OCXO with gate-array (1) and quartz in ceramic package (2).

The quartz ceramic package is fixed on the edge of a hole located on the upper substrate in order to reduce heat losses. Since the gate-array is mounted directly onto the quartz package, intimate thermal contact is provided between the heater and the quartz housing. This arrangement allows to maintain the crystal close to its inversion temperature (typ. 80°C). The IC is electrically connected to the circuit on the upper substrate via wire bonding.

3.2. MEASUREMENTS

The first minature OCXO prototypes built use CX-AT quartz crystals (10 or 16 MHz) together with the gate-array MMC 023 supplied by ATAC-Diffusion (Sèvres, France). The complete device (Photo 2) is encapsulated in a DIL 14 metal case.
Typical current consumption is shown in Fig. 5.
At the lowest temperature (-20°C), the consumed power is less than 500 mW. A frequency stability of +/-0.1ppm can be achieved between -20 and +70°C (Figure 6).
Short term frequency stability (Fig. 7) and phase noise (Fig. 8) have been measured on first oscillator samples at the Laboratoire de Physique et de Métrologie des Oscillateurs (LPMO) at Besançon. These results are

comparable to those obtained earlier (4) with somewhat larger oscillators. Long term stability as measured by CNES on ovenized oscillators with CX-AT 10.150 MHz miniature crystals is 4.10^{-9} /day after 3 to 6 months operation.

The warm-up behaviour has also been tested. A shown in Fig. 9, a stability better than 2.10^{-9} (10 s integration time) is obtained after 10 minutes switch on. The miniature OCXO has a sinusoidal output signal level of 300 mV eff (min) and its frequency can be corrected for aging with a dc voltage (0 to 5 V) in a range of +/- 5 ppm.

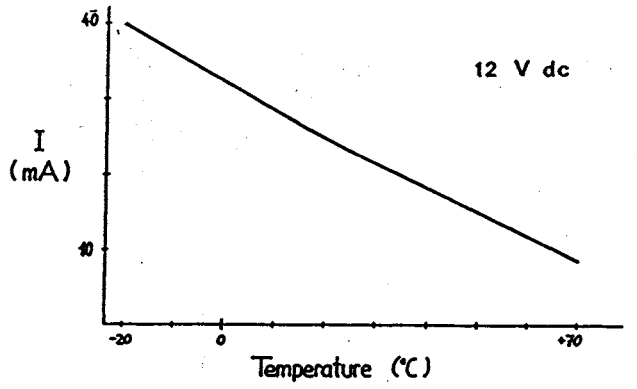


Fig. 5 Current consumption of miniature OCXO as a function of temp.

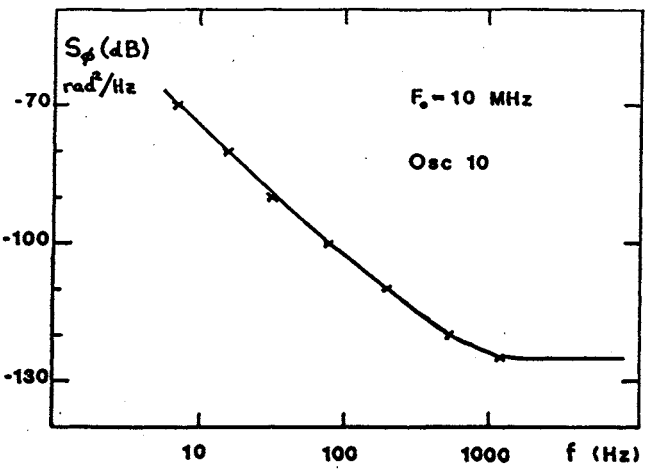


Fig. 8 Phase noise spectral density

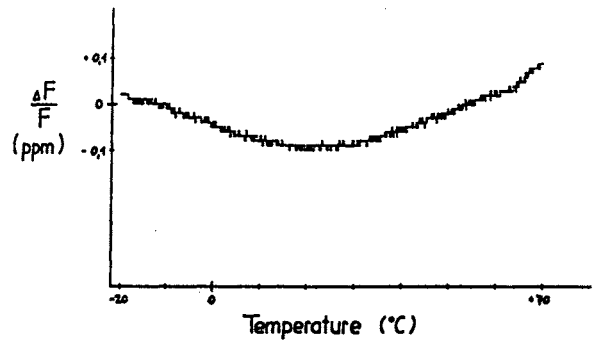


Fig. 6 Temperature stability

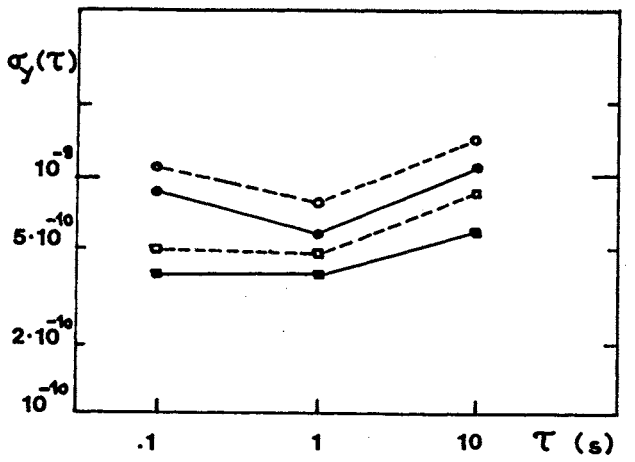


Fig. 7 Short term frequency stability of 10 MHz miniature OCXO

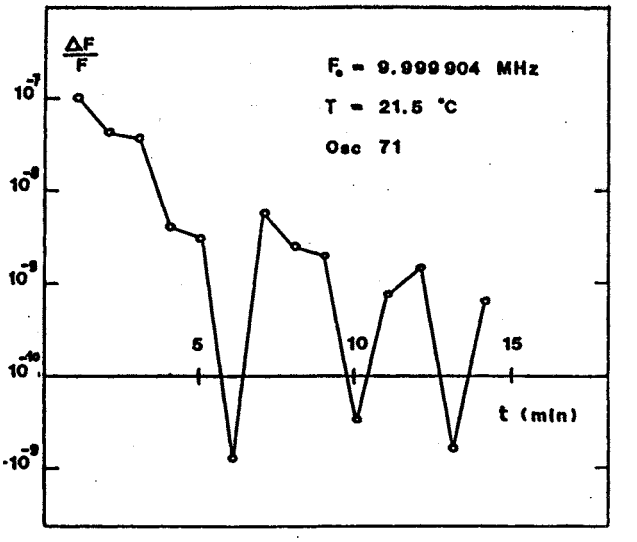


Fig. 9 Warm-up behavior

CONCLUSION

A surface mountable crystal clock oscillator (frequency .1 to 35 MHz) in a ceramic package has been developped. The production started at the end of 1987 and is now soaring. Since all constitutive components are cheap and several manufacturing steps can be readily automatized (e.g. chip wire bonding, frequency tuning, sealing and testing) it is expected to become very competitive. In view of its performance and price, the MCSO is assigned to applications like computers and digital processing equipment.

On the other hand, the miniature OCXO implemented by association of an AT-strip crystal and a gate-array comprising oscillator and stabilized heater (licenced by CNRS/CNES) is aimed to fill a gap between TCXO's and expensive OCXO's. This oscillator assembled in a DIL-compatible metal case is a trade-off with respect to size, frequency stability and power consumption.

ACKNOWLEDGEMENT

The authors are indebted to Mr. G. Marianneau from LPMO, Besançon, for his cooperation and the measurements performed on the miniature OCXO.

REFERENCES

(1) Séminaire sur les Etalons de fréquence, leur caractérisation et leur utilisation (LMPO-Cetehor, Besançon, mars 1985), p. N1 - N15 W. Zingg, Résonateurs et capteurs à quartz subminiatures.

(2) 1er Forum européen Temps-Fréquence (Besançon, mars 1987), p. 31 - 35 W. Zingg, B. Studer et J. Hermann, Fabrication de résonateurs ZT.

(3) 41st Annual Frequency Control Symp. (1987) p. 295 - 302 L.N. Dworsky AT quartz strip resonators.

(4) 40th Annual Frequency Control Symp. (1986), p. 350 - 354 G. Marianneau, D. Hauden and J.J. Gagnepain Integrated oven controlled quartz crystal oscillator.

(5) Licence CNRS / CNES Patent F 85.14141

Materials and Processing

		Chairman
MP-01	Crystallography	J. Rutman
MP-02	Physical Properties	C. Hruska

MP-01

Crystallography

1. Quartz-Homeotypic Crystals
   G.F. Engel, P.W. Krempl
   AVL List Ges. m.b.H, Graz, Austria

2. Berlinite Blank Absolute Orientation Determinations by X-Ray Diffraction
   J.F. Darces, M. El Bellal, H. Merigoux
   U.F.R. des Sciences et des Techniques, Besançon - France

3. A New Approach to Berlinite Crystal Growth
   E. Philippot, A. Goiffon, J.C. Jumas
   USTL, Montpellier - France
   X. Buisson, R. Arnaud
   SICN, Annecy - France
   Y. Toudic
   CNET, Lannion - France

4. Lattice Defects in Lithium Tantalate
   P. Cordier, B. Boulogne, N. Doukhan, J.C. Doukhan
   Université de Lille Flandres Artois, Villeneuve d'Ascq - France
   M. Fragneau, J.P. Aubry
   Compagnie d'Electronique et de Piezo Electricité, Argenteuil - France

5. Evaluation and Application of High Purity Berlinite Crystals
   J. Detaint, J. Schwartzel, R. Lefevre
   CNET PAB/BAG/MCT/CMM Bagneux - France
   J.P. Aubry, L. Michel
   CEPE, Argenteuil - France
   B. Capelle, A. Zarka
   Université P. et M. Curie, Paris - France
   E. Philippot
   Université de Montpellier, Montpellier - France

2 ND EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 16 - 17 - 18 MARCH 1988

Session MP-01 Number 1

QUARTZ-HOMEOTYPIC CRYSTALS

G.F. Engel, P.W. Krempl

AVL LIST Ges.m.b.H., Physics Division, Graz, Austria

Abstract

Among all the crystals which are homeotypic to α-quartz, berlinite ($AlPO_4$) has attracted the most research activities and only few data are available from the others. To estimate the thermal stability of physical properties and the piezoelectric sensitivity of the less known homeotypes, two models were applied:

- By means of an idealized structural model, an order-disorder picture of the prototypic β-phase and a mean field pseudo-spin formalism of tilted tetrahedral units, estimates of the temperature stability of the order parameter were made. From this, a high temperature stability of several physical and structural properties in Ga-compounds was predicted.
  X-ray-data of $GaPO_4$ and $AlAsO_4$ at room temperature, 500°C and 750°C are shown to be in accordance with these predictions.
- A bond additivity model of piezoelectric constants using the same X-ray data as input parameters suggested that the piezoelectric constants of $GaPO_4$ and $AlAsO_4$ show less temperature dependence than those of quartz and berlinite. Furthermore, this structural point of view shows that the longitudinal piezoelectric coefficients in several quartz-homeotypes are greater than in quartz.

As a consequence of these theoretical predictions the most promising crystals have been grown hydrothermally. Preliminary measurements performed for one of these materials confirm the theoretical expectations.

1. INTRODUCTION

The present paper deals with members of the family of crystals having a formula $M^{3+}X^{5+}O_4$ (M^{3+} = Fe,Mn,Ga,Al; X^{5+} = P,As) and a homeotypic structure derived from α- and β-quartz by alternatively substituting Si with M and X. Among these crystals, α-$AlPO_4$ (also known as Berlinite) has currently a prominent position, since it is most intensively studied in many laboratories. This is due to its potentially industrial-scale use in electronic devices, e.g. bulk resonators (BAW-devices, [1]) or surface acoustic wave (SAW-) devices [2]. Relatively few reports on the other crystals of this isomorphous series have appeared. Papers dealing with at least three of these MXO_4-crystals have to be emphasized and we remember recent papers on solubility and growth [3], crystallography [4,5], photoelectron-spectroscopy [6] and radiation hardness [7]. Except for $AlPO_4$, no reports have been found by the present authors about optical, elastic, piezoelectric or dielectric properties and their potential use in electroacoustic devices has been assessed recently only by the analogizing method [8].

The purpose of the present paper is to show evidence for the high temperature stability of properties related with piezoelectricity in some of the less known quartz homeotypic MXO_4 crystals by means of two models concerning the thermal stability of physical properties and the piezoelectric sensitivity.

Several reasons motivate study of thermal stability in crystals other than quartz:

The limits of quartz as the active element in pressure sensors are reached far below the α-β transition point due to ferrobielastic twinning induced by nonuniform mechanical stress. The "critical temperature" is lowered by increasing the applied pressure (fig. 1), well in accordance with results in other laboratories [9].

Another drawback of quartz is the decrease of sensitivity at elevated temperatures due to the temperature dependence of the piezoelectric coefficient d_{11} [10, 11], which becomes dramatic above, say 200°C.

Furthermore, the relatively low value of d_{11} causes the coupling factor to be low which is also unfavourable in some resonant applications.

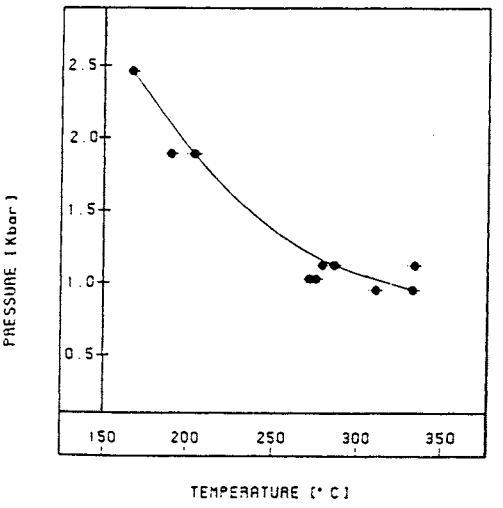


Fig. 1: Onset of ferrobielastic twinning in x-cut quartz-modules consisting of three disks. At a given temperature, the pressure was increased up to values shown by the full circles, where the piezoelectric sensitivity dropped due to twinning. After the experiment, etching revealed twins according to the Dauphiné law.

2. MEAN FIELD α–β TRANSITION DYNAMICS

2.1 Theoretical background

Since the point group of α-quartz and its homeotypes is identical, all symmetry considerations valid for this point group (D_3 or 32) and already made for α-quartz can be applied: An order parameter $\eta(T)$ is considered, which is zero in the β-phase and non-zero in the α-phase. Following Dolino and Bachheimer [12], the functional dependence of $\eta(T)$, given in its simplest theoretical form by Landau-theory, is then related linearly to e.g. the piezoelectric coefficients d_{11}, e_{11}, to the second harmonic generation (SHG-) coefficient d_{11}, other corresponding nonlinear optical 3rd rank tensor components, elastic constants c_{14}, s_{14} and the mean tetrahedral tilt angle of SiO_4-tetrahedra in quartz in first approximation.

An approximately quadratic relationship to $\eta(T)$ follows for suitably renormalized elastic constants c_{14}, s_{14}, c_{66}, s_{66}, specific heat, refractive indices, birefringence, expansion coefficient and other properties. Here, renormalization with respect to the β-phase values is required.

In some cases it may be a coarse approximation to relate the temperature dependence of the above mentioned quantities linearly or quadratically to $\eta(T)$. However, considering quartz as a typical example, the transition at T_t = 846 K still influences the temperature dependence of e.g. the elastic constants at room temperature, because it is still remarkably different there from its saturation value $\eta(T=0)$. An ad-hoc-generalization leads to the corresponding statement for the quartz homeotypes. It is explicitly noted that the crystals are included where β-phases cannot be obtained physically.

To evaluate a measure of the α-phase stability, we have to adopt a common mechanism of the α–β phase change. Even for the α–β transition in quartz some questions have been answered only recently [9]. Of particular interest is the recent revision of the β-quartz structure. In contrast to the previous picture, which is purely displacive, an order-disorder-contribution to the phase transition of quartz is now evident [13].

We therefore make the assumption that any prototype β-phase in the quartz homeotypic series is of this order-disordertype: the potential of a representative formula-unit (MXO_4) has a relative maximum at the midpoint between the "disordered positions", and minima at both sides for each "Dauphiné -"state, i.e. it can be represented by a double well.

This double well potential is modelled by a nonlinear local mode Hamiltonian and processed by mean field pseudo spin theory [14] yielding

$$(1) \qquad \Omega/J = \tfrac{1}{2} \tanh (\Omega/2K_B T_c)$$

as an equation for the stability limit temperature T_c.

Ω and J are the tunneling frequency and the isotropic elastic exchange term, respectively. Since quantum tunneling between the "Dauphinée-"states plays no role at all, we have $\Omega \to 0$ and equation (1) simplifies to

$$(2) \qquad T_c = J/4K_B .$$

Approximate stability limits based on this equation have been reported in [15] and are shown in table 1.
Defining $t = T/T_c$ ($T_c - T_t$, see [16]), the order parameter takes the form

$$(3) \qquad \eta^{\beta\,s} = \tanh(\eta^{\beta\,s}/t)$$

It is clear from table 1 and equation 3 that the Ga-compounds have the highest thermal stability of those physical properties which are associated with the order parameter.

TABLE 1:

Approximate stability limits of quartz-homeotypes with respect to the β-phase in units of T_c of quartz (after ref. [15])

$AlPO_4$	$MnPO_4$	$FePO_4$	$AlAsO_4$	$GaPO_4$	$GaAsO_4$
1.02	1.05	1.12	1.7	2.3	2.1

2.2 Structural stability of $GaPO_4$ and $AlAsO_4$

The results of the previous section predict $GaPO_4$ to be more stable (with respect to their hypothetical β-phases) than $AlAsO_4$ and SiO_2. Since β-quartz analogue phases in $GaPO_4$ and $AlAsO_4$ cannot be physically obtained, evaluation of these results is performed through structural stability considerations.

In the following, the temperature dependences of structural data of quartz and $AlPO_4$ are taken from [17] and [18], respectively, for comparison purposes. In the case of $GaPO_4$ and $AlAsO_4$ data at - 100°C, ambient temperature and 100°C were recently published [19] but no structural data at high temperatures were available. Therefore, studies of their structures at room temperature, 500°C and 750°C were initiated. The results are described in [20] for $AlAsO_4$ and in [21] for $GaPO_4$.

In both cases, position parameters, interatomic distances and angles show only small variations (tables 2 and 3) even over this large temperature region:
The metal-oxygen-metal angles of both crystals are slightly increasing: ~ 5° for $AlAsO_4$ and ~ 3° for $GaPO_4$. Consequently, the effect of temperature on the structure can be considered as the "back-tilting" of the MO_4 and XO_4 tetrahedra.

TABLE 2

Position parameters in $GaPO_4$ and $AlAsO_4$;

Space group $P3_121$; 3 Formula units per unit cell;

	20°C		500°C		750°C	
	$AlAsO_4$	$GaPO_4$	$AlAsO_4$	$GaPO_4$	$AlAsO_4$	$GaPO_4$
a(Å)	5.032	4.901	5.067	4.934	5.11	4.973
c(Å)	11.229	11.048	11.258	11.075	11.30	11.105
c/a	2.2315	2.2542	2.2218	2.2446	2.2114	2.2331
M^{3+}						
x	0.4587	0.4564	0.4606	0.4496	0.4659	0.4634
y	0	0	0	0	0	0
z	1/3	1/3	1/3	1/3	1/3	1/3
X^{5+}						
x	0.4571	0.4564	0.4608	0.4617	0.4658	0.4665
y	0	0	0	0	0	0
z	5/6	5/6	5/6	5/6	5/6	5/6
O1:						
x	0.4094	0.4090	0.4077	0.4072	0.4161	0.4071
y	0.2971	0.3174	0.2894	0.3125	0.2844	0.3046
z	0.3867	0.3927	0.3882	0.3932	0.3915	0.3956
O2:						
x	0.4067	0.4096	0.4075	0.4152	0.4180	0.4224
y	0.2848	0.2730	0.2773	0.2715	0.2735	0.2676
z	0.8808	0.8726	0.8818	0.8730	0.8833	0.8756
R(%)	2.75	2.3	2.67	5.1	3.74	4.4

3.2 Results

Upon calculation of the quantities G_{111} from equation (7) using the same structural data as in section 2, we see from the results (table 5) that

(9) $$G_{111}(M\text{-}O) = G_{111}(X\text{-}O)$$

and

(10) $$\frac{\partial G_{111}(M\text{-}O)}{\partial T} = \frac{\partial G_{111}(X\text{-}O)}{\partial T}$$

From this,

(11) $$\frac{1}{d_{11}}\frac{\partial d_{11}}{\partial T} = \frac{1}{2}\left(\frac{1}{G_{111}(M\text{-}O)}\frac{\partial G_{111}(M\text{-}O)}{\partial T} + \frac{1}{G_{111}(X\text{-}O)}\frac{\partial G_{111}(X\text{-}O)}{\partial T}\right) +$$
$$+ \text{terms depending on } \phi^3 \text{ and } N$$

is obtained as an approximation. A similar but less accurate equation holds for $1/d_{14}\cdot\partial d_{14}/\partial T$.

Equation (11) reflects a splitting of the piezoelectric constants into a structural part and a contribution representing a bond-domain property. The latter depends largely on e.g. effective charges and other bond properties which are nearly temperature independent, and the stiffness of the lattice which will vary with temperature.

Fig. 3 displays the results for the temperature dependence of d_{111} due to the structural contributions alone. $AlPO_4$ and quartz have been included in this theoretical consideration. The details of the calculations for these crystals will be published elsewhere.

For the cases, where $1/d_{11}\cdot\partial d_{11}/\partial T$ is the "figure of merit" of an application of the piezoelectric effect (e.g. the measurement of pressure in an internal combustion engine) we conclude from fig. 3 that the figure of merit increases following the series quartz, berlinite, $AlAsO_4$ and $GaPO_4$.

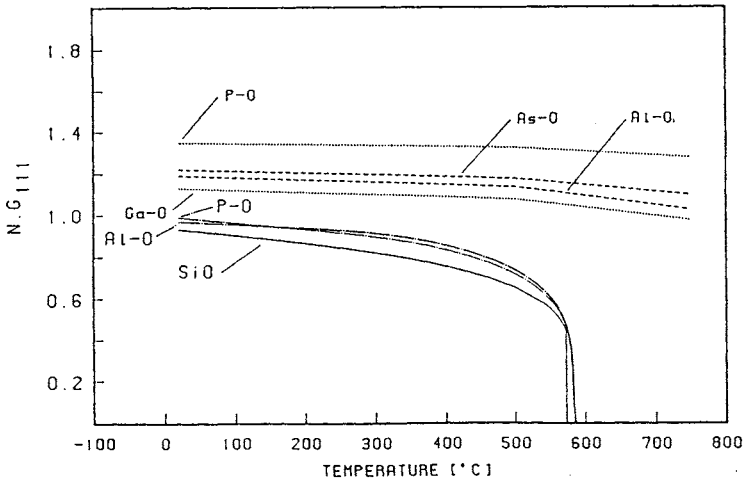


Fig. 3: Structural contributions $N\cdot G_{111}$ to the piezoelectric constant d_{111} depending on temperature. The linetypes are the same as in the previous figure.

TABLE 5:

Temperature dependence of G_{ijk} in $GaPO_4$ and $AlAsO_4$

	$GaPO_4$				$AlAsO_4$				Source of
	Ga-O		P-O		Al-O		As-O		X-ray
T	G_{111}	G_{123}	G_{111}	G_{123}	G_{111}	G_{123}	G_{111}	G_{123}	data
-100°C	1.21	0.28	1.33	0.53	1.20	0.36	1.24	0.18	c
20°C	1.13	0.59	1.35	0.13	1.18	0.37	1.22	0.19	a, b
100°C	1.12	0.57	1.32	0.10	1.19	0.34	1.21	0.16	c
500°C	1.07	0.60	1.32	0.12	1.12	0.30	1.17	0.10	a, b
750°C	0.97	0.53	1.27	0.07	1.01	0.40	1.09	0.04	a, b

a Behmer et al (1983)
b Baumgartner et al (1984)
c Goiffon et al (1986)

In addition to the results shown in table 5, the structural factors G_{ijk} at ambient temperatures were also calculated for $GaAsO_4$ (X-ray data from [5]) and $FePO_4$ (using the data from [26]). The values obtained for all those crystals, where refined structural data were available, are summarized in table 6 and show, that there is a constant increase of the "longitudinal" structural factor G_{111} in the series starting with quartz, $AlPO_4$, $FePO_4$, $AlAsO_4$, $GaPO_4$ to $GaAsO_4$, the latter being nearly 50% higher than the quartz value. Since the bond polarizabilities ϕ are not expected to exhibit the reversed trend within this homeotypic series for the reasons discussed below, it turns out that in addition to the high thermal stability, the Ga-compounds are predicted to have the largest "longitudinal" piezoelectric coefficient d_{11} within the quartzhomeotypic series.
The bond constants ϕ depend primarily on the effective charges Z of the lattice ions and on how external macroscopic stresses reach the bond domains [24]. Thus, equation (7) is rewritten in the form

(13) $$d_{11} = \{G_{M\phi}(M\text{-}O)Z(M)\gamma(M\text{-}O) + G_{X\phi}(X\text{-}O)Z(X)\gamma(X\text{-}O)\}$$

where γ is assumed not to depend on Z.
The effective charges of quartz [27] are $Z(Si) = 2.285e_0$, $Z(O) = -1/2\,Z(Si)$ and those of $AlPO_4$ [28] are $Z(Al) = 2.2e_0$ and $Z(P) = 2.4e_0$. Assuming the trend of Z is proportional to $(M)^{1/2}$ (M is the reduced mass of the bonds). Consequently $Z(Ga) = 2.5e_0$, $Z(As) = 2.7e_0$, $Z(Fe) = 2.5\,e_0$ result from extending this trend to the other quartzhomeotypes.

The order of magnitude of $\gamma(M\text{-}O)$ and $\gamma(X\text{-}O)$ can be obtained from the experimental data of quartz and $AlPO_4$ [29], yielding $\gamma(Si\text{-}O) = 0.54\ pC/e_0N$, and $1/2\{\gamma(Al\text{-}O) + \gamma(P\text{-}O)\} = 0.75\ pC/e_0N$, respectively.

The high strength of the SiO_2-lattice is not likely to be typical for the other quartz homeotypes. Therefore $\gamma = 0.75\ pC/e_0N$ is used to estimate $d_{11}(FePO_4) = 4.3\ pC/N$, $d_{11}(AlAsO_4) = 4.4\ pC/N$, $d_{11}(GaPO_4) = 4.5\ pC/N$ and $d_{11}(GaAsO_4) = 5.2\ pC/N$. However, due to the uncertainties related with the inaccurate knowledge of Z and γ, these values remain largely based on the structural view of the piezoelectric effect in these crytals.

TABLE 3:

Bond lenghts and angles in $GaPO_4$ and in $AlAsO_4$ at 20°C, 500°C and 750°C

	20°C		500°C		750°C	
	$AlAsO_4$	$GaPO_4$	$AlAsO_4$	$GaPO_4$	$AlAsO_4$	$GaPO_4$
d(M-O1)	1.740	1.807	1.731	1.812	1.726	1.810
d(M-O2)	1.744	1.817	1.759	1.816	1.753	1.806
d(X-O1)	1.644	1.529	1.669	1.529	1.652	1.523
d(X-O2)	1.663	1.529	1.651	1.532	1.635	1.527
M–O1–X	136.4	135.3	137.7	135.7	141.5	137.6
M–O2–X	135.5	134.5	136.9	135.7	140.4	138.8

It is clear from table 2 and table 3, that the tetrahedra are not regular. Since, however, the distances X-O1, X-O2, Ga-O1 and Ga-O2 are virtually identical, we have quite small deviations from regularity, which are neglected for the computation of mean tetrahedral tilt angles as a function of temperature.
The results are shown in table 4. While the mean tetrahedral tilt angles of quartz [22] and the berlinite [23] drop to zero, which means that the "back-tilting" is complete, there is only a small change in $AlAsO_4$ and $GaPO_4$.

No other significant structural changes are present in $GaPO_4$ and $AlAsO_4$ in addition to the "back-tilting" of the tetrahedra. This justifies the statement that only α-β transition dynamics are active in these crystals over the considered temperature interval, although β-phases cannot be obtained physically.

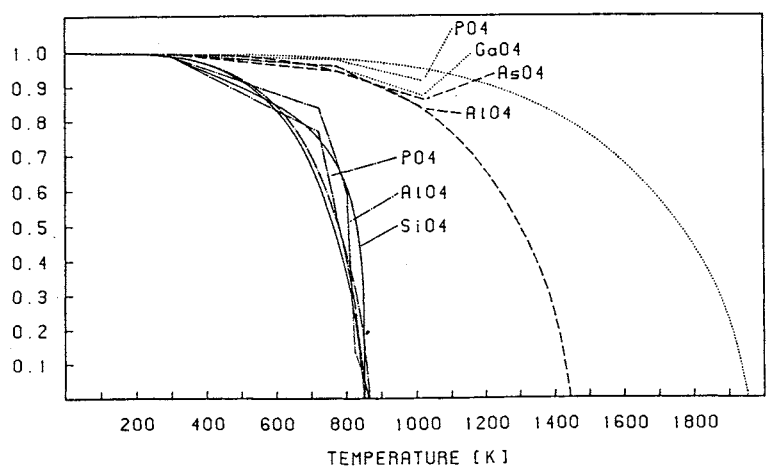


Fig. 2: Normalized tilt angles of tetrahedra in quartz, $AlPO_4$, $AlAsO_4$ and $GaPO_4$. The normalization factor is $\eta(20°C)/\delta(20°C)$. The data for quartz (——) and $AlPO_4$ (— · —) are from Grimm and Dorner [22] and Lang et al [23], respectively, the values for $AlAsO_4$ (----) and $GaPO_4$ (····) were calculated using the data of Behmer et al [20] and Baumgartner et al [21]. The order parameter η according to the mean-field pseudo-spin model is also shown using the identical linetypes.

TABLE 4:

Mean tetrahedral tilt angles in quartz, $AlPO_4$, $AlAsO_4$ and $GaPO_4$ at 20°C, 500°C and 750°C

	Quartz	$AlPO_4$		$AlAsO_4$		$GaPO_4$	
	SiO_4	AlO_4	PO_4	AlO_4	AsO_4	GaO_4	PO_4
20°C	15.6	15.8	19.0	21.8	22.9	20.8	25.6
500°C	11	10.5	10.5	21	21.7	20	25.1
750°C	0	0	0	18.3	19.8	18.2	23.5

In fig. 2, normalized tilt angles of quartz, $AlPO_4$, $AlAsO_4$ and $GaPO_4$ are displayed and compared with the corresponding order parameter η calculated from equation (3). As it is expected from the estimated stability limits, the variations of these quantities with temperature in $AlAsO_4$ are indeed greater than those in $GaPO_4$, but are still lower than the corresponding variations in quartz and berlinite.

The comparison between η and the normalized tilt angles shows that T_c ($GaPO_4$) may be overestimated, and T_c($AlAsO_4$) may be slightly underestimated. However, in view of the approximate calculation of T_c and the fact that η^{β} has no free parameter the agreement is satisfactory.

3. PIEZOELECTRIC COEFFICIENTS

3.1 Theoretical background

Theoretical treatments on the piezoelectricity of ternary compounds can be simplified using bond additivity methods [24, 25]. The concept of bond additivity is introduced by writing the piezoelectric strain coefficients in the form [24]

$$(4) \qquad d_{ijk} = N \sum_{\beta} G^{\beta}_{ir} G^{\beta}_{js} G^{\beta}_{kt} \phi^{\beta}_{rst}$$

G_{ir} are direction cosines relating the bond-coordinate system to the crystal coordinate system, ϕ are microscopic stress polarizabilties and N is the number of unit cells per unit volume. For the linear dipole approximation

$$(5) \qquad \phi \triangleq \begin{pmatrix} \cdots & \phi_{xx} & \cdots \\ \cdots & \phi_{xy} & \cdots \\ \cdots & \phi_{yx} & \cdots \end{pmatrix}$$

From this and the 32-symmetry of the quartzhomeotypic crystals, equ. 4 simplifies considerably to yield

$$(7) \qquad d_{111} = N \sum_{\beta \in \Lambda} G_{111}{}^{\Lambda} \phi_{x}{}^{\Lambda}, \qquad d_{123} = N \sum_{\beta \in \Lambda} G_{123}{}^{\Lambda} \phi_{123}{}^{\Lambda}$$

where

$$(8) \qquad G_{111}{}^{\Lambda} = \sum_{\beta \in \Lambda} (G_{13}{}^{\beta})^3, \qquad G_{123}{}^{\Lambda} = \sum_{\beta \in \Lambda} [(G_{33}{}^{\beta})^2 - (G_{32}{}^{\beta})^2],$$

$$(9) \qquad \phi_{x}{}^{\Lambda} = \phi_{333}{}^{\Lambda} - 2\,\phi_{113}{}^{\Lambda}.$$

and $\Lambda = 1,2$ indexes M-O and X-O-bonds. M-X bonds and other second neighbour interactions will not be considered due to the negligible contribution expected.

TABLE 6:

Ambient temperature data of G_{ijk} in quartz-homotypic crystals

MXO_4	M-O G_{111}	M-O G_{123}	X-O G_{111}	X-O G_{123}	Source of X-ray data
$SiSiO_4$	0.923	0.144	0.923	0.144	Smith & Alexander (1963)
$AlPO_4$	0.97	0.11	0.99	0.30	Schwarzenbach (1966)
$FePO_4$	1.08	0.49	1.28	0.12	Ng & Calvo (1975)
$AlAsO_4$	1.18	0.37	1.22	0.19	Behmer et al (1983)
$GaPO_4$	1.13	0.59	1.35	0.13	Baumgartner et al (1984)
$GaAsO_4$	1.32	0.55	1.37	0.30	Goiffon et al (1986)

4. PRELIMINARY EXPERIMENTAL RESULTS ON PIEZOELECTRICITY IN $GaXO_4$-CRYSTALS

Upon evaluation of the two models described in the previous sections, the Ga-compounds have emerged as the most promising candidates among all quartz homeotypes for the use in piezoelectric devices requiring high temperature stability of physical properties and a sufficiently large piezoelectric coefficient d_{11}.

As a consequence of these theoretical predictions, $GaXO_4$- (X=P,As)-crystals have been grown by different methods.

Growth techniques and basic characterizations will be described elsewhere. Electroded $GaXO_4$ (X=P,As) crystal plates were prepared from hydrothermally grown material, and were piezoelectrically excited to elastic vibrations, thus proving the presence of piezoelectricity in both materials.

A $GaPO_4$-plate was selected further and the temperature dependence of the resonance and antiresonance frequencies of a contour mode driven by d_{11} was recorded by standard methods from room temperature to beyond 500°C. The temperature dependence of the piezoelectric constant d_{11} was calculated from these measurements. The results are shown in fig. 4, together with the theoretical values.

Another $GaPO_4$-crystal plate was subjected to measurements in a "force-calibrating-device". High loads of up to 1 kN - measured by a load measuring cell - were applied to the plate by means of a hydraulics system. The charge that developed on the main faces was fed into a charge-amplifier and recorded by means of a computer, which also served as the controller for the hydraulics-systems and as the recorder of the load measuring cell signals. d_{eff} = 4.2 pC/N resulted from this direct measurement after a suitable correction for the orientation of the sample. Due to fringe effects on the outer regions of the samples surfaces, contributions of transversal and shearing forces to the measured charges cannot be excluded. Based on the experience with quartz plates of similar dimensions in the same device, we expect d_{11} to be greater than d_{eff} by at least 10%.
A thorough investigation of this subject is presently being performed.

5. DISCUSSION

There are six known MXO_4-quartz homeotypic crystals stable at ambient temperature and pressure, each being an a-priori candidate for use in electroacoustic applications.

Since the development of growth techniques for each of the crystals is very elaborate, it is desirable to use theoretical results as guide lines to chose the "best" candidate for use in a specific application.

This paper demonstrates that two different models result in a consistent description of thermal stability of physical properties and indeed provide a valuable help for selection of the "best" candidate for all piezoelectric applications where a high temperature stability and a sufficiently large piezoelectric coefficient d_{11} is required.

It turned out that $MnPO_4$ and $FePO_4$ have intermediate positions concerning T_c and their growth, characterization and determination of material constants is of academic rather than of industrial interest. For $FePO_4$ an intermediate position among the MXO_4 crystals results also from the calculation of G_{111} (table 6).

It turned out further, that $AlPO_4$, although it is the most intensively studied crystal family member, is the one having the lowest capability for high temperature applications. Furthermore, it has the lowest longitudinal structural factors G_{111} of all quartz-homeotypic crystals, for which refined X-ray data are known.
(These data are lacking only for $MnPO_4$). The remaining three members, $AlAsO_4$, $GaAsO_4$ and $GaPO_4$ turned out to be the most promising according to both models, with the Ga-compounds most preferred.
The preliminary experimental evidence for piezoelectricity of $GaPO_4$ shows good agreement with the theoretical temperature dependence (fig. 4), although the $T_c(GaPO_4)$ of table 1 again appears to be overestimated:

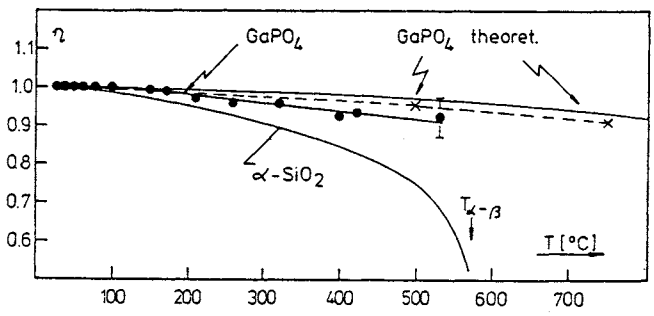


Fig. 4: Temperature dependence of normalized piezoelectric constant d_{111} $GaPO_4$ and quartz.
Measurements: full circles; order parameter η: upper curve; calculation according to equ. (11): -- x --. For comparison, experimental values for quartz (Cook and Weissler [11]) are also shown. All curves have been normalized to 1 at 20°C.

The relative error of the measurements is increasing with temperature and is estimated to be ± 5% at ~ 530°C, so that the agreement is within the error limits. The values derived from the temperature dependence of G_{111} (in fig. 4 shown as X) are lower than the curve derived from the mean field pseudo-spin model and are considered to be more realistic, since the full structural informations have been used in the calculation, including the distortion of the tetrahedra. On the other hand, the temperature dependence of the microscopic stress polarizabilities, which is not included, could cause small but important corrections (e.g. in quartz, the effective charge increases with temperature by an amount of a few percent over the range form ambient temperature to T_t [to be published elsewhere]).

The result quoted above for d_{eff} is accurate to at least 1%. However, there is no one to one correspondence to d_{11}, due to the nonuniform stress distribution on the measured sample. Further work is presently carried out to eliminate this effect, which amounts to ~ 10%.

Another point that is directly connected with measurements of piezoelectric coefficients is the effect of growth twinning.

The presence of only a few microtwins of the "Brazilian law" is able to degrade the measured d_{11}, so that any measurend d_{eff} is to be considered as the minimum value fo d_{11} of $GaPO_4$, in case that a few twin lamellae have been overlooked upon examination of the sample.

However, these restrictions do not affect the conclusion that $GaPO_4$ shows considerably less temperature dependence of d_{11} than $AlPO_4$ and quartz and has a sufficiently large piezoelectric d_{11}, which is considerably larger than d_{11} of quartz.

It has been questioned for a long time [2, 8], whether the existence of an α-β-transition is a necessary and sufficient prerequisite for the existence of temperature compensated cuts in quartzhomeotypic crystals, which rely on a positive temperature coefficient of the elastic constant c_{66} [30] ("c_{66}-stiffening").

In case it is such a prerequisite, $AlAsO_4$, $GaPO_4$ and $GaAsO_4$ would not exhibit c_{66}-stiffening and vice versa.

In our opinion, not the existence of an α-β-transition, but the existence of α-β-transition-dynamics is the necessary prerequisite for the existence of temperature compensated cuts in quartzhomeotypic materials.

Such α-β-transition dynamics are clearly present in $AlAsO_4$ and $GaPO_4$ (see above), so that the c_{66} stiffening is possible and would support the corresponding statement for quartz and berlinite.

We note, that it was necessary to take large temperature intervals (e.g. ambient temperatures up to 750°) for X-ray-analysis, to demonstrate the existence of α-β-transition dynamics in $AlAsO_4$ and $GaPO_4$.

It is quite problable, that in the near future such dynamics will also be found in $GaAsO_4$.

Acknowledgement:

The financial support of the Austrian Science Research Fund (Projects 3714 and 4618) is greatfully acknowledged.

REFERENCES

1. Detaint J., Philippot E., Jumas J. C., Schwartzel J., Zarka A., Capelle B. and Doukhan J. C., Proc. 39th Ann.Frequ.Control Symp. 234-46, (1985).
2. Jhunjhunwala A., Vetelino J. F. and Field J. C., J.Appl.Phys.48, 887-92, (1977).
3. Goiffon A., Jumas J. C., Astier R. and Philippot E., J.Cryst.Growth 71, 763-70, (1985).
4. Kosten K. and Arnold H., Zeit.Kristallogr.152, 119-33, (1980).
5. Goiffon A., Bayle G., Astier R., Jumas J. C., Maurin M. and Philippot E., Rev. Chimie Minerale 20, 338-50, (1983).
6. Klingelhöfer H. and Lehmann G., J.Electron.Spec.&Related Phen.16, 259-66, (1979).
7. Karioris F. G., Ramasami K., Goneda K. A. and Cartz L., Phil.Mag.A52 525-31, (1985).
8. Sherman J.H.,Jr. Proc. 40th Ann.Frequ.Contr.Symp., 91 (1986).
9. Shiau S.M., Anderson T.L., Newnham R.E. and Cross L.E., Mat.Res.Bull. 19, 831, (1984).
10. Mullen A.J., J.Appl.Phys. 40, 1693, (1969).
11. Cook, R.K. and Weissler P.G., Phys.Rev. 80, 712 (1950).
12. Dolino G. and Bachheimer J. P., Ferroelectrics 43, 77-86, (1982).
13. Wright A. F. and Lehmann M. S., J.Solid State, Chem.36, 371-80, (1981).
14. Blinc R. and Zeks B., Soft Modes in Ferroelectrics and Antiferroelectrics,(Amsterdam: North-Holland), (1974).
15. Engel G. F. and Krempl P. W., Ferroelectrics 54, 349, (1984).
16. Engel G.F. and Krempl P.W., to be published in J. Phys. Chem. Solids
17. Young R. A., Defence Doc Center, Rep.No AD 276235, (1962).
18. Ng H. N. and Calvo C., Can.J.Phys.54, 638-46, (1976).
19. Goiffon A., Jumas J.-C., Maurin M. and Philippot E. J. Solid State Chem. 61, 384 (1986).
20. Behmer H. and Preisinger A., private communication, (1983).
21. Baumgartner O., Preisinger A., Krempl P. and Mang H., Zeit. Kristallogr. 168, 83-91, (1984).
22. Grimm H. and Dorner B. J. Phys. Chem. Solids 36, 407 (1975).
23. Lang R., Calvo C. and Datars W.R. Can. J. Phys. 55, 1613 (1955).
24. Krempl P.W. and Engel G.F. Ferroelectrics 40, 159 (1982).
25. Krempl P.W. and Engel G.F. Ferroelectrics 54, 357 (1984).
26. Ng H.N. and Calvo C., Can.J.Phys. 53, 2064 (1975).
27. Gervais F., Solid State Commmun. 18, 191 (1976).
28. Scott J.F., Phys. Rev. B4, 1360 (1974).
29. Landolt - Börnstein tables, New series, group III, vol. 11. Editor: Hellwege K.H., (Berlin: Springer Verlag, 1979).
30. Newnham R.E., Trans.Am.Cryst.Assoc. 11, 87 (1975).

2 ND EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 16 - 17 - 18 MARCH 1988

Session MP-01 Number 2

BERLINITE BLANK ABSOLUTE ORIENTATION DETERMINATIONS BY X-RAY DIFFRACTION

J.F. DARCES, M. EL BELLAL, H. MERIGOUX
Laboratoire de Cristallographie et Chimie Minérale
U.F.R. DES SCIENCES ET DES TECHNIQUES
F - 25030 BESANCON Cedex

ABSTRACT

The absolute orientation determination methods used for doubly or triply rotated Quartz cuts are applied to Berlinite crystals.

Although the Quartz and Berlinite structures are similar, for a cut defined by the same θ, ϕ angles it is not possible to observe the same lattice planes for Berlinite as those for Quartz. However the same method, using comparison between scattered intensities allows us to calculate the θ and ϕ angle values with their signs.

As for Quartz crystals, if an approximate orientation is known, we can compute and draw the expected diffraction patterns in order to choice the lattice planes giving the best accuracy.

Results on several Berlinite blanks are presented.

INTRODUCTION

It seems easy, if we present methods which give the orientation of a blank cut in a Berlinite crystal, to compare with the methods used for Quartz crystals. This comes from the great similitude that exists between their crystal structures.

The Si-Si sequence in Quartz is now an Al-P sequence. The great order in the succession of the aluminium and phosphorus atoms induce that the Berlinite c parameter is nearly twice the Quartz c parameter. So we have : $a_B \simeq a_Q$ and $c_B \simeq 2c_Q$. This very simple connection induces a very simple law with the (hkl) indexes of a lattice plane.

A (hkl) lattice plane, with the $(h_Q\ k_Q\ l_Q)$ values in quartz and $(h_B\ k_B\ l_B)$ values in Berlinite and with $h_B = h_Q$, $k_B = k_Q$, $l_B = 2l_Q$ possesses approximately the same θ and ϕ angles, if we describe this plane in the three axis system recommended by the IEEE conventions, ie $(\theta,\phi)_{(50.4)_B} \approx (\theta,\phi)_{(50.2)_Q}$. This remark is very helpfuf is we compare roughly directions described in both crystals. We cannot use it if we need a very accurate comparison.

There is also an other difference between Quartz and Berlinite crystals : the reflective powers or the scattering factors of the lattice planes are not always similar.

So, according to the knowledge of the blank orientation we must introduce some differences. If we just have to control the orientation during the preparation process, the methods used for quartz[1][2] may be directly adapted to Berlinite crystals. If we do not know at all about the presumed orientation, we propose a new procedure.

In both cases, the piezogoniometer we use is the device we have in our laboratory. It gives either a great accuracy when we measure an angular position or a great facility when we look for the observation of any lattice plane of the crystal space.

ORIENTATION DETERMINATION PRINCIPLES

If it is an orientation control, the already presented methods are still suitable.

The utmost interesting situation occurs when we have no information at all about the Berlinite orientation. Then we obtain the absolute orientation in two times :

- a first set of measures gives us the absolute value of the θ and ϕ angles of the blank[2]

- a second one gives the sign of these angles[3].

Determination of the absolute values of the θ and ϕ angles

We suppose that the blank orientation is defined by θ_o and ϕ_o. From θ_o and ϕ_o we calculate the contravariant coordinates of the blank in the crystal axis system and the ρ_i angle between the blank and the $(hkl)_i$ lattice planes.

We may have, for instance, ρ_1, ρ_2, ρ_3 respectively associated to $(hkl)_1$, $(hkl)_2$, $(hkl)_3$. If these three lattice planes are choosen between equivalent lattice planes according to the crystal symmetry operation, these three planes are observed without moving the X-ray detector on the piezogoniometer. In such a case the θ_o, ϕ_o, ρ_i and $(hkl)_i$ values are well correlated.

This method need the simultaneous observation of three or more lattice planes when the X-ray detector stays unmoved. So the choice of the angular position of the detector must be done very carefuly, because its value commands the number of the observed lattice planes.

This angular position value is twice θ_B, the Bragg angle corresponding to the expected planes.

The θ_B value must be as high as possible, because for a geometrical condition, a lattice plane can be seen only if

$$\rho_{measured} < \theta_B$$

The scattering factors of the expected planes must also have a very high value because it is possible we have a stong attenuation of the X-ray beam intensity[2].

With Berlinite we may use the (33.2) or (50.4) familles respectively seen with $\theta_B = 70.5°$ and $71.0°$[4].

During a measurement if we observe three or more ρ angles, as shown in figures 1 and 2, when the X-ray detector is unmoved we try to find a correlation between the unknown values θ_o, ϕ_o the measured ρ_i values and the $(hkl)_i$ values choosen between all the values predicted by the crystal symmetry. This trial and error method must give a solution if the observed ρ angles are actually associated with plans of the same (hkl) family.

If we have no solution it means that among the three observed ρ angles some of them are seen with the $K\beta$ or $K\alpha_2$ X-ray radiation instead of the expected $K\alpha_1$ radiation.

This kind of calculation gives only the orientation expressed by the absolute values of θ and ϕ.

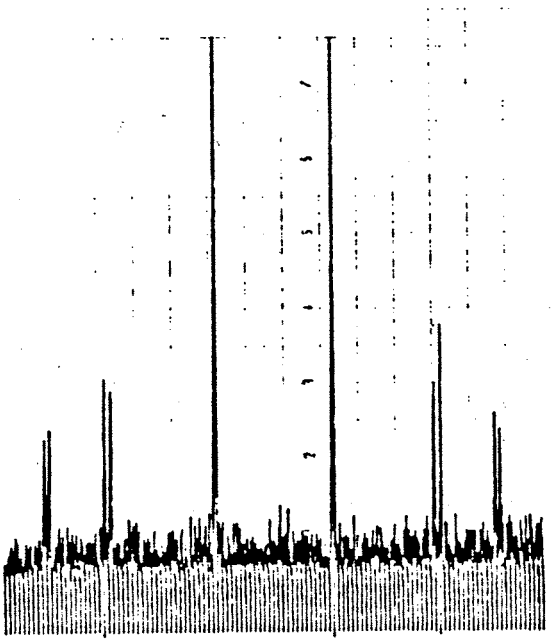


Figure 1 : Observed diagram when the Bragg angle is 71.0°.

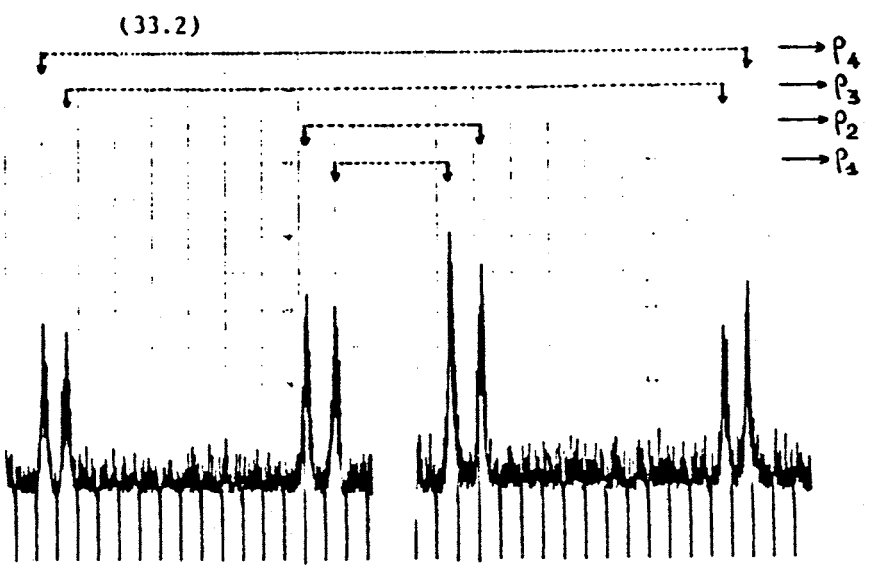


Figure 2 : ρ angle measurements on an observed X-ray diagram.

Table I gives a result obtained with a blank when $\theta_B = 70.5°$. We can notice that the first ρ angle is not linked with the (33.2) expected family.

(33.2) family lattice plane (hk.1) research

measurements	(hk.1)
$\rho_1 = 12.7°$	other family
$\rho_2 = 34.5°$	(33.2)
$\rho_3 = 36.1°$	$(\bar{3}6.2)$
$\rho_4 = 47.6°$	$(33.\bar{2})$

$|\theta| = 30.02°$
$|\phi| = 1.03°$

Table 1 : $|\theta|$ and $|\phi|$ automatic determination. We only introduce the $\rho_{i\,measured}$ and the name of the (hk.1) family, here (33.2)

Determination of the sign of θ and ϕ angles

The θ angle sign is easily obtained using plane families having the same Bragg angle but a very different scattering factor[4]. With crystals as Quartz or Berlinite the (hk.1) or (kh.1) families may be choosen so we have :

$$F(hk.1) << F(kh.1)$$

With Berlinite the two (50.4) and (05.4) families are again very careful.

Where the Bragg angle is fixed we compute the diagrams we would observe when the θ angle (θ IEEE) is affected with both positive and negative sign. Figures 3 and 4.

Then we compare them to the observed diagram. Figure 5.

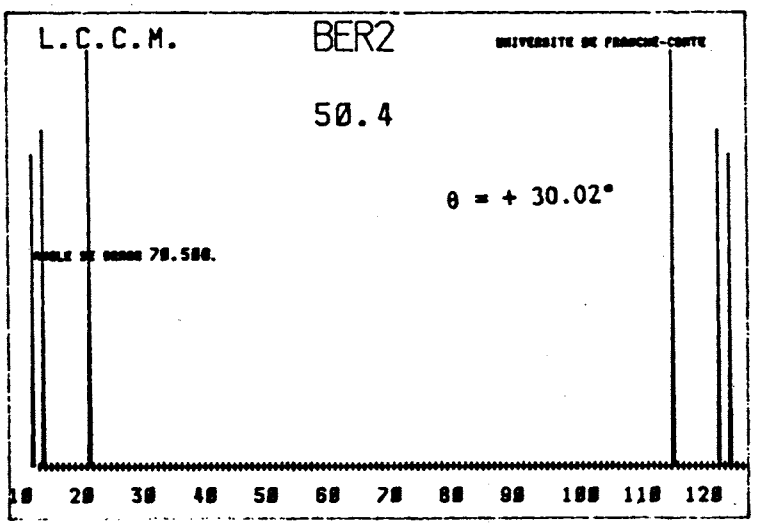


Figure 3 : Calculated diagram when $\theta = + 30.02°$ with (50.4) family ($\lambda = 1.540$ Å)

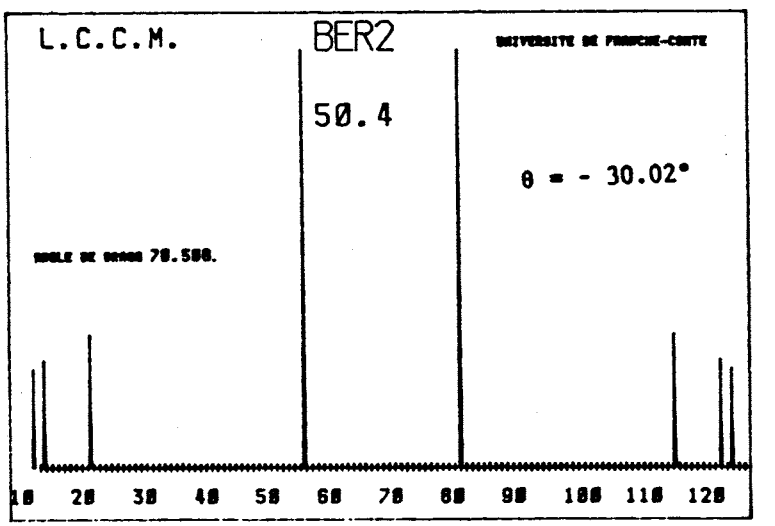


Figure 4 : Calculated diagram when $\theta = - 30.02°$ with (50.4) family ($\lambda = 1.540$ Å)

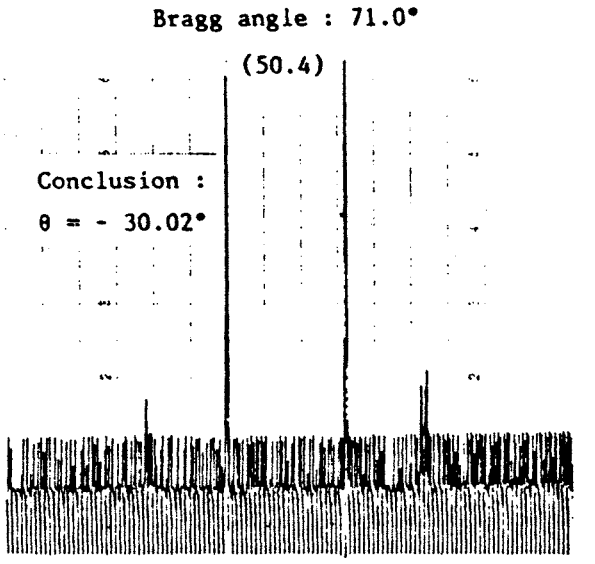


Figure 5 : Observed diagram obtained on a piezogoniometer. The comparison with figures 3 and 4 gives the θ angle sign.

The ϕ angle sign is determined as with Quartz, because Berlinite and Quartz correspond to the same symmetry group. There is the same difference between the two faces of a blank.

If one face is $(\theta, +\phi)$ the other is $(\theta, -\phi)$.

The distinction between both possibilities is linked to the appearance order of the lattice planes when we turn the crystal around the horizontal axis of the piezogoniometer (μ angles).

Results are presented for two blanks in tables 2 and 3.

BLANK N°1

(hkl)	$\theta_{measured}$	$\mu_{measured}$	$\phi < 0$		$\phi > 0$	
			$\mu_{calculated-}$	$\mu_{calculated-} - \mu_{measured}$	$\mu_{calculated+}$	$\mu_{calculated+} - \mu_{measured}$
(04.6)	50.35°	74.3°	163.7°	89.4°	196.3°	122.0°
(04.6)	52.00°	287.4°	16.9°	89.5°	343.1°	55.7°
(20.12)	36.90°	359.8°	89.3°	89.5°	270.7°	270.9°
(04.0)	30.02°	182.5°	272.1°	89.6°	87.9°	265.4°
(22.6)	26.51°	287.6°	17.0°	89.4°	343.1°	55.5°
(22.6)	24.78°	73.5°	163.0°	89.5°	197.0°	123.5°

CONCLUSION : $\phi < 0$

BLANK N° 2

(hkl)	$\rho_{measured}$	$\psi_{measured}$	$\phi > 0$		$\phi < 0$	
			$\psi_{calculated+}$	$\psi_{calculated+} - \psi_{measured}$	$\psi_{calculated-}$	$\psi_{calculated-} - \psi_{measured}$
(50.4)	38.95°	266.5°	90.8°	184.3°	269.2°	2.7°
(50.4)	55.93°	168.9°	353.1°	184.2°	6.9°	198.0°
(50.4)	56.85°	2.9°	187.2°	184.3°	172.8°	169.9°
(02.12)	45.41°	85.3°	169.7°	184.4°	90.3°	5.0°
(42.0)	44.87°	199.0°	23.3°	184.3°	336.7°	137.7°
(42.0)	45.76°	333.2°	157.4°	184.2°	202.6°	229.4°

CONCLUSION : $\phi > 0$

RESULTS

Diverse Berlinite cuts are investigated for a large range of θ angle values. θ and ϕ angle values are calculated from ρ angle measures for different lattice planes.

Blank n° 1

(hk.l)	$\rho_{measured}$
$(\bar{4}6.\bar{6})$	52.00°
$(40.\bar{6})$	50.35°
(04.0)	30.02°
$(22.\bar{6})$	24.78°
$(\bar{2}4.\bar{6})$	26.51°
$(02.\bar{1}\bar{2})$	36.90°

$\theta = - 30.02° \pm 0.005°$

$\phi = - 1.03° \pm 0.02°$

Blank n° 2

(hk.1)	$\rho_{measured}$	
(05.$\bar{4}$)	38.95°	
($\bar{5}$5.4)	55.93°	
(50.4)	56.85°	θ = + 21.52° ± 0.03°
(02.12)	45.41°	ϕ = + 0.50° ± 0.01°
($\bar{4}$6.0)	44.87°	
(42.0)	45.74°	

Blank n° 3

(hk.1)	$\rho_{measured}$	
(24.0)	41.87°	
($\bar{2}$6.0)	42.12°	
(42.0)	53.31°	
($\bar{4}$6.0)	53.72°	θ = - 31.15° ± 0.01°
(02.$\bar{1}\bar{2}$)	28.81°	ϕ = - 0.32° ± 0.01°
($\bar{2}$2.12)	43.86°	
(05.$\bar{4}$)	20.78°	

Some Y cuts are controled. The divergence between the expected orientation and the really realized orientation varies from nearly zero to half a degree.

CONCLUSION

A simple adjustment of the methods which are used for Quartz crystals allow us, in any case, to determine the absolute orientation of Berlinite blanks. Measurements are operated on the piezogoniometer which has been developped for Quartz crystals studies.

ACKNOWLEDGEMENTS

We wish to thank D. HAUDEN (Université de Franche-Comté - Besançon), J. DETAINT (C.N.E.T.), A. ZARKA (Université Pierre et Marie Curie) for providing diverse Berlinite blanks and the D.R.E.T. of the Ministry of Defense who partially supported this work.

REFERENCES

1 - J.F. DARCES, "Problèmes posés par la détermination de l'orientation absolue d'une lame monocristalline : application au quartz", Thèse de Doctorat d'Etat es Sciences Physiques, 1984, Université de Franche-Comté Besançon.

2 - M. MOUSSETAD, "Mesures cristallographiques par rayons X sur monocristaux à applications industrielles, Thèse d'Université, 1987, Université de Franche-Comté Besançon.

3 - J.F. DARCES, J. LAMBOLEY, H. MERIGOUX, "Dissymetry used for ϕ, θ angle sign determination of a piezoelectric crystal blank by a non destructive method", Ferroelectrics, 1982, vol. 40, pp. 245-248.

4 - SCHWARZENBACH, "Die tiefquartz - Modifikation Von Al PO_3", 2. Krist 123-161, 1966, Thesis Zurich, 1965.

2 ND EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 16 - 17 - 18 MARCH 1988

Session MP-01 Number 3

A NEW APPROACH TO BERLINITE CRYSTAL GROWTH

E. PHILIPPOT - A. GOIFFON - J.C. JUMAS. USTL MONTPELLIER UA407 CNRS

X. BUISSON - R. ARNAUD. SICN ANNECY

Y. TOUDIC. CNET LANNION

Since the work of CHANG and BARSCH (1) many works have been devoted to growing and characterizing berlinite crystals and to modelling and experiment with bulk and surface wave devices made with this material. It was successively shown that this material was a close piezoelectric analog of quartz, that possesses similar cuts with zero frequency temperature coefficients, either for bulk or surface wave applications.

In the past ten years, many efforts were made to develop hydrothermal methods that can produce large crystals of high perfection needed for piezoelectric applications (2-9). Unfortunatly all crystals show an excessive amount of water incorporated during growth, fig.1. Recently the importance of reducing the water solubility in crystals was emphasized in two papers (10-11) that show the large influence of this phenomenon on device properties (insertion losses, or Q factor, temperature coefficients,...). Transmission electron microscopy method (12) was used after high temperature precipitation of water to assess the H_2O content of crystal as a function of some growth parameters: temperature and pressure.

According to this "concept", following these investigations, we have demonstrated, last year, the feasability of obtaining very high Q crystals with thermal behaviour superior to this of quartz. These growth cycles have been performed in H_3PO_4 acid at T>240°C. Unfortunatly, a major drawback of this method is the difficulty of transposition at an industrial scale:

- very small change of berlinite solubility in H_3PO_4 acid when T>230°C, fig.2.
- seed dissolving before thermal equilibrium.

Instead of trying to modify drastically growth parameters, we have

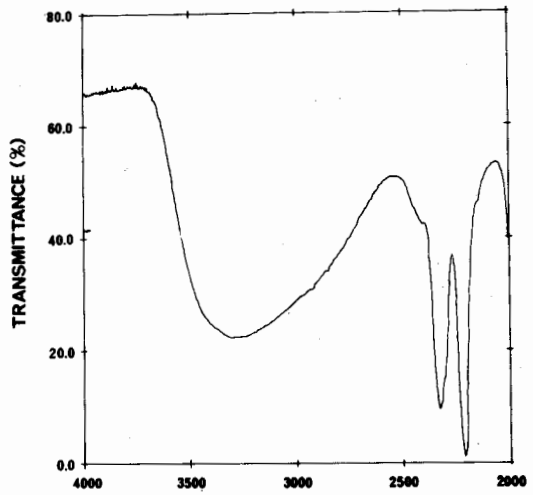


Figure 1 - Infrared absorption of "wet" berlinite. T=170°C, H_3PO_4 6.5M.

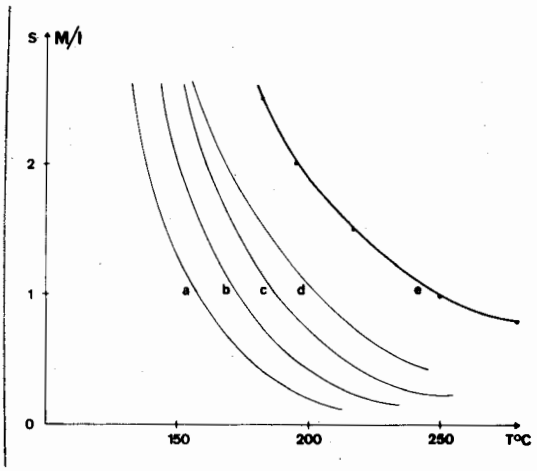


Figure 2 - $AlPO_4$ solubility in H_3PO_4: a=5M b=6.5M c=8M d=9.5M e=12M.

looked for another solvent. After some tests, we have choosen sulfuric acid, H_2SO_4. A preliminary study of berlinite in this medium with different concentrations shows a retrograde solubility, fig.3, just as with previously considered acids.

If we compare this solubility to that in other acids, with the same concentration, H_2SO_4 shows a much higher solubility of berlinite, fig.4.

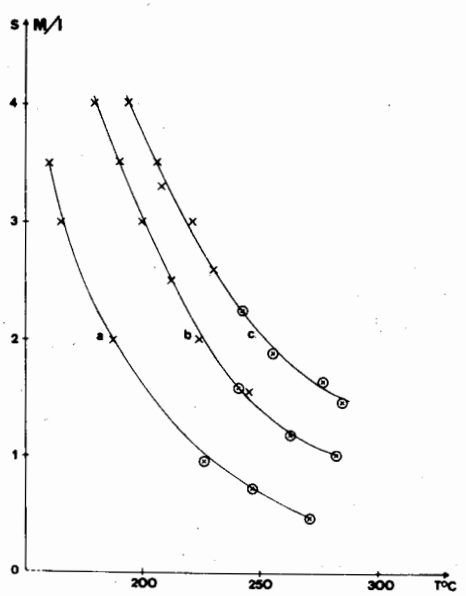


Figure 3 - $AlPO_4$ solubility in H_2SO_4 a=5M b=7.5M c=9M

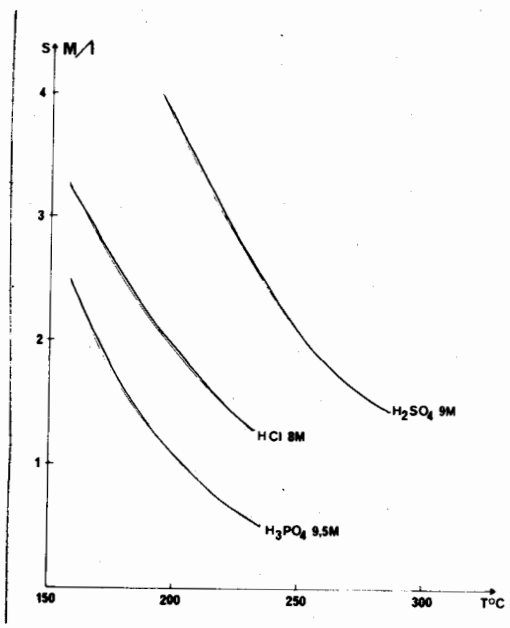


Figure 4 - Compared solubility of $AlPO_4$ in H_2SO_4, HCl and H_3PO_4 acids.

In these conditions, we have made some crystals from seeds with varied orientations (chiefly X, Z and Y). The crystals obtained were clear and of high crystalline perfection. Shape of crystals is quite different of that obtained in phosphoric acid, fig.5. Whatever the seed orientation, crystals obtained in sulfuric acid are chiefly delimited by π and π' faces instead of r and z in phosphoric acid, fig.6. Morphology difference must be explained from different growth rates.

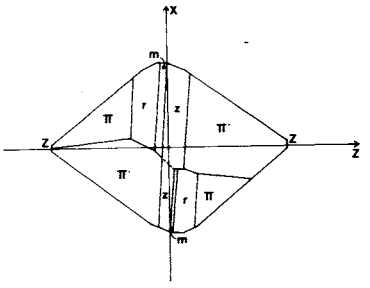


Figure 5 - Shape of berlinite crystals grown in sulfuric acid.

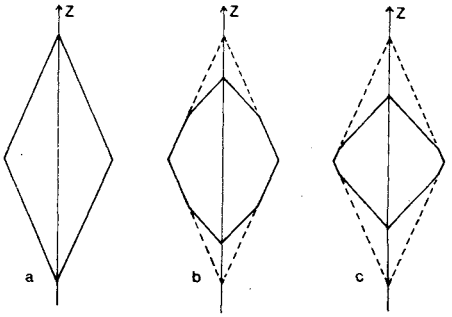


Figure 6 - Compared shapes of berlinite crystals grown in : a=H_3PO_4 b=H_3PO_4/H_2SO_4 c=H_2SO_4

In order to determine the best crystal growth conditions of berlinite in H_2SO_4 solvent, we have undertaken a systematic study in terms of:

- crystal growth temperature, Tc.
- temperature difference between cold and hot zone.
- solvent concentration: H_2SO_4 6 and 9M.

Crystal growth experiments in H_2SO_4 acid

All these experiments were conducted through the reverse temperature gradient method (2,3,6), using two 375cm^3 platinum lined autoclaves with baffle aperture 8.5%. To keep a low pressure during the growth, the filling is only 80% to avoid a "full" autoclave at Tc.

Though experiments are not yet quite completed, it can be emphasized some interesting results, in particular for growth rates Vx and Vz. Fig. 7 shows the Vx and Vz rates as a function of ΔT for a constant temperature, Tc=230°C. In the considered ΔT range, Vx and Vz rate variations versus ΔT are linear and their slopes can be computed: $a_{Vx}=0.14 = 2\star (a_{Vz}=0.07)$. This

difference explains the existence of a point where Vx=Vz≈0.08 mm/d/face. The higher sensitiveness of Vx to ΔT can be seen again when considering the Vx change with respect to Tc, with the same ΔT, fig.8.

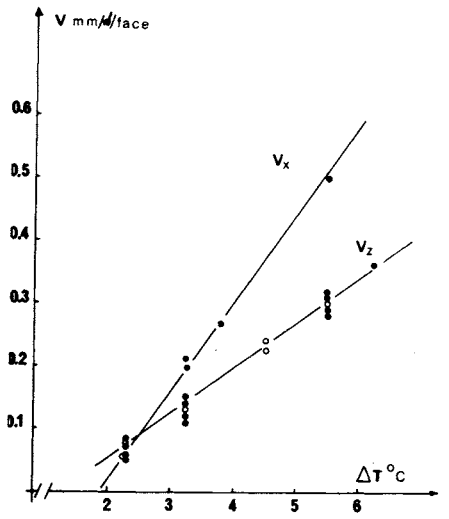


Figure 7 - Growth rates, Vx and Vz, in H_2SO_4 6M at 230°C (●=growth rate normal to the seed(Vx-Xseed), o=growth rate parallel to the seed (Vx-Zseed)

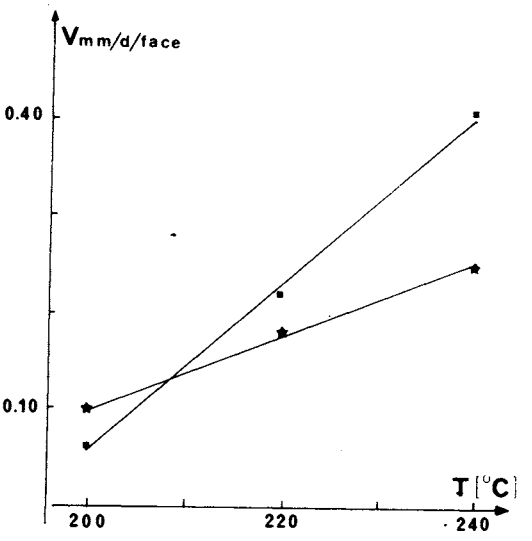


Figure 8 - Growth rates, Vx and Vz in H_2SO_4 6M with constant ΔT=5°C.

We have found the same results in sulfuric acid, H_2SO_4 9M. But in this case the rate variations are not always linear and show a maximum value for

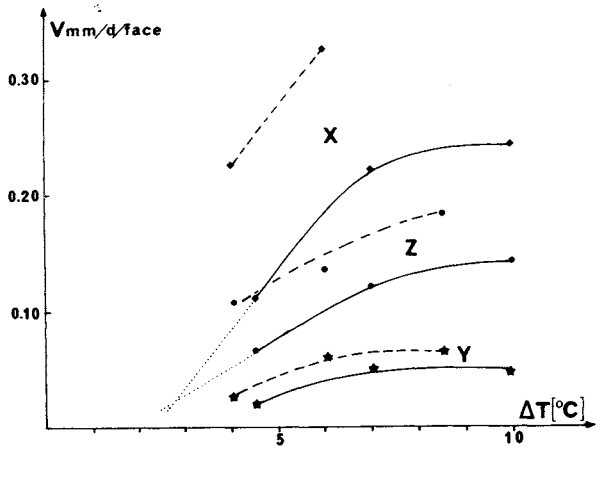


Figure 9 - Growth rates, Vx, Vz and Vy, in H_2SO_4 9M. . Lines for Tc=223°C, dashes for Tc=241°C, dots for extrapolated Vx=Vz for Tc=223°C.

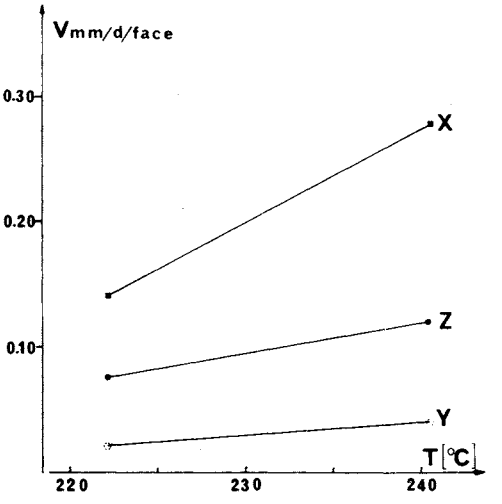


Figure 10 - Growth rates, Vx, Vz and Vy, in H_2SO_4 9M with constant ΔT=5°C.

ΔT=10°C, fig.9. In this case also, slopes can be computed for linear parts and it can be found the same values at 223 and 241 °C: a_{Vx}=0.05 and a_{Vz}=0.025 (a_{Vx}=2* a_{Vz}). It should be observed that increasing the value of Tc, from 223 to 241°C, gives a growth rate twice bigger. The same evolution can be observed if Vx and Vz are plotted versus Tc with constant ΔT=5°C, fig.10.

What conclusions can be drawn from these experiments? They are of several kinds and show a clear difference between crystal growth of berlinite in orthophosphoric and sulfuric acids.

Generally speaking, the growth rates, Vx and Vz, are much more sensitive to temperature in H_2SO_4: growth temperature, Tc, and vertical temperature difference, ΔT. That sensitivity impose a very accurate regulation of these parameters during crystal growth. This sensitiveness had not be encountered in H_3PO_4 solvent and increases with lower acid concentration:

H_2SO_4 6M	a_{Vx} = 0.14	a_{Vz} = 0.07
H_2SO_4 9M	a_{Vx} = 0.05	a_{Vz} = 0.025

but the ratio a_{Vx}/a_{Vz} is constant whatever the acid concentration, a_{Vx}/a_{Vz}=2.

In the other hand, the Vx growth rate is generally higher than the Vz one. It is the opposite of what happens in the orthophosphoric acid where Vz is always much higher than Vx. For exemple:

H_3PO_4 6.5M	Tc≃170°C	ΔT=5-6°C	Vx=0.12mm/d/face	Vz=0.40mm/d/face
H_2SO_4 6M	Tc≃230°C	ΔT=5-6°C	Vx=0.50mm/d/face	Vz=0.30mm/d/face

The growth rate anisotropy is much stronger in the H_3PO_4 solvent and explains the difficulty to obtain good crystal growth from Z-seeds. It can be observed liquid inclusions in the direction perpendicular to the seed plane. This anisotropy explains also different morphologies observed.

Nevertheless, in sulfuric acid, it exists a narrow temperature range where Vz>Vx and then a temperature where Vz=Vx. The corresponding value of Tc can be expressed in terms of ΔT, fig. 11. It can be observed that values are independent of acid concentration, 6 or 9M.

Finally an increasing value of acid concentration decreases strongly the growth rate with the same ratio for Vx and Vz: V_{9M}/V_{6M} = 1/3 fig.12.

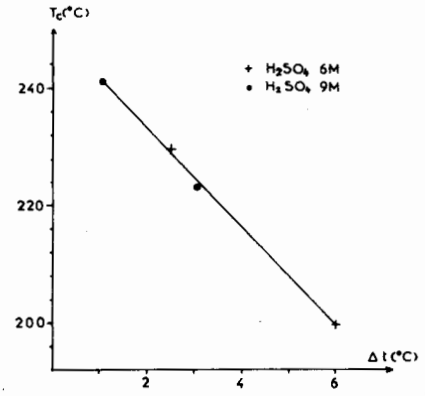


Figure 11 - Growth temperature where Vx=Vz in terms of ΔT.

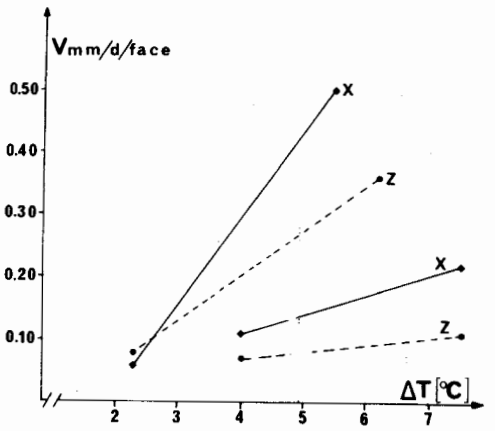


Figure 12 - Growth rates, Vx and Vz, for approximatively the same Tc≈230°C. In both cases $V_{6M} \simeq 3x\ V_{9M}$.

Infrared spectrometry measurements.

As water content, $c[H_2O]$, in the crystals and their acoustic attenuation are drastically related, we have performed a systematic investigation by infrared spectrometry of water content in terms of crystal growth characteristics. Samples were slices taken on both sides of the seed. Results are certainly pessimistic because measurements have been undertaken from small crystals with no more than 2mm growth thickness.

Nevertheless, it can be assessed water content depends of: growth temperature, Tc, growth rate, V, and acid concentration. In fig. 13, we have

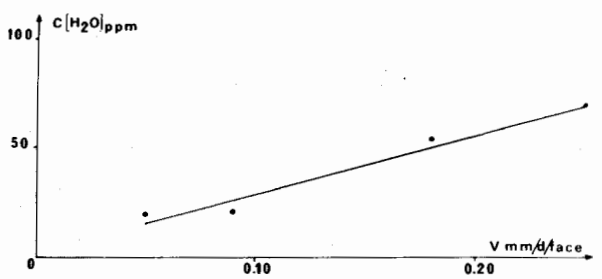


Figure 13 - Water content, c H_2O ppm in terms of growth rate, Tc=240°C, H_2SO_4 6M.

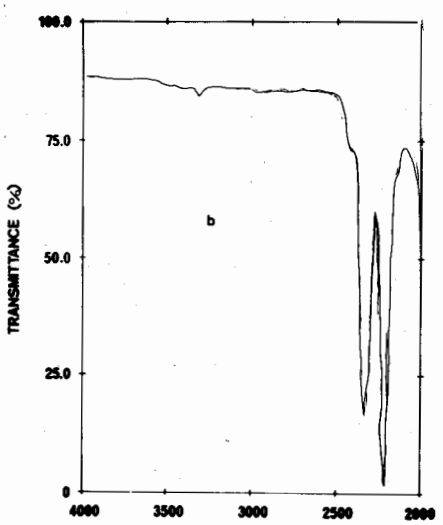
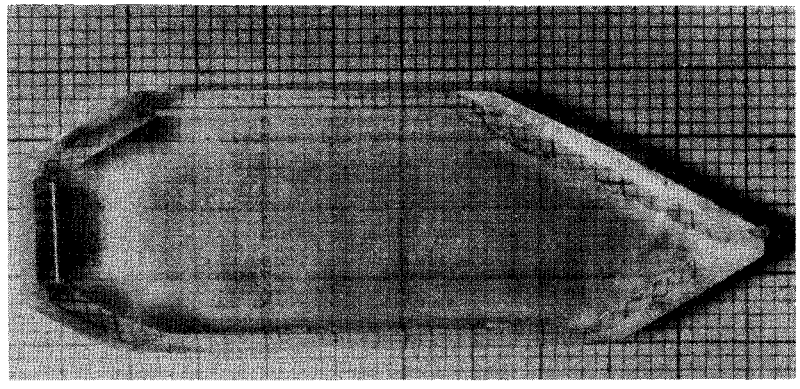


Figure 14 - IR absorption of "dry" berlinite

reported results for Tc≈240°C and H_2SO_4 6M. For weak groth rate, V≈0.10 mm/d/face, crystals contain not much water, in the 20 ppm range, comparable to that of good quartz, fig.14.

The results of bulk resonator measurements will be described elsewhere.

In conclusion, sulfuric acid is a very good solvent for berlinite crystal growth. It presents the highest known solubility of $AlPO_4$ without seed dissolving before thermal equilibrium of autoclaves. Under certain conditions of temperature and acid concentration, it can be obtain "dry" berlinite and industrial crystal growth is now investigated by SICN, Annecy.



Crystal grown at Tc≈230°C, H_2SO_4 6M.

REFERENCES

1 - Y.P. CHANG, G.R. BARSCH, IEEE Trans. Sonics Ultras, 23,127, 1976.

2 - J.C. JUMAS, A. GOIFFON, B. CAPELLE, A. ZARKA, J.C. DOUKHAN, J. SCHWARTZEL J. DETAINT, E. PHILIPPOT, J. Cryst. Growth, 80, 133, 1987.

3 - E.D. KOLB, R.A. LAUDISE, J. Cryst. Growth, 43, 313, 1978.

4 - D.F. CROXALL, T.R.A. CHRISTIE, B.J. ISHERWOOD, A.G. TODD, J. BIRCH, Europ. Conf. on Cryst. Growth, Lancaster Univ., 1979.

5 - B.H.T. CHAI, M.L. SHAND, E. BUEHLER, M.L. GILLEO, Proc. Iee Ultrasonics Symp., 577, 1979.

6 - K. NAGAI, T. OGAWA, J. OKUDA, J. ASAMARA, S. TAKI, Proc. 1st Int. Symp. Hydrothermal Reactions, 496, 1982.

7 - J. DETAINT, H. POIGNANT, Y. TOUDIC, Proc. 34th Freq. Control Symp., 93, 1980.

8 - B.H.T. CHAI, E. BUEHLER, J.J. FLYNN, US Patent, 4 324773, 1982.

9 - E. PHILIPPOT, J.C. JUMAS, A. GOIFFON, R. ASTIER, French Patent, 2559166 1984.

10 - R.S. NARAYANAN, B.H.T. CHAI, J.F. VETELINO, W. SOLUCH, R.S. FALCONER,

B. CHICK, Proc. IEEE Ultrasonics Symp., 262, 1984.

11 - R.F. STEINBERG, M.K. ROY, A.K. ESTES, B.H.T. CHAI, R.C. MORRIS, Proc. IEEE Ultrasonics Symp., 279, 1984.

12 - J.C. DOUKHAN, B. BOULOGNE, P. CORDIER, E. PHILIPPOT, J.C. JUMAS, Y. TOUDIC, J. Cryst. Growth, 84, 167, 1987.

2 ND EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 16 - 17 - 18 MARCH 1988

Session MP-01 Number 4

LATTICE DEFECTS IN LITHIUM TANTALATE

by
P. CORDIER, B. BOULOGNE, N. DOUKHAN and J.C. DOUKHAN
Laboratoire de Structure et Propriétés de l'Etat Solide
Université de Lille Flandres Artois, 59655 Villeneuve d'Ascq Cedex - FRANCE
and
M. FRAGNEAU and J.P. AUBRY
Compagnie d'Electronique et de Piezo Electricité
BP 165 - 95105 ARGENTEUIL - FRANCE

ABSTRACT

Lithium tantalate single crystals are grown by a Czochralski process. To get homogeneous crystals it is necessary to start from a non equimolar mixture of oxides Li_2O and Ta_2O_5 leading to a congruent melt. As a result the crystals are not stoichiometric and this introduces a first kind of lattice defects : the point defects associated to this non stoîchiometry ($[Li]/[Li+Ta] \cong 48\%$). When cooled down from high temperature, $LiTaO_3$ suffers a second order phase transition at 620°C from a paraelectric rhombohedral phase $R\bar{3}c$ to a ferroelectric rhombohedral phase R3c which is the one stable at room temperature. A second type of lattice defects, ferroelectric domains are thus generally nucleated at the transition. They constitute a poison for piezoelectric applications because the polarisation vector is reversed in the domains. However one can in principle prevent the occurrence of such domains by a poling process (cooling under a static electric field). Other as-grown lattice defects are dislocations and twins while the machining operations (sawing, grinding ...) introduce fresh dislocations with a density which can largely overcome the one of as-grown defects. Twins are rare in high quality crystals and the strain fields of the dislocations alone should have a very small effect on the performances of the piezoelectric devices. However the piezoelectric constants of $LiTaO_3$ are so large that the dislocation strain fields can nucleate ferroelectric domains which cannot be removed by a further poling process. Dislocation cores are also preferred diffusion paths for cationic exchange (Li exchanged with other monovalent cations like H, Na, K ...). Therefore any lattice defect, as long as it can be considered as isolated should not induce severe limitations to the piezoelectric performances, or there exists some process to eliminate this defect. In contrast when one considers interacting defects, dramatic damages arise which cannot be suppressed. This paper presents an investigation by transmission electron microscopy and by chemical etching of the defects occurring in as-grown and machined $LiTaO_3$.

I - INTRODUCTION

There is no known natural lithium tantalate. This material has been synthetized since 20 years approximately with a view to technologic applications, especially piezoelectric applications owing to its very strong coefficients of electro mechanical coupling (see for instance a recent review in (1)). Furthermore, $LiTaO_3$ is one of the rare piezoelectric materials exhibiting crystallographic orientations such that the frequency of a resonator is not affected by temperature variations (the temperature dependences of the stiffness constants and of the piezoelectric constants are such that no frequency shift results). The crystal structure of $LiTaO_3$ has been refined by X-Ray and neutron diffraction techniques (2,3). It is isostructural with $LiNbO_3$ and the structural principles and models for the ferroelectric R3c - paraelectric $R\bar{3}c$ phase transition have thoroughly been discussed by Megaw (4,5). The only difference is the transition temperature which is of the order of 620°C rather than 1100°C for $LiNbO_3$. The ferroelectric domain walls correspond to a reversal of the polarity and are thus an electric poison for piezoelectric applications. In principle such domains can be eliminated by a "poling" operation which consists in applying a static electric field when cooling down the crystal from the high temperature paraelectric phase. At room conditions the atomic coordinates are given in Table 1. A simplified structure can be drawn in assuming that the oxygen sublattice is a perfect HCP structure with the cations Li^+ and Ta^{5+} occupying the octahedral sites. As there are as many octahedral sites as oxygen ions, 2/3 of these sites are occupied by the Li and Ta's while the last third is empty. As pointed out by Megaw (5) the rigid framework of the structure is constituted by the skeletton of edge sharing TaO_6 octahedra. This framework is shown on Figure 1 and the configuration of a ferroelectric domain with its polarity reverses is depicted on Figure 2 for a single column of octahedra parallel to the $\vec{c}$ axis. This domain is obtained by jumps of the Li ions from their initial position beneath the Ta ions to the empty octahedral sites above the lower Ta ions. The intermediate configuration corresponds to the paraelectric (high temperature) phase.

We have investigated by chemical etching and by transmission electron microscopy (TEM) the lattice defects occurring in this structure, either the as-grown defects or the ones which can accidentally be introduced during mechanical machining. The above experimental techniques

only allow extended lattice defects (dislocations, stacking faults and twins) to be observed. It is well known that high quality Czochralski-grown $LiTaO_3$ is Li deficient with a ratio Li/(Li+Ta) of the order of 48%. The point defects associated to this non stoichiometry,as long as they are randomly distributed,cannot be detected by TEM ; however if some clustered to form extended defects like crystallographic shear structures these later defects could be in principle unambiguously detected and characterized by TEM (see 6 or 7 for extended reviews on these defects in non stoichiometric oxide compounds).

Among the chemical impurities which can occur in $LiTaO_3$, hydrogen is rapidly dissolved in $LiTaO_3$ and is easily substituted to Li, mostly under high temperature electromigration. We have thus estimated the hydrogen concentration in $LiTaO_3$ crystals from various origins by infrared absorption which is a technique very sensitive to OH vibrations (8). We have simulated the damages due to mechanical machining by two types of experiments i) by constant strain rate plastic deformation at high temperature and ii) by scratching a well polished surface and observing by chemical etching and by TEM the resulting defects. We first report our experimental results, then we discuss their possible influence on the performances of devices like thermally compensated resonators.

II - EXPERIMENTAL RESULTS

2.1 Macroscopic characterizations

Crystals from various origins have been tested, however most of the experiments have been done on two crystals only. One was grown 10 years ago by CNET, Bagneux and the other one has been grown recently by CRISMATEC. Characterization of non stoichiometry (i.e. the ratio Li/(Li+Ta)) can in principle be performed by X-Ray diffraction, the lattice parameters a and c slightly varying with composition but the relative variations $\Delta a/a$ and $\Delta c/c$ are very small (9,10) and reliable measurements would require a very high precision diffractometer. We have used another method based on the variation with the composition of the ferroelectric-paraelectric transition temperature (9 to 12). This temperature is easily measured on very small pieces of material by Differential Scanning Calorimetry (DSC). For fast scans (40°C/min) one records slightly different transition temperatures for experiments done by increasing T from room temperature and the

ones done by decreasing T from high temperature ; therefore one does not obtain absolute compositions (the calibration curve itself is not very precise) but comparisons between different samples are easy. Some typical DSC curves are reported on Figure 3. Choosing the turning point of the heating curve as a measurement of the transition temperature T_t a change ΔT_t of 4° (which is unambiguously detected) corresponds to a change of 0.1% of the composition ratio Li/(Li+Ta) (see 10). As can be seen on Figure 3 the crystal grown by CRISMATEC has a rather homogeneous composition but this composition varies from boule to boule (T_t = 621 ± 3°C in the recent boule, T_t = 628 ± 1°C in an older boule). The japanese material (HITACHI) has a different composition (T_t = 633°C) nearer to stoichiometry. We have also performed a series of annealings in air to test the possibility of reequilibration of the concentration of oxygen point defects and (or) the possibility of Li-H exchange (there are always some traces of H_2O in air). No change of T_t is detected after an annealing of $\approx$ 100 hours at 800°C. However a marked change of the OH content is detected on these samples by infrared absorption spectroscopy. For standardisation we have first characterized by infrared spectroscopy a cube cut in the as-grown CNET crystal with its faces parallel to (0001), $(2\bar{1}\bar{1}0)$ and $(0\bar{1}10)$ respectively. The absorbance for unpolarised light in the range 4000-3000 cm^{-1} is reported on Figure 4. This clearly shows that there is some dissolved hydrogen in the material (characteristic absorption due to the stretching vibration of O-H groups in well defined sites) the largest absorbance occurs for a beam parallel to the $\vec{c}$ direction. This is the geometry adopted for the other tests done on annealed samples. To estimate the H content we have used the same calibration as for wet quartz (13). The concentration H/Li in atomic ppm is approximately equal to (or at least proportional to) the surface in cm^{-2} under the absorbance curve of a basal slab. This leads to a concentration H/Li $\cong$ 30 at ppm in the as-grown CNET crystal. In as-grown CRISMATEC samples this concentration is approximately three times larger and is homogeneous in all the parts of the boule which have been tested. After annealing at $\approx$ 500°C no change is detected but for annealing at $\cong$ 800°C, the H content increases significantly for long annealing times and H/Li is roughly multiplied by a factor 3 after 48 hours annealing (see Figure 4b). The solubility of H in $LiTaO_3$ is already known and the H content is known to affect a number of physical properties, however Li-H exchange at high temperature in the absence of electric field appears to be very low or to have a very low cinetics as compared to the Li-Ag exchange

observed by Bursill and Peng at temperatures as low as 200°C (14). From the few characterizations reported here, it can be concluded that there are significant differences from boule to boule which probably originate in slightly different conditions of growth, however every boule appears to be homogeneous. Further investigations at a microscopic scale are necessary to get a better characterization, especially the lattice defects, dislocations and ferroelectric domains must be characterized.

2.2 Microscopic investigations by chemical etching

The main advantage of this technique is that it is very simple. It just requires some polishing machine to get very well polished surfaces to be chemically etched and an optical microscope to observe the resulting etch pits. $LiTaO_3$ like $LiNbO_3$ is rather chemically inert. The only known chemical etchant is a solution constituted of 1/3 of concentrated HF and 2/3 of concentrated HNO_3 the etching conditions being 1 to 4 hours at 110°C (9, 15 to 18). A variety of samples have been mechanically polished then etched. The best results for the mechanical polishing have been obtained with SYTON, grade W30 (provided by MONSANTO). Figure 5 is a typical result on a basal slab. One first remarks that the etching rate is different on both faces, the etch pits being much more marked on one of the faces (Figure 5b). On this later face the etching process has dug tunnels which emerge at the etch pits. One could be tempted to interpret these tunnels as the result of a preferential dissolution of the dislocation cores, however such a process should probably occur on both faces, it is therefore more probable that the tunnels reveal ferroelectric domains having the shape of pipes. As shown below dislocations or at least some types of dislocations have a stress field such that pipe shaped ferroelectric domains are nucleated around them. The etch pit pattern would thus reveal the dislocation emergences while the tunnel pattern would reveal the dislocation lines themselves owing to their associated domains. The major problem with the etch technique is that one is never sure that all the dislocations - and only dislocations - are revealed. For instance it is well known that very small amounts of impurities can affect dramatically the etching conditions in such a way that only some types of dislocations would be revealed (for instance the so called fresh dislocations while the as-grown dislocations with cores saturated by the impurities would exhibit a different reactivity and could be unrevealed). Whatever the exact mechanism of etching, the defects introduced by severe mechanical grinding are easily detected as

shown on Figure 6. The corresponding surface has been scratched with low grade SiC abrasive (large grains) then perfectly polished with SYTON W30 in such a way that all scratches had disappeared. After etching, the scratches are again clearly visible;further polishing with SYTON shows a decrease of the pit density along the scratches and it is necessary to remove a layer of the order of 10 μm with SYTON to suppress the damages induced by the scratches. It is reasonable to assume that the induced defects are half loops of dislocations nucleated by the SiC abrasives grains acting as mobile indenters.

2.3 Characterization of lattice defects by Transmission Electron Microscopy

Tantalium is a highly absorbant element for the electron beam and very thin and well polished lamellas have thus to be prepared by ion thinning. Thin samples have been mechanically polished on both faces with SYTON down to a thickness of the order of 30 μm then ion thinned. We have remarked that ion thinning with a moderate accelerating voltage (4 to 4.5 kV) at usual incidence (15°) induces a marked preferential thinning at the emergence of the remaining defects especially the scratches as shown on Figure 7.With an accelerating voltage of 6 to 7 kV,the samples become dark. Annealing under vacuum induces a similar darkening which has been attributed to the nucleation of point defects associated with a loss of oxygen because a further annealing in air restores the initial transparency (9). It is not impossible that the argon beam accelerated by 6 to 7 kV induces locally an appreciable increase of temperature ; it could also induce irradiation damage. Finally, an accelarating voltage of 5.5 kV and an incident angle of 15° have appeared as a good compromise. Two types of defects have been observed in TEM : dislocations and ferroelectric domains.

By heating a crystal above its transition temperature and cooling it without electric field one introduces a large density of ferroelectric domains. The domain boundaries have any orientation (Figure 8). Similar domains but with a much lower density are also systematically observed in materials which have been poled. They are sometimes connected to a dislocation (Figure 8b) but one also observes more or less cylindrical domains without dislocations (Figure 8c). It is not clear if these later domains have resisted to the poling operation or if they have been nucleated at room temperature by the mechanical stresses developed during machining

(sawing and (or) grinding). Whatever their origin they must induce severe limitations to the piezoelectric performances if they become pervasive, thus leading to a reversed polarity in an appreciable volume.

A few dislocations are also observed, some are surrounded by ferroelectric domains (see Figure 8b) but some are not (we come back to this important point in the discussion). Furthermore some dislocations are apparently not dissociated while others are clearly bounded by a stacking fault (Figure 9). One also notices in the vicinity of this dislocation several cylindrical domains with a diameter of the order of 1 μm. The diffraction contrast experiments are consistent with a Burgers vector $\vec{b} = \frac{1}{3}|10\bar{1}1|$. We come back to this point in the discussion. We have also been able to observe some of the dislocations nucleated by a scratch by ion thinning the sample on one of its faces only (of course the thinning time is twice longer). Figure 10 shows these dislocations in an approximately edge-on configuration. They all are surrounded by a ferroelectric domain with the shape of a butterfly. These dislocations are distributed along the initial direction of the scratch and the dislocation lines themselves are roughly parallel to the | | direction. It has not been possible to examine their core structure (dissociation) because of the contrast of the surrounding domain which hides the one of the dislocation core. For the same reason it has not been possible to characterize the corresponding Burgers vectors. However it becomes now evident that at least some types of dislocations (i.e. some types of Burgers vectors and line orientations) have a stress field such that a ferroelectric domain can be spontaneously nucleated around them at room temperature. During their glide, as long as the dislocation direction remains unchanged they probably move with their ferroelectric domain attached to them ; if the line orientation changes during glide it is not impossible that the ferroelectric domains would be left in their wake. Such domains could thus slightly rearrange their configuration to decrease their surface energy. They would thus tend to a cylindrical shape similar to the ones observed on Figure 8c. The above hypothesis is a possible mechanism for the formation of these cylindrical domains.

2.4 The case of twins

No as-grown twins have been observed by TEM although the occurrence of twins in the $\{10\bar{1}2\}$ planes has already been reported in $LiTaO_3$ as well as in $LiNbO_3$ (5,9,14,19). Furthermore Bursill and Peng (14) have

motion of these dislocations they would move together with their surrounding domain at least as long as their direction remains unchanged during glide. If the dislocation reaches the end of the crystal and escapes, the accompanying domain should also vanish. One can wonder what happens if an external stress induces dislocation multiplication by a Frank-Read source mechanism , the dislocation directions change into their glide plane and if all the dislocations were surrounded by a domain whatever their orientation, the crystal should become progressively pervaded by ferroelectric domains. One can imagine that because of their spatial extension some new domains should grow into older ones thus restoring the initial polarity. It is clear that such a situation would dramatically affect the piezoelectric performances of the crystal. It is however to be remembered that apparently not all dislocations are surrounded by a ferroelectric domain. The occurrence of a domain, or it nucleation probably depends on the dislocation characteristics, Burgers vector and line direction i.e. on the orientation and magnitude of its stress field or strain field. It could thus happen that during its motion a dislocation changes its line orientation and does no more generate a domain around its new line ; such a dislocation would thus move and would let a cylindrical domain in its wake. This mechanism if it does occur would render account of the numerous small cylindrical domains without dislocations into them like the one shown on Figure 8c. It is thus important to characterize the various types of dislocations which can be generated into $LiTaO_3$, especially the ones which are nucleated by mechanical grinding and to estimate the ability of these dislocations to nucleate ferroelectric domains around them. Such a study is actually in progress, the various dislocation types and their electrostatic properties are not yet completely elucidated. We develop here some simple theoretical considerations on the crystallographic aspects (Burgers vectors, glide planes, dissociations ...) and the associated elastic and electric fields.

3.1 Dislocations properties related to the crystal structure

The crystal structure R3c is schematically represented on Figure 1, the smallest lattice repeats are the basal vectors of the hexagonal cell $1/3\langle 2\bar{1}\bar{1}0\rangle$ with a modulus a = 5.15 Å ; the basic vectors of the rhombohedral cell $1/3\langle 10\bar{1}1\rangle$ have almost the same modulus r = 5.47 Å. The next shortest lattice repeats are $1/3\langle 20\bar{2}1\rangle$ with a modulus of 7.51 Å then $\langle 10\bar{1}0\rangle$ with a modulus of 8.92 Å and finally c (13.78 Å). The dense planes

observed "chemical" twins in the $\{10\bar{1}4\}$ planes (for the concept of "chemical twinning" see Anderson and Hyde (20)). We have thus generated mechanical twinning by deformation (twinning is known to occur at moderate temperature under relatively high stress). We have chosen a compression orientation parallel to $\vec{c}$ which gives a large Schmid's factor for both kinds of twins, the ones in $\{10\bar{1}2\}$ as well as the ones in $\{10\bar{1}4\}$. At room temperature or moderate temperature, $LiTaO_3$ exhibits a purely elastic behaviour up to fracture even for the lowest strain rates available with an Instron machine $\dot{\varepsilon} \cong 2 \times 10^{-6}\ s^{-1}$. Fracture occurs for compressive stresses $\sigma \cong 400$ MPa. For temperatures of at least $\tilde{\sim}$ 800°C, and applied stresses above 100 MPa, slight discontinuities superimpose on the elastic deformation. The stress-strain curves appear as a succession of discontinuities with various amplitudes (see Figure 11) which are typical of twinning deformation. This is confirmed by the examination of thin sections (30 μm) observed in optical microscopy between crossed nicols. Three directions of twins are clearly visible (Figure 12) ; they correspond to the three equivalent planes $(10\bar{1}\bar{2})$, $(0\bar{1}1\bar{2})$ and $(1\bar{1}0\bar{2})$. In TEM these twins are unambiguously characterized (Figure 13), they correspond to the usual twins of the hexagonal close packed structures like for instance Cd, Zn, Mg (see any standard book on defects in crystals like (21)). No twin in $\{10\bar{1}4\}$ has been found. It is also to be noted that almost no twin dislocations are found except at the intersections of two twin bands. This shows that the twin dislocations are very mobile ; as soon as nucleated they propagate up to the end of the crystal (or to another twin). The geometry of the twins $K_1 = (10\bar{1}2)$, $\eta_1 \,//\, [\bar{1}011]$ is consistent with a shear produced by partial dislocations with a Burgers vector resulting from the dissociation of the lattice translation $1/3[\bar{1}011]$ in $(10\bar{1}2)$ this point as well as other models of dislocations with or without ferroelectric domains are now discussed with a view to get a more coherent view of the above experimental results.

III - DISCUSSION

The above observations have shown that in some cases the dislocations are surrounded by ferroelectric domains which are probably nucleated by the stress field of the dislocations themselves. This means that such ferroelectric domains cannot be removed by a poling operation because they would be automatically renucleated at low temperature when the poling field is removed. One can also presume that if an external stress induces a glide

which could operate as glide planes are the basal plane (0001), the rhombohedral planes $\{10\bar{1}2\}$ which are also the clivage and twinning planes and the $\{10\bar{1}4\}$ planes which have been characterized as composition planes for the chemical twins (14). $\vec{a}$ glide in the basal plane is one of the most usual glide systems of hexagonal structures and dissociation of the $\vec{a}$ dislocations into Shockley partials is also quite usual. It would locally change the HCP sublattice of the O^{--}'s into a FCC one and the distance between the Ta cations would be changed leading to an increase of their electrostatic energy of interaction. An associated synchroshear of the Ta cations could perhaps minimize the increase of this energy. As some other tantalates like $NaTaO_3$ and $KTaO_3$ have the perovskite structure (22), one could look for dissociation modes which would generate in the staking fault a lamella with the perovskite structure.

Dislocations with a Burgers vector $1/3[10\bar{1}1]$ can also be dissociated into their glide plane $(10\bar{1}2)$. Such a dissociation could account for the mechanical twinning in $(10\bar{1}2)$ and furthermore the skeletton of TaO_6 octahedra is not affected by this dissociation. The glide system which appears to be the easiest one to activate would thus be $\{10\bar{1}2\}1/3$ $<10\bar{1}1>$; this is consistent with our TEM observations (Figure 9) but it is clear that more experimental work and more precise TEM characterizations have to be done to confirm these crystallographic considerations. Assuming that the dislocations occurring in $LiTaO_3$ have the possible Burgers vectors $1/3<2\bar{1}\bar{1}0>$ or $1/3<10\bar{1}1>$ we now consider their ability to generate ferroelectric domains.

3.2 Electrostatic properties of dislocations

We consider in this section the electric field generated by dislocations of the type defined above. Let us consider a system constituted by a cylinder of unit length of $LiTaO_3$ material with a single straight dislocation along its axis which is by convention the z axis. To know if a ferroelectric domain can be nucleated around the dislocation, one has to compute the Gibbs energy (G) of the corresponding system. It is the sum of three terms i) the elastic energy associated to the dislocation, ii) the electrostatic energy associated to the electric field induced by the strain and iii) the surface energy of the domain boundary. The exact shape of the domain can be computed, at least in principle, by minimizing G in considering various boundary shapes. To compute G it is

necessary to solve the well-known electro-elastic equations which describe an insulating piezoelectric material (23 to 29)

$$\sigma_{ij} = \sum_{k\ell} C_{ijk\ell} \frac{\partial u_k}{\partial x_\ell} + \sum_m e_{mij} \frac{\partial V}{\partial x_m} \qquad (1a)$$

$$D_m = \sum_{k\ell} e_{mk\ell} \frac{\partial u_k}{\partial x_\ell} + \sum_m \chi_{mj} \frac{\partial V}{\partial x_m} \qquad (1b)$$

$c_{ijk\ell}$, $e_{mk\ell}$ and χ_{mj} are the components of the stiffness tensor, of the piezoelectric tensor and of the dielectric tensor respectively. σ_{ij}, u_k and V are the components of the stress tensor of the displacement and the electrostatic potential. Within a ferroelectric domain $c_{ijk\ell}$, $e_{mk\ell}$ and χ_{mj} are constant but when one crosses a domain boundary they change abruptly. σ_{ij}, u_k and V are generated by the dislocation lying along the cylinder axis and do not vary with the corresponding coordinate of this axis. D_m is the component of the electric excitation field.

The equilibrium conditions are

$$\sum_j \frac{\partial \sigma_{ij}}{\partial x_j} = 0 \qquad \text{(mechanical equilibrium)} \qquad (2a)$$

$$\sum_j \frac{\partial D_j}{\partial x_j} = 0 \qquad \text{(electrical neutrality)} \qquad (2b)$$

The boundary conditions also have to be considered. The external surfaces of the cylinder are stress free. On the lateral surface with unit normal n_j, one has

$$\sum_j \sigma_{ij} n_j = 0 \qquad (3a)$$

Furthermore the Burgers circuit theorem imposes the following condition

$$\oint_\Gamma du_i = b_i \qquad (i = 1 \text{ to } 3) \qquad (3b)$$

where the b_i's are the components of the Burgers vector and Γ is any closed circuit enclosing the dislocation line.

This electro elastic problem is completely solved if one knows

(30) in $LiNbO_3$. Their results can be used here as a correct approximation of the expected electric fields around dislocations in $LiTaO_3$. These authors find that $\vec{E}$ can become huge in the immediate vicinity of some dislocation cores ($|\vec{E}|$ reaches values as high as 10^7 V cm^{-1} at distances of the order of 1000 Å in $LiNbO_3$). Similar values are probably obtained in $LiTaO_3$ for the same dislocation types and we suppose that the cases with the largest values of $\vec{E}$ and with $\vec{E}$ opposite to the spontaneous polarisation orientation $// \vec{c}$ are the ones with ferroelectric domains around the dislocations. The examination of Figure 2 of Cerva et al. (30) shows that such a case happens for edge dislocations with b = 1/3<1$\bar{1}$01> and the domain boundaries defined as the locus of equal E modulus would exhibit a butterfly shape somewhat similar to the one observed here (Figure 10). In contrast $\vec{E}$ around a screw dislocation with the same Burgers vector is more symetrically distributed and its modulus is lower. One can suppose that no domain would be nucleated around such dislocations. It is clear that the above considerations remain extremely rough and qualitative. A more quantitative approach is necessary but such a study starts by a detailled and unambiguous characterization of the dislocations in $LiTaO_3$, as grown dislocations as well as the ones introduced by mechanical machining.

IV - CONCLUSION

The present investigation of lattice defects in $LiTaO_3$ shows that some types of defects could dramatically affect the performances of piezoelectric devices and a detailled study with carefull characterizations is necessary before being able to predict the effect of such or such defect. Such a knowledge could probably allow improvements of the actual elaboration processes. The first step is probably a control of the composition and the homogeneity of the starting material. However the actual author's feeling is that the most dangerous defects would be the dislocations as far as they are able to nucleate around them ferroelectric domains which cannot be removed by a poling operation.

the displacement vector $\vec{u}$ and the electric potential V at any point. The stress tensor $\overline{\overline{\sigma}}$ and the electric excitation vector $\vec{D}$ can thus be calculated with eqs(1). The elastic energy part of the total Gibbs energy is the following integral computed in the volume of the cylinder

$$G_{elast} = \iiint_V \frac{1}{2} \sum_{ij} \sigma_{ij} \frac{\partial u_i}{\partial x_j} \, dv \qquad (4a)$$

This integral is in fact divided into two parts, one corresponding to the "normal" material which occupies the volume V_n while the other one which occupies the remaining volume $V-V_n$ corresponds to the ferroelectric domain. The electrostatic energy is also given by a volume integral over the total volume V (which is to be divided into V_n and $V-V_n$)

$$G_{e\ stat} = \iiint \frac{1}{2} \sum_{j} D_j \frac{\partial V}{\partial x_j} \, dv \qquad (4b)$$

Finally the surface energy of the domain wall is more difficult to evaluate because its energy density can vary with the surface orientation. A rough approximation consists in multiplying the total surface by some constant representing a mean density of surface energy. All the terms of the total Gibbs energy G being computed, G should be minimized. It is to be noted that in the above calculation an unknown quantity q is introduced at some level. q represents the linear charge density on the dislocation line and G should also be minimized versus q to get the true equilibrium configuration.

The complete calculation would be very complicated and would not be very precise because the surface energy density of the domain boundary is not known (it is certainly anisotropic and it probably varies appreciably with temperature). There exists however a simplified method. One assumes that there is no domain for the computation of the field generated by the dislocation, $\overline{\overline{\sigma}}$, $\vec{u}$ and V are thus computed by the method already developed by previous authors (24 to 29), and the electric field $\vec{E} = -$ grad V is easily computed at any point around the dislocation. Then, if in some region, $\vec{E}$ is found to be large and roughly opposite to the spontaneous polarisation, one assumes that in the corresponding region a ferroelectric domain can be spontaneously nucleated whitout changing appreciably the stress and the strain field of the dislocation. Computations of $\vec{E}$ around various types of dislocations have already been performed by Cerva et al.

REFERENCES

(1) L.E. HALLIBURTON and J.J. MARTIN (1985),Properties of piezoelectric materials in E.A. GERBER and A. BALLATO (Eds) Precision Frequency Control, Vol.I, Academic Press, p.1-45.

(2) S.C. ABRAHAMS and J.L. BERNSTEIN (1967), J.Phys.Chem.Solids 28, 1685

(3) S.C. ABRAHAMS, W.C. HAMILTON and A. SEQUEIRA (1967), J.Phys.Chem.Solids. 28, 1693.

(4) H.D. MEGAW (1968), Acta Cryst.A 24, 583

(5) H.D. MEGAW (1973), Crystal Structures : a working approach ; Saunders p.233

(6) L. EYRING and M. O'KEEFFE (1970), The chemistry of extended defects in non metallic solids, North Holland

(7) R. DE RIDDER, J. VAN LANDUYT and S. AMELINCKX (1972), Phys.Stat.Sol. (a) 9, 551

(8) R.D. AINES and G.R. ROSSMAN (1984), J.Geophys.Res. 89, 4059

(9) M. NOLF, R. COQUILLE, A. LEBRET, P. AUVRAY, R. LANCON and J. DETAINT (1976), Technical report CPM/PMT/39, CNET

(10) R.L. BARNS and J.R. CARRUTHERS (1970), J.Appl.Cryst. 3, 395

(11) MIYAZAWA and IWASAKI (1972), Review Elec.Comm.Lab. 21, 374

(12) A.A. BALLMAN, H.J. LEVINSTEIN, C.D. CAPIO and H. BROWN (1967),J.Amer. Ceram.Soc. 50, 657

(13) M.S. PATERSON (1982), Bull.Mineral. 105, 20.

(14) L.A. BURSILL and J.L. PENG (1982), Philos Mag. 46, 1

(15) K. NASSAU, H.J. LEVINSTEIN and G.M. LOIACONO (1965), Appl.Phys.Lett. 6, 228

(16) K. NASSAU, H.J. LEVINSTEIN and G.M. LOIACONO (1966), J.Phys.Chem. Solids 27, 983

(17) N. OHNISHI and T. LIZUKA (1975), J.Appl.Phys. 46, 1063

(18) J.D. VENABLES (1974) Appl.Phys.Lett. 25, 254

(19) A.A. BLISTANOV, I.V. NOSOVA and M.M. TAGIEVA (1975), Soviet Phys. Crystallogr., 20, 409

(20) S. ANDERSON and B.G. HYDE (1974), J.Solid State Chem., 9, 92.

(21) A. KELLY and G.W. GROVES (1970),Crystallography and Crystal Defects, Longman p.291

(22) F.S. GALASSO (1969), Structure, properties and preparation of perovskite type compounds, Pergamon Press

(23) W.G. CADY (1946),Piezoelectricity, Mc Graw Hill

(24) L. MERTEN (1964), Phys.Kondens.Mater. 2, 53 ; (1964), Ibid 2, 66 ; (1964),Z.Naturf. 19, 788

(25) A.M. KOSEVICH, L.A. PASTUR and E.P. FELDMAN (1968), Soviet Phys.Crystallogr. 12, 797

(26) G. SAADA (1971), Phys.Stat.Sol. (b) 44, 717

(27) G. FAIVRE & G. SAADA (1972),Phys.Stat.Sol.(b) 52, 127

(28) I.D. ESHELBY, W.T. READ and W. SHOCKLEY,Acta Metall. 1, 251

(29) L. TREPIED (1978), J.Physique 39, L 433

(30) H. CERVA, P. PONGRATZ and P. SKALIKI (1986),Philos Mag. 54, 199

	Li_I	Li_{II}	Li_{III}	Li_{IV}	Li_V	Li_{VI}
x	0	0	1/3	1/3	2/3	2/3
y	0	0	2/3	2/3	1/3	1/3
z	0.02783	1/2 + z	2/3 + z	1/6 + z	1/3 + z	-1/6 + z

	Ta_I	Ta_{II}	Ta_{III}	Ta_{IV}	Ta_V	Ta_{VI}
x	0	0	1/3	1/2	2/3	2/3
y	0	0	2/3	2/3	1/3	1/3
z	0	1/2	2/3	1/6	1/3	5/6

	O_I	O_{II}	O_{III}	O_{IV}	O_V	O_{VI}	O_{VII}	O_{VIII}	O_{IX}
x	0.0498	1-y	y-x	1-y	x	y-x	1/3 + x	4/3 - y	1/3 + y-x
y	0.3436	x-y+1	1-x	1-x	x-y+1	y	y - 1/3	2/3 + x-y	2/3 - x
z	0.0688	z	z	1/2 + z	1/2 + z	1/2 + z	2/3 + z	2/3 + z	2/3 + z

O_X	O_{XI}	O_{XII}	O_{XIII}	O_{XIV}	O_{XV}	O_{XVI}	O_{XVII}	O_{XVIII}
4/3 + y	1/3 + x	1/3 + y-x	2/3 + x	2/3 + y	2/3 + y-x	2/3 - y	2/3 + x	2/3 + y-x
2/3 - x	2/3 + x-y	y - 1/3	1/3 + y	1/3 + x-y	1/3 - x	1/3 - x	1/3 + x-y	1/3 + y
1/6 + z	1/6 + z	1/6 + z	1/3 + z	1/3 + z	1/3 + z	5/6 + z	5/6 + z	5/6 + z

Table 1 : Atomic coordinates in hexagonal axes $\vec{a}_1$, $\vec{a}_2$ and $\vec{c}$. The lattice parameters are a = 5.15 Å and c = 13.78 Å and the atomic positions are given by $\vec{R}_i = x_i\vec{a}_1 + y_i\vec{a}_2 + z_i\vec{c}$. The x_i, y_i, z_i values are listed below for the 30 atoms (6 $LiTaO_3$ units in the $\vec{a}_1$, $\vec{a}_2$, $\vec{c}$ cell) at room conditions.

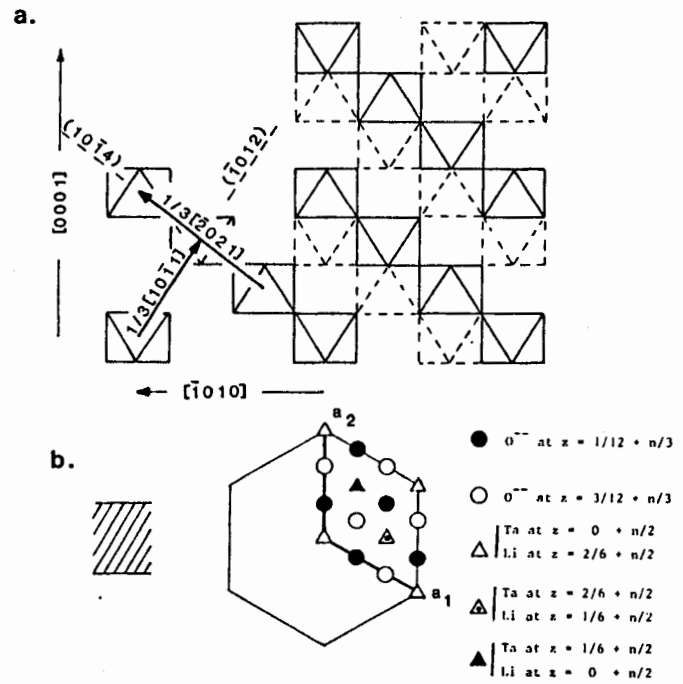


Figure 1 : Simplified structure of $LiTaO_3$. The oxygen sublattice is assumed to be a perfect HCP lattice. The Ta ions occupy the octahedral sites drawn in full lines. The Li ones occupy the octahedra just below (broken lines). A few small lattice repeats which could be Burgers vectors are indicated

a/ projection on $(1\bar{2}10)$. For the sake of simplicity only a slab of the structure corresponding to the width of an octahedron is represented.

b/ projection on (0001). The width of the above slab is indicated on the left (hatched region). The atomic coordinates along the $\vec{c}$ axis are indicated by symbols explained on the right of the figure.

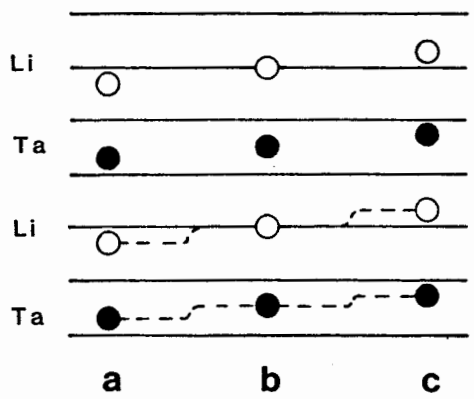


Figure 2 : Schematic representation of the formation of a ferroelectric domain by jump of the Li ions in the empty octahedral sites. The HCP oxygen sublattice in indicated by the sequence of planes ABAB ...

a/ starting configuration

b/ intermediate configuration corresponding to the high temperature paraelectric phase

c/ domain with reversed polarisation.

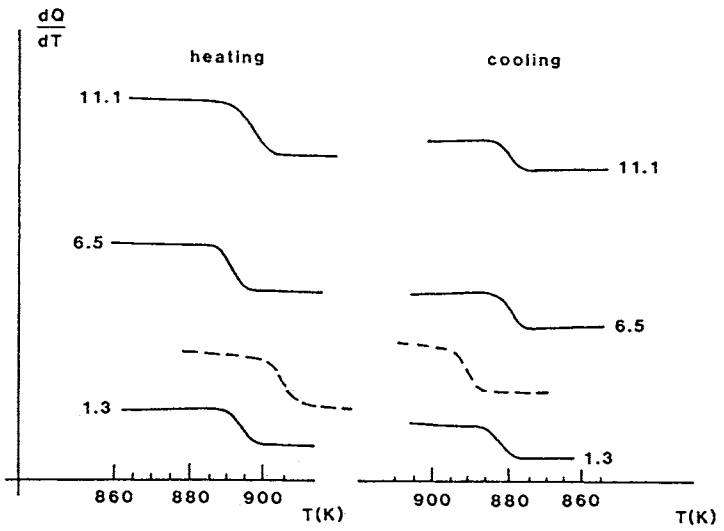


Figure 3 : Typical DSC curves obtained with a scan rate ≅ 40°C/min for increasing and decreasing temperature. Labels 11.1, 6.5 and 1.3 correspond respectively to the bottom, the middle and the top of a CRISMATEC boule. The hatched curves correspond to a HITACHI crystal with a slightly different stoichiometry.

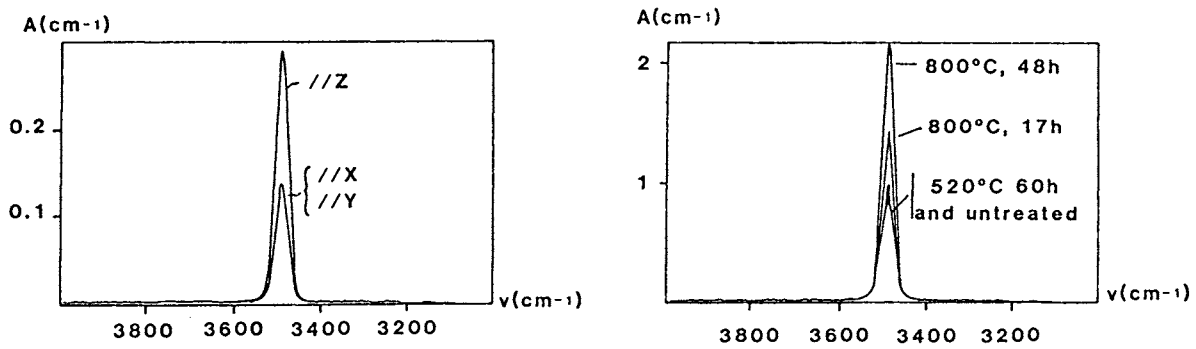


Figure 4 : Infrared absorbance spectroscopy
a/ influence of the crystallographic orientation ; as grown CNET crystal tested with the incident beam parallel to Z (c axis), X (a_1 axis) and Y respectively
b/ influence of annealing in air ; CRISMATEC crystal, beam parallel to Z, the annealing conditions are reported on each curve.

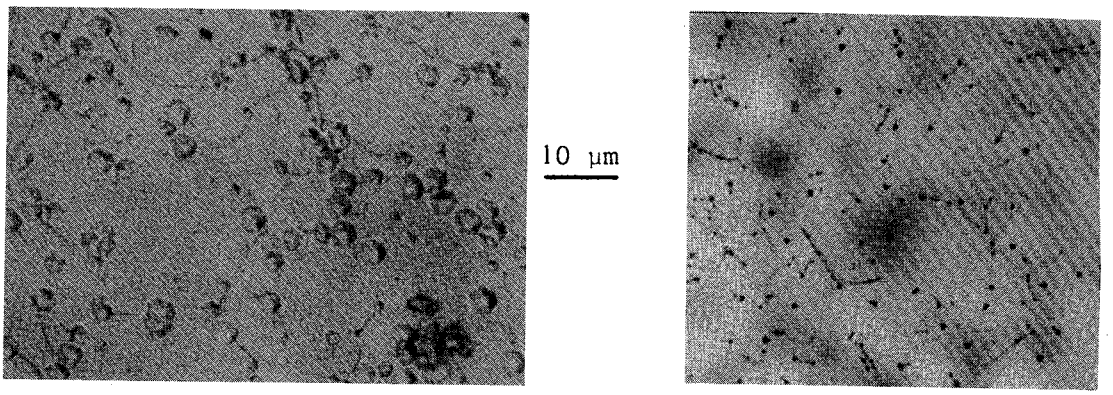


Figure 5 : CNET crystal, basal slab mechanically polished with SYTON W30 then etched 2h at 110°C in 1/3 HF + 2/3 HNO_3. a and b are the two opposite faces which are etched at different rates. On face b one clearly distinguishes the channels emerging at the etch pits.

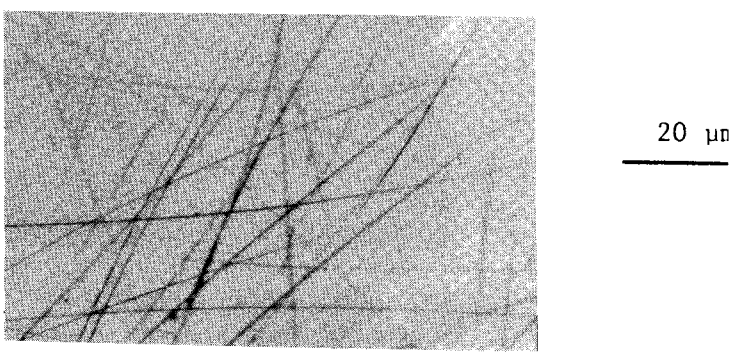


Figure 6 : Effect of scratches made with low grade SiC abrasive. The surface has been polished with SYTON up to a good polishing, the scratches being no more visible but the defects induced are revealed by etching. $(10\bar{1}0)$ face, CRISMATEC crystal.

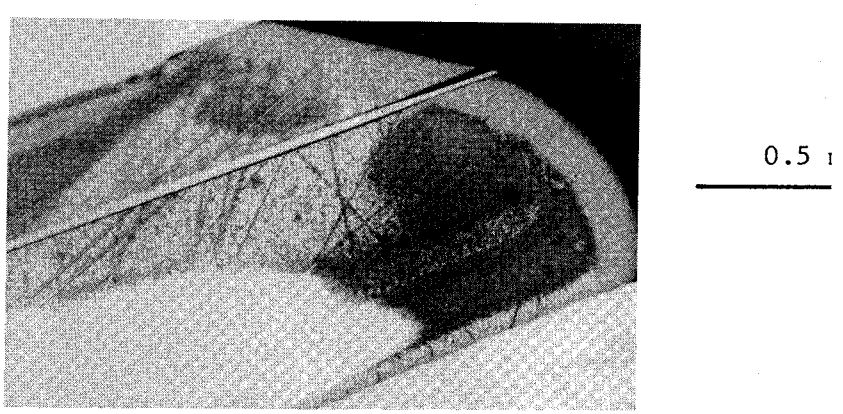
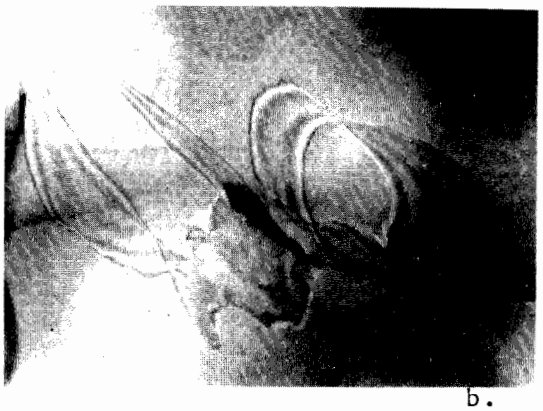
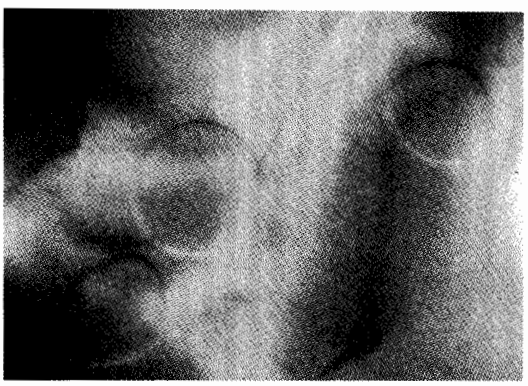


Figure 7 : Defects revealed by preferential ion thinning with an accelerating voltage 4.5 kV

a.



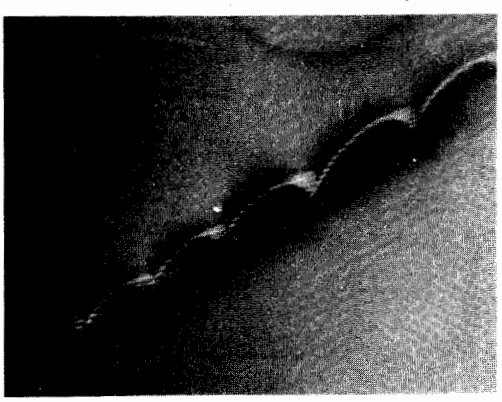
b.



1 µm

c.

Figure 8 : Ferroelectric domains observed in TEM
a/ after heating above T_t and cooling without electric field
b/ in a poled material, ferroelectric domain connected to a dislocation line
c/ roughly cylindrical domains not connected to dislocations in a poled material. The cylinder axes are approximately parallel to a_1.



0.5 µm

1 µm

Figure 9 : Poled CRISMATEC crystal;$(2\bar{1}\bar{1}0)$ thin foil
a/ scalloped dislocation bounded by a stacking fault;weak beam
b/ the dislocation is out of contrast for $g = \bar{1}2\bar{1}0$. One can see the pinning defects which could be very small inclusions with a diameter <500Å.

Figure 10 : Dislocations introduced by a scratch at room temperature. Note that all are surrounded by a ferroelectric domain nucleated by the dislocations CRISMATEC crystal; $(10\bar{1}0)$ foil
a/ the dislocations are in contrast and the domains out of contrast
b/ same area, dislocations and domains are in contrast.

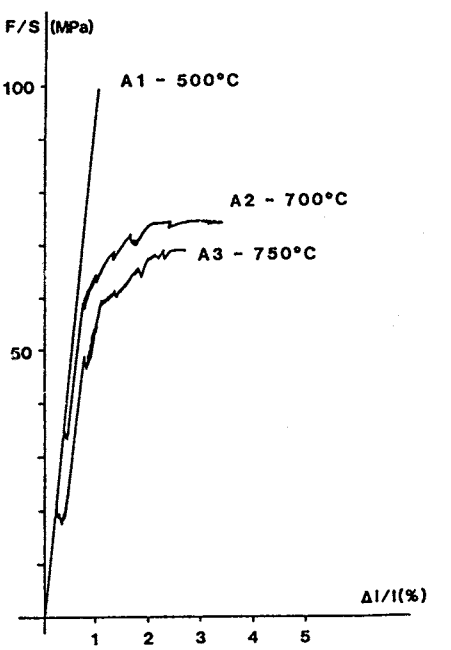
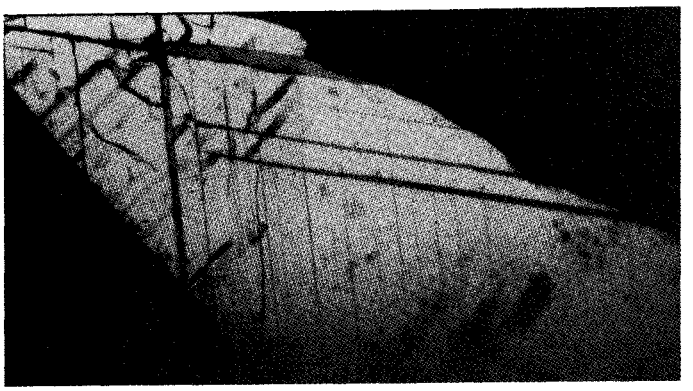
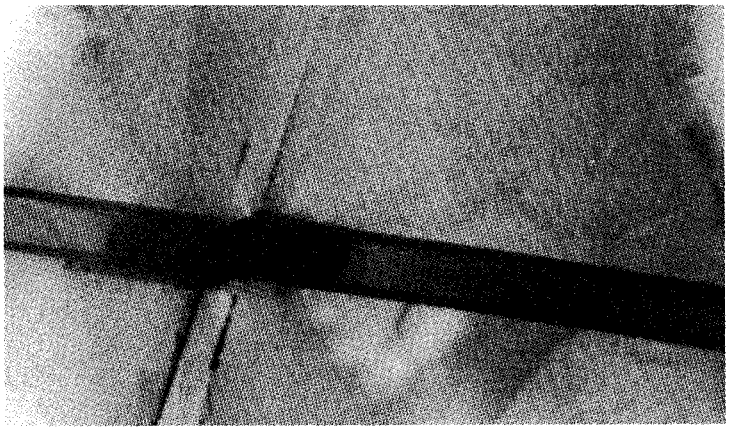


Figure 11 : Stress-strain curves for constant strain rate compressive tests. For all the samples the compression orientation is // c and the strain rate is $\dot{\varepsilon} = 2 \times 10^{-6} s^{-1}$. The temperature of the test is indicated on each curve



1 mm

Figure 12 : Mechanical twins generated by compression along the c direction at T = 800°C.Thin section (30 μm) observed between crossed nicols. The three equivalent twinnings in $\{10\bar{1}2\}$ are present; $(1\bar{1}00)$ lamella.



1 μm

Figure 13 : Mechanical twin produced by uniaxial compression along the c axis at 750°C;$(11\bar{2}0)$ thin foil ; note the irradiation damage on the twin boundary.Note also the ferroelectric domains due to the fact that this sample has been heated above T_t.

Session MP-01 Number 5

EVALUATION AND APPLICATIONS OF HIGH PURITY BERLINITE CRYSTALS

J. DETAINT[1], J.P. AUBRY[2], B. CAPELLE[3], A. ZARKA[3]
J. SCHWARTZEL[1], L. MICHEL[2], R. LEFEVRE[1], E. PHILIPPOT[4]
(1) CNET PAB/BAG/MCT/CMM BAGNEUX - FRANCE
(2) CEPE 44 Avenue de la Glacière Argenteuil - FRANCE
(3) Université P. et M. Curie Laboratoire de Minéralogie et de Cristallographie PARIS - FRANCE
(4) Université de Montpellier - Laboratoire de Physico-Chimie des Solides Montpellier FRANCE

INTRODUCTION

Berlinite is a compound a isostructural to quartz. The interest of this material results mostly of enhanced piezoelectric properties (1). It was demonstrated that this materials possesses thermally compensated orientations (2) (3) and that it was possible to obtain large crystals (4). In the recent years the feasibility of crystals with high Q factors was demonstrated (5) (6) using high temperature growth in the phosphoric acid. Very recently, (7) a new growth method was set up to obtain high Q factor crystals. In this method, a new solvent (H_2SO_4) is used with growth conditions much more favorable for an industrial development of the material (8). In this paper, we report the results of the characterizations made during the development of this new growth method. The results of crystallographic investigations and of electrical measurements made on resonators are discussed. Then, we present results concerning some of the compensated cuts of berlinite, and examples of devices made with these cuts.

For these studies, we have used crystals grown at the Laboratoire de Physico-chimie des Matériaux (Université de Montpellier) and at the Société Industrielle des Combustibles Nucléaires (Annecy-France).

I - CHARACTERIZATION BY X-ray TOPOGRAPHY

This method was used to determine the nature and the density of extended defects which can exist in synthetic berlinite and then to assess the crystal quality as a function of the growth parameters and the seed quality.

The most common extended defects in berlinite, are, as in quartz, the dislocations. However in the first growth experiments, with the new solvent, solid inclusions and related defects were also observed. In figures 1 to 4 are given typical topographs of the differents features observed during the optimization of the growth conditions in H_2SO_4 solvent. The sample of figure 1 is a Y plate obtained from a crystal grown on a Y seed. The contrasts of this topograph are quite intricated but one can observe that most of them are related to small solid inclusions. The sample presented in figure 2 is an X plate grown from a Z seed. The dislocations are nearly the only observable defects in this plate. Their density is still quite high. Many of the dislocations result of the propagation of those existing in the seed but others are initiated at the seed interface. This confirms the importance of the selection of seeds with high crystalline perfection and to proceed with good initial growth conditions.

In figure 3 is presented a topograph of a crystal grown on a seed containing very few dislocations. Although, the initial growth conditions were not then optimized, the density of dislocations is comparable with that of a synthetic quartz of electronic grade. In this sample quite all the dislocations have their origin at the seed interface. The topograph of figure 4 is typical of a high quality crystal obtained from a high quality seed. It has a highly reduced dislocation density ; this density appears to be even lower than that usually observed in the higher grades of synthetic quartz.

III - PIEZOELECTRIC CHARACTERIZATION OF CRYSTALS

Many bulk wave resonators were made with the crystals obtained during the crystallogenesis studies. Several electrical properties were measured : the frequencv constants, the temperature coefficients, and the Q factors.

A previous study (13) have demonstrated that these quantities are quite sensitive to the impurities of the crystals. In berlinite, as in quartz, water (OH) is the most common impurity, it limits the Q factor and products a modification of the temperature coefficients of the resonance frequencies (RFTC).

In Table I are displayed the results obtained with Y cut resonators.

Table I : Properties of Y cut resonators (crystals grown on Y seeds)

CRYSTAL (solvent)	Q(5MHz)	NF (KHz.mm)	CTF1 $10^{-6}\,°C^{-1}$	CTF2 $10^{-9}\,°C^{-2}$
H3P04 180°	3500	1700	+55 à ±65	-60 à -180
H2S04 200°	48400	1714	78.1	24.8
H2S04 210°	53300	1710	79.0	45.4
H2S04 220°	82500	1712	80.9	47.
H2S04 230°	51500	1714	81.3	47.
CALCUL	---	1688	63.5	---
QUARTZ Y	---	----	92.	58.

In this table we can notice that the Q factors obtained with the new solvent are much higher than those obtained from crytals grown in phosphoric acid at low temperature. The values measured for plan resonators using, here, the fundamental mode are rarely higher than 75000 at 5 MMz. The frequency constants (which are mostly dependant of C_{66}) are also quite sensitive to the [OH] concentration. The temperature coefficients of the resonance frequencies are very dependant of this concentration. All these facts were already observed with the crystal grown in H_3PO_4 at high temperature (13), but it must be remarked that with the new solvent, the results obtained for T_c=260° to 280°C in H_3PO_4 are now obtained for T_c near 210°C, this demonstrate the fundamental interest of H_2SO_4.

The electrical response of a Y cut resonator using the fundamental mode is given in Figure 5. This resonator was cut in a crystal grown at 220°C in H_2SO_4. The Q factors measured with the fundamental mode of plane Y plates are much lower than the intrinsic value caracteristic of the material ; a better approximation can be obtained using higher overtones. In Table 2 are given the Q factors measured with the 5th and 7th overtones.

Table 2 : Q factors (at 5 MHz) for overtones of Y resonators.

CRYSTAL	FUNDAMENTAL	OVERTONE 5	OVERTONE 7
H2SO4 210°	38200	840000	650000
H3PO4 265°	65400	450000	--
H3PO4 240°	32000	300000	--

In this table we have the confirmation of the interest of the solvent. In figure 6 is represented the phase of S_{21} for the 5th overtone of a Y plate obtained at 207°C in H_2SO_4.

In table 3 the results obtained for plates of other orientations are given.

We can observed that the AT cut of crystals with low OH concentrations is situated near Y-32°, whereas for water rich crystals, it was found near Y-30°. In table 1 and table 3 it can be noticed that the 1st and 2nd order R.F.T.C. of berlinite are lower than those of the same orientations for quartz. In figure 7 the electrical response of a Y-32° resonator is given.

Table 3 : Properties of Y rotated resonators.

CUT	CRYSTAL	Q (5MHz) $Q_{fund.}^{tal}$	CTF1 $10^{-6}\,°C^{-1}$	CTF2 $10^{-9}\,°C^{-2}$
Y+21°(r)	220° / Y	75000	61.	21.
Y-21°(z)	210° / Y	50000	47.	17.
Y-30°	230° / Z	17000	10.	*
Y-32°	230° / Z	22000	5.4	+2.6
Y-38°	230° / Z	15000	-21.	*
Y+21°	QUARTZ	---	64.	26.
Y-21°	QUARTZ	---	58.	49.

* Experimental growth conditions that were optimized after the growth of the corresponding crystals.

IV - RESEARCH OF COMPENSATED CUTS

IV.1 - At cut

In figure 8 are represented the computed variations of the 1st order RFTC near the AT cut. The data used for this computation were measured (14) with crystals having large OH concentrations. With such crystals, the AT cut was experimentally found (3) (11) for a cut close to

the computed value (Y-30°). In figure 9 is represented the frequency-temperature behaviour of such a resonator. A new research of this cut was made with the crystals obtained in sulfuric acid. In figure 10 is represented the frequency-temperature relation measured for two resonators of orientation near Y-32°. The exact cristallographic orientations of these two plates were carefully measured with a double rotation goniometer that give the absolute orientation of the normal to the plate. The curves on figure 10 have a cubic behaviour, as for AT quartz, with an inflexion point near 50-55°C. This suggests that we can, as for quartz define a "Beckman number", B, for berlinite by :

$$\Delta\theta = \left(\frac{\partial f}{\partial T}\right)_{inf} = \theta - \theta o$$

where θ is the exact orientation

θo is the orientation for a curve with horizontal inflexion tangent.

For the curves on figure 10 the values are $B = -2{,}9\ 10^{-6}\ C^{-1}\ mm^{-1}$

$\theta o = 31°52'$

VI.2 - Extensional bars

Bars with the thickness normal to the x direction and length along y were made. On figure 11 the electrical response of such a resonator is given. We can notice that the Q factor is comparable to that of a quartz resonator of similar design. The measurements made on these resonators have permitted new measurements of some material constants. Their values are given table 4.

Table 4 : Value of S_{11} and d_{11}

	This study	Wang 1986 (15)	Ilraev 1986 (16)	Chang et (1) Barsch
$S_{11}10^{-12}$ MKSA	16.24	16.95	17.3	17.31
$d_{11}10^{-12}$ MKSA	- 3.21	- 3.30	-----	- 5.29

The measured coupling coefficient is k^2_{31} = 0.0134 ; this value is to be compared with an experimental value of k^2_{31}=0.00987 for a X+5° quartz resonator. The frequency temperature relation for such a resonator is given on figure 12. In this case, the inversion temperature is 2°C and the second order T.C. at this temperature is : $-27.10^{-9}\,°C^{-2}$ (this value is lower than that measured for a X+5° Quartz resonator). During this study it was observed that the sensitivity of the resonance frequency to the excitation level was much lower than what is observed for X+5° quartz. On figure 13 we can observe that the resonance curve present very little deformation even with the higher excitations. The frequency shift (one Hertz for the maximal excitation) has much probably a thermal origin.

V - EVALUATION OF BERLINITE DEVICES

This material has for orientations near AT a coupling coefficient k^2 which is nearly twice that of AT quartz (11). Berlinite is thus particularly adapted to the obtention of filters with larger bandwidth and oscillators with enhanced shifts.

V.1 - Discrete filters (AT cut)

On figure 14 is given the response curve of a 2 resonators Jaumann filter. The band pass of the adapted filter (R_a=5KΩ) is 250 kHz at 3dB. This value for a center frequency near 20 MHz situates the interest of berlinite. However, the distorsion in the band pass and in the transition band indicates that further optimizations of the resonators design are needed.

On figure 15 is displayed the transmission response of a 4 poles Jaumann filter with no broadening of the band pass by the inductances. Though the response is very satisfactory, the Q factors of the resonators were still a little insufficient to obtain exactly the theoretical response (Chebyschev with 0.1 dB ondulation (17)). The relative band pass (0.58 %) is much higher than what can be obtained with quartz in the same conditions. The phase response is also very satisfactory.

During the experiments with AT resonators we have observed that they have a reduced sensitivity to the excitation level and also that with high driving voltages they present a behaviour quite different of that of AT quartz resonators. In figure 16 is represented a typical example of the evolution of the amplitude response of a Y-32° resonator for high levels of excitation.

V.2 - Monolithic filters

Two-poles monolithic cells are now well-known for their low-cost and small-size in design of monolithic crystal filters. Coupling between the two resonators depends on geometrical characteristics (such as resonator-electrode size, resonator spacing...) and acoustical characteristics (such as plate-back, cut-off frequencies...) of the plate.

To evaluate the capabilities of rotated Y-cut Berlinite plates in the monolithic filter domain, we have plotted bandwidth versus frequency lowering for both x (figure 17) and z' (fig. 18) coupling axis. The electrode length along the coupling axis is equal to 0.8 mm while the resonator spacing is equal to 0.24 mm. These curves show that for low plate-back (# 10^{-3}) the obtained bandwidth is about twice that of a monolithic quartz filter designed in the same conditions.

On figure 19 and 20 is plotted the frequency response of a two poles monolithic filter with z' as coupling axis and loaded with 12,5 Ω or with its matching impedances (2.7 KΩ). From these curves, it appears that in spite of an already large plate-back (about 1.5 %), the in band and stopband characteristics of the filter are rather good ; the rejection of the unwanted modes being larger than 40 dB).

To improve the bandwidth of the monolithic filter a strip electrode was deposited on one surface of the crystal between the two resonator electrodes. For the monolithic cell designed here with x as coupling axis and a plate back of about 0.3 %, a bandwidth of 60 KHz is obtained without the strip ; the strip electrode increases this bandwidth up to 90 KHz (figure 21) for a center frequency of about 20 MHz. These characteristics are achieved without damaging the electrical behaviour. The frequency-response of such a filter is shown on figure 22 when it is loaded with 12.5 Ω and with its matching impedances on figure 23. The 3dB bandwidth is then of about 120 KHz. This value is far beyong from that achievable with AT cut quartz.

V.3 - Surface wave filters

Filters with a center frequency of 496,5 MHz were studied. A typical example of frequency response is given on figure 24. The corresponding apparent velocity is 2774. ± 4 m/s. We can notice on figure 24 that the transmission response (non-adapted) is exactly similar to the computed one. In this case, the response is mostly dependant on the transducers geometry (400 unappodized fingers), so that, it is quite independant on the nature of the material. However the perfection of the phase and amplitude frequency responses shows that we have obtained simultaneously good surface qualities and good materials properties. This material can thus be easily applied to obtain surface wave devices. Further work is needed to confirm the new possibilities permitted by this material and among them those concerning the obtention of filters with large bandwidth (5 %).

VI - CONCLUSION

The crystallographic studies have indicated that the new growth method using H_2SO_4 as solvent permits to obtain a very high crystalline quality provided that high quality seeds are used.

The electrical measurements have demonstrated that the obtained crystals display high Q factors together with good thermal properties. The filters made during this study have permitted to confirm the interest of berlinite to obtain much larger bandwidth than what is possible with quartz.

ACKNOWLEDGEMENTS

The authors are indebted to the research group of SICN directed by X. Buisson for many of the crystals used in this study and to MM R. Chenebault, C. Joly, A. Jeanne Michaud, Y. Bernard for experimental work. They acknowledge the assistance of M.J. Ph. Michel for the design of the surface wave filters and of Y. Toudic for expert advices concerning the optimization of the growth conditions. Special thanks are due to Mrs. M.P. Louis for the preparation of this manuscript.

Several part of this study were realized under CNET and DAII contracts.

REFERENCES

(1) Z.P. Chang, G.R. Barsch
IEEE Trans. Sonics UltraSon. SU23 p 127 (1976)

(2) A. Ballato ; G.J. Iafrate
Proc. 30th Ann. Symp. on Freq. Control p 141 (1976)

(3) J.P. Hou, B.H.T Chail
Proc. 1987 IEEE UltraSonics Symp. (1987)

(4) A. Ballato, E.R. Match, M. Mizan, B.H.T Chai
Proc. 38th Ann. Symp. on Freq. Control (1984)

(5) E. Philippot
C.R. 1er Forum Européen Temps Frequence p 205 (1987)

(6) B.H.T. Chai, J.P. Hou
C.R. 1er Forum Européen Temps Fréquence p 204 (1987)

(7) E. Philippot, A. Goiffon, J.C. Jumas, X. Buisson, A. Arnaud
Y. Toudic. A new approach to Berlinite Crystal Growth.
These proceedings (1988)

(8) Brevet Français n° 8603009

(9) A. Zarka, B. Capelle, J.C. Jumas
C.R. Acad Sc. Paris t 302 série II n° 8 523, 526 (1986)

(10) A. Zarka, Liu Lin, X. Buisson
J. Cryst. Growth vol 54 p 394 (1981)

(11) J. Detaint, E. Philippot, J.C. Jumas, J. Schwartzel, A. Zarka,
B. Capelle, J.C. Doukhan
Proc. 39th Ann. Symp. on Freq. Control p 234 (1985)

(12) B. Boulogne, P. Cordier, J.C. Doukhan
C.R. 1er Forum Européen du Temps-Fréquence p 209 1987

(13) J. Detaint, E. Philippot, B. Capelle, Y. Toudic, B. Boulogne,
J.C. Jumas, A. Zarka, J.C. Doukhan
Proc. Symp. on Freq. Control and Synthesis p 314 (1987)

(14) D.S. Bailey, J.C. Andle, D.L. Lee, W. Soluch, J.F. Vetelino,
B.H.T. Chai
Proc. 1983 Ultrasonics Symp p 335 (1983)

(15) H. Wang, B. Xu, X. Liu, J. Han, S. Shan, H. Li
J. of Cryst. Growth vol 79 p 227 (1986)

(16) A. Ilraev, V.E. Iopolitov, B.C. Imarov
Kristallographia vol 31 n° 1 p 201 (1986)

(17) J. Detaint, J. Schwartzel, C. Joly, E. Philippot
Proc. 41st Ann. Freq. Control Symp p 101 (1986)

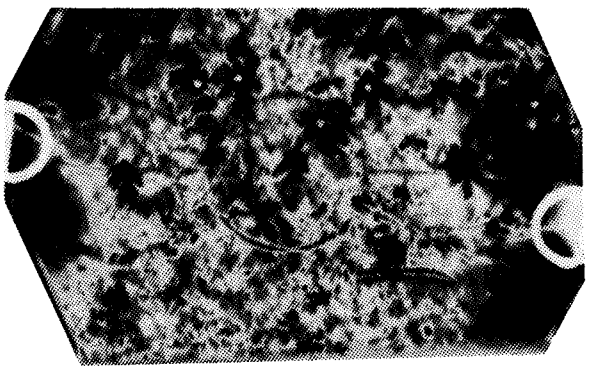


Figure 1: Topograph of a Y plate.

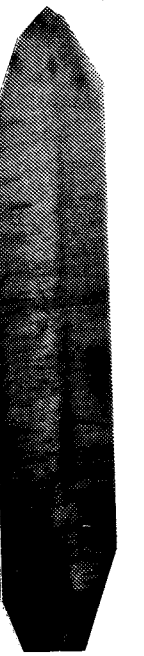


Figure 2: Topograph of a Z plate.

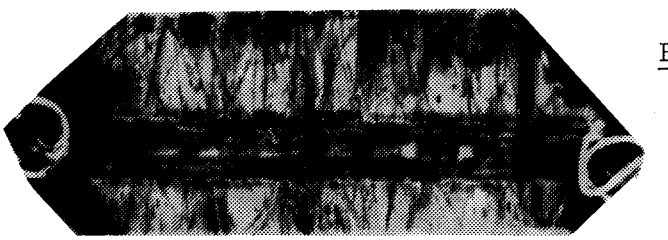


Figure 3: Topograph of a Y plate.

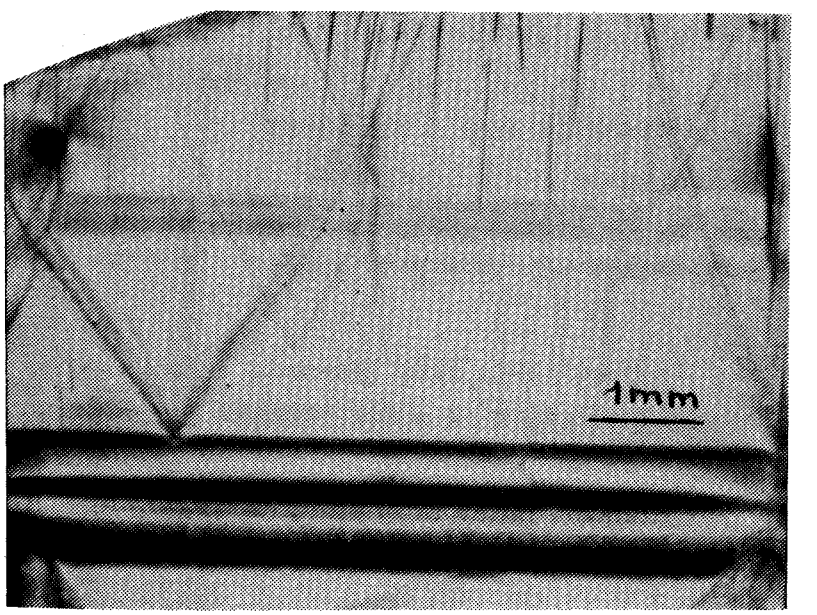


Figure 4: Topograph of a Z plate.

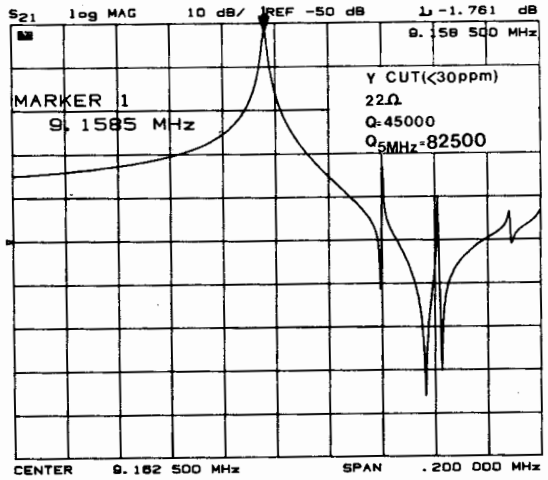


Figure 5: Electrical response of a Y cut resonator.

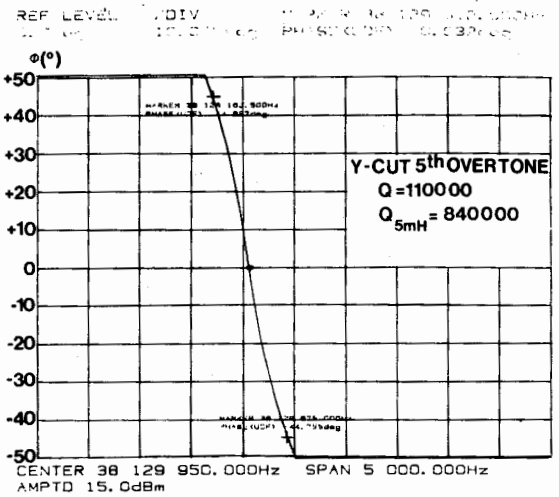


Figure 6: Phase response of the 5_5th overtone of a Y plate.

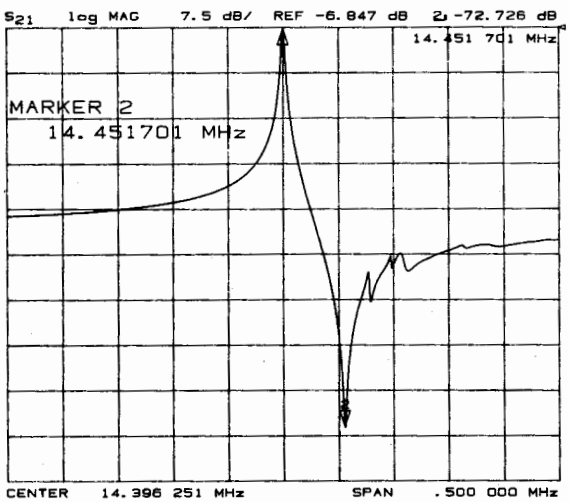


Figure 7: Electrical response of a Y-32° resonator.

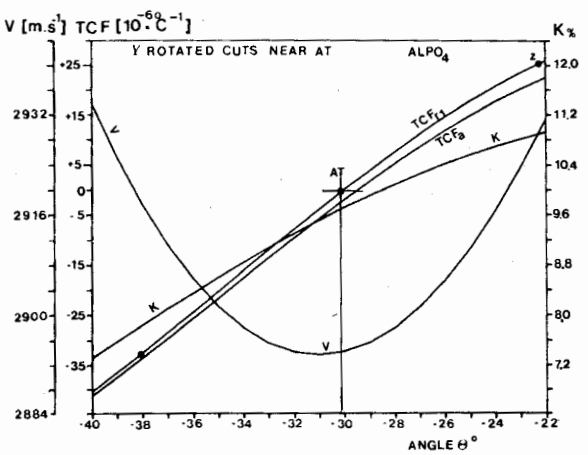


Figure 8: Computed properties of the Y rotated plates near the AT cut.

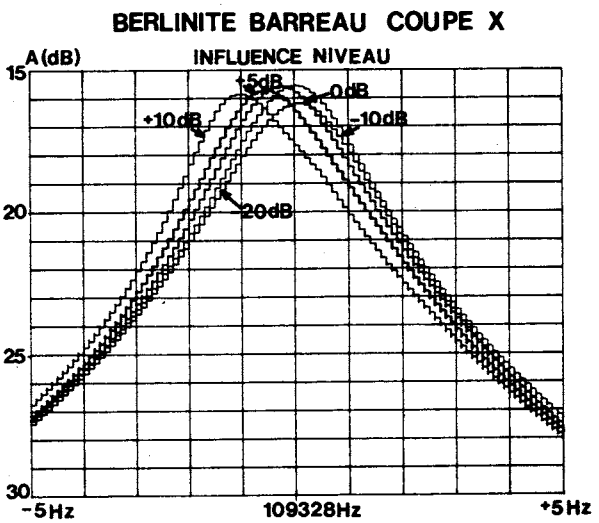


Figure 13: Effect of the excitation level on the response of an X bar.

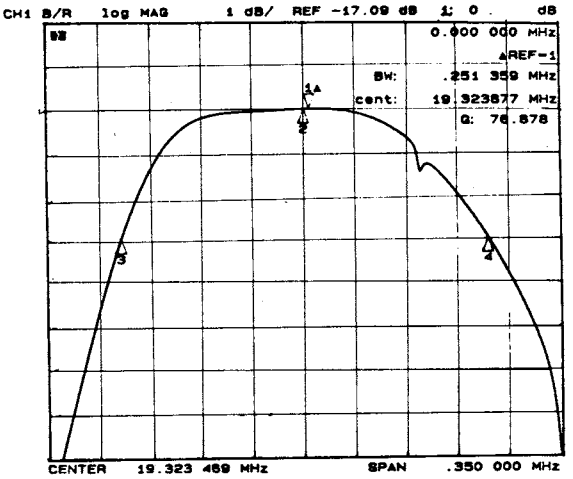


Figure 14: Two poles Jaumann Filter.

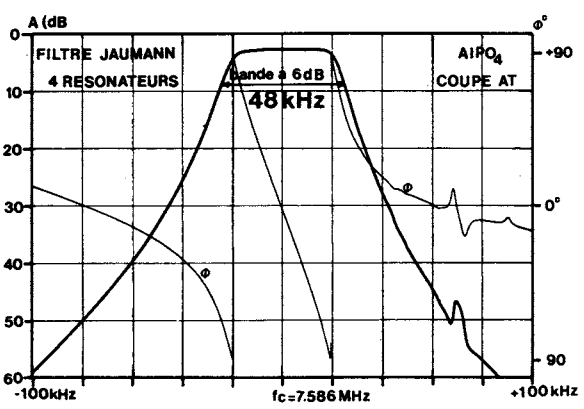


Figure 15: Four poles Jaumann filter

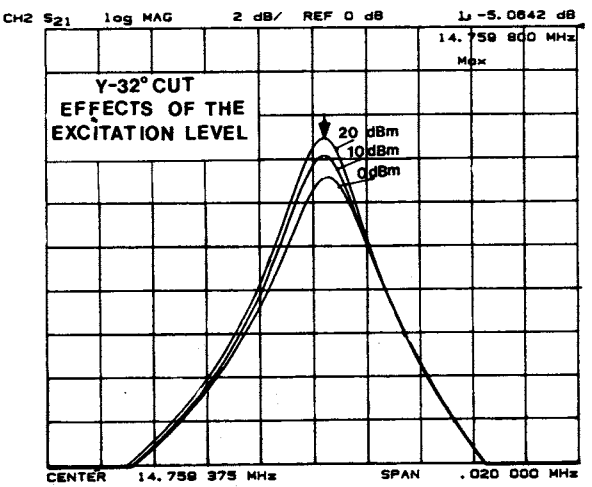


Figure 16: Influence of the excitation level on AT cut.

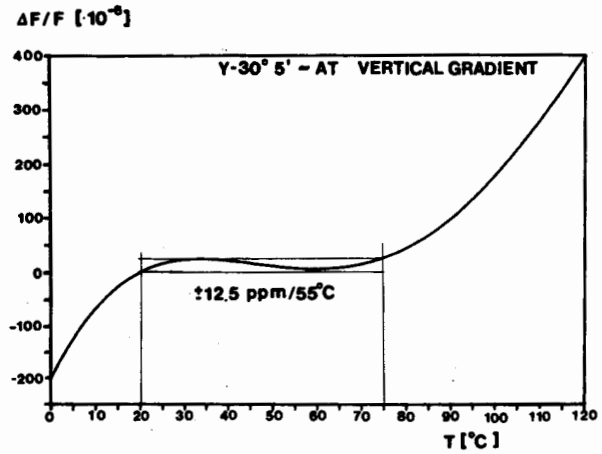


Figure 9:Frequency-Temperature behaviour of the AT cut of a crystal with a large OH concentration.

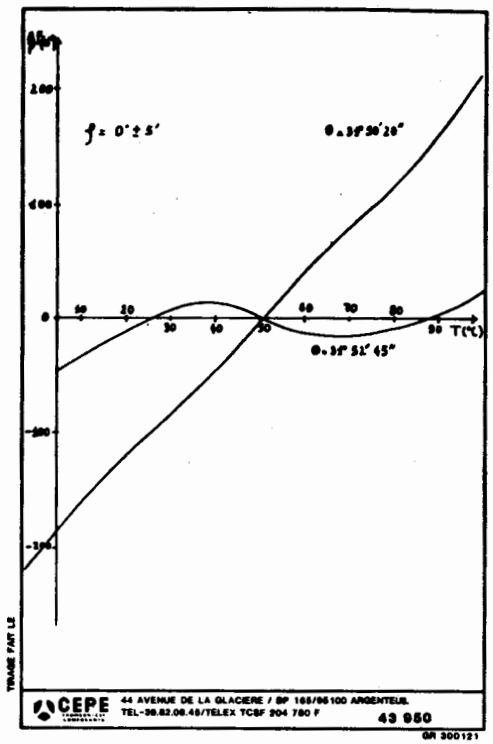


Figure 10: Frequency-Temperature behaviour for the AT cut of a crystal with a reduced OH concentration.

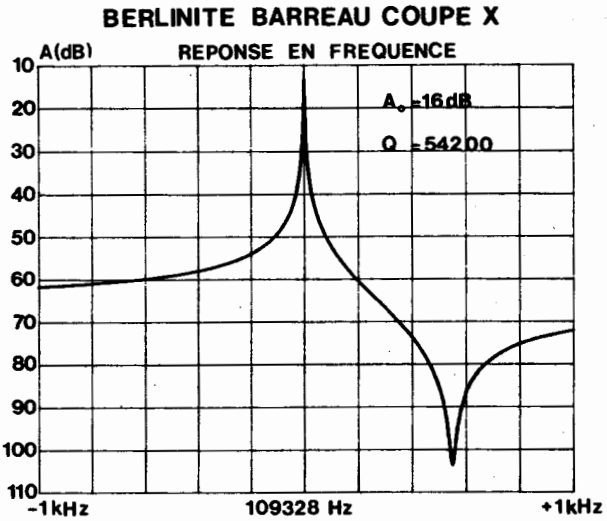


Figure 11: Electrical response of an X+0° bar resonator.

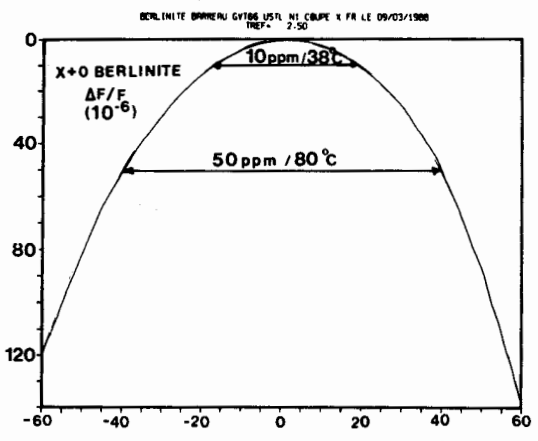


Figure 12: Frequency-temperature behaviour of a X bar resonator.

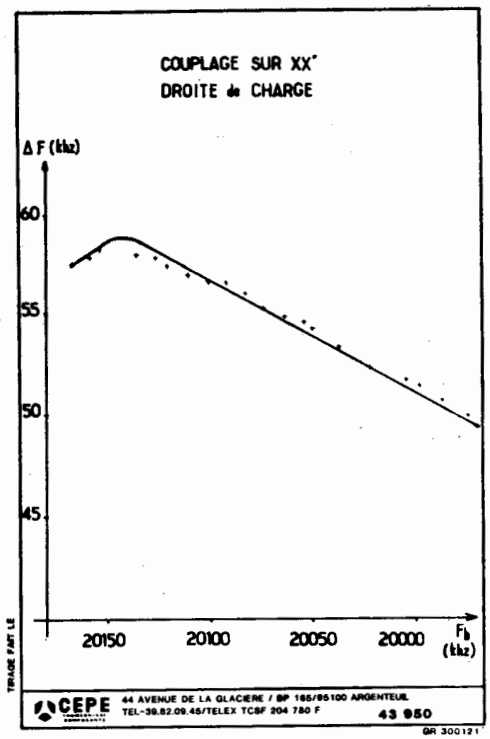


Figure 17:Influence of the mass-laoding on a x coupled monolithic cell.

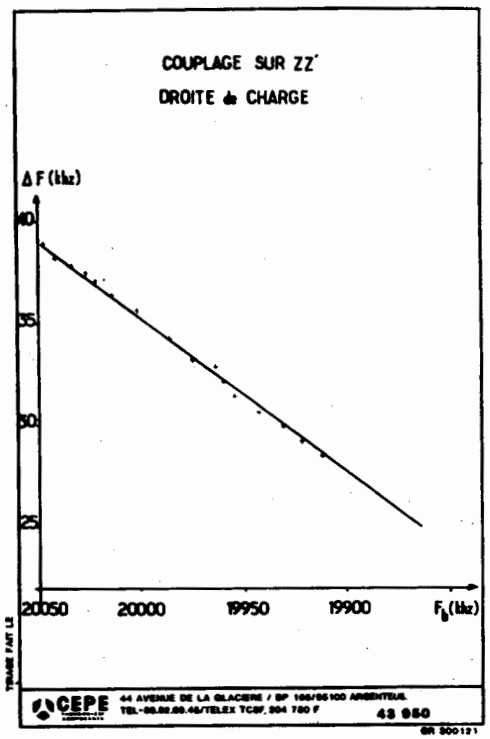


Figure 18:Influence of the mass-loading on a z coupled monolithic cell.

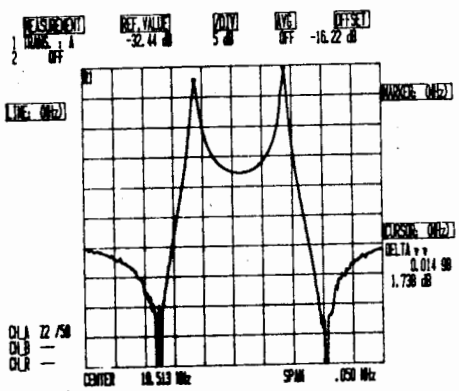


Figure 19:Transmission response of a z coupled monolithic cell (loaded by Z=12.5Ω)

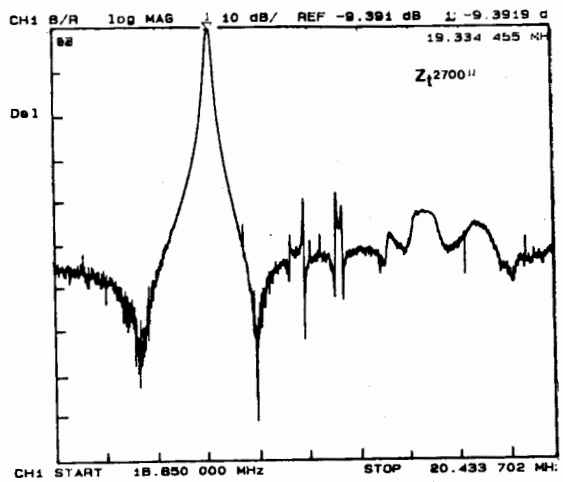


Figure 20:Transmission response of a z coupled monolithic cell adapted with Z_a=2.7 kΩ .

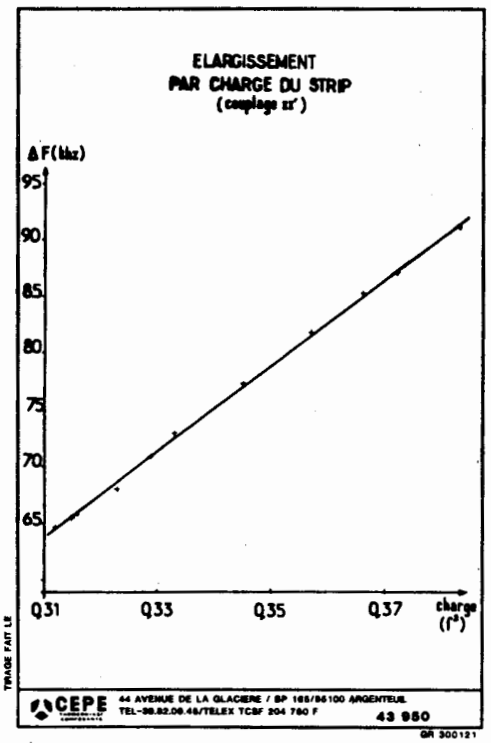


Figure 21: Influence of the mass-loading of a strip in the coupling zone of a x coupled monolithic cell.

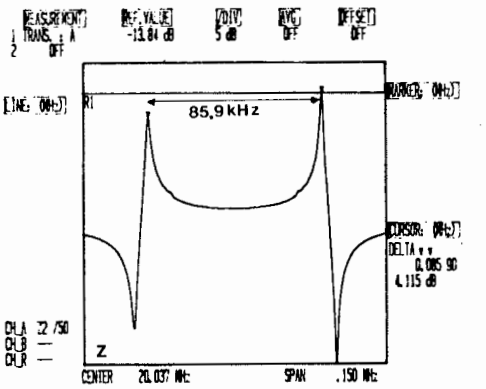


Figure 22: Response of a cell with a metallized strip in the coupling region (Z=12.5Ω).

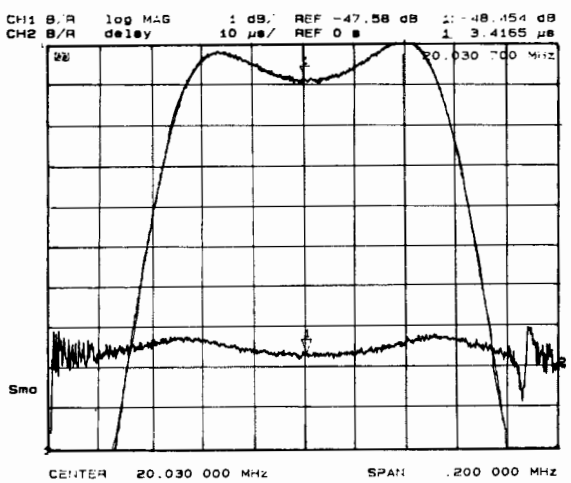


Figure 23: Adapted transmission response of the cell of Fig.22.

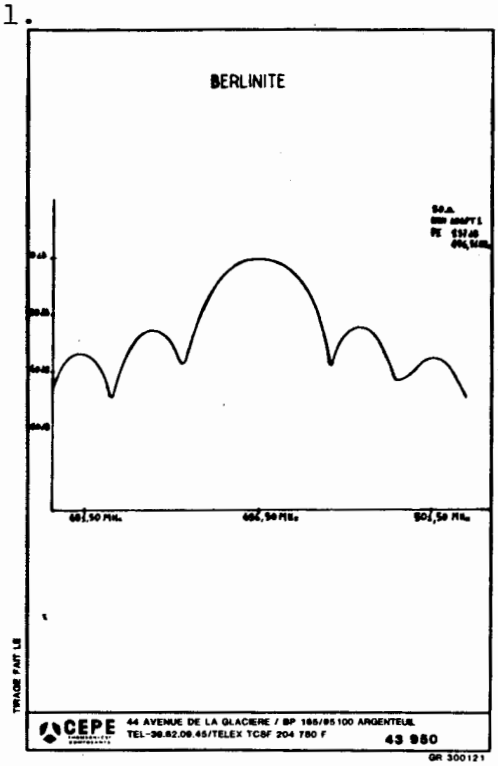


Figure 24: Transmision response of a SAW filter (Y cut, Xpropagation°).

MP-02

Physical Properties

1. A Model for the Dependence of the Q-Factor of Quartz Resonators on the Surface Texture: Application to Etched Quartz Plates
   T. Leblois, R. Bourquin, C.R. Tellier
   Ecole Nationale Supérieure de Mécanique et des Microtechniques, Besançon - France

2. Nonlinear Properties of Piezoelectric Semiconductors
   N. Daher, J.J. Gagnepain
   CNRS, Besançon - France

3. Growth Defects in Cultured Quartz, a Microscopic Investigation
   P. Cordier, J.C. Doukhan
   Université de Lille Flandres Artois, Villeneuve d'Ascq - France

4. The Orientation Dependence of the Dissolution Rate of Singly Rotated Quartz Plates Derived from Combined Numerical and Digital Techniques
   C.R. Tellier, P. Blind, D. Jozwick
   Ecole Nationale Supérieure de Mécanique et des Microtechniques, Besançon - France

2 ND EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 16 - 17 - 18 MARCH 1988

Session, MP-02 Number 1

A MODEL FOR THE DEPENDENCE OF THE Q-FACTOR OF QUARTZ RESONATORS ON THE SURFACE TEXTURE : APPLICATION TO ETCHED QUARTZ PLATES

T. LEBLOIS, R. BOURQUIN and C.R. TELLIER
Laboratoire de Chronométrie, Electronique et Piézoélectricité
ECOLE NATIONALE SUPERIEURE DE MECANIQUE ET DES MICROTECHNIQUES
Route de Gray - La Bouloie - 25030 BESANCON CEDEX - FRANCE

ABSTRACT

A new model for the calculation of the Q-factor of a quartz crystal resonator is proposed assuming that the plane wave which propagates in the bulk crystal is attenuated in two different ways. The first attenuation process is due to the viscosity of the quartz material. The second attenuation results from the reflection of the wave at the external rough surfaces. We suppose that the surface is composed by a succession of plane triangular faces whose heights follow a Gaussian law. The wave diffracted by each face is calculated in the case of an isotropic propagation medium. However for a face of given orientation the reflection coefficient takes the same value that the reflection coefficient for a plane wave reflected in an anisotropic medium by this face. The attenuation caused by the surface roughness is determined from the coherent part of the total wave diffracted by all the triangular faces.

This model allows us to follow the evolution of the Q factor either with the r.m.s. roughness parameter Rq or with geometrical parameters, a and b, describing the spatial arrangement of the triangular faces.

Experimental results on singly rotated quartz plates with respectively θ = 37°5 (AT-37 plate) and θ = - 41° (BT-41 plate) are presented. Differences in the surface texture for a given cut are obtained by repeated etchings in a $NH_4F.HF$ solution. The surface texture after each etching is characterized by two different ways : average values of the a and b parameters are evaluated from SEM micrographs and average values of the roughness parameters Rq and R_{max} are determined from profilometry traces. From these characterization of the geometrical aspect of the etched surface theoretical plots of the changes in the Q-factor of BT-41 and AT-37 plates with either the Rq or the a parameters are presented : they give evidence for opposite variations of the Q factor with increasing values of the Rq and the a parameters. Measurement of the Q-factor of etched BT-41 plates are performed in high vacuum. The observed variations of Q-factor can be easily interpreted in terms of the above Q-factor vs Rq and Q-factor vs a plots showing that the proposed model gives satisfactory description and explanation for the dependence of the Q-factor on the surface texture of etched quartz plates.

1 - INTRODUCTION

It is well known **(1,2)** that the Q factor of a resonator depends on the bulk properties of the crystal and of some fondamental parameters such as its resonance frequency, overtone, conditions of energy trapping but also on the geometrical properties of the crystal surfaces. Moreover as the resonance frequency increases (i.e. as the resonator becomes thinner and thinner), the influence of the crystal surfaces becomes predominant. The crystal surface can be prepared either by mechanical lapping and polishing or by chemical etching. Each type of preparation leads to specific geometrical properties of the surface, in particular mechanical processing causes a superficial degradation of the crystal with the formation of a disturbed surface layer.

Here, we are concerned with plane plates vibrating in thickness mode, specially in thickness shear mode. But it should be pointed out that the model proposed in this paper can be also applied to plate vibrating in compression thickness mode. To relate the Q factor to the geometrical properties of the surface, it is necessary to determine the dissipation of energy induced by the roughness of the active surfaces of the resonator (the lateral surfaces remain inactive for a resonator correctly trapped).

For this purpose, we suppose that the resonator can be identified with a plane plate of infinite dimensions and we consider a plane wave propagating along the thickness direction of the resonator. As a result of the surface roughness, this wave is not totally reflected by the resonator surface, but in reality is partly diffracted in such a way that the reflected wave can be decomposed in a coherent plane wave and in a incoherent diffracted wave **(3)**. This diffracted wave which is totally dissipated in the resonator cannot be involved for the resonance.

The diffracted wave can be calculated in the framework of various methods in the case where as in this study the maximum peak to valley height takes small values compared with the value of wavelength **(4,5)**. In the other hand it appears to be necessary to characterize the geometrical surface not only by the root-mean-square parameter Rq but also by some other parameters such as the maximum peak-to-valley height, R_{Max}, and parameters which describe the spatial arrangement of the surface.

When the surface imperfections are caused by a chemical etching process, the etched surfaces are found to be covered by pits or hillock of typical shape **(6)**. Since these dissolution figures are generally composed of a lot of intersecting planes which can be assimilated to triangular faces , we assume in this paper that at a given etching time, an etched surface can be described by a succession of triangular faces whose dimensions in the ground plane can be considered as statistically constant and whose distribution heights supposed to exhibit a Gaussian form.

In the calculation, the parameter R_{Max} has an important role to play ; hence, here we extend a method previously used **(7)** to evaluate the influence of the surface roughness on the Q factor of the AT-cut resonator.

2 - THE Q FACTOR OF A RESONATOR

The triangular face corresponding to a dissolution figure is shown in Figure 1. The magnitudes of geometrical parameters a, b, y_o, h and h' are small in comparison with the wavelength of the incident plane wave arriving on a face. Moreover we consider that the triangular faces make a small angle with the reference surface of the resonator that is to say $h < a$ and $h' < b$. The plane incoming wave, $\vec{u}_I$, of wave vector, $\vec{k}_I$, propagates in the direction $-\vec{x}_2$. The coherent part, u_{coh}, of the total wave diffracted by an infinity of triangular faces is given by **(3)**.

$$u_{coh} = \iiint_{-R_{max}/\ell}^{R_{max}/\ell} u(P)\, R(\ell_1, \ell_3)\, P(z)\, P(z+h)\, P(z+h+h')\, dz\, dh\, dh' \quad (1)$$

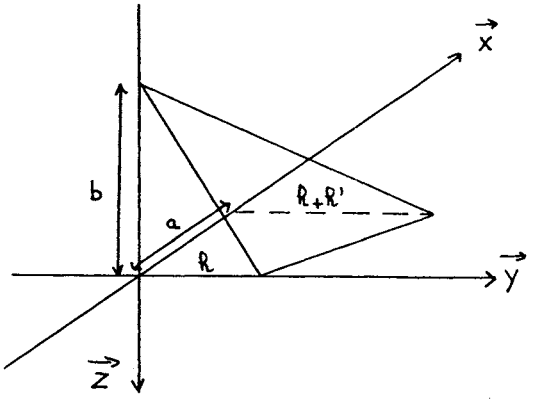


Figure 1 : The geometry of a triangular face.

In this equation the magnitude of the diffracted wave is at a point P lying at an infinite distance of the face given by $u(P).R(\ell_1, \ell_3)$ is the reflexion coefficient of the incoming plane wave for a reflecting flat plane of same orientation that the triangular face unit with outward normal $\vec{\ell}\ (\ell_1, \ell_2, \ell_3)$. The function P(z) corresponds to the probability to have a point positionned at a corner of the triangular face lying at the height z with respect to the reference surface. This probability function is described by a symetric gaussian form in such way that z remains in the range $(-R_{max}, R_{max})$.

The term u(P) is expressed by following the mathematical treatment which is usually performed in the special case of isotropic medium. Although this procedure is not rigorous but can be reasonably applied to materials which, as the quartz crystal, are slightly anisotropic. Thus in the framework of the Fraunhoffer approximation we get

$$u(P) = \frac{Aab}{4\pi s'h}\left(1 - \frac{h'^2}{2a^2} - \frac{h^2}{2b^2}\right)\left(\frac{\sin kh'}{kh'} - e^{ikh}\,\frac{\sin k(h+h')}{k(h+h')}\right)$$

$$\times \exp ik\left(s' - \frac{2hh'}{a^2} - \frac{2h^3}{b^2} - h'\right) \qquad (2)$$

where A is the amplitude of the plane incoming wave with wave number k. Note that in this equation only the terms up to the third order are retained.

We take into account the anisotropy of the quartz crystal to calculate the reflexion coefficient $R(\ell_1, \ell_3)$. As usually this reflexion coefficient is derived assuming that the boundary condition of zero stress at free the reference surface is satisfied **(8)**

$$\sum_{X=L,T_1,T_2} C_{ijk\ell}\,\ell_j\,k^x_\ell\,u^x_k + c_{ijk2}\,\ell_j\,k^I_2\,u^I_k = 0 \qquad (3)$$

Here we use a system of orthogonal axes (x_1, x_2, x_3) which corresponds with the rotated system related to the crystal cut (singly or doubly rotated crystal). In this system of axes, the incoming wave propagates along the x_2 axis : $k^I_2 = k^I$. The mechanical displacement associated with the incoming wave is $\vec{u}^I$. The wave vector $\vec{k}^x = (\omega/v_x)\vec{n}^x$ and the mechanical displacement field $\vec{u}^x$ related to the three scattered waves are calculated starting from the Christoffel equation.

$$C_{ijk\ell}\,n^x_j\,n^x_\ell\,u^x_k - \rho v^2_x\,u^x_i = 0 \qquad (4)$$

where n^x_j, u^x_i and v_x are functions of the components of the normal $\vec{\ell}$ at the scattering surface. If we suppose that the angle that the normal $\vec{\ell}$ makes with the displacement $\vec{u}^I$ is very close to $\pi/2$ we can expand these functions as ascending powers of ℓ_1 and ℓ_3. Now with the usual laws for the scattering of acoustical waves we get

$$k^x n^x_1 = -\mu\ell_1 \;;\; k^x n^x_2 - k^I = -\mu\ell_2 \;;\; k^x n^x_3 = -\mu\ell_3 \qquad (5)$$

where $\vec{n}$ (n_1, n_2, n_3) denotes a unit vector in the direction of the scattered waves and μ is a proportionality factor. Taking $\vec{n}$ as a normalized vector gives

$$\mu = k^I\left(1 + \frac{k^x}{k^I} - \frac{\ell_1^2 + \ell_3^2}{2}\left(1 + \frac{k^I}{k^x}\right)\right) \qquad (6)$$

If the directions of the vectors $\vec{\ell}$ and $\vec{n}$ are close together, we can write

$$v_x = v_x^o \left(1 + \frac{\delta^x}{\rho (v_x^o)^2}\right)^{\frac{1}{2}} \tag{7}$$

where v_x^o is the velocity of propagation along the x_2 axis.

For a direction of propagation which lies in the vicinity of the x_2 axis (i.e. n_1 and n_3 take small values but differ from zero) the deviation δ^x for the eigenvalue of equation (4) reduces to

$$\delta^x = \Delta_1 n_1 + \Delta_2 n_3 + \Delta_3 n_1^2 + \Delta_4 n_3^2 + \Delta_5 n_1 n_3 + \ldots \tag{8}$$

Combining equation (8) with equation (5-7) gives :

$$n_1^x = \left(1 + \frac{v_I}{v_x^o}\right)\left(-\frac{v_x^o}{v_I}\ell_1 + \frac{\Delta_1}{2\rho v_I^2}\ell_1^2 + \frac{\Delta_2}{2\rho v_I^2}\ell_1\ell_3\right)$$

$$n_2^x = -1 + \frac{\ell_1^2 + \ell_3^2}{2}\left(1 + \frac{v_x^o}{v_I}\right)^2$$

$$n_3^x = \left(1 + \frac{v_I}{v_x^o}\right)\left(-\frac{v_x^o}{v_I}\ell_3 + \frac{\Delta_2}{2\rho v_I^2}\ell_3^2 + \frac{\Delta_1}{2\rho v_I^2}\ell_1\ell_3\right)$$

and

$$k^x = \frac{v_I}{v_x^o} k^I \left\{1 + \left(1 + \frac{v_x^o}{v_I}\right)\frac{\Delta_1}{2\rho (v_x^o)^2}\ell_1 + \left(1 + \frac{v_x^o}{v_I}\right)\frac{\Delta_2}{2\rho (v_x^o)^2}\ell_3 \right. \tag{9}$$

$$+ \left[\frac{\Delta_1^2}{8(\rho v_x^o v_I)^2}\left(1 + 4\frac{v_I}{v_x^o} + 3\frac{v_I^2}{(v_x^o)^2}\right) - \frac{\Delta_3}{2\rho v_I^2}\left(1 + \frac{v_I}{v_x^o}\right)^2\right]\ell_1^2$$

$$+ \left[\frac{\Delta_2^2}{8(\rho v_x^o v_I)^2}\left(1 + 4\frac{v_I}{v_x^o} + 3\frac{v_I^2}{(v_x^o)^2}\right) - \frac{\Delta_4}{2\rho v_I^2}\left(1 + \frac{v_I}{v_x^o}\right)^2\right]\ell_3^2$$

$$\left. + \left[\frac{\Delta_1 \Delta_2}{(2\rho v_x^o v_I)^2}\left(1 + 4\frac{v_I}{v_x^o} + 3\frac{v_I^2}{(v_x^o)^2}\right) - \frac{\Delta_5}{2\rho v_I^2}\left(1 + \frac{v_I}{v_x^o}\right)^2\right]\ell_1\ell_3 \right\}$$

Expressions of the eigenvectors in terms of ℓ_1 and ℓ_3 can be also extracted from equation (4). For example, for the longitudinal mode, equation (4) yields

$$u_i^L = \{\Gamma_{1i}^L - \Gamma_{2i}^L \left(1 + \frac{v_L^o}{v_I}\right) \ell_1 - \Gamma_{3i}^L \left(1 + \frac{v_L^o}{v_I}\right) \ell_3 + \left[\Gamma_{1i}^L \frac{\Delta_1}{2\rho v_I^2} \left(1 + \frac{v_I}{v_L^o}\right) + \Gamma_{4i}^L \left(1 + \frac{v_L^o}{v_I}\right)^2\right] \ell_1^2 + \left[\Gamma_{3i}^L \frac{\Delta_2}{2\rho v_I^2} \left(1 + \frac{v_I}{v_L^o}\right) + \Gamma_{5i}^L \left(1 + \frac{v_L^o}{v_I}\right)^2\right] \ell_3^2$$

$$+ \left[\Gamma_{2i}^L \frac{\Delta_2}{2\rho v_I^2} \left(1 + \frac{v_I}{v_L^o}\right) + \Gamma_{3i}^L \frac{\Delta_1}{2\rho v_I^2} \left(1 + \frac{v_I}{v_L^o}\right) + \Gamma_{6i}^L \left(1 + \frac{v_L^o}{v_I}\right)^2\right] \ell_1 \ell_3 \} u_2^L \quad ; \; i = 1 \text{ ou } 3 \tag{10}$$

where we have chosen to express the u_1^L and u_3^L components in terms of the major solution u_2^L. Similar relations stand for the transverse modes provided that the u_1^{T1} and u_2^{T1} components, and the u_2^{T2} and u_3^{T2} components are respectively written in terms of the major components u_3^{T1} and u_1^{T2}.

Introducing equations (9) and (10) into equation (3) we get the following set of linear equations which yields the major components of the three scattered waves.

$$\begin{bmatrix} A_1 \\ A_2 \\ A_3 \end{bmatrix} = \begin{bmatrix} a_{ij} \end{bmatrix} \begin{bmatrix} u_2^A \\ u_3^B \\ u_1^C \end{bmatrix} \tag{11}$$

in which the components of the vector A and the elements of the matrix a_{ij} are second order polynomial functions of the terms ℓ_1 and ℓ_3. Moreover the vector A depends also on the components of the displacement field associated with the incoming wave. Solving equation (11) gives the reflection coefficient $R_x(\ell_1, \ell_3) = |u^x| / |u^I|$ for any mode of the incoming wave.

Omitting terms of order greater than two leads to

$$R_x = 1 + \alpha_x \ell_1 + \beta_x \ell_3 + \gamma_x \ell_1^2 + \delta_x \ell_3^2 + \chi_x \ell_1 \ell_3 \tag{12}$$

Introducing equation (2) and (12) into equation (1) give the real part of u(P) as

$$\text{Real } \{u(p)\} = \frac{Aab\, k\, \text{Erf}^3(\frac{R_{Max}}{\sigma\sqrt{2}})}{4\pi s'} \left(1 + \frac{\sigma^2}{b^2}(2\gamma - 1) + \frac{\sigma^2}{a^2}(2\delta - 1) - \frac{\sigma^2}{ab}\chi - 7k^2\sigma^2\right)\left(1 - \frac{2R_{Max}}{\sqrt{2\pi}\sigma\, \text{Erf}(\frac{R_{Max}}{\sqrt{2}\sigma})} \exp\left(-\frac{R_{Max}^2}{2\sigma^2}\right)\right) \tag{13}$$

and the imaginary part as

$$\mathrm{Im}\{u(p)\} = \frac{Aab\ k\ \mathrm{Erf}^3(R_{Max}/\sqrt{2}\sigma)}{4\pi\ s'}\left(\frac{2\alpha\sigma^2}{b} - \frac{\beta\sigma^2}{a}\right) \times 2k$$

$$\times \left(1 - \frac{2R_{Max}}{\sqrt{2\pi}\sigma\ \mathrm{Erf}(R_{Max}/\sqrt{2}\sigma)} - \exp\left(-\frac{R_{Max}^2}{2\sigma^2}\right)\right) \qquad (14)$$

in which Erf denotes the Error function and σ the root-mean square rugosity.

As the coherent part of the total wave scattered by all the triangular faces is defined by

$$|u(p)| = \sqrt{(\mathrm{Im}\{u(p)\})^2 + (\mathrm{Real}\{u(p)\})^2} \qquad (15)$$

the attenuation, $\mathcal{A}$, caused by the texture of the surface emerges from equations (13) and (14)

$$\mathcal{A} = \frac{|u(p)|}{|u(p)_{\sigma=0}|} = \left[4k^2\left[\left(\frac{2\alpha\sigma^2}{b} - \frac{\beta\sigma^2}{a}\right)\left(1 - \frac{2R_{Max}}{\sqrt{2\pi}\sigma\ \mathrm{Erf}\left(\frac{R_{Max}}{\sqrt{2}\sigma}\right)}\right.\right.\right.$$

$$\left.\left.\left. \times \exp\left(-\frac{R_{Max}^2}{2\sigma^2}\right)\right)\right]^2 + \left[1 + \left(\frac{\sigma^2}{b^2}(2\gamma_x - 1) + \frac{\sigma^2}{a^2}(2\delta_x - 1) - \frac{\sigma^2}{ab}\chi\right) \times \right.\right.$$

$$\left.\left. - 7\,k^2\,\sigma^2\right)\left(1 - \frac{2R_{Max}}{\sqrt{2\pi}\sigma\ \mathrm{Erf}\left(\frac{R_{Max}}{\sqrt{2}\sigma}\right)} \exp\left(-\frac{R_{Max}^2}{2\sigma^2}\right)\right)\right]^2\Bigg]^{1/2} \qquad (16)$$

3 - EXPERIMENTAL RESULTS

In this section, we are concerned with two successive types of results. Firstly, we can determine experimentally the average values of the geometrical parameters Rq, R_{Max}, a and b at different stages of etching of a given cut and then draw the theoretical plot which illustrates the variations in the Q factor induced by changes in these geometrical parameters. Secondly, to establish the validity of our model, experimental measurements of the Q factor of some etched singly rotated quartz resonator plates are performed in high vacuum ($\simeq 10^{-6}$ Torr) in order to undertake a complete comparison of the theoretical predictions with the experimental results.

3.1/ The theoretical plot for an etched BT-41 plate

A great number of BT-41, 5 MHz, third overtone quartz resonator plates* were initially lapped with a 5 µm abrasive. These plates were etched at a constant temperature for at least ten successive periods of time. The etchant used was a concentrated $NH_4F.HF$ solution. After each period of time a depth of etch, Δds, of about 1 µm was reached. After the various stages of etching the surface topography data such as R_{Max} and Rq were measured using a microcomputer based profilometer : in particular, the average values of the R_{Max} and Rq parameters were evaluated from several profilometry traces made along various directions on the surface. Moreover, since the description of the surface texture was achieved using a microcomputer and a program **(10)** which offers the possibility to extract the distribution of heights , the assumption of a heights distribution of Gaussian form was verified after each period of etching. Then, the etched BT-41 plates were examined by scanning electron microscopy (S.E.M.) in order to determine average values of the parameters a and b which reflect the mean dimensions of the dissolution figures (i.e. the triangular faces) in the respective X and Z' directions.

In figures 2, 3 and 4 are displayed experimental results for a given BT-41 cut surface etched down to 3 µm. Figures 5, 6 and 7 provide informations on the surface texture of the same BT-41 cut for a depth of etch of about 7 µm. It should be pointed out that all the various BT-41 plates behave in the same way so that here we restrict the discussion of data to a single BT-41 plate.

These figures merit some comments :

(i) The distributions of heights (Figures 2 and 5) at any stage of etching can be effectively qualified of Gaussian even if in Figure 5 the heights distribution appears as relatively peaky.

(ii) Magnifications of the Z' roughness profiles (figures 3 and 6) show that value of the maximum peak-to-valley height, R_{Max}, increases with the depth of etch : The observed changes in R_{Max} (from 0,8 µm to 1,2 µm for the BT-41 plate etched respectively to about 3 µm and to 7µm) agree well with the average values of the R_{Max} and Rq parameters as evaluated from several traces. Moreover figures 3 and 6 reveal that successive etching causes an enlargment of the spatial arrangment along the Z' direction.

* Here, for convenience a singly rotated plate with $\theta = -41°$ is denoted BT-41.

1 In the first stage of etching (i.e. for a depth of etch Δd smaller than 2 μm corresponding to a $\Delta f < 10$ kHz for fifth-overtone, 5 MHz resonator) we observe an anhencement of the Q-factor which can be attributed to the progressive etching of the surface disturbed layer which is created by the mechanical lapping and whose depth can be estimated to be close to 2 μm **(11)**.

2 In the intermediate stage of etching ($2\mu m < \Delta d < 4 \mu m$) stable dissolution figures develop on the BT-41 surface as evidenced by the SEM micrographs in figures 4 and 7. However measurements of the average value of geometrical parameters (a, b, R_q, R_{max}) have revealed simultaneous increases in these four parameters. Thus as shown in the previous section we are now concerned with two opposite variations in the Q-factor.

3 In the last stage of etching ($\Delta d > 5 \mu m$) the increase in the peak-to-valley height becomes less pronounced and in the same time typical etch pit, with large dimentions, probably connected with the presence of dislocations emerging at the crystal surface **(12)**, cover more and more the surface. Thus the consecutive decrease in the Q-factor can be easily understood in terms of an increasing number of etch pit whose average roughness depth is large as compared with that of the usual dissolution figure.

Owing to the preceding features it appears that to establish without ambiguity the validity of our model, it is more convenient to compare quantitatively the theoretical values of Q-factor as evaluated from experimental values of the geometrical parameters with the experimental ones (Table 1).

	Theoretical Q	Mesured Q	$\frac{\Delta Q}{Q}$
Initial	3 453 000	3 181 450	7,9 %
2nd etching	4 178 000	4 195 015	0,5 %
4th etching	4 400 400	4 878 200	10,9 %
7th etching	4 795 600	5 089 850	6 %
9th etching	3 028 500	3 596 470	18,8 %

Table 1 : Comparison of experimental and theoretical values of the Q-factor at various stages of etching.

(iii) The S.E.M. micrographs (figures 4 and 7) show that elongated dissolution figures all uniformly oriented in the Z' direction develop on the BT-41 surface and that the enlargment of the dissolution figure with prolonged etching is more marked along the Z' direction than along the crystallographic X axis.

Repeating this experimental procedure at any stage of etching allows us to plot the theoretical variations of the Q factor with the depth of etch. But since all the geometrical parameters a, b, Rq vary in different ways with the chemical attack it seems more appropriate, here, to present and to discuss the two following theoretical plots :

1 We assume that R_{Max} takes a constant value of 1 µm and Rq a constant value of 0,25µ m and that changes on a and b parameters with the depth of etch correspond to those evaluated from the successive S.E.M. micrographs. The resulting theoretical Q-factor vs the geometrical parameter a plot indicates that as expected the Q factor increases with the enlargment of the spatial arrangment of the dissolution figures (see figure 8).

2 Now we are concerned with dissolution figures with constant sizes a = 5 µ m, b = 50 µ m and we suppose that the etching produces only increases of the R_{Max} and Rq parameters with the depth of etch. Figure 9 shows the evolution of the Q factor with the Rq parameter : as commonly observed **(7)** the Q factor falls rapidly to a minimum with increasing values of the r.m.s. roughness. In view of these two theoretical plots we observe that simultaneous increases in the one hand a and b parameters and in the other hand the roughness parameters, Rq and R_{Max}, induce opposite variations in the Q factor.

3.2/ The Q-factor of the etched BT-41 resonator

Measurements of the Q-factor of the various BT-41 plates were performed at room temperature. The electrodeless resonators were mounted between two brass electrodes in such a way that an airgap of about 5-10µ m was formed. Since it remains relatively difficult to obtain an excellent reproductibility on measurements of the Q-factor, data related to at least ten resonators were collected. The curves displayed in figure 10 show the changes in the Q-factor with the change Δf in resonance frequency owing to the chemical attack. The behaviour of the various BT-41 plates does not depart very much from the general behaviour whose main features can be summarized as follows :

We observe that the theoretical values are not far from the experimental values. Typically the deviation is less than 10 %. Hence our model gives satisfactory explanation of the influence of the texture of the surface on the quality factor.

3.3/ Theoretical plot in the case of AT-37 plate

As chemical etching of AT-37 plate (θ = 37°5) gives rise to dissolution figures (figure 11) whose shapes differs to the shape of the dissolution figures formed on BT-41 plates and since correlated variations of the geometrical parameters a and b are observed for the etched AT-37 plate, it may be of interest to give the theoretical plots which illustrate the influence of the geometrical parameters on the Q-factor.

Hence in this section we repeat the procedure descraibed in section 3.1. Confining our attention to the theoretical plots of figures 12 and 13, we remark that the main conclusions presented in the above remain valid.

CONCLUSION

Measurements of the Q-factor of quartz resonators at different stages of chemical etching have been done and experimental values compared to computed data. Variations of the Q-factor of a quartz resonator cannot be simply related to the root mean square Rq of the roughness of its surface. It is necessary to describe more precisely the surface texture by taking account of the size of the dissolution figures. It is also necessary to introduce the anisotropy of the propagating media.

* * * * * *

ACKNOWLEDGEMENTS

The authors gratefully acknowledge financial support from the Direction des Recherches Etudes et Techniques which depends on the French Department of Defense.

* * * * * *

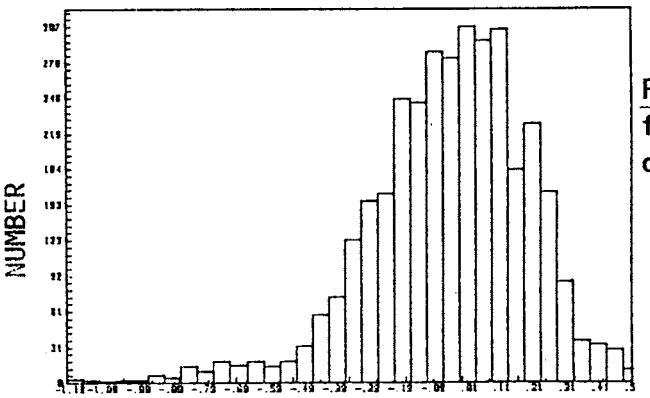
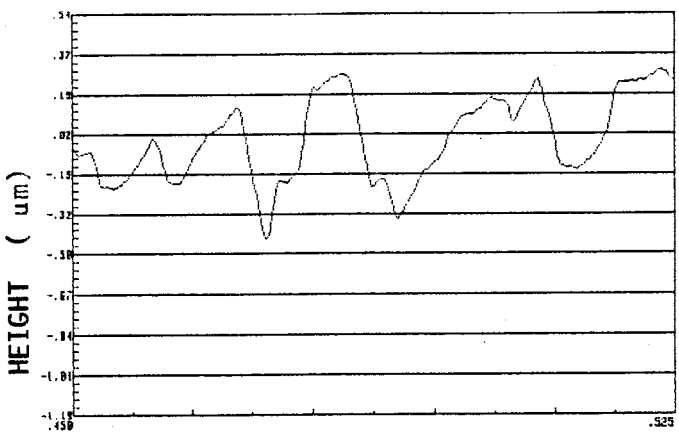


Figure 2 : Histogram of heights for a BT-41 plate with a depth of etch, Δ d, of 3 μ m.

Figure 3 : The Z' surface profile of a BT-41 plate with Δ d of 3 μ m. Here : R_{Max} = 0,2 μ m and Rq = 0,8 μ m.



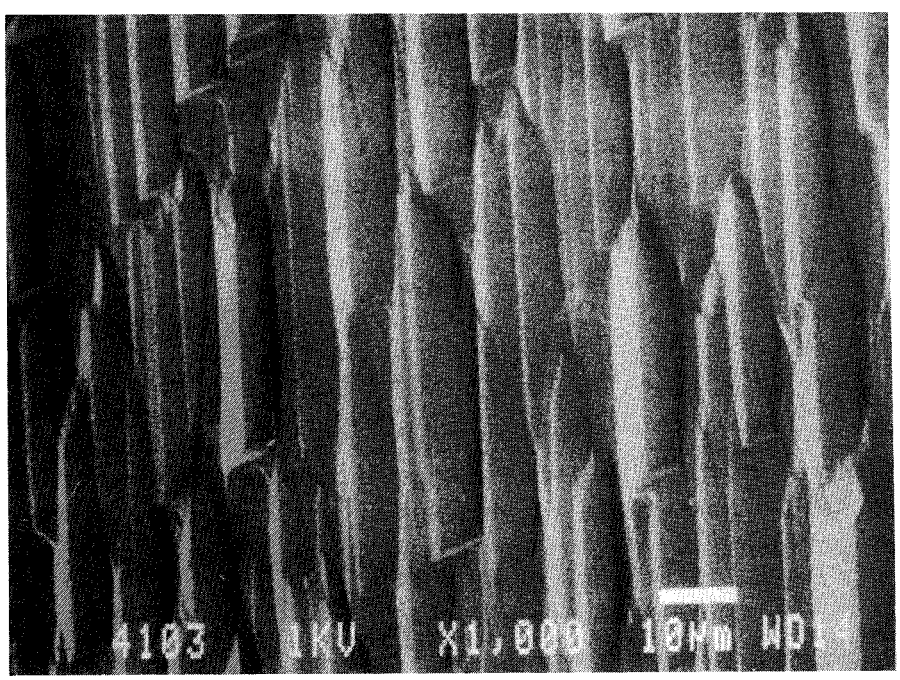


Figure 4 : SEM micrograph of a BT-41 plate with Δ d of 3 μ m. The enlargment of the dissolution figure is more marked in the Z' direction than in the X direction. The average value of a and b are evaluated a = 8 μ m, b = 15 μ m.

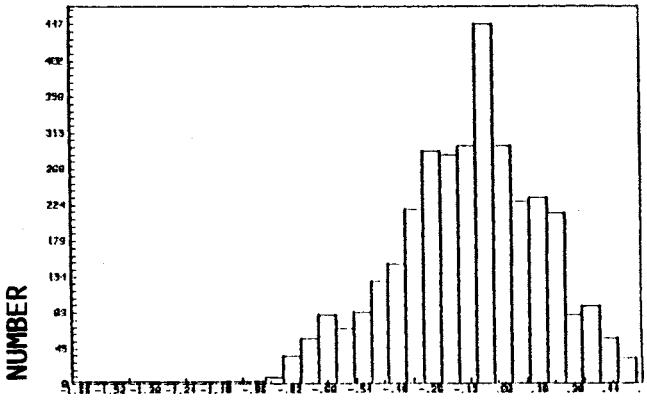
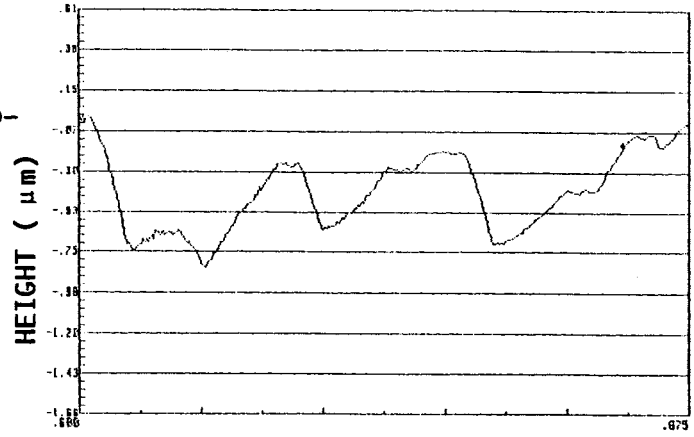


Figure 5 : Histogram of heights for a BT-41 plate with a depth of etch, Δ d, of 7 μm.

Figure 6 : The Z' surface profile of a BT-41 plate with Δd of 7 μm. Here : R_{Max} = 1,2 μm and Rq = 0,30 μm.



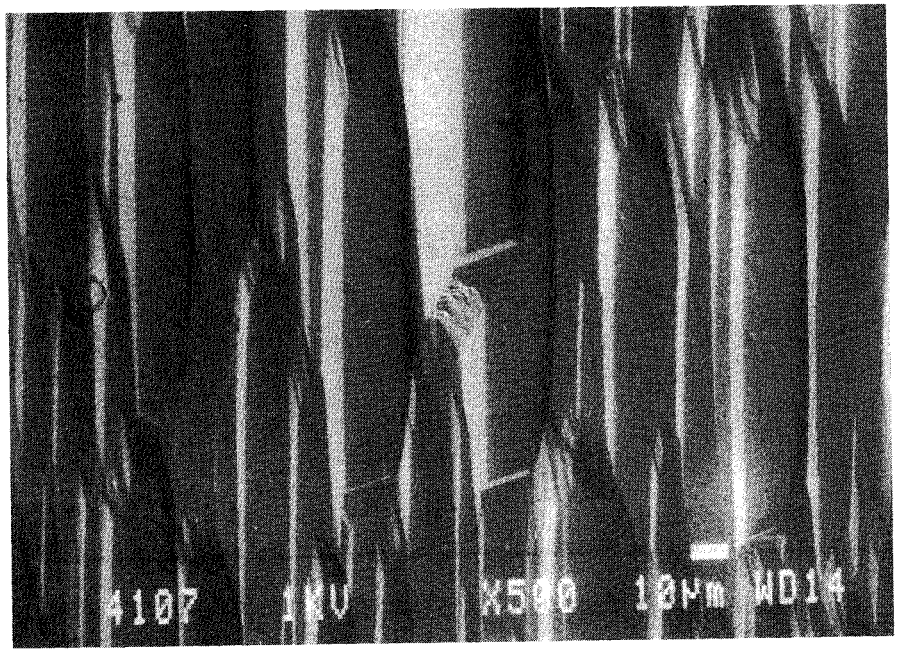


Figure 7 : SEM Micrograph of a BT-41 plate with Δ d of 7 μm. The enlargment of the dissolution figure is more marked in the Z' direction than in the X direction. The average value of a and b are evaluated : a = 15 μm, b = 190 μm.

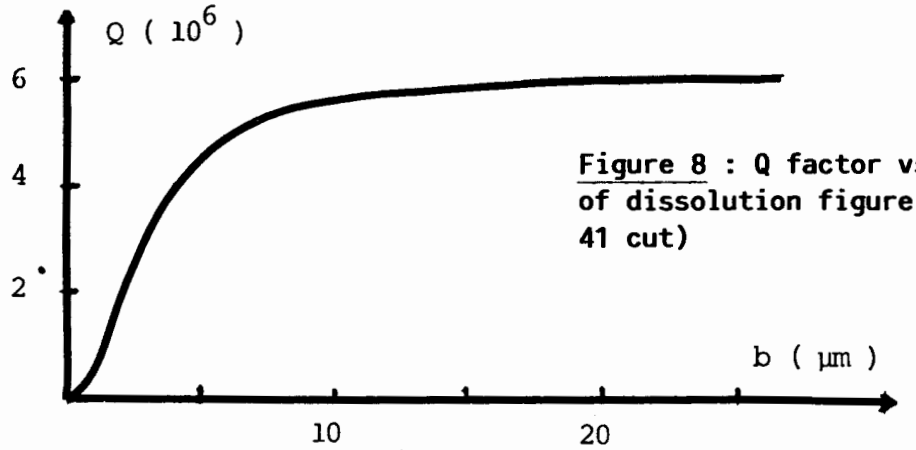


Figure 8 : Q factor vs size of dissolution figure (BT-41 cut)

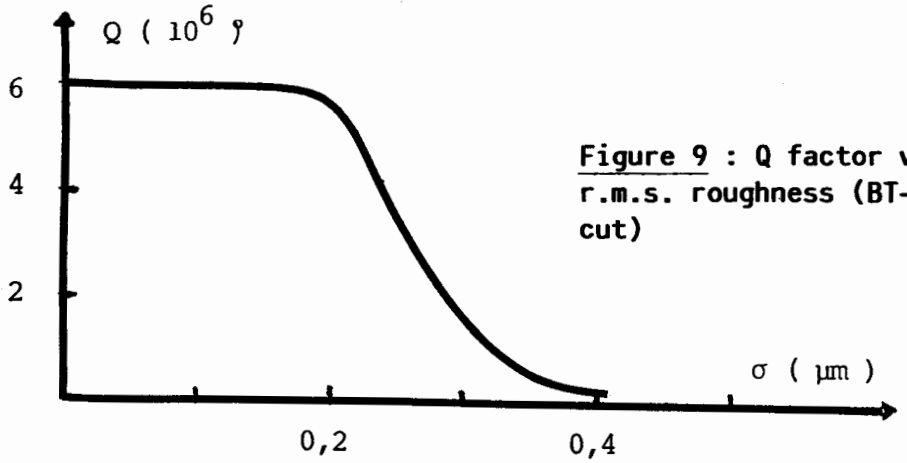


Figure 9 : Q factor vs r.m.s. roughness (BT-41 cut)

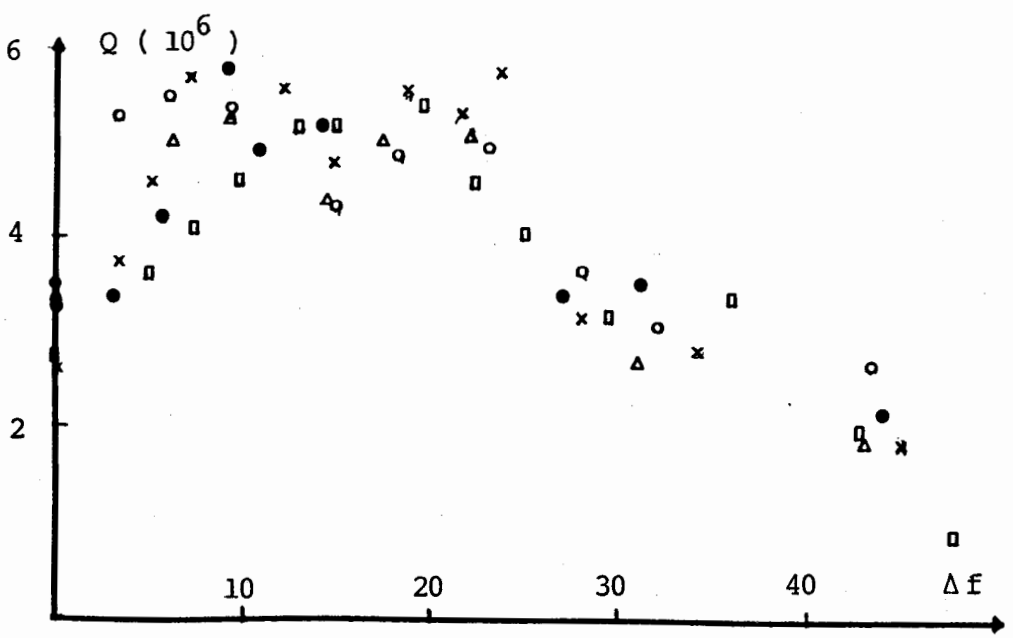


Figure 10 : Measured Q factors in function of the frequency variation arising from chemical etching (5MHz, 5th overtone, BT-41 cut).

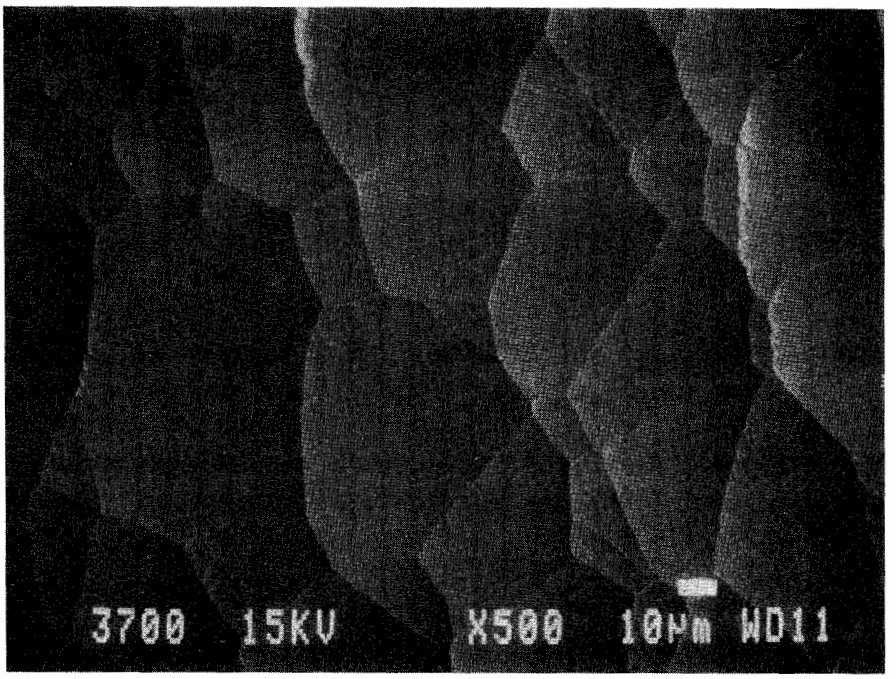


Figure 11 : SEM micrograph of a AT-37 plate with Δd of 10 µm.

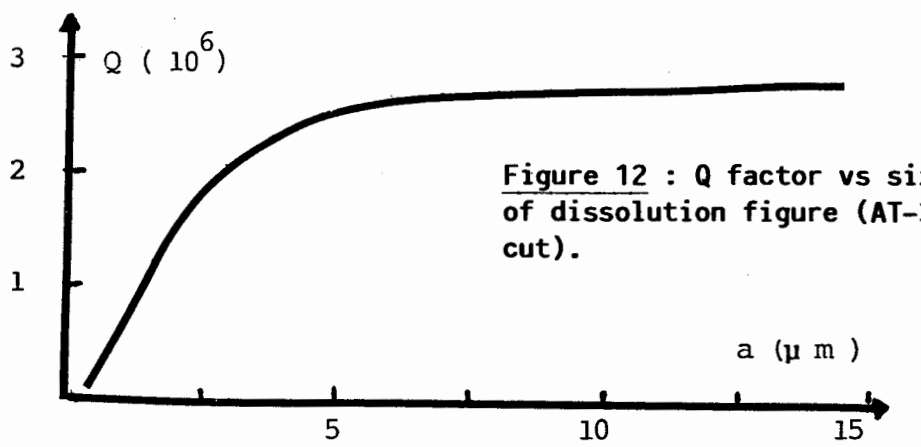


Figure 12 : Q factor vs size of dissolution figure (AT-37 cut).

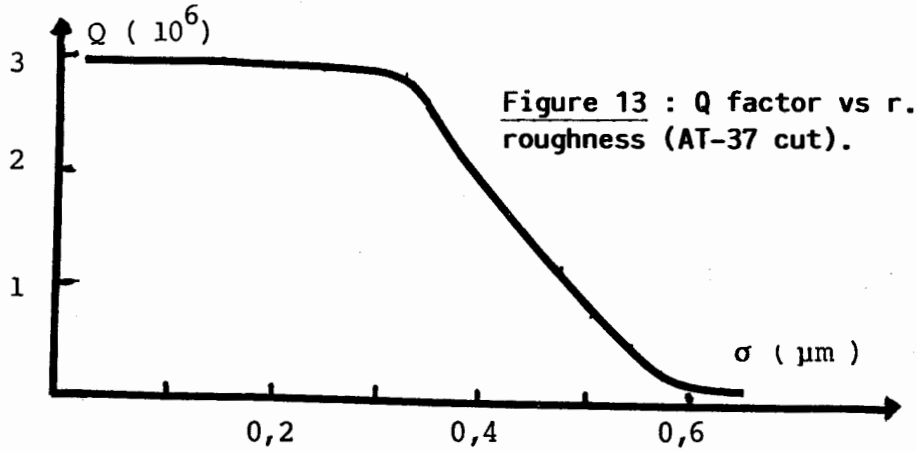


Figure 13 : Q factor vs r.m.s. roughness (AT-37 cut).

REFERENCES

(1) **R.N. CASTELLANO, T.R. MEEKER, R.C. SUNDAHL, J.C. JACOBS,** "The relationship between quartz surface morphology and the Q of high frequency resonators". Proc. of the 31st. Ann. Symp. on Freq. Cont., p 126-130 (1977).

(2) **J.R.HUNT, R.C. SMYTHE,** "Chemically Milled VHF and UHF Resonators". Proc. of the 39th Ann. Symp. on Freq. Cont., p 292 (1985).

(3) **P. BECKMAN, A. SPIZZICHINO,** "The scattering of electromagnetic waves from rough surfaces". Mac Millan, New York (1963).

(4) **L. FORTUIN,** "Survey of Litterature on reflection and scattering of sound waves at the sea surface". Journ. Acoust. Soc. Am., 47, 5, p 1209-1228, (1970).

(5) **P.J. WELTON,** "The potential method formulation of acoustic wave scattering by rough surfaces", Journ. Acoust. Soc. Am., 54, p 66-73 (1973).

(6) **C.R. TELLIER, N. VIALLE, J.L. VATERKOWSKI,,** "Simulation numérique des profils de surface obtenus par usinage chimique", 1er Forum Européen Temps-Fréquence, Besançon, p 159-169 (1987).

(7) **T. DEGUIN, R. BOURQUIN,** "Influence de l'état de surface sur le coefficient de qualité des résonateurs à quartz", 1er Forum Européen Temps-Fréquence, Besançon, p 184-188 (1987).

(8) **E. DIEULESAINT, D. ROYER,** "Ondes élastiques dans les solides", Masson, Paris, (1974).

(9) "I.E.E.E. Standard on Piezoelectricity -Std 176- 1978, I.E.E.E. pub, New York, (1978).

(10) **C.R. TELLIER, P. BLIND and D. JOZWICK,** 2nd European Forum Frequency and Time, Neuchâtel, Switzerland, 16-18 march 1988 (to be published).

(11) **C.R. TELLIER and J.L. VATERKOWSKI**, Surf. Technol., 26 (1985) 275.

(12) **R.B. HEIMANN**, in J. GRABMAIER (Ed.), "Silicon Chemical Etching" (Springer, Berlin, 1982) p 197.

2 ND EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 16 - 17 - 18 MARCH 1988

Session MP-02 Number 2

NONLINEAR PROPERTIES OF PIEZOELECTRIC SEMICONDUCTORS

N. DAHER and J.J. GAGNEPAIN

Laboratoire de Physique et Métrologie des Oscillateurs du CNRS
associé à l'Université de Franche-Comté-Besançon
32, avenue de l'Observatoire - 25000 Besançon - France

The present paper deals mainly with nonlinear effects in electroelastic semiconductors and more particularly in Gallium Arsenide crystal (GaAs). The understanding of such phenomena allows one to benefit from their advantages and avoid their inconveniences in technical applications. The basic continuum equations are deduced from a previously given phenomenological nonlinear theory of deformable semiconductors. The study of harmonic generation and anisochronism is performed with the help of an iterative procedure known as "Poincaré's Method". This work follows the same line as a recent study concerning purely elastic and piezoelectric dielectrics where a new formulation of the iterative procedure has been proposed. The latter clarified and unified the inconsistent results derived through different nonlinear resolution techniques. Therefore, we start by recalling briefly the above-mentioned theoretical features then the new version of "Poincaré's Method" is used to analyze the nonlinear electronic behaviour in semiconductors and the influence of conduction, diffusion and applied initial electric field. The attention is focused on the importance of electronic nonlinearities as compared to elastic, electric or electroelastic nonlinear effects. In particular, it is shown that the application of a bias electric field permits the alteration of the wave characteristics at will. Moreover, unlike dielectric materials relative velocity and frequency variations of order 10^{-3} are possible in semiconductors while keeping a stable resonant structure. This could answer need for variable and controllable frequency oscillators in the high frequency range.

INTRODUCTION

The study of nonlinear effects in solids has greatly developed in the last two decades. The most commonly studied form of the nonlinear behaviour is second harmonic generation [1-8] where one accounts for only one order of nonlinearity. However, in some cases such as the evaluation of anisochronism or amplitude-frequency effect, two orders of nonlinearity are required. This fact has been firstly pointed out by Tiersten, Thompson and Baumhauer [8-9] then followed by Planat, Théobald and Gagnepain [10-11]. In particular, two formulations known as "the method of characteristics" and the "Poincaré's expansion" showed their efficiency in the treatment of the subject matter but the solutions derived from both procedures were not in full agreement. Recently, a new version of "Poincaré's expansion" has been presented [12]. The latter, reconciliated the different results encountered in previous works and showed to be consistent with recent experimental verifications [13]. In the present contribution, the "Poincaré's expansion" is used to describe elastic, electric and electronic nonlinear effects in semiconductor devices and more particularly Gallium Arsenide. Comparisons between linear and nonlinear descriptions are made in a particular case where diffusion processes and applied electric fields are disregarded. Such a simplified situation allows one to place in evidence analytically the importance of electronic or charge carrier non linearities with respect to elastic and electroelastic ones. Finally, a numerical evaluation is performed in GaAs and in the general case where diffusion as well as an applied electric field are taken into account.

1 - NONLINEAR ELECTROACOUSTIC EQUATIONS FOR PIEZOELECTRIC SEMICONDUCTORS

Basic equations

The differential equations which govern the wave propagation in a piezoelastic semiconductor are deduced from Refs [12, 14, 15] as follows

- Equation of motion

$$\rho_o \frac{\partial^2 x_i}{\partial t^2} = T^T_{iK,K} \qquad (1)$$

Gauss' equation

$$D_{K,K} = Q \tag{2}$$

• Equation of conservation of charges

$$\frac{\partial Q}{\partial t} + \mathcal{T}_{K,K} = 0 \tag{3}$$

where we have set

$$\begin{aligned} T^T_{iK} = \delta_{iL} \{ {}_oT^T_{KL} &+ C^T_{KLMN}\, U_{M,N} - B^T_{PLK}\, \overline{W}_P \\ &+ \frac{1}{2} C^T_{KLMNPQ}\, U_{M,N}\, U_{P,Q} - \frac{1}{2} E^T_{PRKL}\, \overline{W}_P\, \overline{W}_R - H^T_{PLKMN}\, \overline{W}_P\, U_{M,N} \\ &+ \frac{1}{3!} C^T_{KLMNPQRS}\, U_{M,N}\, U_{P,Q}\, U_{R,S} - \frac{1}{3!} E_{PRSLK}\, \overline{W}_P\, \overline{W}_R\, \overline{W}_S \\ &- \frac{1}{2} H^T_{PLKMNPQ}\, \overline{W}_P\, U_{M,N}\, U_{P,Q} - \frac{1}{2} F^T_{PRLKMN}\, \overline{W}_P\, \overline{W}_R\, U_{M,N} \} \end{aligned} \tag{4}$$

$$\begin{aligned} D_K = {}_oD_K &+ \varepsilon_{KP}\, \overline{W}_P + B^T_{KMN}\, U_{M,N} + \frac{1}{2} \chi_{KPR}\, \overline{W}_P\, \overline{W}_R \\ &+ \frac{1}{2} H^t_{KMNPQ}\, U_{M,N}\, U_{P,Q} + E^t_{KPMN}\, \overline{W}_P\, U_{M,N} \\ &+ \frac{1}{3!} \chi_{KPRS}\, \overline{W}_P\, \overline{W}_R\, \overline{W}_S + \frac{1}{3!} H^t_{KMNPQRS}\, U_{M,N}\, U_{P,Q}\, U_{R,S} \\ &+ \frac{1}{2} E_{KPRMN}\, \overline{W}_P\, \overline{W}_R\, U_{M,N} + \frac{1}{2} F^t_{KRMNPQ}\, \overline{W}_R\, U_{M,N}\, U_{P,Q} \end{aligned} \tag{5}$$

$$\mathcal{T}_K = {}_o\mathcal{T}_K + \Sigma^o_{KP}\, \overline{W}_P - D_{KP}\, \overline{Q}_{,P} + M_{KP}\, \overline{Q}\, ({}_oW_P + \overline{W}_P) \tag{6}$$

with

$$x_i = \delta_{iK}\, [X_K + U_K] \qquad T^T_{iK} = \delta_{iL}\, [T^E_{LK} + T^F_{LK}]$$

$$\begin{aligned} {}_oD_K &= {}_oW_K + {}_o\mathcal{P}_K \qquad & {}_oT^T_{KL} &= {}_oT^E_{KL} + {}_oT^F_{KL} \\ {}_o\mathcal{T}_K &= \Sigma^o_{KL}\, ({}_oW_L - \mu_{o,L}) \end{aligned} \tag{7}$$

where X_K and x_i denote respectively the position of a material particle in the reference and current configurations, while ${}_o\underset{\sim}{W}$, μ_o, ${}_oT^E_{KL}$, ${}_oT^F_{KL}$, ${}_o\mathcal{P}_R$, ${}_o\mathcal{T}_R$ and ${}_oQ$ are respectively the electric field, the chemical potential, the elastic stress tensor, the Maxwell tensor relative to electromagnetic fields, the polarization, the conduction current and the charge density. All these parameters are evaluated at the initial state. The

effective material constants relative to C_{KLMN}, B_{NKL}, ε_{LN}, Σ_{KL}, D_{KL}, M_{KL}, C_{KLMNPQ}, H_{QKLMN}, E_{NQKL}, χ_{LNQ}, $C_{KLMNPQRS}$, $H_{SKLMNPQ}$, F_{QSKLMN}, E_{NSQKL} and χ_{LNSQ} denote respectively, the second order elastic, piezoelectric, second order dielectric, conductivity, diffusion, mobility, third order elastic, first odd electroelastic, electrostrictive, third order electric permeability, fourth order elastic, second odd electroelastic, first even electroelastic, third odd electroelastic, and fourth order electric permeability tensors. As to the upperscripts "t" and "T", they indicate the deviation of a quantity from its thermodynamical definition. This deviation is due either to convection terms or to the presence of an initial state, or to dynamical electromagnetic contributions (see Ref. [14] for details). As an example we give the expression of B^{T}_{KRL} as follows

$$B^{T}_{KPQ} = B_{KPQ} + \mathcal{B}_{PQKL}\ {}_{o}W_{L} \tag{8}$$

with

$$\mathcal{B}_{PQKL} = \mathcal{B}_{KLPQ} = \mathcal{B}_{PQLK} = \delta_{KL}\,\delta_{PQ} - \delta_{PL}\,\delta_{QK} - \delta_{PK}\,\delta_{QL} \tag{9}$$

Before proceeding further, it is to be noted that in the above-mentioned equations many simplifying hypotheses have been implicitly assumed.

More precisely, in the nondissipative constitutive equations (4)-(5) the interactions between the semiconducting and dielectric continua have been disregarded, thus accounting only for elastic, electric and electroelastic contributions. As to the electronic nonlinearity, it appears in the last term of the dissipative equation (6).

These simplifications are necessary to allow for analytical and numerical treatment. Indeed, Refs [12, 14, 15] clearly show the complexity of the general case even in the framework of a linearized theory about an initial state, and, most of all, no experimental support is presently available for such new contributions placed in evidence theoretically only in recent years. In conclusion, we may say that although the phenomenological theory exhibits various linear and nonlinear cross-effects, elastoresistance etc. it seems enlightening to start with the study of the analytically and numerically exploitable features to allow for practical applications. By the way, we recall that even in simpler situations such as in pure elasticity and piezoelectric dielectrics, we often have either a bad knowledge or a complete ignorance of some of the nonlinear elastic, electric and electroelastic coefficients. Finally, it is useful to recall that the applied electric field must not exceed a certain limit beyond which Ohms law (6)

does not apply anymore since the concentration of carriers as well as the mobility are both modified in the presence of an intense bias field.

Following the same procedure as in Refs [10,12] we set

$$(U_K, \bar{\Phi}, \bar{Q}) = \varepsilon\,(\overset{1}{U}_K, \overset{1}{\Phi}, \overset{1}{Q}) + \varepsilon^2\,(\overset{2}{U}_K, \overset{2}{\Phi}, \overset{2}{Q}) + \varepsilon^3\,(\overset{3}{U}_K, \overset{3}{\Phi}, \overset{3}{Q}) \qquad (10)$$

where ε is an infinitesimally small parameter and $\bar{\Phi}$ the electric potential such that $\bar{W}_P = -\bar{\Phi}_{,P}$. On carrying eq. (10) into (4)-(6) and identifying with respect to ε eqs. (1)-(3) yield four order differential systems. If one does not account for an initial state, the zeroth-order system vanishes and we are left with linear, quadratic and cubic equations that we do not reproduce here (see Ref. [12] for details).

In the present context we consider a constant initial state so that the material coefficients become homogeneous and eqs. (1)-(9) greatly reduce.

It is to be noted that the interest of the above-mentioned procedure known as "Poincaré's Method" lies in the fact that it may be applied to involved problems such as surface wave propagation. However, in the last two decades a controversy surrounded the results obtained in a purely elastic case by the above-mentioned procedure and by another formulation known as the method of characteristics to which one applies a development in series because of the smallness of the amplitudes. In particular, insofar as anisochronism is concerned the relative difference between the two solutions could reach 50 %. The disagreement is due to the omission of a constant of integration in the Poincaré's expansion by assuming a vanishing radiation stress associated with the steady state solution. Recently, the situation was clarified and the results reconciliated by the development of a new version of the Poincaré's expansion [12]. The latter is consistent with recent experimental results [13]. It is precisely this new version that we use here.

2 - BULK WAVE PROPAGATION

We recall that the "Poincaré's method" consists first in the study of the linear system then nonlinear corrections are introduced progressively through various iterations that account for the different nonlinear orders.

Therefore, as usual in the study of linear wave propagation problems we look for a plane harmonic wave solution so that the linear contributions in eqs. (1)-(6) lead to the following dispersion relation

$$\det \left[(\underset{\sim}{S}^{P} + \xi_o \, \omega_c \, \underset{\sim}{S}^{E})^2 + (\eta_o \, \omega_c \, \underset{\sim}{S}^{E})^2 \right] = 0 \tag{11}$$

where we have set

$$\underset{\sim}{S}^{P} = V^2 \underset{\sim}{I} - \underset{\sim}{\tilde{\Gamma}}^{T}/\rho_o \quad , \quad \underset{\sim}{S}^{E} = V^2 \underset{\sim}{I} - \underset{\sim}{\Gamma}^{T}/\rho_o \tag{12}$$

with

$$\tilde{\Gamma}^{T}_{KM} = \Gamma^{T}_{KM} + \frac{B^{T}_{K} \, B^{T}_{M}}{\varepsilon} \tag{13}$$

$$\Gamma^{T}_{KM} = C^{T}_{KLMN} \, b_L \, b_N \quad , \quad B^{T}_{K} = B^{T}_{MKP} \, b_M \, b_P \quad , \quad \varepsilon = \varepsilon_{KL} \, b_K \, b_L \tag{14}$$

and

$$\xi_o = \frac{1}{\omega_D} \frac{1}{1 + \left(\frac{\omega_E}{\omega_D}\right)^2} \quad , \quad \eta_o = \frac{1}{\omega_E} \frac{1}{1 + \left(\frac{\omega_E}{\omega_D}\right)^2} \tag{15}$$

$$\omega_c = \Sigma^{o}/\varepsilon \quad , \quad \omega_D = V_E^2/D \quad , \quad V_E = \gamma V = \omega_E/k \quad , \quad \gamma = 1 + \frac{V_d}{V} \tag{16}$$

$$\Sigma^{o} = \Sigma^{o}_{KL} \, b_K \, b_L \quad , \quad D = D_{KL} \, b_K \, b_L \quad , \quad V_d = -M_{KL} \, {}_{o}\dot{W}_L \, b_K \tag{17}$$

V and k are respectively the phase velocity and the wave number, ω_c is the dielectric relaxation frequency, ω_D is the diffusion frequency and ω_E is an effective frequency which depends on the applied electric field. Obviously, in the absence of conduction ($\omega_c = 0$) we obtain the piezoelectric solution

$$\det \left(\rho_o V^2 \, \delta_{KM} - \Gamma^{T}_{KM} - \frac{B^{T}_{K} \, B^{T}_{M}}{\varepsilon} \right) = 0 \tag{18}$$

As to great conductions ($\omega_c \to \infty$) we are left with

$$\det \left(\rho_o V^2 \, \delta_{KM} - \Gamma^{T}_{KM} \right) = 0 \tag{19}$$

which is none other than the purely elastic solution. This clearly delineates the important role played by the dielectric relaxation frequency in the coupling between elasticicy, piezoelectricity and semiconduction.

This fact still holds for the quadratic and cubic solutions that we do not give here explicitly. The procedure is performed as follows : on carrying the linear solution into the second order system and accounting for the addition for products of trigonometric functions one deduces the second order solution containing second harmonics. The latter solution is carried into the third order system and the same procedure is applied again. Thus, it is shown that the third order solution contains third harmonics as well as corrective nonlinear contributions at the fundamental frequency. These corrective contributions are essential in the evaluation of the relative velocity change due to nonlinear effects which is our main concern. In a first step we shall compare the linear and nonlinear descriptions in elastic, piezoelectric and semiconducting crystals in the absence of an applied electric field (i.e. the initial state is nil) and where diffusion processes are disregarded. The general case, which requires a numerical evaluation is discussed separately in Section 4.

3 - COMPARISON BETWEEN LINEAR AND NONLINEAR DESCRIPTIONS

In order to compare the relative velocity change obtained in the present nonlinear theory with respect to the one derived in the framework of a linear description [12], it is useful to set the following self explanatory notation

$$\left(\frac{\Delta V}{V}\right)^{L,N}_{S,P,E} = \frac{V^{L,N}_{S,P,E} - V^{L}_{E}}{V^{L}_{E}} \tag{20}$$

where the upperscripts L and N stand for linear and nonlinear respectively while the subindexes S, P, E stand for semiconduction, piezoelectricity and elasticity respectively. On account of the above-mentioned definition (20) and of Ref. [12] we obtain

$$\left(\frac{\Delta V}{V}\right)^{L}_{E} = 0 \quad , \quad \left(\frac{\Delta V}{V}\right)^{L}_{P} = \frac{K^2}{2} \quad , \quad \left(\frac{\Delta V}{V}\right)^{L}_{S} = \left(\frac{\Delta V}{V}\right)^{L}_{P} - \frac{K^2}{2}\,\frac{\beta_c^2}{1+\beta_c^2} \tag{21}$$

insofar as the linear theory is concerned where K^2 is the electromechanical coupling coefficient and $\beta_c = \omega_c/\omega$. As to the nonlinear description, after some manipulations accounting for Ref. [12] one may write

$$\left(\frac{\Delta V}{V}\right)^{N}_{E} = \frac{1}{8}\left[\left(\frac{\Gamma}{\lambda}\right)^2 - \frac{1}{2}\,\frac{\Lambda}{\lambda}\right]k^2 U_o^2 \quad , \quad \left(\frac{\Delta V}{V}\right)^{N}_{P} = \left(\frac{\Delta V}{V}\right)^{N}_{E} + \delta \tag{22}$$

where λ, Γ, Λ and U_o are respectively second, third and fourth order elastic coefficients and the amplitude of vibrations.

The parameter δ contains different linear and nonlinear electroelastic coefficients that we do not reproduce here.

As to semiconducting crystals, it is shown that the nonlinear elastic and piezoelectric coefficients may be neglected with respect to electronic nonlinearities for intermediate conduction such that

$$\beta_c^{min} < \beta_c = \omega_c/\omega < \beta_c^{max} \tag{23}$$

where β_c^{min} and β_c^{max} depend on the material in consideration. It is clear that if the ratio of β_c^{max} over β_c^{min} is not great enough, the electronic nonlinearities can never be predominant. This is the case in dielectric media where conductivities are limited in a narrow range. As to semiconductor devices a great number of situations may correspond to $\beta_c^{max}/\beta_c^{min} \gg 1$ so that for intermediate conductions the electroelastic nonlinearities may be negligible as compared to electronic ones. More precisely, we have the following situation

$$\left(\frac{\Delta V}{V}\right)_S^N = \frac{3}{2} K^2 \left(\frac{eM}{\epsilon V}\right)^2 \frac{\beta_c^2}{(1+\beta_c^2)^3 (1+4\beta_c^2)} k^2 U_o^2 + \Delta \tag{24}$$

where Δ is a nonvanishing contribution only when the electronic nonlinearities are smaller or of the same order as the electroelastic ones. e and M are respectively the piezoelectric constant and the mobility coefficient.

Since Δ does not seem to be of a great interest for practical applications and the determination of its explicit expression requires very long and tedious calculations, a special attention is devoted merely to limit situations. After some calculations we get

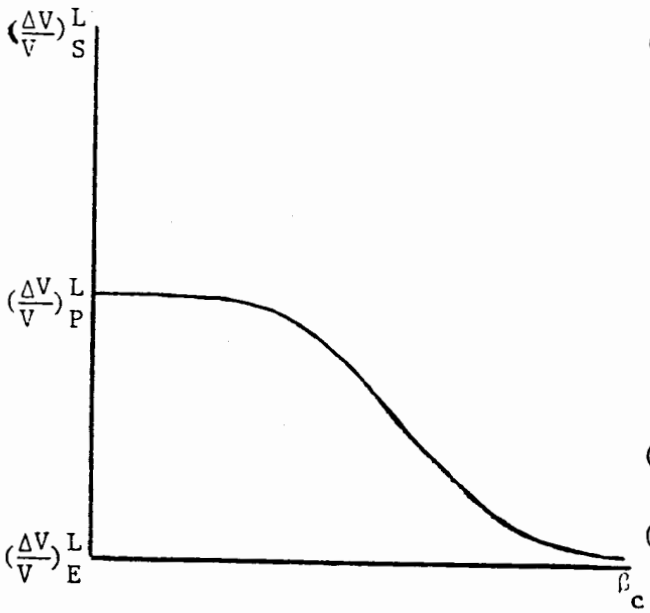
$$\Delta = \left(\frac{\Delta V}{V}\right)_P^N \qquad \text{when} \quad \beta_c \to 0 \tag{25}$$

and

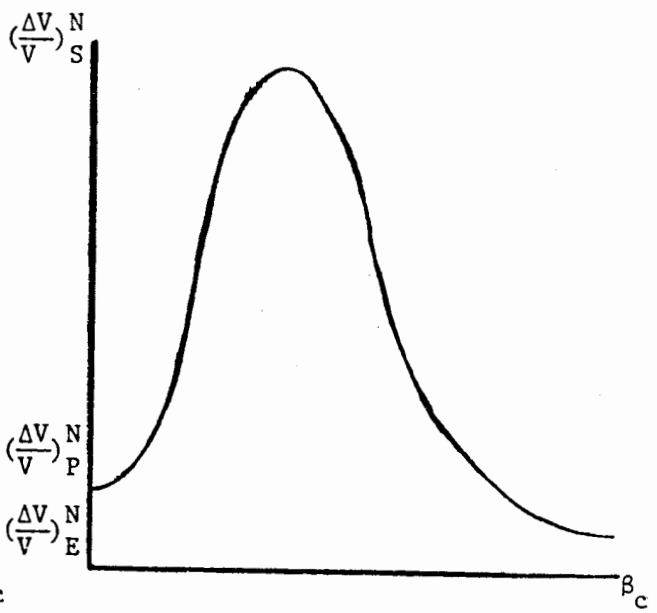
$$\Delta = \left(\frac{\Delta V}{V}\right)_E^N \qquad \text{when} \quad \beta_c \to \infty \tag{26}$$

Eq. (24) subject to eqs. (25)-(26) clearly shows that for a vanishing β_c we obtain the piezoelectric solution while for an infinite value of β_c we are led to the purely elastic one. This fact has been already pointed out and discussed in the linear solution.

In order to place in evidence the similarities and differences between the linear and nonlinear behaviours eqs. (21)-(26) allow one to exhibit the variation of the relative velocity changes with respect to β_c. After some calculations we obtain the following qualitative behaviour (not to scale !)



Linear behaviour



Nonlinear behaviour

It is easy to note the similar behaviour at limit conditions since

$$(\frac{\Delta V}{V})^{L,N}_S = (\frac{\Delta V}{V})^{L,N}_P \qquad \text{when} \quad \beta_c \to 0 \tag{27}$$

and

$$(\frac{\Delta V}{V})^{L,N}_S = (\frac{\Delta V}{V})^{L,N}_E \qquad \text{when} \quad \beta_c \to \infty \tag{28}$$

As to the situation corresponding to intermediate values satisfying (23) a quite different behaviour occurs between the linear and nonlinear descriptions. More precisely, we have

$$(\frac{\Delta V}{V})^L_S < (\frac{\Delta V}{V})^L_P \qquad \forall \quad \beta_c \tag{29}$$

according to the linear theory while the nonlinear description leads to

$$(\frac{\Delta V}{V})^N_S > (\frac{\Delta V}{V})^N_P \qquad \text{for} \quad \beta_c^{min} < \beta_c < \beta_c^{max} \tag{30}$$

On performing a numerical evaluation one deduces that electronic or charge carrier nonlinearities may exceed one hundred times the nonlinear effects due to elastic or electroelastic contributions.

This confirms experimental results showing that the electronic nonlinearities are predominant as compared to the electroelastic ones in many practical and useful situations and justifies previous theoretical treatments concerned with only second harmonic generation and disregarding elastic and electroelastic nonlinearities from the start [4,5].

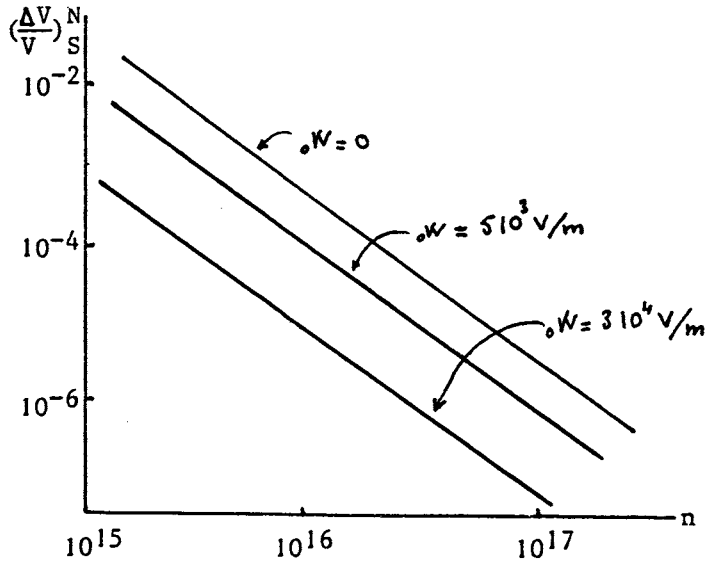
4 - INFLUENCE OF DIFFUSION AND APPLIED DC ELECTRIC FIELD

Although, in Section 3, we placed in evidence the similarities and differences between linear and nonlinear descriptions, the neglect of diffusion process is not realistic in some materials such as GaAs. Moreover, the application of a bias electric field allows one to alter the wave characteristics at will leading to the possibility of obtaining variable and controllable frequency oscillators.

Consequently, this Section deals with numerical evaluations based on a general relation derived in the same manner as eq. (24) but the latter is too complex to be given here or to be analytically exploitable (see Ref. [12] for details). Here the attention is focused on the following realistic conditions

$$10^{15} < n < 10^{17}\ \text{cm}^{-3}\ ,\quad \omega = 10^{9}\ \text{s}^{-1} \Leftrightarrow f \sim 160\ \text{MHz}\ ,\quad 0 < {}_{o}W < 3.10^{4}\ \text{V/m}$$

where n denotes the number of charge carriers per unit volume. After numerical calculations we are led to



Variation of the relative velocity change due to electronic or charge carrier nonlinearities

In order to evaluate the influence of a DC electric field, the phase velocities in the absence then the presence of such a field are denoted by V_o and V respectively. Thus, for $n \simeq 10^{16}$ cm^{-3} we obtain

$$0 < \left|\frac{V - V_o}{V_o}\right| < 7.10^{-4} \quad , \quad 0 < {}_oW < 3.10^4 \text{ V/m or } 0 < \Phi < 10 \text{ V}$$

For a fixed DC electric field such as ${}_oW \simeq 5.10^3$ V/m and variable concentration of carriers one deduces

$$10^{-2} < \left(\frac{\Delta V}{V}\right)^N_S < 10^{-6} \quad \text{when} \quad 10^{15} < n < 10^{17} \text{ cm}^{-3}$$

This clearly shows that unlike dielectrics where $(\Delta V/V)^N_P \sim 10^{-6}$, relative velocity changes or frequency variations of order 10^{-3} which are needed in high frequency oscillators are possible in semiconductors according to the present theoretical approach.

REFERENCES

[1] R.W. Dixon, "Acoustic nonlinear frequency mixing detected using optical Bragg diffraction", Appl. Phys. Lett., 11, 340-344 (1967).

[2] R.B. Thompson and C.F. Quate, "Acoustic parametric oscillations in $LiNbO_3$", Appl. Phys. Lett., 16, 295-298 (1970).

[3] R.B. Thompson and C.F. Quate, "Nonlinear interaction of microwave electric fields and sound in $LiNbO_3$", J. Appl. Phys., 42, 907-919 (1971).

[4] B. Tell, "Piezoelectric ultrasonic harmonic generation in Cadmium sulfide", Phys. Rev., 136, 3A, 772-775 (1964).

[5] P. Palanichamy and S.P. Singh, "Acoustic second harmonic generation in piezoelectric semiconductors : effect of nonuniform electric field intensity", J. Appl. Phys., 54(7), 3958-3964 (1983).

[6] D.F. Nelson, "Acoustic wave mixing in piezoelectric crystals", J. Acoust. Soc. Am., 64(2), 652-657 (1978).

[7] D.F. Nelson, "Three-field electroacoustic parametric interactions in piezoelectric crystals", J. Acoust. Soc. Am., 64(3), 891-895 (1978).

[8] H.F. Tiersten and J.C. Baumhauer, "Second harmonic generation and parametric excitation of surface waves in elastic and piezoelectric solids" J. Appl. Phys., 45 (10), 4272-4287 (1974).

[9] R.B. Thompson and H.F. Tiersten, "Harmonic generation of longitudinal elastic waves", J. Acoust. Soc. Am., 62(1), 33-37 (1977).

[10] M. Planat, G. Théobald et J.J. Gagnepain, "Propagation non linéaire d'ondes élastiques dans un solide anisotrope. I - Ondes de volume", L'Onde Electrique, vol. 50, n° 11 (1980).

[11] M. Planat, "Propagation non linéaire des ondes acoustiques dans les solides", Thèse d'Etat, Université de Franche-Comté (1984).

[12] N. Daher, "Principe des puissances virtuelles étendu aux discontinuités et interfaces. Application à l'acousto-électronique", Thèse d'Etat, Université de Franche-Comté (1987).

[13] J.H. Cantrell, W.T. Yost and P. Li, "Acoustic radiation-induced static strains in solids", Phys. Rev. B, 35(18), 9780-9782 (1987).

[14] N. Daher and G.A. Maugin, "Nonlinear electroacoustic equations in semiconductors with interfaces", Int. J. Engng. Sci., 26(1), 37-58 (1988).

[15] N. Daher and G.A. Maugin, "Deformable semiconductors with interfaces", Int. J. Engng. Sci., 25, 1093-1129 (1987).

2 ND EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 16 - 17 - 18 MARCH 1988

Session MP-02 Number 3

GROWTH DEFECTS IN CULTURED QUARTZ
A MICROSCOPIC INVESTIGATION

P. CORDIER and J.C. DOUKHAN
Laboratoire de Structure et Propriétés de l'Etat Solide
Université de Lille Flandres Artois, 59655 Villeneuve d'Ascq Cédex FRANCE

J. BIGNON and X. BUISSON
SICN, 4 rue du radar, 74000 Annecy FRANCE

ABSTRACT

Although cultured quartzes have now reached a high degree of purity and crystalline perfection, there remains, at least in some crystals, various types of defects which can induce limitations to the performances of piezoelectric devices. One reports here a detailled investigation at a microscopic scale of the various defects occurring in synthetic quartz. Optical microscopy, X-Ray Lang topography, analytical transmission electron microscopy as well as infrared spectroscopy have been used. Small solid (acmite and other silicates) and fluid inclusions as well as water point defects are thus characterized and related to the various growth parameters. It is shown that water which is generally assumed to be the most cumbersome impurity can enter into the quartz lattice under two distinct forms : i) very tiny clusters of ≅ 20 Å diameter in an otherwise quite dry and pure material and ii) isolated and randomly distributed point defects with a concentration much larger than the equilibrium one at the growth conditions or at the standard use conditions (room temperature, atmospheric pressure). Migration and precipitation of these defects can thus occur slowly and progressively, thus inducing an aging of the device. Mobility and equilibrium concentration of point defects are tentatively characterized.

I - INTRODUCTION

Since twenty years numerous studies of lattice defects in cultured quartz have been published (for recent reviews 1 to 3). The best characterized defects are probably dislocations. Since the first studies of Land and Miuskow (4,5) they are currently visualized by X-Ray Lang topography. Their Burgers vectors are most often the lattice repeat units $\underset{\sim}{a}$ (4.9 Å) and $\underset{\sim}{c}$ (5.4 Å) while their line directions are frequently approximately perpendicular to the growth surfaces. The density of as grown dislocations strongly depends on the crystalline perfection of the seed because any dislocation contained in the seed is necessarily prolongated in the crystal. Dislocation densities lower than 10^3 cm^{-2} are currently obtained ; in some cases densities as low as 1 cm^{-2} have been obtained with dislocation free seeds (6 to 9). X-Ray Lang topography also shows that several growth sectors are to be distinguished in a cross section perpendicular to the seed (4,5,10 to 13), some sectors exhibiting a lower dislocation density while others seem to have a lower impurity content. Fluid inclusions have also been frequently observed in the immediate vicinity of the seeds (14 to 17). They are most usually interpreted as the result of preferential etching of the seed along its dislocation lines at the very beginning of the growth process. The occurrence of such elongated fluid inclusions does not imply that far from the seed the crystal is wet. Tiny solid inclusions are also found from time to time. They are generally assumed to be small crystals of acmite Na Fe^{3+} Si_2O_6 resulting from the slow corrosion of the non lined steel autoclave by the Na rich aqueous solution at high temperature and high pressure. There have been very few systematic studies of these solid inclusions (17 to 19). In contrast chemical impurities dissolved in the lattice as point defects have received considerable attention since a number of years (20 to 23) because their concentration seems to influence at a time the quality factor Q and the irradiation hardness of the material. Among the chemical impurities, water or OH radicals play a special role and the water (or OH or H) content of cultured quartz has been investigated by numerous authors. The standard method for the characterization of the OH content is infrared spectroscopy (24 to 26). It is to be noted that mineralogists are also interested by the phenomenon of water incorporation in quartz. Their technique for the characterization of the OH content is slightly different [27 to 29]. A number of mineralogists actually believe that there is certainly a certain amount of water which is truly dissolved in the lattice as point defects but the corresponding solubility would be very small at least at the growth conditions ($P < 2kbar$, $T \cong 360°$).

We report in this paper our own investigations by infrared and near infrared spectroscopy (IRS and NIRS respectively), optical microscopy, X-Ray Lang topography and analytical transmission electron microscopy. The first section is devoted to what can be

called macroscopic defects i.e. the solid and fluid inclusions. Several types of solid inclusions have been characterized in cultured quartzes (acmite as well as other silicates). Chemical impurities like Al can also induce another type of macroscopic defects because their heterogeneous distribution in various growth sectors induces slight changes of the lattice parameters which in turn induce internal stresses, misorientations and changes of the growth rate. The second section deals with water in quartz ; the water content as well as the mode of water incorporation has been characterized in various specimens. The equilibrium concentration of water point defects as well as their mobility are tentatively experimentally characterized as a function of temperature and pressure.

Finally it should be noted very clearly that all the defects analysed in this study are rarely present all together in high quality quartzes grown commercially. This report should not give the impression that the availiable PQ quartzes still are of poor quality and their growth still needs serious improvement.

II - MACROSCOPIC DEFECTS

As noted above three types of macroscopic defects can be distinguished i) fluid inclusions, ii) solid inclusions and iii) internal strains due to heterogeneous concentrations of chemical impurities. Fluid inclusions often occur in the vicinity of the seed and they probably result from preferential etching of the dislocation cores in the seed at the very beginning of the growth process (30). Very tiny fluid inclusions are also found in the bulk when the growth rate becomes large (31) or when the nutrient or the solution are impure (32). Fluid inclusions have also been found in some natural quartzes and they have also been interpreted as the result of perturbed growth (33). Figure 1a is an optical micrograph at moderate magnification of elongated fluid inclusions in a synthetic quartz grown in normal conditions and in a chemically pure environment. The tunnels underline the seed and penetrate into it over a length of approximately 50 µm. There is no other inclusion in the crystal and the ones detected here are probably irrelevant for the qualities of the material far from the seed ; they could even improve these qualities because such inclusions are discontinuity surfaces where the dislocations of the seed end up, i.e. the material grown on deeply etched seeds could exhibit a lower dislocation density. The situation is completely different when numerous very tiny fluid inclusions occur in the bulk. They can only be observed by TEM at relatively large magnifications (x 10^4 to x 10^5). The TEM micrograph Figure 1b corresponds to a quartz grown at normal conditions but in an impure environment, the solution having deliberately been doped with iron oxide. One clearly observes a large density of very tiny bubbles of ≅ 500 Å diameter and dislocations which seem to be pinned by the bubbles. The small bubbles could not be detected by optical microscopy (the crystal is perfectly clear) but the IRS signature of this

quartz indicates that it contains a large amount of water. Other quartzes grown in solutions enriched in various metal oxides (FeO, CaO, Al_2O_3, TiO_2...) also exhibit the same phenomenon even when the growth rate is extremely low. They all are very wet and contain very tiny bubbles with a diameter comprised between 100 and 500 Å (but the minimum bubble size detected in TEM is ≅ 80 Å diameter i.e. smaller undetected bubbles could occur). A very similar phenomenon occurs when in a chemically pure environment the growth rate is increased. For instance an incident has led to an abrupt change during a short time of the growth rate of the crystal shown on Figure 1c. The affected zone observed in TEM shows a configuration very similar to the one of Figure 1b, numerous very tiny bubbles and dislocations are visible but this configuration is now restricted to a narrow band 10 to 50 μm thick while it was pervasive in the previous case. Except for this narrow band the crystal is dry and pure. It is thus clear that either impure solutions or high growth rates induce the formation of tiny water bubbles (detected by their IRS signature) which dramatically affect the performances of the material. We have not elucidated the mechanism of formation of such bubbles and we do not know if they are accompanied by other defects, for instance chemical impurities dissolved in the lattice as point defects (metal ions and/or water point defects).

Large solid inclusions occur in some crystals and they are easily detected by optical microscopy. Figure 2a is a typical example. The current interpretation for the occurrence of such solid inclusions is the following. The steel vessel is slowly corroded by the Na rich solution at high temperature and high pressure. In non lined autoclaves small crystals are known to grow on their inner surfaces (31). Some small crystals could thus leave the surface and be incorporated in the growing quartz. Various Na Fe silicates have been indentified in the coating which covers the autoclave inner surfaces. Acmite, a clinopyroxene with the chemical formula $NaFeSi_2O_6$ (34) is the most usually cited and a number of authors when they observed in their samples very tiny dark spots too small to be conveniently analysed have assumed that this was inclusions of acmite. In fact the formation mechanism proposed should lead to a random distribution of the solid inclusions inside a given crystal and among the various crystals grown at a time in any given run. One rather observes that numerous crystals are free of inclusions while in some others one observes clusters of needle shaped cristalline inclusions. They may be not visible at the naked eye, they show a weak contrast in optical microscopy but they are clearly identified in TEM as shown on Figure 2b. The needles are a few microns long with a roughly square cross section of ≅ 0.05 μm^2. X-Ray microanalyses performed by energy dispersive spectroscopy (EDS) indicate that some needles do are Na Fe silicates in agrement with the hypothesis cited above but some other inclusions are calcium silicates (see the EDS spectra Figure 2c). Such inclusions have necessarily grown in the solution from the impurities contained in the water or in the nutrient. The purity of the starting products is clearly an

important parameter (difficult to control systematically in the case of commercial production). Furthermore although the inner surface of the autoclave is a preferential site of nucleation of Na Fe silicates, some cristalites can be nucleated and grow in the solution, the previous model of acmite needles grown on the surface and leaving it to be incorporated in the growing quartzes is not the only possible mechanism of formation of the solid inclusions. Finally it can be said that, like for the case of fluid inclusions, the mechanism which induces the occurrence of solid inclusions in cultured quartzes remains largely unknown.

The third type of macroscopic defects is illustrated by Figure 3. In optical microscopy one observes a hole in the crystal which seems to have started from the external (final) surface of the crystal and which could be more or less healed at its bottom. This hole is somewhat similar to an etch tunnel which would habe been deeply etched at the end of the growth process by the growing solution itself. The hole lies at the boundary between two growth sectors. This is more evident on the X-Ray Lang topogram which shows that the hole ends a misoriented region (the one out of contrast on the Lang topogram). This region starts at the bottom of the seed. The misoriented area as well as the misorientation angle increase altogether from the seed to the external surface of the crystal. Iwasaki and Kurashige (31) have already made similar observations and they have performed X-Ray mappings in a scanning electron microscope with a view to characterize the concentration of impurities on a cross-section of the crystal. They find a much larger concentration of Aluminium in the misoriented zone. One cas thus imagine the following mechanism for the formation of this later macroscopic defect. For some reason related to the growing process itself the chemical impurities tend to concentrate in the boundary between two sectors. Because of this large Al content the lattice parameters of quartz are slightly changed and matching both lattices, the pure and the Al rich ones induces elastic strains which in turn induce a slight misorientation visible on the X-Ray topograms. The growing rate is affected by the strain and reduced, leading to the final shape of the cross section. At the end of the growth, when the crystal is slowly cooled down, the impure and elastically strained region would be selectively redisolved by the solution thus leading to the observed holes. Such an etching process would relax the internal stress and allow partial healing with pure material ; of course such a scenario remains highly speculative and is to be considered just as a working hypothesis. This model underlines again one time the importance of a low concentration of impurities in the nutrient. Not only these later ones can affect the purity of the final product but they can induce lattice imperfections, solid inclusions and internal strains as well.

III- WATER-RELATED DEFECTS

The high performances and the reliability required by some applications of piezoelectric devices point out the role played by the remaining impurities even if their concentration is very low. Among these defects, the most important are probably aluminium, alkali ions and water-related defects. Alkali ions are mostly present as charge compensators of substitutionnal aluminium $(Al)_{Si}$, and affect the irradiation hardness of resonators. They can be removed by sweeping. Water defects are known to decrease the quality factor Q and a good correlation has been found between Q and the amount of OH measured by IRS at 3500 cm^{-1} (35). In fact, water absorption is not restricted to one wavelength but extends as a broad band between 3800 and 3000 cm^{-1} with some sharper peaks superimposed (especially in the vicinity of 3600 cm^{-1}). The amount of water incorporated during growth is strongly dependent on the growth rate and small fluctuations of this parameter (the so-called breathing of crystal growth) gives rise to striation along the c axis. They are easily observed by light scattering tomography after heat treatment (36). Although IRS is very usefull to get a quantitative measurement of the water content (26) (27), it provides very few information about the mode of incorporation of water defects. Several models have been proposed for the location of hydrogen (water) in the quartz lattice.

i) The first one is probably the one of Griggs who imagined that the water molecules would hydrolyse the strong Si-O-Si bonds (37). ii) an other possibility is, in analogy with hydrogarnets (38), a substitutionnal defect where four H's take the place of one silicon. This defect noted $(4H)_{Si}$ or $(2H_2O)_{SiO_2}$ has been identified by ESR in quartz by Nutall and Weil (39), iii) More recently, Hagon et al. (40) have proposed that interstitial water molecules would be located inside the $\underset{\sim}{c}$ channels and they would be very mobile in this $\underset{\sim}{c}$ direction (activation energy of the coefficient of diffusion : 0.2 eV), iv) detailled TEM investigations on very wet quartzes clearly show very tiny water bubbles. This suggests that in dryer quartzes some water could be incorporated as still smaller bubbles not detected even by TEM, their radius being of the order of 30 Å or even less (one should probably speak of clusters or agregates of water molecules rather than bubbles).

The influence of such defects on the piezoelectric properties of resonators can be discussed as follows : tiny bubbles are expected to scatter the accoustic waves while point defects are rather supposed to induce an accoustic loss peak by their oscillating jumps (while the device is working) between different possible sites. This process is very similar to the Snoeck effect in carbon steels. Whatever the kind of defect, as long as the system quartz-water is not in equilibrium conditions, a slow and continuous shift of the characteristics with the working time can be expected. We have therefore tried to get more information about the kind of defects occurring in our synthetic crystals. As noted above, conventionnal

IRS is not a convenient technique for this purpose. In contrast NIRS between 7000 and 4000 cm^{-1} has been successfully used on various minerals ((41) to (43)) and on silicate glasses (44) to distinguish OH's from water molecules. The disadvantage of this method is its lower sensitivity as compared to IRS, this has two main consequences ; i) samples ten times thicker (than the ones for conventionnal IRS) are required to record unnoisy spectra. ii) it is not possible to investigate the defects present in the blanks into which commercial devices are machined because their water content is too low. We have thus used wet crystals (refused by the production) and extrapolated to the case of dryer ones.

All the spectra have been recorded with a grating monochromator with an unpolarized beam parallel to the Y direction. The sample labelled QB is an old synthetic quartz grown 20 years ago by Bell Tel.+. Its water content varies in the sample studied here from 600 atomic ppm near the seed to 150 at. ppm far from it. Figure 4 shows two near infrared spectra recorded in these two extreme regions. The peak located at 5200 cm^{-1} is due to water molecules while the one in the vicinity of 4500 cm^{-1} is attributed to point defects X-OH (see (44),(45)). Two distinct types of defects are distinguished on these spectra : clusters of water molecules (no bubbles can be detected by carefull examination on transmission electron micrographs) and point defects. Another information is provided by the spectra : as long as growth proceeds, the crystal incorporates more and more point defects and less and less clusters. These results are in good agreement with those found by Aines et al (45) on very similar samples. We have also examined two more recent crystals grown by SICN (labelled QAI and QA2). Although they are much dryer than the QB one these crystals have been rejected because their water content was considered as prohibitively high : respectively 100 at. ppm and 180 at. ppm (it is to be noted that in these samples, the water content is very homogeneous all along the $\underset{\sim}{c}$ axis, i.e. during the whole growth process). The NIR spectra are represented on figure 5. The molecular water content is somewhat similar into both crystals and is approximately equal to 30 at ppm (see (45) for quantitative analysis of the NIR spectra), while the main part of water is incorporated as point defects. These point defects are probably not interstitial water molecules located for example into the $\underset{\sim}{c}$ channels because such a situation would lead to a strongly anisotropic peak and that is not observed. At the present time we can only notice the occurrence of three distinct peaks in the vicinity of 4500 cm^{-1}, the precise origin of these peaks is completely unknown. One also remarks at 5000 cm^{-1} a peak which has already been detected (45) but has never received any interpretation. The prominent result is that water enters in these quartzes mostly as isolated point defects. Does it reflect a high solubility at the growth conditions or even at the working conditions ? This point has major implications for aging phenomena. It is well known since a number of years that at atmospheric pressure the

---

+ Sample kindly provided by prof. Chaklader, Vancouver.

water solubility is very low. Annealing a wet crystal induces water precipitation and/or bubble coalescence. The mechanisms of water precipitation have been studied in detail by Cordier et al. (32). The main results are briefly summerized below. When annealing at room pressure wet quartzes, water defects precipitate in tiny bubbles the internal pressure of which increases. To allow precipitation to continue, this pressure must be relaxed by nucleating a sessile dislocation loop. At the very beginning, it is more favourable to nucleate a partial dislocation loop enclosing a stacking fault. Then, a second partial is rapidly nucleated in the same plane with the complementary Burgers vector to form the perfect dislocation loop. Figure 6 shows some TEM micrographs characteristic of the water precipitation pattern. It is also possible to estimate the diffusion coefficient D(T) associated to the precipitation process in measuring the mean distance X between precipitates for various annealing conditions. Assuming a random process, one has

$$X/2 = \sqrt{2D(T)t}$$

D(T) has been determined by this method in the temperature range 350-1000°C which extends over both sides of the α-β transition (displacive transition at 573°C at 1 bar). The results fit an Arrhenius law whatever the stability field α or β.

$$D = D_0 \exp - E/RT \text{ with } D_0 = 10^{-12} \text{ m}^2 \text{ s}^{-1} \text{ and } E = 95 \text{ KJ mole}^{-1}$$

Water precipitation requires the transport of two species i) water from the matrix to a bubble or from a bubble to another bubble and ii) SiO_2 from the bubble surface to the extra half plane of the sessile dislocation loop. Assuming that this later process is rapid owing to pipe diffusion in the dislocation core. D(T) must correspond to the bulk diffusion of the water point defects.

This study of precipitation phenomenon shows that, at least at room pressure, wet quartzes are always supersaturated. There is almost no reliable experimental data about water solubility versus T and P. This problem has received some attention recently from a theoretical point of view (46,47).

The predicted solubility versus P and T is represented on figure 7. The general trend is the following. A constant T the solubility increases with pressure from almost zero then it reaches a maximum value then it decreases. The given numerical data may be over or under estimated by one or two orders of magnitude because of the lack of reliable calibration data. We have tried to experimentally determine the solubility at 700°C. The method is based on the observation of water precipitation. Quartzes containing a given and known amount ot water have been annealed at 700°C under various confining pressures. If the inner concentration of water is larger than the solubility at the conditions (700°C,P)

of the experiment, the supersaturated water precipitates in tiny bubbles easily detected by TEM. Figure 8 shows three micrographs corresponding to 3 different annealing conditions made on samples cut in the same region containing 150 ppm of water (this sample is part of the quartz labeled QB). Whatever the pressure (< 1400 MPa) one always observes unambiguous precipitation at 700°C. But, it is clear that precipitation is far less important at 1200 MPa than at room pressure where after only 90 minutes, numerous bubbles with negative crystal shapes are clearly visible on dislocation segments. On the contrary, at 1200 MPa and after 240 minutes, the occurrence of precipitation is only detected by the presence of little white spots in weak-beam conditions. We can conclude from these experiments that at 700°C, the maximum solubility occurs for a pressure of the order of 1200 MPa and the solubility is lower than 150 ppm (probably between 50 and 100 ppm). Quartz growth occurs at a markedly lower temperature and pressure (typically 360°C, 200 MPa). The corresponding solubility can be obtained by extrapolation and it is certainly very low (probably below ten ppm). This is consistent with the fact that very dry quartzes are usually produced by SICN. The important conclusion of this study is that when during growth, quartz incorporates a significant amount of water, this water can be either dissolved thus forming a supersaturated solution of point defects or as tiny fluid inclusions or clusters. Unfortunately we are not able, at the moment, to elucidate the structure of these point defects.

IV - CONCLUSION

Some growth defects are characterized at a microscopic scale in the present study. If fluid inclusions are most often visible at the naked eye in the vicinity of the seed, one reports here the occurrence of tiny bubbles (only detectable at high magnification) restricted to a narrow band corresponding to an incident in the growth process. Among solid inclusions, not only acmite, but also calcium silicates are detected by X-Ray microanalysis. One also finds in z sector and near the final surface, some kind of etched tunnels. X-Ray Lang topography reveals a strong misorientation between two growth sectors (Z+ and Z-) which end on these tunnels. This misorientation is attributed to a preferential location of impurities, probably aluminium, in the boundary. The NIRS study in the second sector shows that water occurs in synthetic quartz crystals either as point defects or as water molecules. The experimental determination of water solubility versus P and T shows that this solubility is very low (below 100 at. ppm at 700°C, 1200 MPa). The water related point defects have thus a supersaturated concentration at the working conditions. A slow aging of the device can thus be expected as long as the concentration of these point defects is not the equilibrium one.

ACKNOWLEDGEMENTS

The authors gratefully thank Dr J. DI PERSIO for the X-Ray Lang Topograms and Dr D. VIGNAUD for the Near Infrared experiments.

REFERENCES

(1) BRICE J.C. (1985) Crystals for quartz resonators. Reviews of Modern Physics 57, 105-146.
(2) HALLIBURTON L.E., MARTIN J.J. and KOEHLER D.R. (1985) Properties of piezoelectric materials in E.A. GERBER and A. BALLATO (Eds) Precision Frequency Control Vol 1 Academic Press p 1-45.
(3) BESSON R.J. (1987) Quartz crystal resonators, Proc 1st Eur. time Freq. Forum p 20-30.
(4) LANG A.R. and MIUSKOW (1967) J. Appl. Phys. 38, 2477.
(6) BARNES R.L., FREELAND P.E., KOLB E.D., LAUDISE R.A. and PATEL J.R. (1978) J. Cryst. growth 43, 676.
(7) CROXALL D.F., CHRISTIE I.R.P., HOLT J.M., ISHERWOOD B.J. and TODD A.G. (1982), 36th Ann. Freq. Control Symp., 62.
(8) ARMINGTON A.F. and LARKIN J.J. (1985) J. Cryst. Growth 71, 799.
(9) CAPELLE B. and ZARKA A. Private communication.
(10) PARPIA D.Y. (1976) Phil. Mag. 33, 715.
(11) PARPIA D.Y. (1976) Phil. Mag. 37, 375 and 401.
(12) YOSHIMURA J. and KOHRA K. (1976) J. Cryst. Growth 33, 311.
(13) ZARKA A., LIN L. and BUISSON X. (1981) J. Cryst. Growth 54, 394.
(14) SPENCER W.J. and HARUTA K. (1966) J. Appl. Phys. 37, 549.
(15) AUVRAY P. and REGRENY A. (1973) Bull Mineral 96, 267.
(16) BARNS R.L., KOLB E.D. and LAUDISE R.A. (1976) J. Cryst. Growth 34, 189.
(17) JOHNSON G.R. and IRVINE R.A. (1987) 41st Ann. Freq. Control Symp., 175.
(18) HOMMA S. and IWATA M. (1973), J. Cryst. Growth 19, 125.
(19) FLICSTEIN J. and SCHIEBER M. (1974) J. Cryst. Growth 24, 603.
(20) CHAKRABORTY D. (1976) J. Cryst. Growth 36, 188 ; (1977) J. Cryst. Growth, 41, 177 and (1978) J. Cryst. Growth 44, 599.
(21) BALASCIO J.F. and ARMINGTON A.F. (1985) 38th Ann. Freq. Control Symp. p 5.
(22) KAHAN A., EULER F.K. and LIPSON H.G. (1987) 41th Ann. Freq. Control Symp. p 216.
(23) LIPSON H.G. and KAHAN A. (1984) 38th Ann. Freq. Control Symp. p 10 ; (1985) J. Appl. Phys. ; 58, 963.
(24) KATS A. and HAVEN Y. (1960) Phys. Chem. Glasses 1, 99.
(25) KATS A. (1962) Philips Res Report 17, 133 and 201.
(26) BRICE J.C. and COLE A.M. (1978) 32nd Ann. Freq. Control Symp. p 1 ; (1979) J. Appl. Phys., 12, 459.
(27) PATERSON M.S. (1982) Bull. Mineral. 105, 20.
(28) AINES R.D., KIRBY S.H. and ROSSMAN G.R. (1984) Phys. Chem. Minerals 11, 204.
(29) AINES R.D. and ROSSMAN G.R. (1984) J. Geophys. Res. 89, 4059.
(30) IWASAKI F. (1977) J. Cryst. Growth 39, 291.
(31) IWASAKI F. and KURASHIGE M. (1978) Jap. J. Appl. Phys. 17, 817 ; (1982) Ferroelectrics 43, 43.
(32) CORDIER P., BOULOGNE B. and DOUKHAN J.C. (1988) Bull. Mineral, in the press.
(33) SER A., BIDEAU J.P., CLASTRE J. and ZARKA A. (1980) J. Appl. Cryst. 13, 50.
(34) CAMERON M., SUENO S., PREWITT C.T. and PAPIKE J.J. (1973) Amer. Mineral. 58, 594.

(35) DODD D.M. and FRASER D.B. (1965) J. Phys. Chem. Solids 26, 673.
(36) MORIYA K. and OGAWA T. (1978) J. Cryst. Growth 44, 53 ; (1982) J. Cryst. Growth 58, 115.
(37) GRIGGS D.T. and BLACIC J.N. (1965) Science 147, 292.
(38) COHEN-ADDAD C., DUCROS P. and BERTAUT E.F. (1967) Acta Cryst. 23, 22a.
(39) NUTALL R.H.D. and WEIL J.A. (1980) Solid State Commun. 33, 99.
(40) HAGON J.P., STONEHAM A.M. and JAROS M. (1987) Philos. Mag B 55 (2), 211.
(41) LANGER K. and FLORKE O.W. (1974) Fortschr. Miner 52(1), 17.
(42) FLORKE O.W., KOHLER-HERBERTZ B., LANGER K. and TONGES I. (1982) Contrib. Mineral. Petrol 80, 324.
(43) HUNT G.R. and SALISBURY J.W. (1970) Modern Geology 1, 283.
(44) STOLPER E. (1982) Contrib. Mineral. Petrol. 81, 1.
(45) AINES R.D., KIRBY S.H. and ROSSMAN G.R. (1984) Phys. Chem. Minerals 11, 204.
(46) DOUKHAN J.C. and PATERSON M.S. (1986) Bull. Mineral. 109, 193.
(47) PATERSON M.S. (1986) Phys. Chem. Minerals 13, 245.

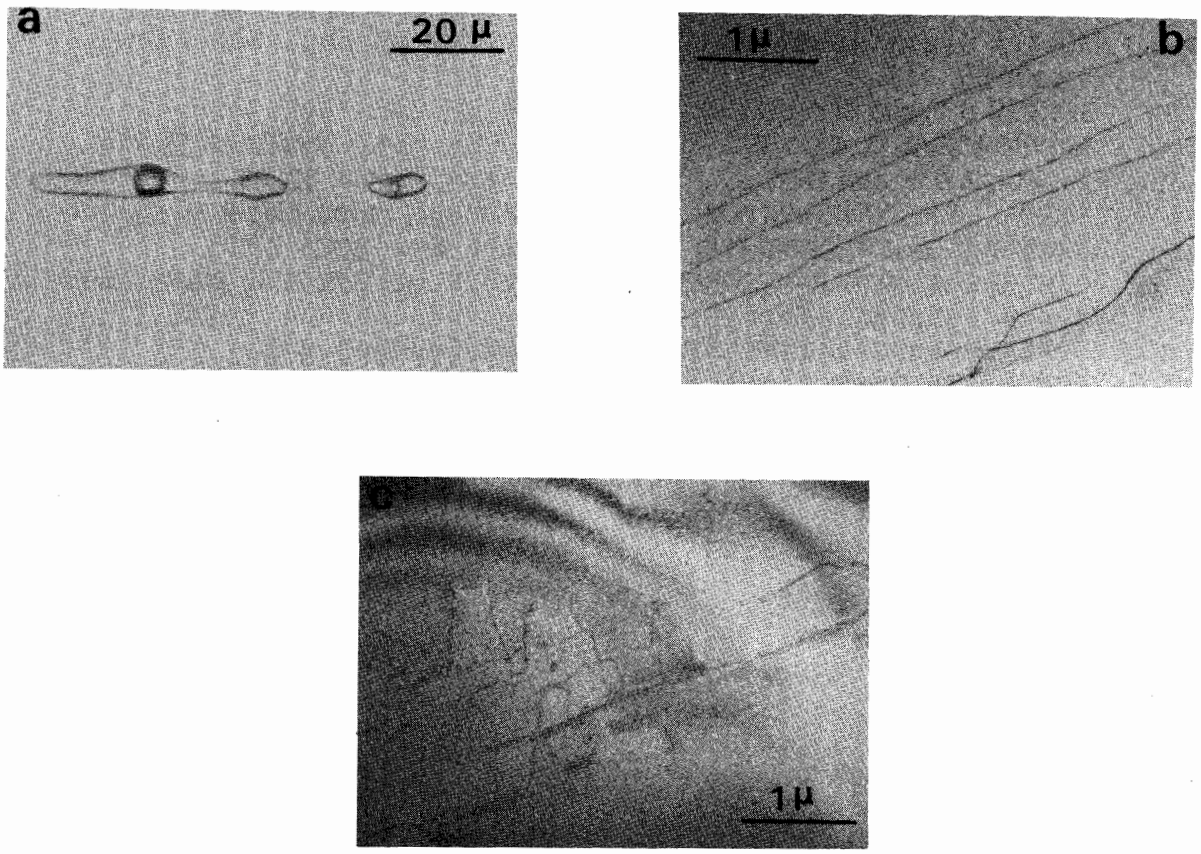


Figure 1
a) Fluid inclusions around the seed ; section observed in optical microscopy.
b) Tiny fluid inclusions and dislocations in quartz grown in a solution enriched in Fe_2O_3, TEM micrograph.
c) Similar defects in a quartz grown in a chemically pure environment but an incident has led to an abrupt increase of the growth rate during a short time. TEM micrograph of the narrow affected zone.

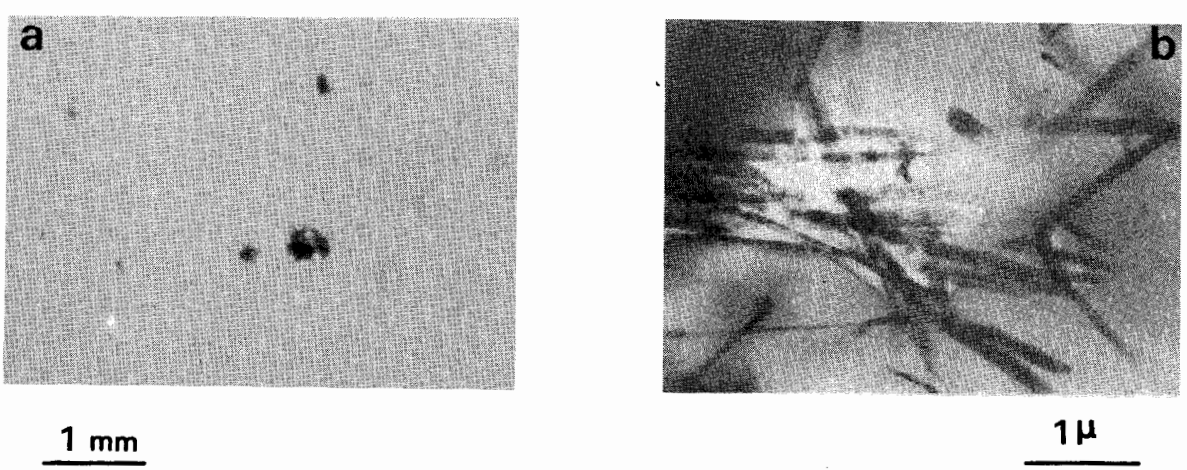


Figure 2
a) Solid inclusion observed at low magnification in optical microscopy
b) Cluster of needle shaped silicate inclusions ; TEM micrograph

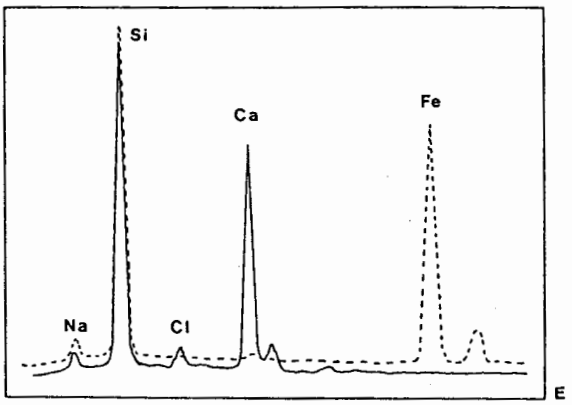
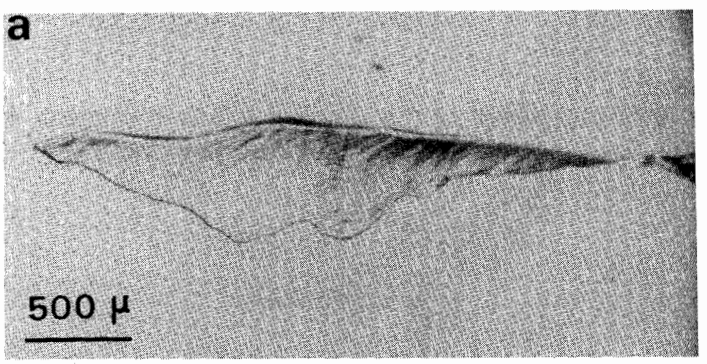


Figure 2
c) EDS microanalyses of some inclusions



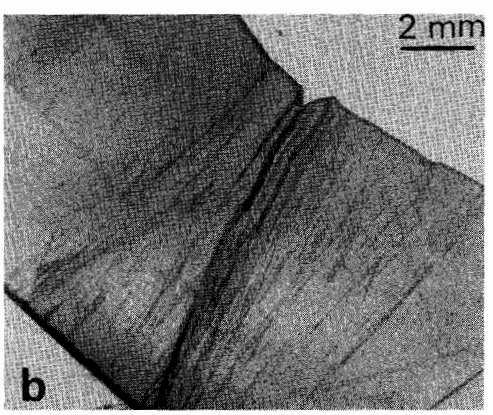


Figure 3
Hole assumed to be etched at the end of the growth process
a) optical micrography
b) X-Ray Lang topography showing the misoriented area.(OITI) reflexion.

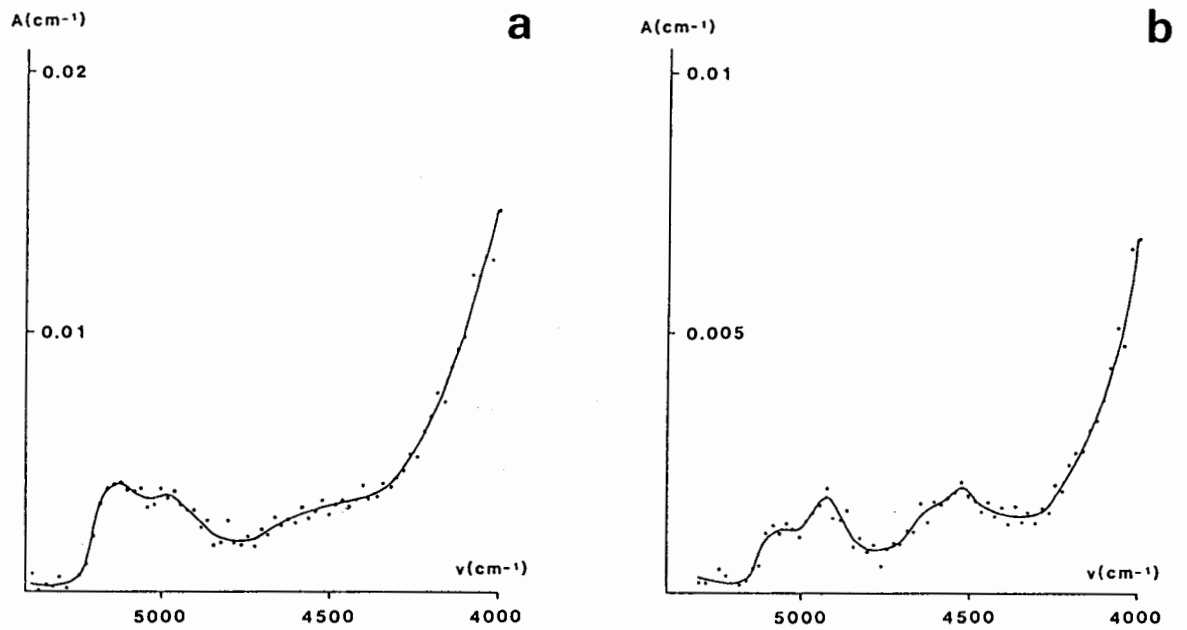


Figure 4
Near infrared spectra of the Bell Tel quartz (thickness 4.5 cm)
a) near the seed (water content 600 at. ppm)
b) end of growth (≅ 150 at. ppm).

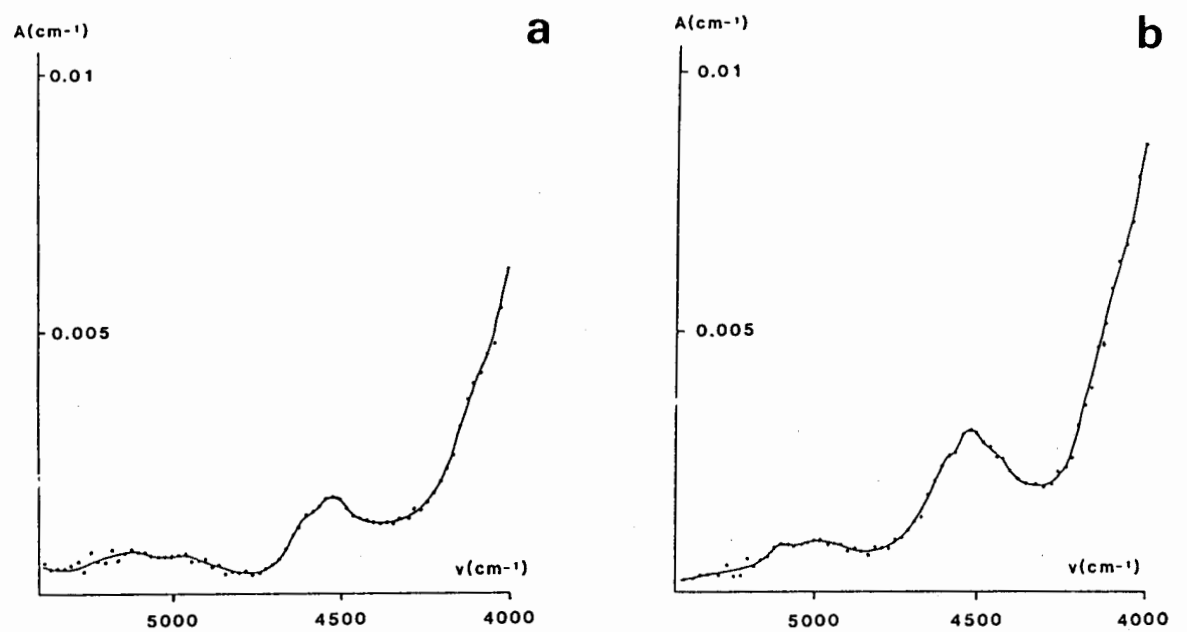


Figure 5
Near infrared spectra of SICN crystals (thickness ≅ 10 cm)
a) QA1 - water content 100 at. ppm
b) QA2 - water content 180 at. ppm.

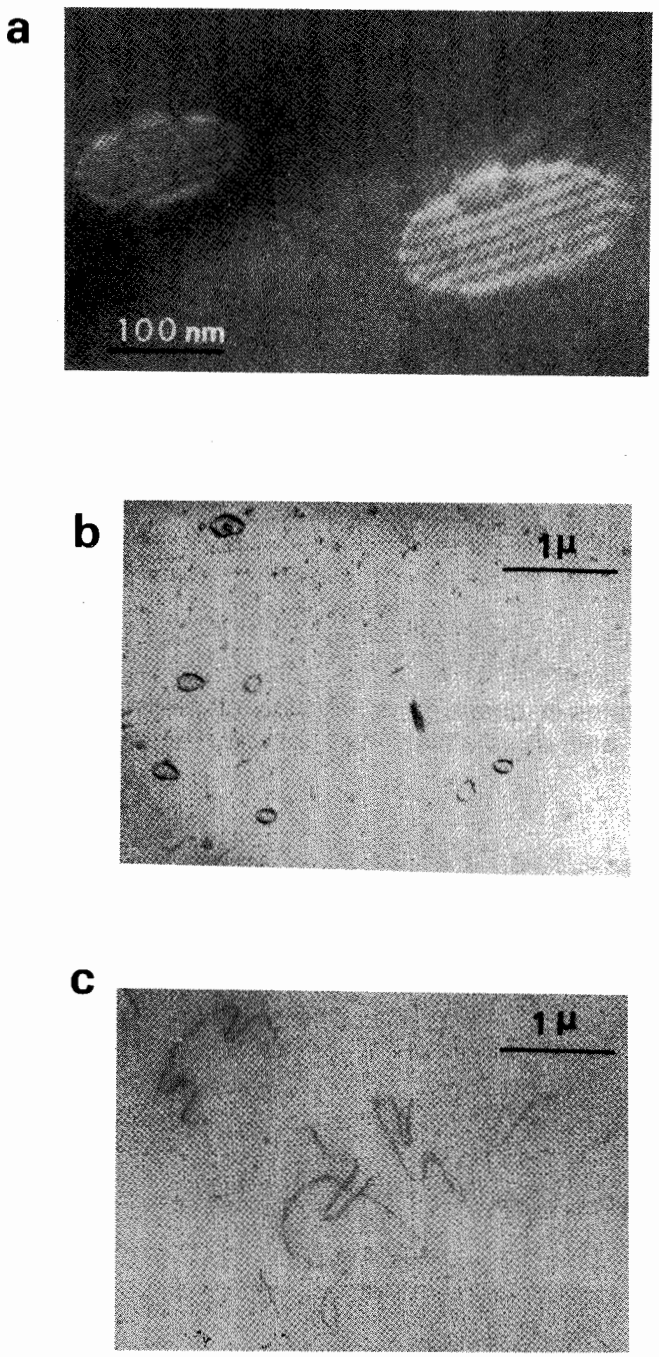


Figure 6
T.E.M. micrographs showing the characteristic features of water precipitation. (0001) thin foil
a) early stage : a partial dislocation loop enclosing a stacking fault is nucleated to relax the inner pressure in the bubble
b) isolated perfect dislocation loops associated to water bubbles
c) third stage - large bubbles and network of dislocations in climb configuration.

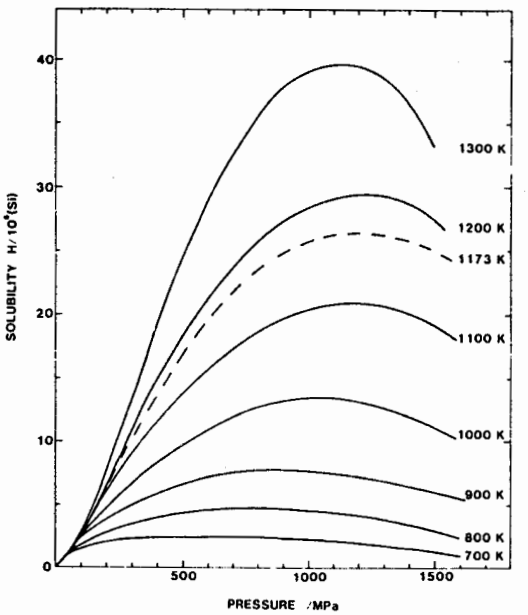


Figure 7
Calculated solubility of water in quartz as a function of pressure at a given temperature (marked on the curves) and at relatively high oxygen fugacity according to Paterson (reference 47).

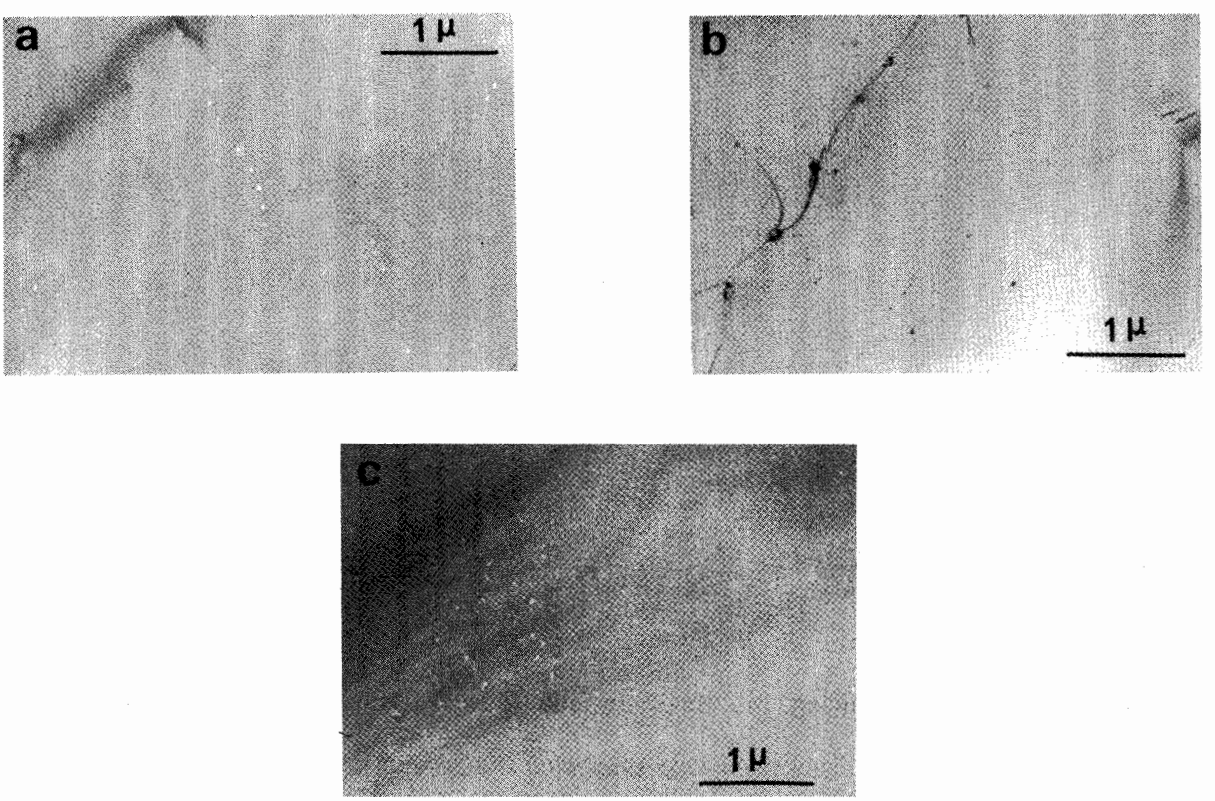


Figure 8
TEM micrographs of a sample containing 150 ppm of water after annealing at various pressures (scale bar = 1 micron)
a) after 1 h 30 at 700°C - room pressure - Bright field
b) after 4 h at 700°C - 300 MPa - Bright field
c) after 4 h at 700°C - 1200 MPa - Dark field

2 ND EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 16 - 17 - 18 MARCH 1988

Session MP-02 Number 4

THE ORIENTATION DEPENDENCE OF THE DISSOLUTION RATE OF SINGLY ROTATED QUARTZ PLATES DERIVED FROM COMBINED NUMERICAL AND DIGITAL TECHNIQUES

C.R. TELLIER, P. BLIND and D. JOZWICK

Laboratoire de Chronométrie, Electronique et Piézoélectricité
ECOLE NATIONALE SUPERIEURE DE MECANIQUE ET DES MICROTECHNIQUES
Route de Gray - La Bouloie - 25030 BESANCON CEDEX - FRANCE
Tél : 81.50.36.55

ABSTRACT

A new method for the graphical simulation of the dissolution profiles of quartz surfaces is proposed. This extended method which uses a polar diagram representation for the dissolution slowness and new equations for the trajectory of a moving surface profile element is applied to randomized initial surface profiles. Starting from a polar diagram which represents the experimental variations of the dissolution slowness with the orientation, θ_o, of singly rotated quartz plates a close agreement is observed between the forms of the Z' etch profiles as revealed by repeated etchings in $NH_4F.HF$ solutions and the predictions of the numerical simulation. Digital techniques for the description of surfaces profiles which can be applied either to a real surface profile of etched quartz crystals or to a numerical etch profile derived from the simulation method are presented. This description is achieved using a microcomputer and the program offers the possibility to characterize the surface roughness and to extract the distribution of slopes, the distribution of peak heights and the correlation length of surface profiles. In particular, attention is concentrated on the experimental and theoretical changes in the distribution of slopes with the etching time, t, for various orientations, θ_o, and for various polar diagrams. The theoretical evolutions of the distributions of slopes are discussed in terms of the form of relatively complex polar diagrams of the dissolution slowness ; precise correlations are thus established. Comparing some theoretical and experimental results, corrections in some regions of the experimental curve illustrating the variations of the dissolution rate of singly rotated quartz plates with the orientation θ are suggested. In conclusion, it appears that combining the graphical simulation with these digital techniques constitutes an efficient "help" procedure for a precise determination of the experimental dissolution rate versus orientation curves for quartz crystals.

1 - INTRODUCTION

Previous studies **(1-6)** have been made of the influence of the crystal orientation on the final surface texture formed on differently oriented quartz plates after prolonged etchings in a concentrated $NH_4.HF$ solution. They have given evidence for a marked correlation of the shape and of the stability of the etch patterns with the orientation, α, of the crystal plane **(2-6)**. We have recently developed a two dimensional model **(7)**, based on the Frank model **(8)** for a numerical simulation of the etched profiles. The simulation procedure uses analytical relations $v_N(\alpha)$ for the variation of the dissolution rate, v_N, with the orientation α of surface profile elements and thus presents the disadvantage to be concerned with a lot of relationships $v_N(\alpha)$ whose ranges of applicability remain small.

For these reasons, the first part of this paper is devoted to the presentation of a new method **(9)** for the graphical simulation of the dissolution profiles which consists of combining theoretical equations of the Frank dissolution model with a geometrical construction of the trajectories of moving profile elements. The main advantage of this method is to use a polar diagram representation for the dissolution slowness (i.e. for the reciprocal of the normal velocity v_N) which permits firstly to extend the simulation to large variations in the orientation α and to treat surfaces of any reference orientation, θ_o.

2 - NEW EQUATIONS FOR THE GRAPHICAL SIMULATION

Starting from the kinematic model of dissolution proposed by Frank **(8)** several years ago we suppose that in a two dimensions space (x, y) a surface profile $y = f(x, t)$ can be described by a succession of linear profile elements $\Delta r(\alpha)$ (Figure 1). The velocity, v_y, of a profile element of orientation α , or slope p, along the inward normal, n_o, at the reference surface of orientation, θ_o, is defined as

$$v_y = \left(\frac{\partial y}{\partial t}\right)_x \qquad (1)$$

at a fixed value of x.

This model is based on the assumption that the velocity, v_y, depends only on the orientation α (i.e. on the slope p) ; an hypothesis which is sustained by a great number of experimental results in the case of chemically etched quartz surfaces **(1-6, 10-13)**.

It follows that a profile element with a fixed slope $p = (\partial y/\partial x)_t$ propagates into the bulk crystal along a linear trajectory called a "characteristic" whose slope is expressed **(8, 9)** as

$$\frac{dy}{dx} = p - v_y \cdot \left\{ \frac{\partial v_y}{\partial p} \right\}^{-1} \qquad (2)$$

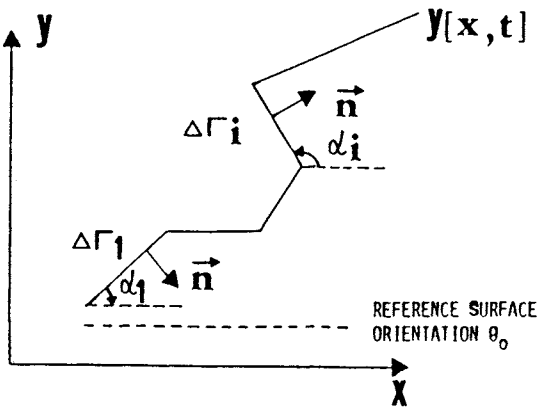


Figure 1 : Geometry of the surface profile

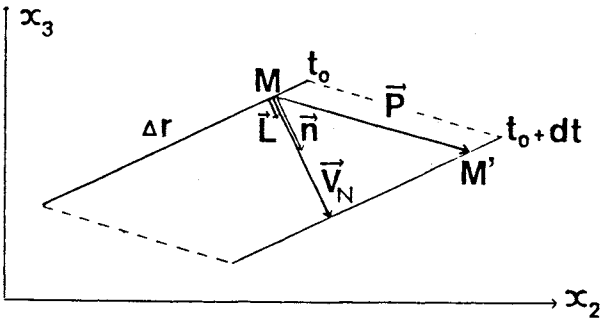


Figure 2 : Definition of the dissolution slowness vector $\vec{L}$

In these conditions after a chemical attack of time dt, we can associate with a profile element of slope p, a propagation vecteur $\vec{P}$ whose components in the coordinates axes (x,y) are readily identified with dx and dy. In a recent paper **(9)** we have shown that we can determine completely the propagation vector $\vec{P}$ by considering the properties of the dissolution slowness vector, $\vec{L}$, whose magnitude is the reciprocal, $1/v_N$, of the dissolution rate, v_N, measured normal to a surface profile element and whose direction coïncides with that of the normal unit vector $\vec{n}$ (Figure 2). In a bi-dimensional model **(9)** the extremity of the slowness vector, $\vec{L}$, moves along a path which corresponds to its polar diagram. Since in a polar representation of $\vec{L}$, the components L_x and L_y can be written as

$$L_x = F(\beta) \cos \beta \quad (3)$$

$$L_y = F(\beta) \sin \beta \quad (4)$$

with $L^2 = \vec{L} \cdot \vec{L} = \{F(\beta)\}^2$ (5)

it appears that the problem is now to express dx and dy in terms of the function $F(\beta)$ and the polar angle, β.

After some geometrical calculations **(9)** we obtain the following set of equations

$$dx = \pm B(\beta) \{F(\beta)\}^{-2} . dt \quad (6)$$

$$dy = \mp A(\beta) \{F(\beta)\}^{-2} . dt \quad (7)$$

since

$$\frac{\partial v_y}{\partial p} = \mp B(\beta) \{F(\beta)\}^{-2} \quad (8)$$

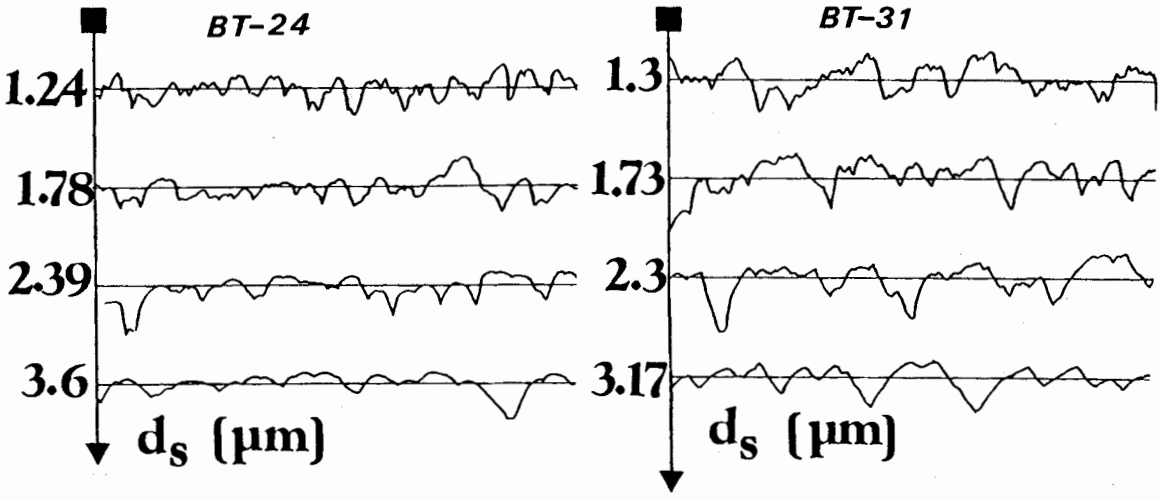
Here, $A(\beta)$, $B(\beta)$ are the components of a vector, $\vec{\mathscr{C}}$, which is the derivative of the rotating vector, $\vec{L}$, with respect to the polar angle β. The choice between the "plus" and "minus" signs results from the value of the angle ϕ that the vector $\vec{\mathscr{C}}$ makes with the vector $\vec{L}$.

Thus we can now track the displacement of any moving profile element within the bulk crystal and we can follow the changes in shape of a surface profile provided the polar diagram of the slowness vector $\vec{L}$ be known.

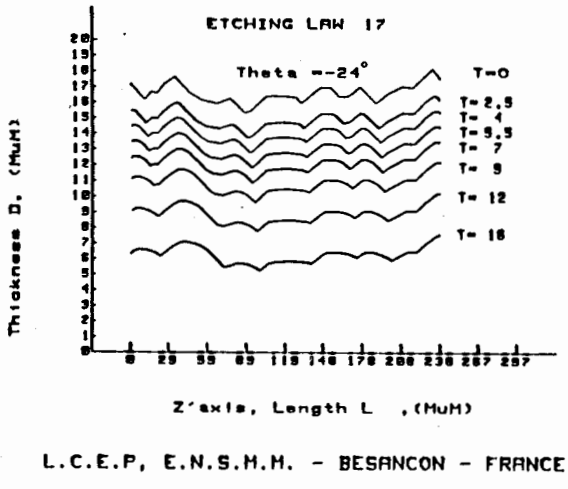
3 - THE GRAPHICAL SIMULATION

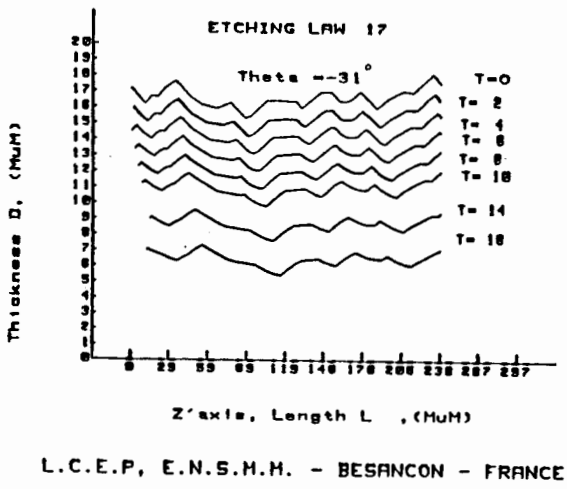
The numerical simulation which has been described elsewhere **(7,9)** is carried out using the new equations derived in the above section and is applied to surface profiles with an initial randomized shape. Starting from a polar diagram which represents crudely the experimental variations of the dissolution slowness with the orientation, θ_o, of singly-rotated quartz plates a close agreement is observed between the forms of the Z' etch profiles of quartz crystals as revealed by repeated etchings in $NH_4F.HF$ solutions and the predictions of the numerical simulation. Here, we have chosen for illustrating this agreement to examine the results concerning two particular cuts : the cuts with an angle, θ_o, of - 24° and - 31° called respectively the BT-24 and BT-31 cuts.

In figures 3 and 4 are shown the experimental changes in the shape of the Z' profiles related respectively to the BT-24 and BT-31 plates with the depth, Δd_s, of etch. A stable convex background structure develops on the BT-24 plate with prolonged etching. Repeated etchings on



Figures 3 and 4 : Changes in the shape of the Z' profilometry traces with the depth of etch of respectively a BT-24 and a BT-31 plate.





Figures 5 and 6 : The theoretical etched profiles for a reference surface of respective orientation $\theta_o = -24°$ and $\theta_o = -31°$.

the BT-31 plate cause the formation of Z' dissolution profiles of more complex shape : the Z' profiles are effectively composed of a succession of curved element alternatively convex and concave ; this type of pattern is said of alternate shape. It becomes evident by comparison of the experimental etch profiles of Figures 3 and 4 with the respective theoretical Z' dissolution profiles of figures 5 and 6 derived from the graphical simulation that the proposed polar diagram certainly represents accurately the dissolution slowness polar parth for orientations, θ_o, in the range (- 46°, - 9°). Thus, it may be of interest to illustrate (figures 7 and 8) the "etch-rate-orientation" relationships which give rise to these two different surface profile features. Figure 7 agrees with the qualitative predictions made by Irving **(14)** several years ago. The alternate shape of the dissolution profiles seems to be correlated with the etch rate distribution shown in figure 8 where the etch rate passes successively through a minimum (for negative slopes) and through a maximum (for positive slopes). An alternate surface profile results effectively from such a typical asymmetric etch rate distribution **(15)**.

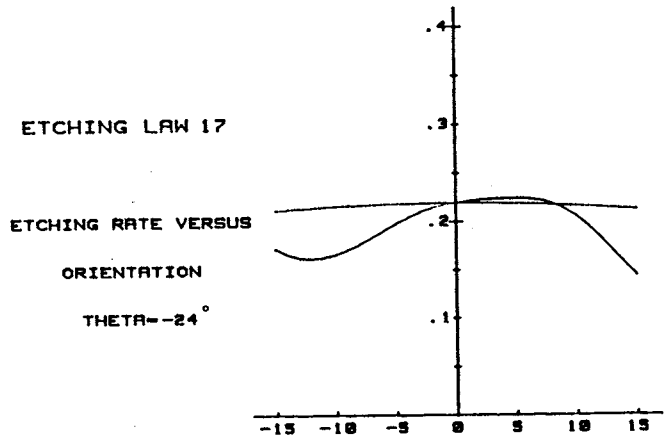


Figure 7 : The etching rate versus orientation plot for a reference orientation, θ_o, of - 24°

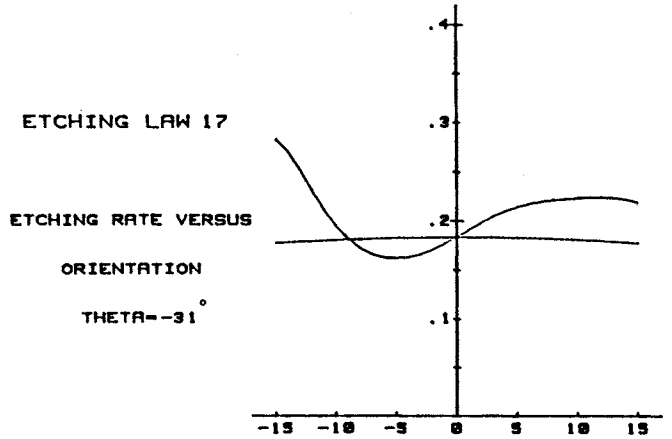


Figure 8 : The etching rate versus orientation plot for a reference orientation, θ_o, of - 31°.

4 - DIGITAL TECHNIQUES FOR THE DESCRIPTION OF SURFACE PROFILES

4.1 - A statistical approach of the surface profile

In this section we are concerned with the development of digital techniques for the description of surface profiles which can be applied either to a real surface profile of etched quartz crystals or to a numerical etch profile theoretical profile derived from the preceding simulation method. This description is achieved using a microcomputer. The real surface profile is split into sampling lengths in such a way that in the first step we can opt for a method of graphical filtering by length of sample. The roughness characterization of the modified surface profile is performed in a second step after a correct meter cut off has been selected. In particular two widely used statistical roughness parameters such as the centre line average, Ra

$$Ra = \frac{1}{L} \int_0^L |z| \, dx \qquad (9)$$

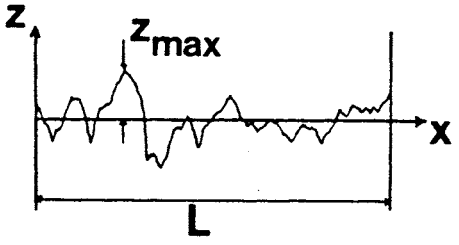


Figure 9 : Definition of statistical roughness parameters.

and the root mean square, Rq, defined (Figure 9) as

$$Rq = \left(\frac{1}{L} \int_0^L z^2 \, dx\right)^{1/2} \qquad (10)$$

are evaluated. Moreover the program offers the possibility to extract the distribution of slopes from the real surface profile and the distribution of peak heights from the modified profile emerging from filter. Since obviously two surfaces with the same roughness parameters can exhibit very different spatial arrangements of surface heights the program includes also the study of the autocorrelation function **(16)** in order to obtain informations on the evolution in the spatial distribution of peak heights with the orientation, θ_ρ, of the reference surface. It should be noticed that the processing of data involves the determination of the reference line. Here the centre line is positioned by either a least-squares fit (first order, method RMS1) or the method of "sampling lengths" (method RMS3) **(16)**. A second filtering where the long-wavelength cut off, λ_c, of the filter is made shorter than the sampling length, L_s, is then performed. To evaluate the roughness parameters of the etched quartz surfaces a sampling length, L_s, of **80** µm and a meter cut off of

10 µm have been selected. But to quote the openness or the closeness of a real surface profile a longer meter cut off is required. A great number of testings on surface profiles of various shape and spatial arrangements have given evidence that a meter cut off of 0.1 µm is appropriate.

4.2 - Experimental results

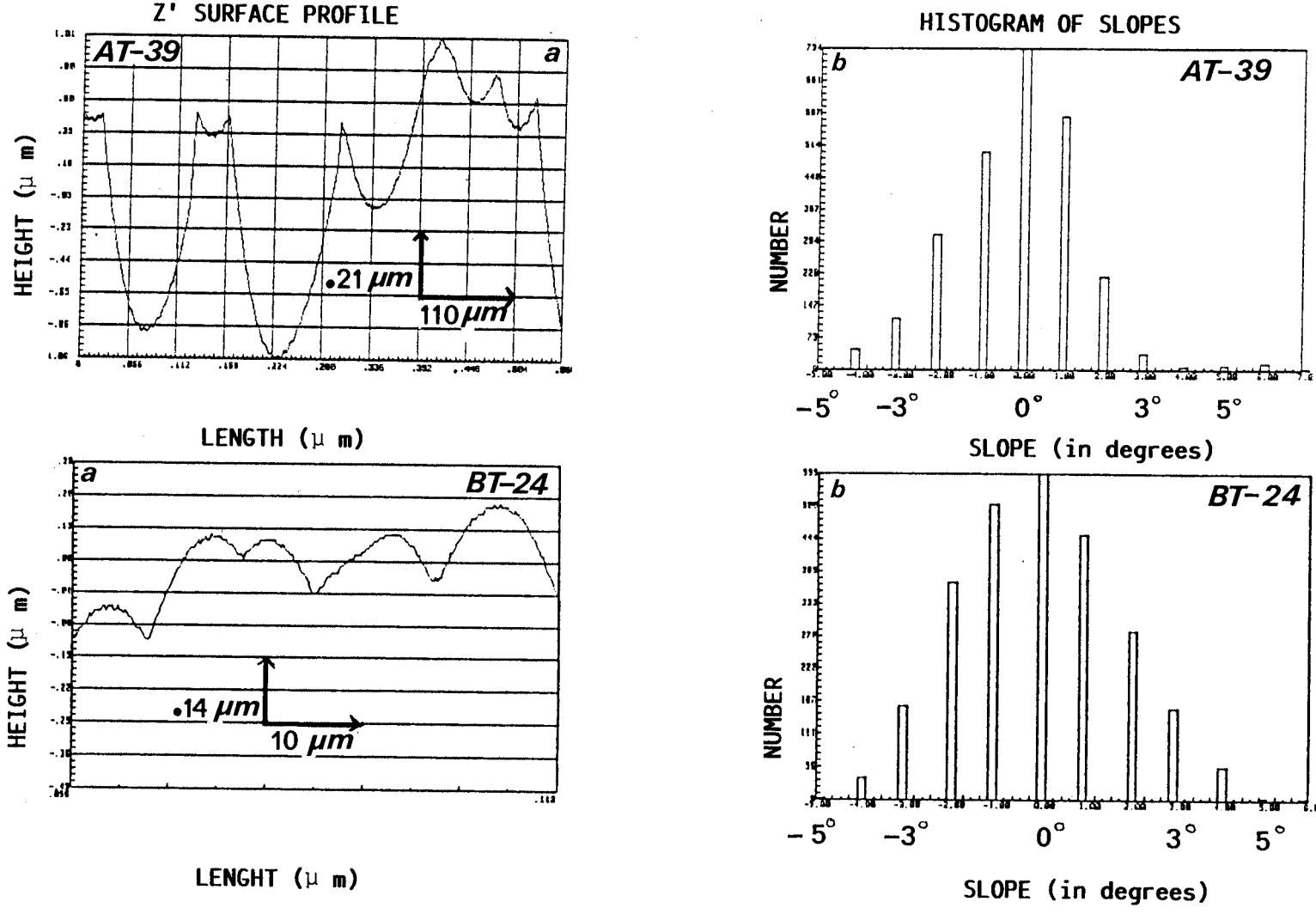
To illustrate these procedures of numerical assessment of the profile topography some results related to deeply etched quartz plates of various orientations are displayed in Figures 10 to 11. The AT-39 plate exhibits (Figure 10a) a stable concave background structure. Magnifications of the Z' dissolution profiles of BT-24, BT-31 and BT-46 plates reveal that

(i) a stable convex background structure and a profile of alternate shape develop respectively on etched BT-24 (Figure 11a) and BT-31 (Figure 12a) surfaces in close agreement with the theoretical results presented in the above section.

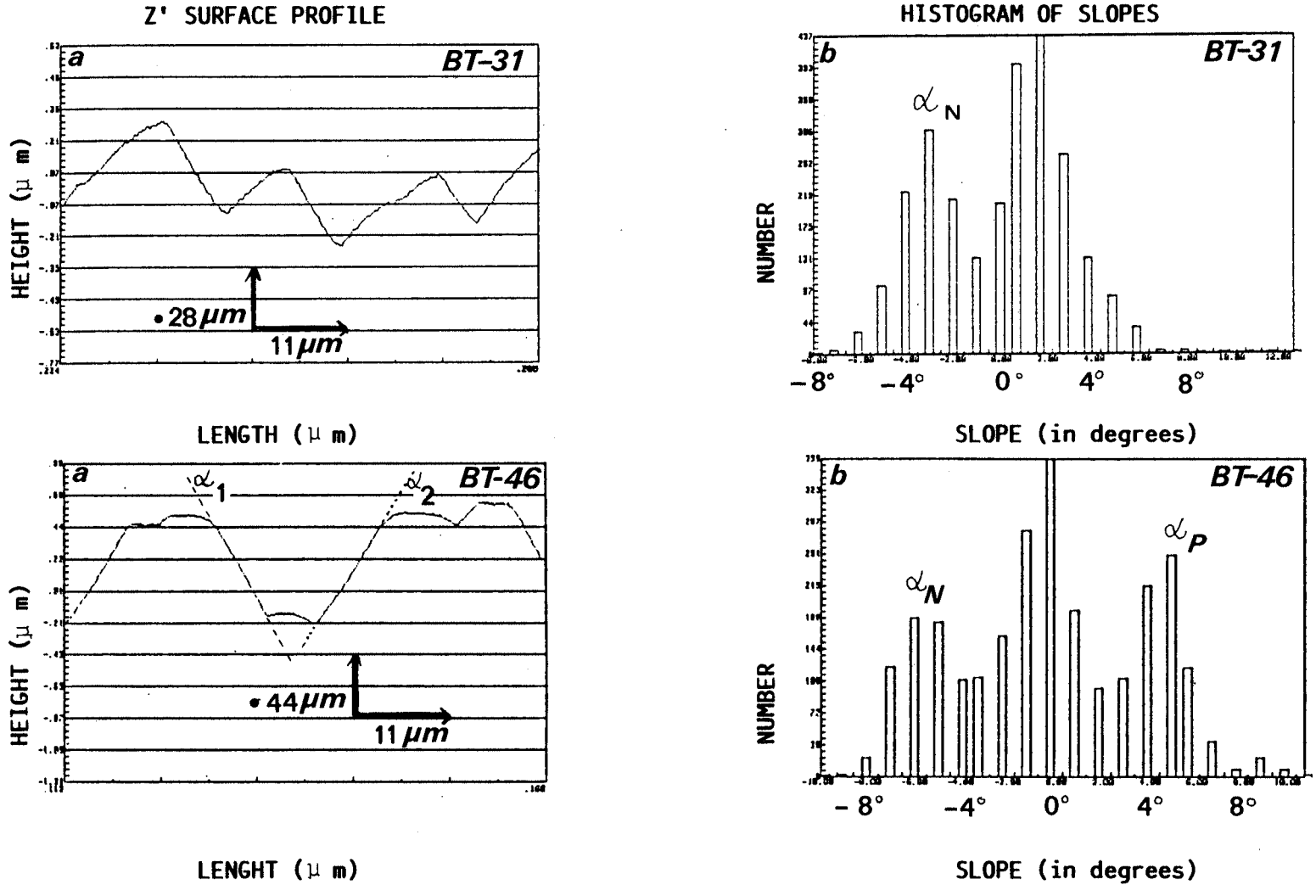
(ii) the Z' convex section of BT-46 plate shows two planes whose orientations denoted α_1 and α_2 (Figure 13a) remain unchanged along the Z' direction.

The histograms of profile slopes related to the stable background structures of AT-39 (Figure 10b) and BT-24 (Figure 11b) plates seem to be of Gaussian form since the distributions are peaky for an orientation which corresponds to the orientation of the reference surface. The histograms related to the BT-31 plate (Figure 12b) and the BT-46 plate (Figure 13b) clearly show a symmetrical distributions : the histogram related to the Z' profile of BT-31 plate crudely corresponds to a form that the distribution takes when negatively skewed but we remark that a second peak occurs at $\alpha_N \simeq -3.5°$. Moreover the form typified by Figure 13b cannot be simply qualified of platykurtic distribution : in reality two secondary peaks can be clearly distinguished at angles $\alpha_N \simeq -6°$ and $\alpha_P \simeq 5°$. Turning to Figure 13a it appears that the angles α_N and α_P can be identified with the respective angles α_1 and α_2. We thus observe that the distribution of slopes of the real profile is evidently characteristic to the shape of the etched Z' profile. In the case of asymmetric profiles the skewness of the distribution indicates the extent of the asymmetry but also gives some precious indications on slopes which remain virtually unaltered on etching.

Typical values of the roughness parameters Ra and Rq are reported in table 1 : as expected for a short meter cut off the roughness parameters decreases with the extent in the spatial arrangement of the profile.



Figures 10 and 11 : Characterization of the etched AT-39 and BT-24 plates.



Figures 12 and 13 : Characterization of the etched BT-31 and BT-46 plates.

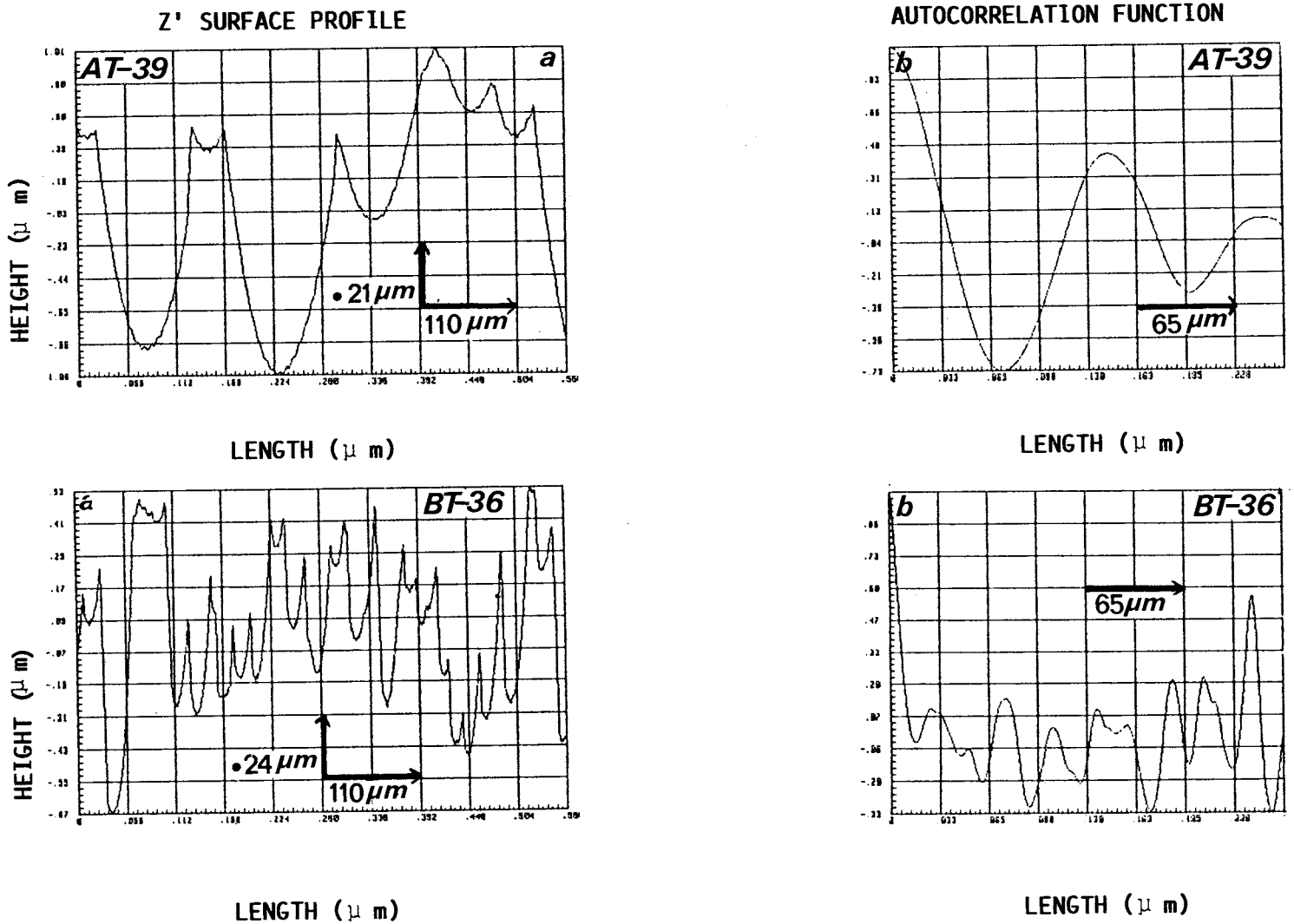
Orientation θ_o	39°	- 24°	- 31°	- 46°
Method RMS1	Ra = 0.030 µm Rq = 0.042 µm	Ra = 0.052 µm Rq = 0.065 µm	Ra = 0.085 µm Rq = 0.112 µm	Ra = 0.137 µm Rq = 0.171 µm
Method RMS3	Ra = 0.040 µm Rq = 0.068 µm	Ra = 0.055 µm Rq = 0.068 µm	Ra = 0.089 µm Rq = 0.118 µm	Ra = 0.138 µm Rq = 0.174 µm

Table 1 : Values of the roughness parameters for the Z' profiles displayed in Figures 10a, 11a, 12a and 13a.

To show that the autocorrelation function (ACF) contains informations on the spatial properties of the Z' profiles we have chosen to study two Z' profiles of similar shape (Figures 14a and 15a) but of different spatial distributions. Figures 14b and 15b show the ACF representation of these two Z' profiles : we remark that the ACF cannot be fitted by a simple exponential form **(16)**. But in fact this exponential form has been derived to fit the ACF of many random surfaces for which one can expect to discriminate different spatial structures by the decay property of the exponential ACF **(16)**. However examination of Figures 14 and 15 reveals firstly that the decay rate at the orign remains very sensitive to the spatial arrangement along the Z' direction and secondly, that the values of the correlation length, ℓ_c, reported in table 2 are in qualitative agreement with the average values of the waviness spacing, ℓ_w. Thus the ACF method can be used as a way of representing spatial variations.

Orientation θ_o	39°	- 36°
Correlation length ℓ_c	120 µm	27 µm
Waviness spacing ℓ_w	125 µm	30 µm

Table 2 : Values of the correlation length ℓ_c as estimated from the ACF plots of figures 14b and 15b. For comparison average values of the waviness spacing, ℓ_w, as evaluated from the Z' profiles represented in Figures 14a and 15a are also indicated in the table.



Figures 14 and 15 : The spatial arrangment of the etched AT-39 and BT-36 plates.

5 - TREATMENT OF THEORETICAL PROFILES

Here, among the different possibilities of characterization of a surface texture we have chosen to concentrate our attention on the distribution of slopes. In particular we follow theoretically the changes in the distribution of slope with the etching time, t, for various reference orientation, θ_o, and for various etching laws $v_N(\alpha)$ when the chemical attack concerns an initial profile of triangular shape with $\alpha_o = \pm 15°$. The evolution of the slope histograms with the etching time t (i.e. the depth of etch Δd_S) are displayed in Figures 16, 17 and 18 for three different etching laws (Figures 19a, 20a and 21a) whereas the changes in the shape of the surface profile with the depth of etch are depicted in Figures 19b, 20b and 21b. These selected results merit some comments :

① **Etching law 16 and reference surface of orientation $\theta_o = 44°$:** At relatively small depth of etch ($\Delta d_S \simeq 0.75$ µm) two secondary peaks situated at $\alpha_N = -8°$ and $\alpha_P = 7°$ are initiated. These peaks develop with further etching but as deeper depths of etch are reached (Figure 16d) slopes of small absolute value become favoured. These successive histograms can be readily interpreted in terms of the $v_N(\alpha)$ relationship depicted in Figure 19a. Elements of profile with large positive or negative slopes which possess relatively high dissolution rate disappear rapidly whereas elements with slope close to $\alpha_1 \simeq -8°$ and $\alpha_2 \simeq 7°$ of lowest etch rate predominate during a long etching time. As soon as slopes corresponding to $\alpha > \alpha_2$ and $\alpha < -\alpha_1$ have been removed we are concerned with a simple etch rate vs orientation plot characterized by a single maximum occuring at $\alpha = 0°$ which is known **(14)** to give rise to a convex background structure. As this maximum is sharp flat convex sections are rapidly formed. But the major fact emerging from this discussion is that $\alpha_1 \simeq \alpha_N$ and $\alpha_2 \simeq \alpha_P$. Hence for this type of etching law the distributions of slope allow us to identify the exact positions of the two minima.

② **Etching law 16c and reference surface of orientation $\theta_o = -1°$:** Figures 17 and 20b indicate what happens when we consider the converse case (Figure 20a). It should be pointed out that the two minima α_1^* and α_2^* (Figure 20a) have no influence on the shape of the etched profile. This can be understood in terms of planes which dissolve less rapidly so that the shape of the profiles will be primarely determined by the features of the $v_N(\alpha)$ law for planes which lie in the vicinity of the reference surface. For this reason we observe (Figure 20b) that a quasi convex section which is formed by two planes of orientation $\alpha_P \simeq 5°$ and $\alpha_N \simeq -6°$ (Figures 17a and 17b) develops in the first step. It seems (Figures 17c, 17d and 20b) that stable hillocks finally result and that no amount of etching will produce a surface without them.

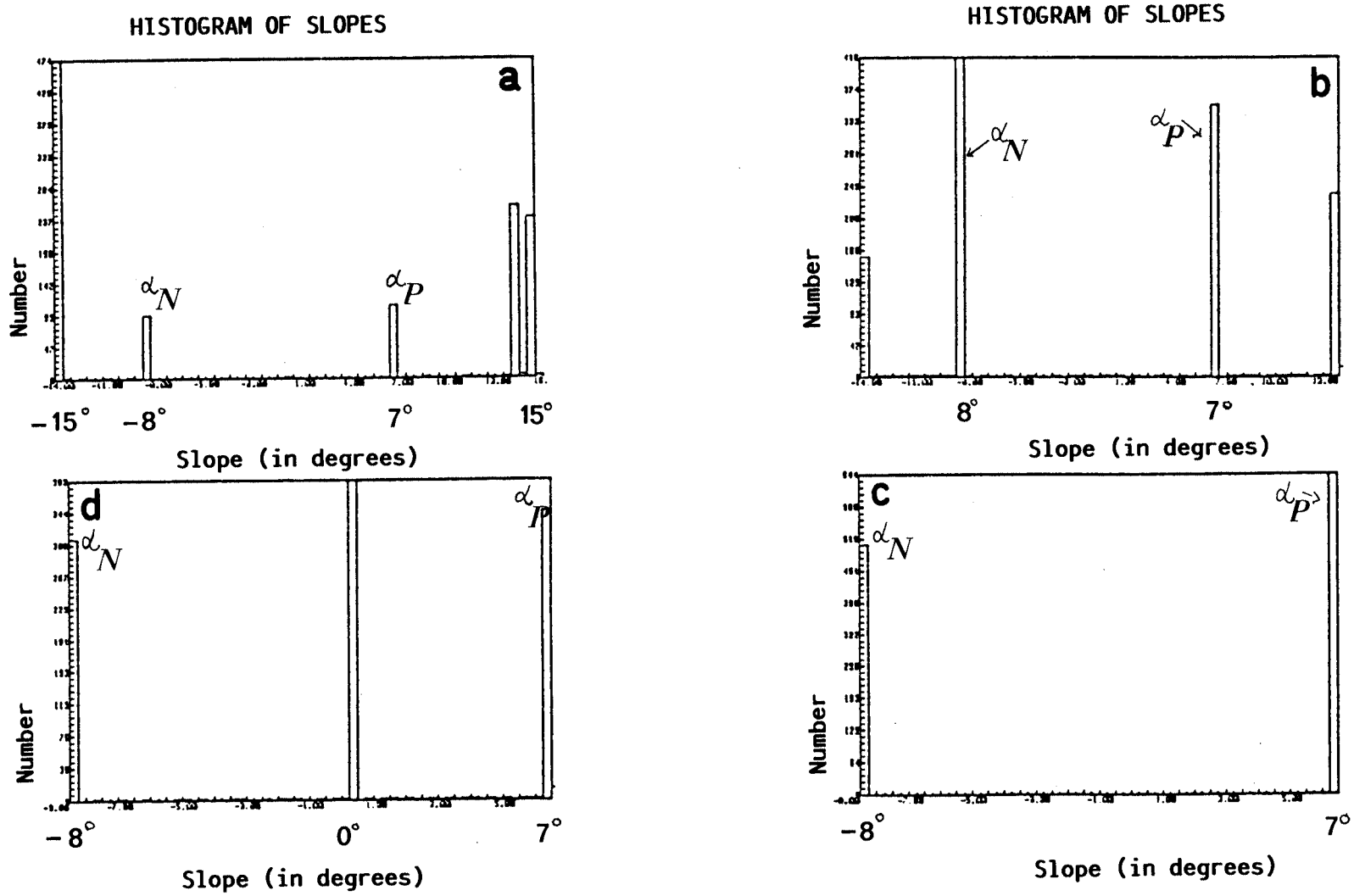


Figure 16 : Evolution of the slope histograms with the depth of etch (a) : $\Delta d_s = 0.75 \mu m$, (b) : $\Delta d_s = 2.5 \mu m$, (c) : $\Delta d_s = 4.2 \mu m$, (d) : $\Delta d_s = 5.9 \mu m$.

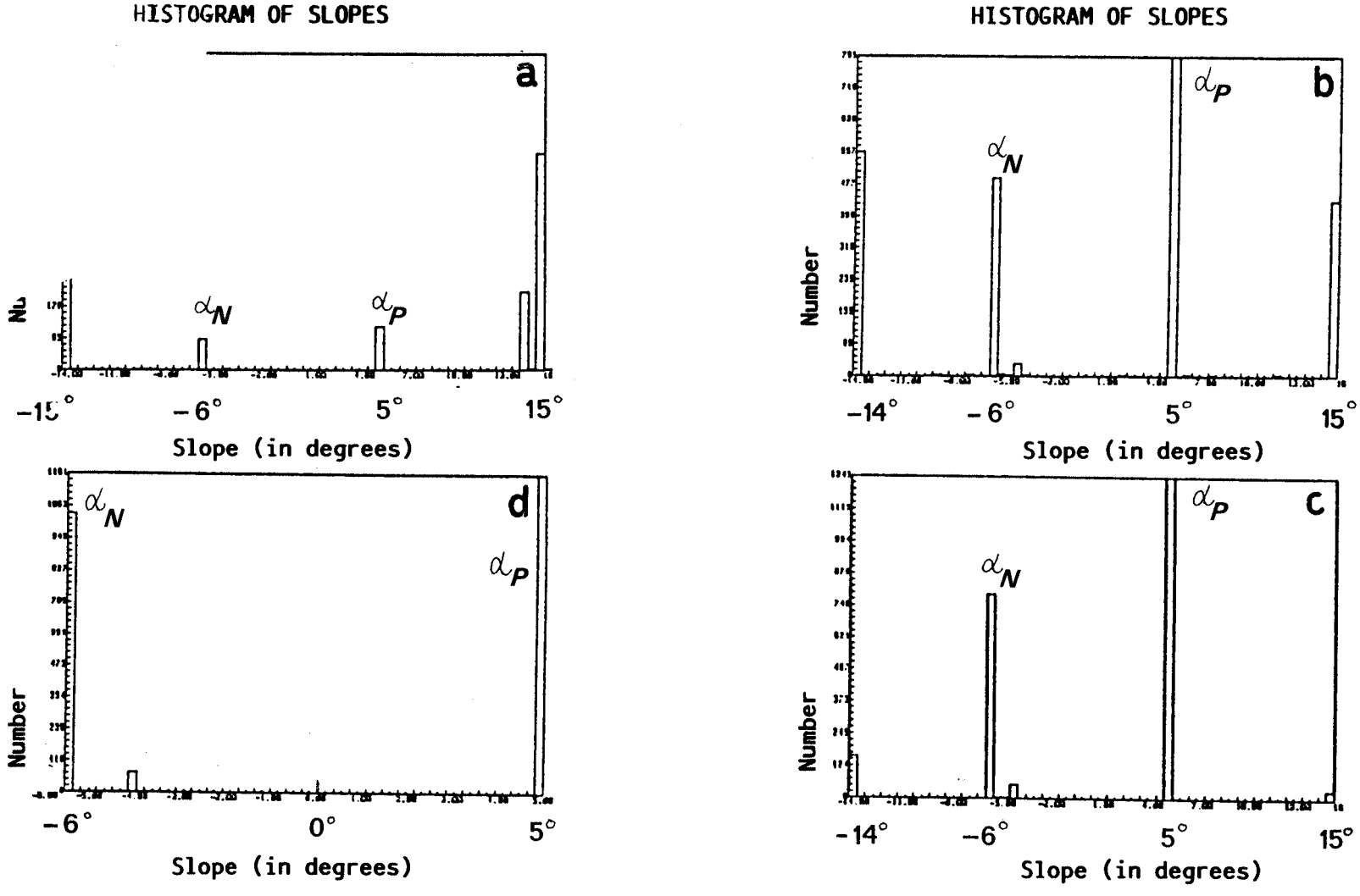


Figure 17 : Evolution of the slope histograms with the depth of etch (a) : Δd_S = 0.73 µm, (b) : Δd_S = 2.6 µm, (c) : Δd_S = 6.4 µm, (d) : Δd_S = 7.9 µm.

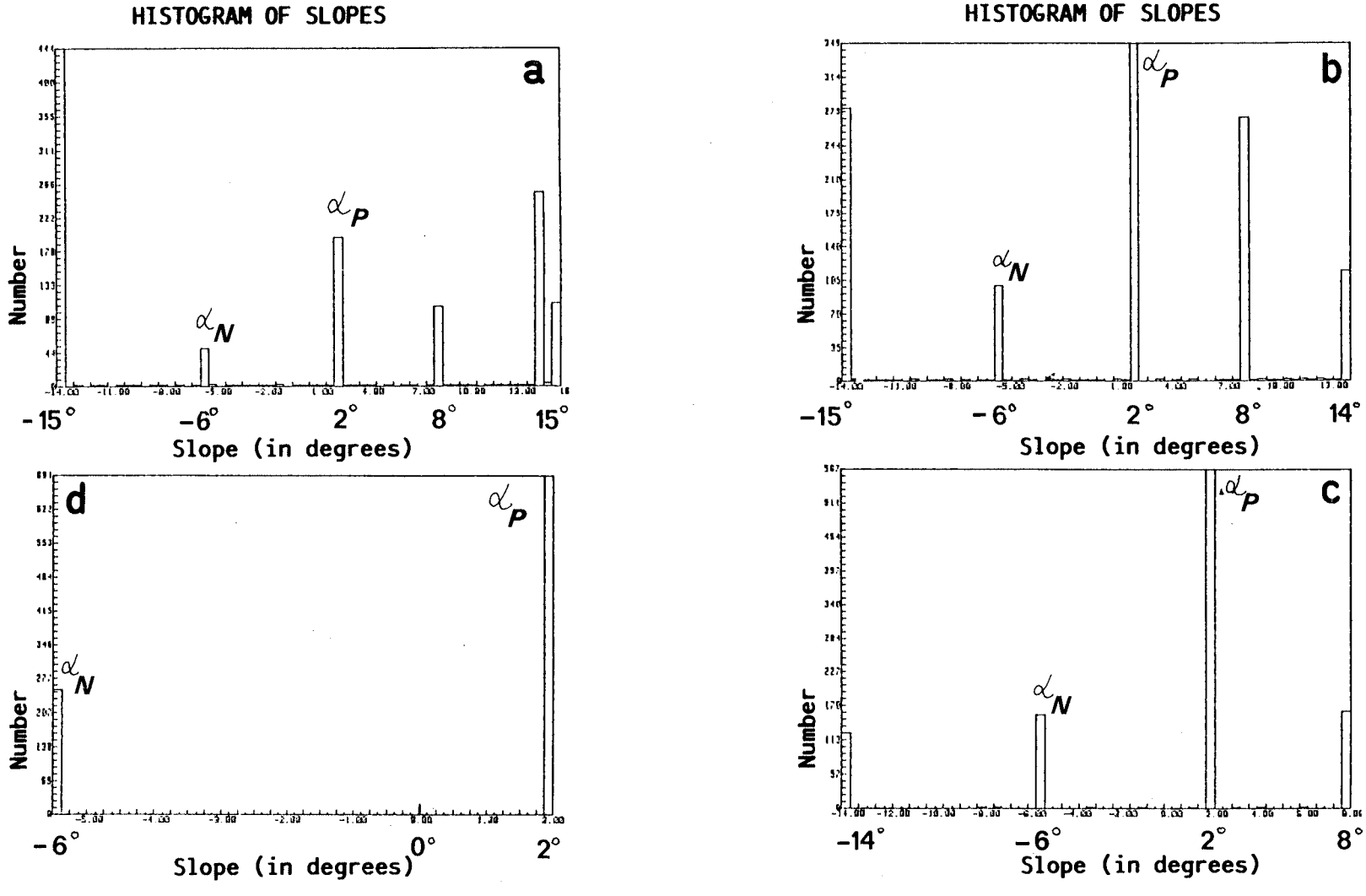


Fig 18 : Evolution of the slope histograms with the depth of etch (a) : Δd_S = 0.94 µm, (b) :Δd_S = 2.2 µm, (c) :Δd_S = 3.7µ m, (d) :Δd_S = 7 µm.

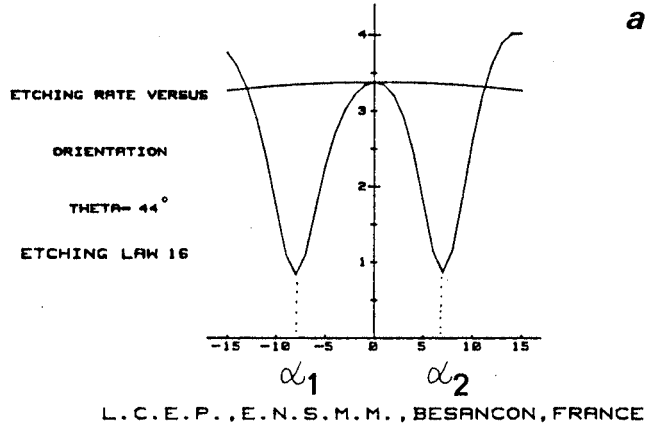
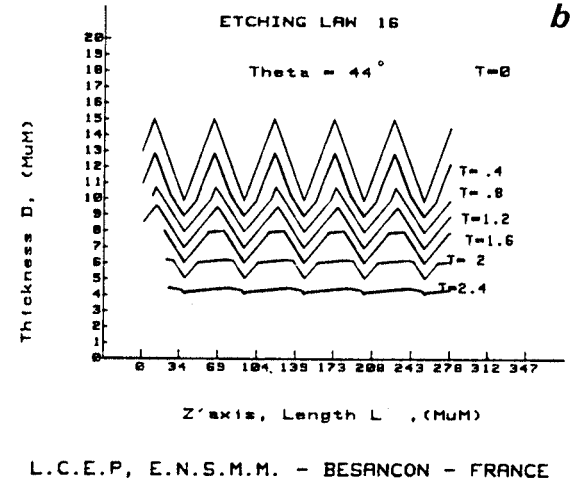


Figure 19



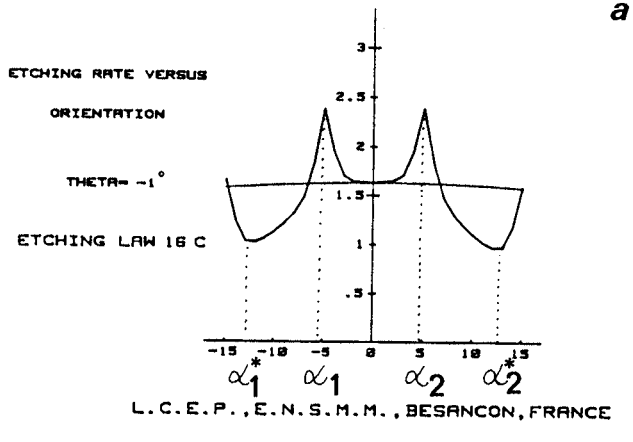
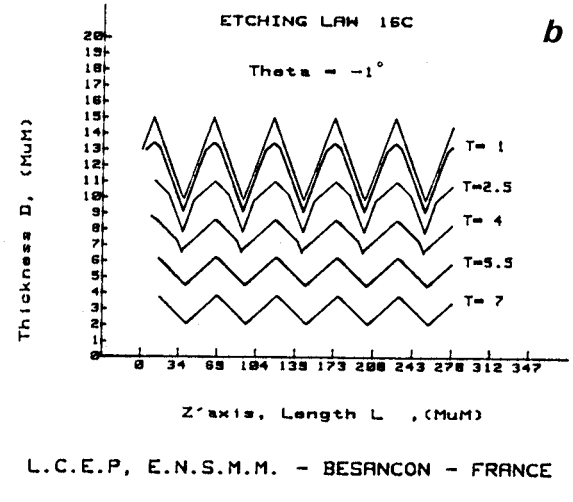


Figure 20



Figures 19 and 20 : Theoretical etched profiles (b) for the respective etching laws (a) 16 and 16C.

Thus at the first sight one can argue that the flat minimum that the etch rate presents at $\alpha = 0°$ does not initiate a concave background structure. In reality if the etching is prolonged to $\Delta d_s > 14\ \mu m$ concave intersection planes of small contrast are slowly formed. But evidently in a large range of depth of etch (i.e. $6\ \mu m < \Delta d_s < 14\ \mu m$) the hillocks can be qualified of stable and since the successive intersections are composed of two planes whose orientations α_N, α_P are identified with the two maxima α_1 and α_2 (Figure 20a) of the etch rate, here again the values of α_1 and α_2 can be worked out from the distribution of slopes.

③ Etching law 16 B and reference surface of orientation $\theta_o = 26°$: The etching law illustrated in Figure 21a is now rather complex, hence several peaks are generated in the slope distribution (Figures 18a, b, and c) for medium depths of etch. But as the depth of etch reaches $7\ \mu m$ (Figure 18d) only two peaks, α_N and α_P, corresponding to the orientations α_2 and α_1^* (Figure 21a) are maintained. Since here two maxima of the etch rate are involved the geometrical aspect of the resultant section will appear as convex in a large range of Δd_s and finally the profile will be composed of the two planes α_2 and α_1^* only (Figure 21b). We must remark that the minimum α_1 close to the reference orientation does not affect apparently the final shape of the surface profile ; effectively we never observe a concavity for the small negative slopes. To interpret this behavior we have to consider the nature of the extrema and the succession of diverging and converging trajectories. Turning to figure 21a we see that the two maxima α_1^* and α_2 will initiate convex sections for elements of profile with slope close to α_1^* and α_2 while the two minima α_1 and α_2^* will give rise to concave sections formed by planes of low etch rate. But since the apparent trajectories of planes α_1^* and α_2^* converge rapidly at a concave intersection and since the intervening planes close to α_1 dissolve very slowly the concavity of the surface profile is masked for planes with small negative slopes and as a consequence apparent stable hillocks develop. But care must be taken that when we are concerned with orientations in the range (α_1^*, α_2) a profile of alternate shape will certaintly form after a prolonged etching. Hence the shape of the dissolution profile varies continuously even if for some stages of the dissolution the etch rate vs angle distribution of figure 21a results in a profile of "stable" geometrical aspect. But even if care is needed in deciding the final appearance of the etched surface when complex etching laws are involved we note that the study of the changes in the distribution of slopes with the depth of etch which indicates the role played by the intervening extrema of the dissolution rate behaves as an "help procedure" to understand the successive dissolution processes.

6 - EXPERIMENTAL DETERMINATION OF THE EXTREMA OF THE ETCH RATE VS ORIENTATION PLOTS

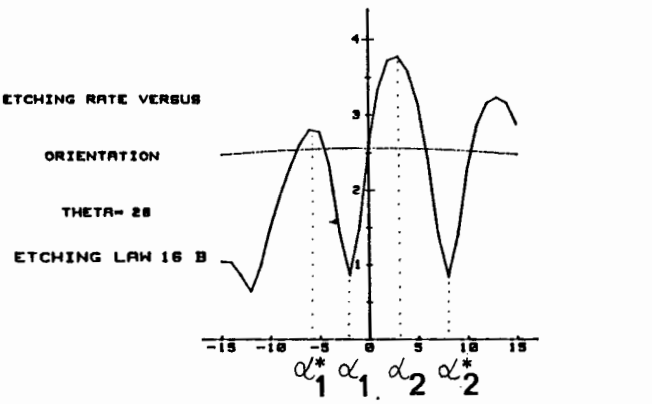


Figure 21 : Theoretical etched profiles (b) for the etching law 16 B (a).

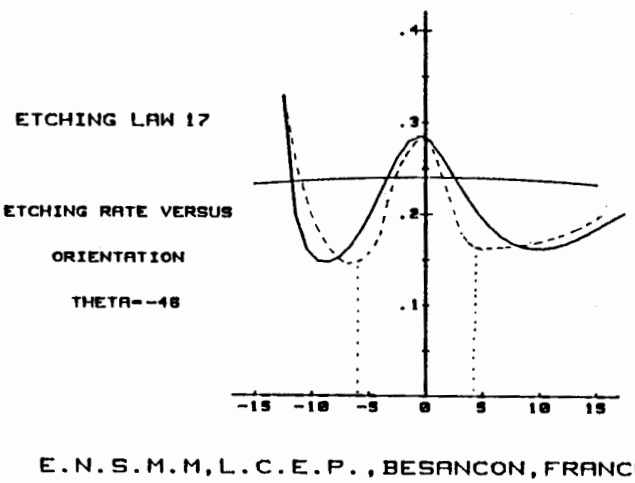


Figure 22 : A corrected dissolution rate vs orientation plot.

In the previous section we have shown that the extrema of the etch rate vs angle distribution can be exactly positioned by tracking some correlated features of the surface topography such as the shape of the profile shape and the slopes distribution. Since these quantitative informations can be easily extracted from experiments an interesting application of the above statistical approach consists of comparing some theoretical predictions derived from the numerical simulation with the probable behavior deduced from experiments in order to test the adequation of the proposed polar diagram of the dissolution slowness.

Let us for example concentrate our attention on the BT-46 plate. The histogram of slopes (Figure 13b) gives evidence of two singular slopes corresponding to $\alpha_N = -6°$ and $\alpha_p = 4.5°$ which are indicative of the positions of extrema of the dissolution rate. Furthermore comparison of the final Z' profilometry trace of the BT-46 plate (Figure 13a) with the dissolution profiles (figure 19b) derived from the numerical simulation for the etching law depicted in figure 19a shows that certainly the real $v_N(\alpha)$ plot will be almost similar to the plot in figure 19a but for the BT-46 plate we believe that the dissolution falls to a minimum for θ around - 52° (i.e. for $\alpha \simeq -6°$), increases to a maximum for around - 46° and falls again to a minimum for θ around 42° (i.e. $\alpha \simeq 4.5°$). Starting from the polar diagram of the dissolution slowness related to singly rotated quartz plates we can draw the thoeretical v_N versus α plot (Figure 22) for a reference surface of orientation $\theta_o \simeq -46°$. We observe that the expected features are crudely satisfied since the variations in the dissolution rate with the orientation α present two minima at $\alpha_1 \simeq -6°$ and $\alpha_2 \simeq 4.5°$ and a maximum at $\alpha \simeq 0°$. However from the experimental study we can propose a corrected version shown in broken line in Figure 22 which ensures that at medium etching times the histogram of slopes (Figure 13b) will stand. One may repeat this procedure for differently oriented AT- or BT-cut plates to obtain a more sharply defined etch rate vs orientation plot and to verify the agreement between theory and experiments.

7 - CONCLUSION

Previous experiments **(1-7, 9-11)** on quartz plates of various orientations have revealed that the etch rate is governed by the crystal orientation and that the morphology of the final etch figures depends on the orientation. Thus a vectorial analysis of the dissolution is proposed in which analytical equations for the polar diagram of the dissolution slowness vector, $\vec{L}$, are derived under the assumption that the magnitude, L, of the vector $\vec{L}$ depends only on the orientation. These new equations allow us to provide a two-dimensional numerical simulation for surface profiles with large variations in the orientation of successive surface elements. A satisfactory agreement is observed between the theoretical moving profiles and the Z' profilometry traces related to some singly-rotated quartz surfaces.

Since this good agreement is primarely bound to the use of a polar diagram which represents accurately the variations of L with its polar angle, some digital methods for a more complete characterization of the surface texture of etched quartz crystals are presented. In particular the changes in the distribution of slopes with the depth of etch give evidence for the possibility to obtain valuable informations on the precise positions of the extrema of the dissolution slowness (i.e. the dissolution rate). This possibility is illustrated for several more and less complicated dissolution rate vs orientation plots. Applying this procedure to some etched quartz surfaces allows us to predict without ambiguity the main features of the etch rate distribution and to give a corrected drawing of the etch rate vs orientation plot. Thus it is now possible to characterize completely the polar diagram of the dissolution slowness for singly rotated quartz plates and to verify the adequation of a future theoretical description of the dissolution slowness based on the symmetry of the quartz crystal.

REFERENCES

(1) C.R. TELLIER and F. JOUFFROY, J. Mater. Sci., 18 (1983) 67.

(2) C.R. TELLIER, Surf. Technol., 21 (1984) 83.

(3) C.R. TELLIER, **Proc. 38th Annual Symposium on Frequency Control, Philadelphia, Pa, 1984** (I.E.E.E., New York, 1984) p 105.

(4) C.R. TELLIER and C. BURON, Surf. Technol., 22 (1984) 287.

(5) C.R. TELLIER, **Proc. 39th Annual Symposium on Frequency Control, Philadelphia, Pa, 1985** (I.E.E.E., New York, 1985) p 282.

(6) C.R. TELLIER, Mater Chem. and Phys., 14 (1986) 25.

(7) C.R. TELLIER, N. VIALLE, J.L. VATERKOWSKI, **Proc. 40th Annual Symposium on Frequency Control, Philadelphia, Pa, 1986** (I.E.E.E., New York, 1986) p 76.

(8) F.C. FRANK in R.H. DOREMUS et al. (Eds.), **Growth and Perfection of Crystals**, John Wiley, New York (1958) pp 411-419.

(9) C.R. TELLIER, N. VIALLE and J.L. VATERKOWSKI, Surf. & Coatings Technol. (to be published).

(10) C.R. TELLIER and F. JOUFFROY, J. Mater. Sci., 18 (1983) 3621.

(11) M. CASTAGLIOLA, C.R. TELLIER and J.L. VATERKOWSKI, J. Mater. Sci., 21 (1986) 3551.

(12) M.W. WEGNER and J.M. CHRISTIE, Phys. Chem. Minerals, 9 (1983) 67.

(13) F. AUGUSTINE and D.R. HALE, J. Phys. Chem. Phys., 13 (1960) 344.

(14) B.A. IRVING in P.J. HOLMES (Ed.), **The Electrochemistry of Semiconductors** ; Academic Press, London, (1962), pp 256-289.

(15) C.R. TELLIER, N. VIALLE and J-L VATERKOWSKI, **Proc. 1st European Time and Frequency Forum, Besançon, France, 1987** (Imprimerie du Conseil Général du Doubs, Besançon, 1987) p 159.

(16) T.R. THOMAS, **Rough Surfaces**, Longman, London, (1981), Chapter 4.

* * * * * *

ACKNOWLEDGMENTS

The authors gratefully acknowledge financial support from the Direction des Recherches Etudes et Techniques which depends on the French Department of Defense.

* * * * * *

AUTHOR INDEX

R. Arnaud 843
J.P. Aubry 687, 873
C. Audoin 385

H. Bahadur 657
J.H. Balbi 421
P. Banerjee 243
A.G. Bates 679
A. Bauch 425
E. Bava 557
J. Beaussier 803
E. Benes 573
L.G. Bernier 125
J. Bignon 921
P. Blind 937
B. Bonnotte 687
B. Boulogne 851
R. Bourquin 893
D.E. Bower 731
H. Brandenberger 231
Ch. Breant 537
X. Buisson 921, 843
G. Busca 409, 445, 513

S. Calisti 739
R.D. Calvert 71
M. Carru 601
V. Candelier 483
B. Capelle 651, 873
O. Ceppi 271
P. Cerez 483
A. Clairon 499, 537
A. Clinton 275
J.M. Cloeren 679
F. Cordara 469
P. Cordier 851, 921
J.C. Craveur 687

A. D'Almeida 739
M. D'Amico 317
P. Dachel 275
N. Daher 909
B. Dahmani 537
J.F. Darces 833
H. de Boer 425
E. de Clercq 499
M. de Labachelerie 547
A. de Marchi 437
R. Delaite 665
J. Deng 555
M. Desaintfuscien 399
J. Detaint 601, 651, 873
E. Detoma 275
F. Deyzac 687
E. Dieulesaint 629
N. Dimarcq 547
K. Diomande 547
J.C. Doukhan 851, 921
N. Doukhan 851

R. Drullinger 531
J.Y. Duquesnoy 739

M. El Bellal 833
R. El Cheikh 795
G. Engel 827

B. Fischer 425
B. Fleischmann 779
M. Fragneau 851
G. Freon 213

H.G. Gaede 365
J.J. Gagnepain 909
A.H. Gérard 499
D. Gillet 421
V. Giordano 483
A. Godone 557
G.K. Goel 243
A. Goiffon 843
R. Gounon 195
M. Granveaud 213
J. Groslambert 151
P. Grudler 213
B. Guinot 187

A. Hamel 483
A.K. Hanjura 243
W. Hanrahan 275
D.W. Hanson 251
W.N. Hardy 383
D. Hauden 739
T. Heindorff 425
H.W. Hellwig 13
P. Hetzel 351
K. Hilty 253
L. Hollberg 531
C.K. Hruska 621
H.X. Huang 661
G. Huebner 419
M. Hürlimann 383
D. Janiaud 687
M. Jardino 399
G. Jenni 701
J.L. Jespersen 251
LU Jia-Fu 661
G.X. Jiang 661
C. Joly 601
D. Jozwick 937
J.C. Jumas 843

D.J. Knight 23
P. Koch 3
Y. Koga 531
B. Komiyama 521
P. Krempl 827
N. Kuramochi 645
P. Klushmeider 275

J.B. Laporte 163
H.A. Laue 99
Ph. Laurent 537
P. Leblois 739
T. Leblois 893
R. Lefèvre 873
J.N. Lelong 195
P. Lesage 385
W. Lewandowski 187
C.F. Lin 661
G. Lomaglio 421
W.H. Luo 661

A. Mahmoud 139
B.S. Mathur 243, 657
D. Mazerolle 629
C. Meis 399
J.P. Mellana 253
H. Merigoux 833
L. Michel 873
Th. Migeot 665

Y. Nakadan 531
C. Novero 557
H. Nowotny 573

S. Ohshima 531
Y. Ohta 521
A. Okada 645
T.N. Oliver 731
M. Olivier 151
N. Oura 645

L. Pan 409
P. Parcelier 213
R. Parshad 657
U.R. Peier 701
L. Penavaire 739
V. Pettiti 469
E. Philippot 843, 873
O. Piller 3
M. Planat 767
J. Ponsonby 341
T. Pratt 45
R. Price 275
L. Prost 231

J. Rabian 461
G. Riha 779
P. Rochat 461
D. Royer 629
E. Rubiola 263, 333

H. Saitho 521
E. Sappl 307
W. Schaefer 715
H. Schild 293
M. Schmid 573
R. Schröder 425
K. Schulthess 817
K. Schwaller 817
J. Schwartzel 601, 651, 873
H. Schweda 445

A. Sengupta 243
D.V. Shick 753
R.C. Smythe 637
R. Spercer 117
S. Starker 307
J.J. Suter 679

Y. Tan 555
P.N. Taneja 243
P. Tavella 469
H.R. Telle 523
C.R. Tellier 893, 937
J.G. Theobald 421
P. Thomann 445
C. Thomas 499
H.F. Tiersten 753
Y. Toudic 843
R. Tourde 213

Jun Umezu 521

M. Valentin 795
M. Vincent 213

F.L. Walls 719
G. Werth 35
H.W. Woerz 779

Y. Xiang 555

N. Yahyabey 385

A. Zarka 651, 873
Z.C. Zhai 661
Y. Zhang 767
X. Zhu 555
W. Zingg 817