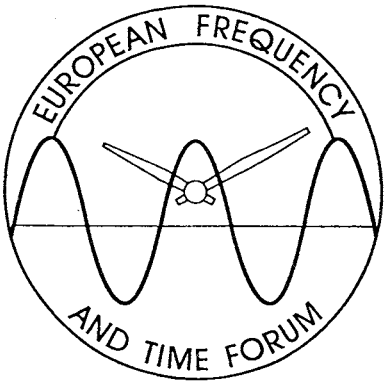
8TH EUROPEAN FREQUENCY AND TIME FORUM WITH EXHIBITION

PROCEEDINGS
Volume I
Verbal Contributions



EFTF 94

MARCH 9-11, 1994

TECHNICAL UNIVERSITY MUNICH
WEIHENSTEPHAN, GERMANY

8th European Frequency and Time Forum

Technical University Munich

Weihenstephan, Germany

March 9-11, 1994

Conference organizer and proceeding publisher:

VDI-PROJEKT UND SERVICE GMBH
Postfach 10 11 39
D-40002 Düsseldorf

Phone: int. +49/211/6214-583
Fax: int. +49/211/6214-164

CONTENTS

Volume I

page

Introduction

Local Organizer, Acknowledgement III

Executive Committee III

Scientific Committee IV

Verbal Contributions

Opening Session 3

A1: New Developments in Atomic Frequency Standards 19

B1: Resonators 73

A2: Telecommunications and Teletiming 149

B2: Materials 203

A3: Time Transfer 275

B3: Atomic Frequency Standards I 359

A4: Time Scales 407

B4: Oscillators 439

A5: Atomic Frequency Standards II 497

B5: SAW Devices 533

A6: Interrogation Signals in Atomic Frequency Standards 627

B6: Applications with Space Systems 683

Volume II

Poster Contributions 749

Local Organizer

VDI-PROJEKT UND SERVICE GMBH, Düsseldorf, Germany
Lutz Zühlke und Angelika Bohnen

1994 Chairman of the Scientific Committee

Dr. Klaus Dorenwendt, Physikalisch-Technische Bundesanstalt, Braunscheig, Germany

Acknowledgement

We wish to thank The Robert Bosch Foundation GmbH for the kind support which made it possible for scientists from Central and East Europe to participate.

EFTF Executive Committee

Audoin C.	C.N.R.S.-L.H.A., Orsay	France
Besson R.	E.N.S.M.M.-L.C.E.P., Besançon	France
Ecabert M.	FSRM, Neuchâtel	Switzerland
Gagnepain J.-J.	C.N.R.S., Paris	France
Kartaschoff P.	Cormondrèche	Switzerland
Schlueter B.	Keres Engineering Ltd., Neuchâtel	Switzerland

Honorary member of the Executive Committee

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

EFTF Scientific Programme Committee

Aubry J.P.	Thomson CEPE, Argenteuil	France
Audoin C.	C.N.R.S.-L.H.A., Orsay	France
Ballato A.	US Army Research Laboratory Fort Nonmouht	U.S.A.
Bauch A.	PTB, Braunschweig	Germany
Besson R.	E.N.S.M.N. - L.C.E.P., Besançon	France
Boloix R.	Real Observarorio Armada San Fernando	Spain
Brunet M.	C.N.E.S., Toulouse	France
Busca G.	Observatoire Cantonal, Neuchâtel	Switzerland
Cutler L.	Hewlett-Packard Co., Palo Alto	U.S.A.
De Jong G.	Van Swinden Laboratory, Delft	The Netherlands
De Marchi A.	Politecnico di Torino, Torino	Italy
Defranould Ph.	Thomson Sintra ASM, Valbonne	France
Deyzac F.	ONERA, Châtillon-sous-Bagneux	France
Dorenwendt, K.	PTB, Braunschweig	Germany
Dowsett J.	C-MAC Quartz Crystals Ltd. Harlow Essex	U.K.
Duchaussoy H.	D.R.E.T., Paris Armées	France
Feltham St.	ESA/ESTEC, Noordwijk	The Netherlands
Graf E.	Oscilloquartz S A., Neuchâtel	Switzerland
Granveaud M.	LPTF-Observatoire de Paris, Paris	France
Handrek H.	Quarzkeramik GmbH, Stockdorf	Germany
Hartl Ph.	Universität Stuttgart	Germany

Hauden D.	L.P.M.O. - C.N.R.S., Besançon	France
Heighway J.	Advanced SAW Products S.A., Bevaix	France
Hilty K.	Swiss Telecom PTT, Bern	Switzerland
Hruska C.K.	York University, Toronto	Canada
Kalliomäki K.	Technical Research Centre of Finland, Espoo	Finland
Kartaschoff, P.	Cormondrèche	Switzerland
Kihara M.	NTT Transmission Systems Laboratories, Yokosuka	Japan
Kirchner D.	Technische Universität Graz, Graz	Austria
Knight D.J.E.	National Physical Laboratory, Teedington	U.K.
Krempl P.	AVL List GmbH, Graz	Austria
Leschiutta S.	Polytecnico di Torino, Torino	Italy
Nakagiri K.	Radio Research Laboratory, Tokyo	Japan
Neubig B.W.	Tele Quarz GmbH, Neckarbischofsheim	Germany
Nieuwejaar P.	Norwegian Naval Materiel Command Haakonsvern	Norway
Nottarp K.	Institut für Angewandte Geodäsie Rodgau	Germany
Prost L.	Offive Fédéral de Métrologie Wabern	Switzerland
Quinn T.J.	Bureau International des Poids et Mesures, Sèvres	France
Reid R.	British Telecom, London	U.K.
Rutmann J.	FIEE Féd. Ind. Electrique et Electroniques. Paris	France
Starker S.	DLR, Weßling	Germany

Steele J.	National Physical Laboratory Teddington	U.K.
Thomann P.	Observatoire Cantonal, Neuchâtel	Switzerland
Uljanov A.	Kvarz Institute of Electronic Measurements, Nizhni Novgorod	Russia
Unverdross R.	Unverdross Technik, Wörthsee	Germany
Vanier J.	National Research Council Canada, Ottawa	Canada
Walls L.	National Institute of Standards & Technology, Boulder	U.S.A.
Winkler G.	US Naval Observatory, Washington	U.S.A.
Yannoni N.	Hanscom AFB Bedford	U.S.A.
Zampetti G.	Telecom Solutions, San José	U.S.A.
Zingg W.	Microcrystal div. of ETA, Grenchen	Switzerland

Opening Session

Welcoming Address

Klaus Dorenwendt, Physikalisch-Technische Bundesanstalt

Ladies and gentlemen, dear colleagues,

I have the great pleasure to welcome you to the 8th European Frequency and Time Forum at Weihenstephan. For the second time in the eight years old - or may I say eight years young - history of our meeting the venue is not Neuchâtel in Switzerland or Besançon in France. We are very pleased that the Executive Committee's choice for 1994 fell on Germany. And we hope that the scientific programme as well as the town of Weihenstephan will not disappoint you.

Time and frequency measurements are closely related, they are different aspects of the same thing. But whereas time measurements have been done since the very beginning of our civilisation, frequency measurements became only important with the development of modern technology. The first invited talk of this morning, which leads us back to the beginning of our science, is, therefore entitled: "The History of Time Measurements", omitting the frequency.

In this context it is perhaps of interest to realize that time measurements always differed in some aspects from the determination of other physical quantities. So the precision that could be achieved exceeded that of other measurements by several orders of magnitude. Already 200 years ago, chronometers for navigation were able to keep the time for many months with deviations smaller than a minute. In modern terms this means, that the relative uncertainty of these instruments was already at the order of 10^{-6} and there was still an urgent need for higher accuracy. Length or mass measurements were less accurate at that time and there was no real demand to improve them. Moreover, the units of time, the day and its fractions, were the same all over the world long before the implementation of the International System of Units (SI), and they have always been accessible to everybody, as they were derived from natural constants like

the revolution of the earth in former days or the separation of energy levels in the Cs-atom today. It is worthwhile remembering that this ideal situation has not yet been attained for all units of the SI. The kilogram for instance is still defined by a prototype, carefully protected in a cellar of the International Bureau for Weights and Measures. But in spite of a well defined and easily accessible unit time measurements in the past still needed clocks for interpolation of the long time intervals that could be deduced from astronomical observations. In the first contribution we shall see some of the beautiful instruments that were developed for this purpose.

The second invited talk of this morning with the title "Frequency Measurement of Visible Radiation" will confront us with one of the modern challenges of metrology. Coherent frequency measurements in the visible means to count the cycles of such a radiation without losing one. If this could be done reliably, an immediate application would be the development of optical clocks less sensitive to external fields and with reduced instability in comparison to the existing devices. At the same time our knowledge of the structure of atoms and of the absolute value of fundamental constants like that of the Rydberg constant would be improved. More generally it can be assumed that the most stringent tests of general relativity, quantum dynamics or other modern theories in physics will in future be based on time and frequency measurements due to the unrivaled accuracy that can be reached in this field. Actually it is of the order of 10^{-14} and devices are investigated which will perhaps enlarge the margin of five orders of magnitude which exist at least with respect to other physical quantities.

The last contribution of this morning is devoted to "Telecom Applications", a modern key technology with growing economic importance which, together with microelectronics, has changed our life. But it is only one example of the many technical applications in science or industry which are based on precise time and frequency measurements. Astronomers, navigators and geodesists all essentially depend upon this technique. The different sessions of our meeting will highlight all these fields as well as the tools for it like crystal oscillators and atomic clocks.

Prof. Dr. Richard Mühe, Furtwangen

History of Time Measurement

1. Anthropological Time Measurement

1.1. Dimensions of Time

Primary ideas of time and time circles have been taken from nature. In all cultural fractions, since milleniums, the day and the year have been derived from mutual motion of earth and sun. Shorter times, such as fractions of the second and longer times have been diverted from these fundamental dimensions.

We are aware that time measurement methods and conceptions of time and world are more static than dynamic. In the 5th century Augustin answered to the question about time:
"If nobody is asking me, I will know it. But shall I explain it to anyone, I don't know it." In the Middle Ages time regulation was exemplified in the Christian Occident by the Benedictine rule: "Ora et labora". Later, in the thirteenth century, Thomas Aquinas concluded: "Time belongs to God". Men didn't reflect upon their influence on running time. Contendetness in this world determined human life. The human situation seemed not to be changing. Therefore an alternation was not desired by men.

1.2. Invention of the Mechanical Clock.

The Benedictine principle of regulation produced more precision and more accuracy in time measurement, although coming from a static world. Different devices for time measurement were developed. The invention of mechanical clockworks seems to have been nearly inevetable. Such clockworks allowed better obedience to the Benedictine rules. Time regulation by human influence seemed to be possible. Obedience was forced to different times for labour and for prayer. Human doing was impressed by the rhythm of clocks and machines, although machines had not yet been invented. In our days, Lewis Mumford says:

"Not the steame-engine, but the clock is the code-word of modernism."

1.3. Natural Scientific Research.

The discoverors of the new world, as there are Nicolaus Copernikus or Christopherus Columbus, caused dramatic changes in world conceptions. Leonardo da Vinci was the first man to begin technical development. Clocks allowed measuring techniques with unexpected accuracy. Importance of these techniques for human being was growing. The Age of Enlightenment was conditioned by scientific research in the 17. and 18. century. Classical time measurement was founded by Galileo Galilei's scientific observations as well as Isaac Newton's research. These were pioneer functions for other precision techniques. In this context Isaac Newton regarded time as an independent force.

1.4. Clocks as Toys for Kings

Emperors and kings were engaged with complicated clocks showing astronomical indications, moving automatons in combination with clockworks. Such miracle objects of handicraft stimulated phantasy and further development. About 1300, early mechanical timepieces which had no clock-faces signaled time by striking bells. With the Roman idea of "horen" or hours, they are parts of our time-system. In this context the relationship of the German word "Glocke" and the English "clock" seems to be obvious.

Already in the fourteenth century minutes and seconds were known. They only went into human mind two hundred years later at the earliest. Quick motion of clock-hands generated human feeling for running time. Quarter hours on clock-faces appeared not earlier than in the sixteenth century. In the seventeenth century precision of timekeeping increased by the use of pendulum escapements. First the early eighteenth century brought the minute-hand to our well known clock-face. Since that time we find central hour- and minute-hands. The identity between the second-beat of mechanical clocks and our life rythm, the heart-beat, appears to be significant. The third division of the hour after minutes and seconds, "Tertian" was not very important, perhaps corresponding to impossible adaption to human customs.

1.5. Basis of Precision Time Keeping.

Time tables and time-control by means of clocks became more actual. The clocks divide time-spaces

into small units, visible at the clock-face. The day is portioned into a series of planned activities, with the result of more perfect and precise modification. Thus, time was unbounded from biological connections. Jeremy Rifkin precises in 1987: "Time was arrested in the interior of wheels from an automatic machine, the clock."

1.6. Industrial Clockmaking.

Time measuring techniques have influenced mankind. Clocks were made by craftsmen under strong regulations of their guild. Clockmaking grew to industrial production in the late eighteenth century. Popularity of clocks and time measurement increased just as the need for them. That may be demonstrated by Black Forest Clockmaking. After beginning in the early eighteenth century a mass production of timepieces came up. The clocks were exported nearly into all of the world. One hundred years later, this manu-facturing was continued by industrial production. The traditional work of craftsmen was growing into the industrial period during the nineteenth century.

2. Technical Time Measurement, Nanoseconds-Culture.

2.1. Distribution of Time.

A new chapter began with the distribution of time by electrical circuits, what we may call a "Mass Production of Time". In the middle of the nineteenth century, precision time keeping with time pulses and communication systems was introduced. The Historian Daniel J. Boorstin (1983) thinks, "Time is the first real discovering and the most important social instrument of mankind.

For the American contemporary Jeremy Rifkin, some ideas of whom we are following, six dimensions of time are combined by an infinite number of different ways:

-sequential structure,
-time,
-planning,
-rate of repeating,
-synchronisation and
-time perspectives.

Simultaneously they are the main categories of social experience. Also in these days, the sociologist Emile Durkheim distinguishes four mechanisms of time distribution:

-biotic rituals,
-astronomic calendars,
-clocks and
-computer programs with nanoseconds-structure.

2.2. Computer Time

Time measurement becomes more important and precise. Time is substantial for regulation of processes and substitutes the use of clocks for single men. The nanosecond of the computer is used for time-fixing and time-control of human life.

Precision measuring technique and control should have substituted the dictation of clocks in human world. Yet as a result we have no time in spite of highly precise measuring with smallest time intervals. The computer brings new perspectives: Instead of day, hour, minute and second we find the smaller time units which are important for men, since computers are ruling. The computer works in nanosecond-steps. We cannot feel them, but they are important for modern life. Before, biological and social time were identical for men. However, the smaller parts of seconds (micro- and nanoseconds) are breaking our usual feeling of time.

2.3. Time and Information

In the industrial age the slogan "Time is money" was known. In capitalism time was sold by the employees. Control-clocks had an important coordination function. Craftsmen and farmers had their own control about their working rhythm. Today people think, everybodies time has the same value.

With computer time we get kinds of time for different matters, thus we will get an individualisation of time in future. Computer time does not rely on nature and real space. It is bound, neither to the rhythms of day, past or future. Usual clocks produce a series of identical time-spaces (seconds, minutes or hours). The computer changes information to nanoseconds and nanoseconds to information. Time is information, and thus time-spaces can be substituted by different information, that means by different times. Life corresponds to a big number of information bits, which can be programmed by infinite different ways.

3. New Dimensions of Time.

3.1. Changes in Ideas of Time

In thirteenth century, Thomas Aquinas regarded God as the master clockmaker of the world and as a dissociated arbiter. In 1377 Nicole d'Oresme defined the clockwork-universe parallel to a master clock by the principles of celestic motions. Invention of clocks generated a new class of time-consciousness. Newton and others isolated time from eventualities and defined time as an independent force. God was the first sacrifice of clockwork-universe. Time is ruling about all. Universe is seen to be a big clockwork, perhaps requiring to be wound up once.

By industrial work and common use of clocks, machines are dictating velocity in life. Punctuality of Nanoseconds-culture had already become a moral imperative. Such ideas and their consequences are based upon:
-increasing precision of time measurement,
-development of techniques for precision timing and frequency measurement,
-electric regulation of clocks,
-quartz crystal clocks,
-atomic frequency standards, including radio controlled clocks.

3.2. Time and Relativity

From Einsteins theory of relativity we know, time is not constant under all circumstances. Relativity of time and longitude only are evident in extrem conditions, near the velocity of light. We don't mention this under normal conditions. But it is reality in universe, either in micro- or macrocosms.

3.3. Time and Universe

Since Albert Einsteins theory, 1905, there is no absolute time. We must reflect all traded ideas on time:

-atomic time is defined by frequency,
-(1 sec = 9 192 631 770 Hz)
-sophisticated technology, such as earth satellits,
-radiation measurement from white dwarfs,
-neutron stars or black holes.

Our imagination does not include four-dimensional space. How-ever, constancy of light-velocity or other scientific facts, we cannot explain, may be handled successfully by this thinking.

3.4. Expectations.

These ideas enforce revisions of real life-experience as:
-time yield,
-time dependence of moving objects,
-turning back of time
-equivalence of energy and mass.

Further, we may expect conclusions from these nearly unreflected formulations. Man is used to reconsidering them in run of the world. For instance, to non specialised people, atomic energy is thought to be questionable due to reflected difficulties.

Such conceptions, as there may be:
-imaginary time,
-indefinite sizes,
-time as fourth dimension,
will become common for men in future. Perhaps by such definitions we shall recognize our own ties to time and our infinity.

Synchronization for the Global Telecommunication Network: a Fundamental Requirement

Stefano Bregni, Luca Valtriani

SIRTI S.p.A, R&D - Network Technologies Division, Via G. Vida 19, 20127 Milano, ITALY,
Tel. +39-2-6677.8375 Fax +39-2-6677.8358

Abstract

Timing and synchronization features are becoming very critical requirements for present complex telecommunications networks since digital systems have been introduced. The synchronization goal is to carry out the distribution of a very precise clock signal throughout the whole network, in order to suppress phase noise and avoid increasing buffer memory size and transfer delays. Moreover, the new SDH transmission technique requires very tight specifications about the timing signal stability.

In this paper, the impact of synchronization on innovative digital networks and services is presented, with the example of the synchronization network designed and deployed for the Italian long distance operator (IRITEL). The paper points out the main features of this network, including the general system architecture, the management and control characteristics, the node structure and clocks performances.

1. Introduction

During the last few years, telecomunication networks went through a considerable number of upgrades: in fact due to the increase of new services demand and to a bit rate which is becoming higher and higher, development plans for new digital exchanges have been sped up, and as far as the transport network is concerned, bandwidth, quality and availability request have reached such a level that cannot be simply fulfilled by using ordinary copper cables.

Meanwhile, the evolution towards optical fibre and high speed systems has also led to the standardization of the new transmission technique SDH (Synchronous Digital Hierarchy) [1].

In this paper, the impact of synchronization on innovative digital networks and services is presented, with the example of the synchronization network designed and deployed for the Italian long distance operator (IRITEL). The synchronization network described in this paper is based on a hierarchical master-slave architecture with redundancies in order to guarantee high reliability. The paper points out the main features of this network including the general system architecture, the management and control characteristics and the node structure and performances. Finally, some clock performance results, measured with a computer based testing system, are also shown.

2. SDH Network Synchronization

A SDH network is characterized by high bit rates, typical of transmission on optical fibre cables, ranging from 155 Mbit/s to 2.5 Gbit/s [1]. Basically, the most striking technological difference between SDH and PDH (Plesiochronous Digital Hierarchy) [2] networks is that the SDH network operates by using a synchronous multipexing technique, instead of the traditional asynchronous one, in a fully manageable environment in terms of performance configuration and alarm surveillance.

For this reason, the network manager is now able to either insert or drop signals more easily by using special equipment specified by SDH standards. These features, as well as other advantages such as, for example, the possibility of transporting a huge amount of data in addition to the payload, have by now brought SDH up to the point of being commonly regarded as the most promising standard in optical fibre networks.

Alongside with such evolution in transmission techniques and standards, it has also come up the need of having in each single office a synchronization signal source that has to be good enough to synchronize the equipment connected to the network.

The most important reason, for having only one synchronization signal uniformly distributed, is that of reducing to the lowest level possible the misbehavings in network operations. This leads to either a better quality of the transmitted signal or a remarkable improvement in the overall network quality [3].

Therefore in a perfectly synchronized, noiseless network there are no slips [4] at all, but since having an ideal network is not realistic, the target to be achieved is to improve the short and medium term stability in SDH equipment internal clocks, so as to mantain the presence of jitter and wander on the transmission lines, which are used to distribute synchronization signal through the network, at a limited amount.

If we account for SDH networks, as far as the stability of the synchronization signal is concerned, very tight requirements have to be fulfilled. In fact, the SDH frame structure consists of Virtual Containers (VC) [1] into which tributary signals are mapped. For each of these VCs there is a pointer which is located in a fixed position and allows the VC to float in the frame.

For this reason, in every single node, from time to time the Container, that has to be retransmitted from every SDH equipment must be re-synchronized to the local oscillator. Therefore, frequency shifts among clocks located in different nodes plus jitter and wander collected during the transmission of synchronization signals may cause some pointer adjustments for VCs within the equipment buffer [5].

These adjustments cause some wander in the tributary signals when passing through the interface between the Synchronous and the Plesiochronous environment where they are demultiplexed before leaving the Synchronous environment [6].

Thus, the most important goal of an efficient SDH synchronization network is that of minimizing pointer adjustments in order to get significant improvements in the overall performances, particularly on the boundaries with the existing Plesiochronous networks.

The quality of the distributed synchronization signal, in terms of short and long term stability, depends both on the performances of the oscillators located in the synchronization network nodes and on the network architecture itself. Therefore, the overall design will have to be carried out by keeping in mind the performances that have to be achieved in all the different applications.

In the next section, an example of this kind of synchronization network is shown. The network described was carried out by SIRTI for the Italian long distance operator.

3. Synchronization of the Italian Long Distance Network

3.1. Network Architecture

The reference architecture for the Italian long distance synchronization network is shown in fig.1. This is in accordance with the trend in the International Standard Committee [7][8][9][10][11] and allows the operator to introduce the new Synchronous Digital Hierarchy (SDH) in the long distance telecommunication network.

This architecture has a tree-like structure according to the Hierarchical Master Slave architecture [12], with four hierarchical levels. This architecture is the most indicated for networks where the nodes which have to be synchronized are located in sites that, from the strategic point of view, have different levels of importance. This structure allows, moreover, to obtain the highest performances with the lowest costs.

Fig. 1 depicts a detailed view of the various hierarchical levels. The first is level 0. The Primary Reference Clock (PRC) consists of a Cesium beam clocks complex, it is managed by the PTT *Istituto Superiore delle Poste e Telecomunicazioni* (ISPT) and it is the reference clock for every Italian telecomunication network operator, viz. SIP, IRITEL, ITALCABLE. It is strategically placed in a central position (Rome) with regard to the Italian telecommunications network, which is basically longitudinally developed to match the shape of the Italian territory.

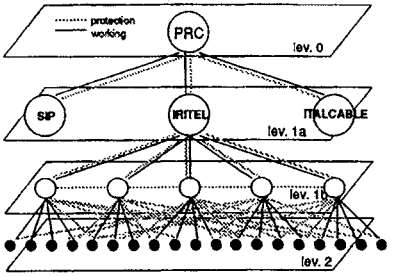


Fig. 1: Italian long distance synchronization network architecture

The second is level 1 which is splitted in level 1a and level 1b. The clocks at level 1 have a Rubidium oscillator, so that a free-run stability in the order of 10^{-11}/month can be guaranteed. The level 1a clocks, both located in Rome, are directly locked to the PRC by means of a duplicated analog optical link. The meaning of the level 1a clock is that being the Master Clock Distributor for the overall synchronization network.

As a certain level of protection has to be assured, a direct link to the *Istituto Galileo Ferraris* of Torino (the Italian station of the Universal Time Coordinated) will be made available in order to have a reference signal with the same quality, as minimum, of the signal provided by the PRC itself.

Each of the five nodes at level 1b is locked to the level 1a by using a couple of links coming from this upper level. In order to obtain a high network reliability, each of the two streams that carry the reference synchronization signal to a single level 1b node, is transmitted on two different paths, thus realizing a protection by means of path diversity.

These nodes have been strategically placed throughout the Italian territory, so that each of them may be able to supply the synchronization signal to five different regions which have more or less the same number of level 2 nodes. Moreover, these nodes carry out the analog (2 MHz) distribution of the synchronization signal to every single equipment within the office.

There are 46 nodes at level 2 that are uniformly distributed throughout the country and are placed in the same long distance synchronization offices of SDH

DXC 4/4, used for the automatic network protection. The level 2 clocks will distribute the reference synchronization signal to every equipment in the office. These nodes are provided with duplicated Quartz oscillators and free-run stability performances of about $5 \cdot 10^{-10}$/day are guaranteed.

The architecture deployed matches International Standards, in particular ITU-T Rec. G.803 [7], as shown in fig. 2. A reduced mesh Master-Slave configuration is adopted since only the level 1b nodes are linked together by means of an existing mesh. This allows a reliable distribution of the synchronization signal to every node in the network with the lowest cost in the use of network resources.

The synchronization signal has the same long term stability as the Network Master itself, while the short term stability depends on the performances of the clocks which are located in the network nodes.

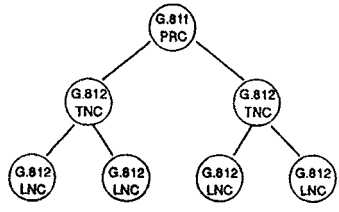


Fig. 2: G.803 synchronization network architecture (TNC: Transit Node Clock - LNC: Local Node Clock)

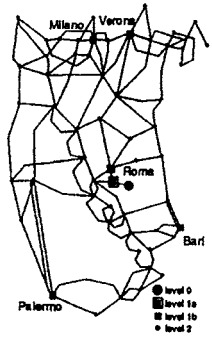


Fig. 3: Topology of the Italian synchronization network

The basic fault sensitivity that affects either equipment or transmission lines has been overcome by introducing both equipment and transmission lines redundancy. This requirement is particularly pressing in the challenging scenario of new telecommunication networks: the synchronization network will be playing a role which is becoming more and more important and in this way its reliability has to be very strong.

The architecture of the Italian network actually used to distribute the synchronization signal and the nodes placement are shown in fig. 3. The most important task, that every synchronization node has to carry out, is to filter the synchronization signal coming from the upper level node and then distribute it to the lower level nodes.

In the network presented, the synchronization signals are distributed by means of the existing plesiochronous network, where the 2 Mbit/s traffic signals, used to distribute the clock, are directly multiplexed into 565 Mbit/s signals. The synchronization signals, in the future, will be transported by using the SDH links when these will be activated. At the moment, the long distance SDH network is in a pre-operating phase.

As the synchronization network has been designed in this way, flexible rerouting will not be allowed. This fact is mainly a consequence of the cost effectiveness and the high reliability of the solution adopted, especially for the long distance network.

3.2. Node Architecture

According to ETSI Standards [8] this application is a sound example of the concept of Synchronization Supply Unit (SSU), as it is shown in fig. 4. The SSU is the equipment that contains both the slave clocks and the hardware which enables the synchronization clocks to be managed and the synchronization signal to be distributed inside the office and through the network.

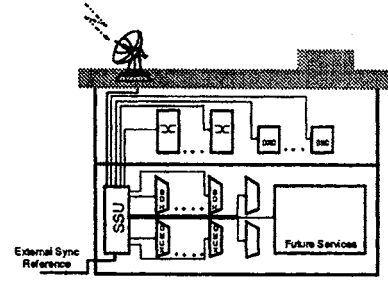


Fig. 4: Architecture of a typical synchronized office

The most important function of the SSU is that of providing equipment that are located in every single node with the most suitable synchronization signal. Therefore, by using SSUs, every equipment in the office can be supplied with low complexity synchronization cards, thus leading to an optimization both in performaces and in costs.

Furthermore, each node can be provided with a flexible number of inputs and outputs which can be increased by adding single new cards to the equipment (see fig. 5).

Every input interface can hold either analog (2048 KHz) or a digital (2048 Kbit/s) signal, fully compliant with the Recs. G.703 [13] and G.704 [14]. The tracking oscillator unit contains two identical oscillators which are Rubidium for levels 1a and 1b, and Quartz for level 2.

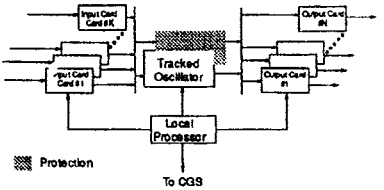


Fig. 5: SSU block diagram

Both these oscillators are locked to a valid reference that has been selected among those available at the input, but only one of these two is normally active while the other one works as "hot spare". This means that it starts operating only if a failure occurs to the oscillator that is actually working.

Specific hardware has been designed in order to continuosly monitor the phase of the signals coming from the two oscillators. In case a swap between either the active references or the oscillators occurs, this hardware takes care to minimize the phase hits at the output. Phase hits on the output signal, by the presence of fault on the inputs, are avoided as well.

It has been estimated that the probability for a node to enter in hold-over mode is lower than 10^{-5}, with a MTTR (Mean Time To Repair) which is guaranteed to be shorter than 24 hours. Moreover, the overall probability to have a total node failure is about 10^{-25}, and such a small value can be easily disregarded (obviously, with the same MTTR).

In the SSU there is even a CPU which has the task either of processing faults coming from the various unities or of providing the local control interface. The local control is able to monitor the basic SSU performances and has functions both of fault filtering and processing. The results are then sent to the *Centro di Gestione della Sincronizzazione* (CGS) of the synchronization network. By means of the remote interface it is also possible to monitor the intra office links and thus to collect the information concerning quality to the CGS.

3.3. Network Management and Control

The network architecture described is able to meet the quality requirements issued by International Standard Committees [5][7][8][10][11]. In order to check how this goal is achieved a functional monitoring of the network elements is done by using a suitable control and management system.

The equipment which the SSU consists of are suitable to be locally monitored either for their performances or for their fault conditions. Such equipments, as shown in fig. 5,6, are linked to the control and management system CGS by means of a packed switched network and a remote interface.

The CGS is able to process inefficiency and quality information regarding the synchronization network, to provide the required statistics of the network nodes together with the inventory and trouble ticketing of the hardware. In addition, it can actively manage the synchronization distributors.

Furthermore, the SSUs installed are able to carry out some measurements on the synchronization signal by means of a suitable measurement procedure. Such measurements continuously monitor the synchronization quality as far as the long term phase error between the upper level nodes and their lower nodes is concerned. The quality of the 2 Mbit/s synchronization links is also evaluated according to the ITU-T Rec. G.821 [15].

In accordance with fig. 6, phase measurements of the synchronization signals are made possible by using loops on the synchronization signals themselves. The short term stability will be also monitored by means of future hardware upgrades in the SSUs.

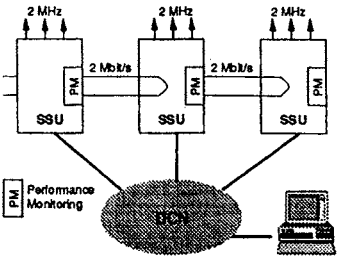


Fig. 6: Performance monitoring logical scheme

3.4. SSU Performances

Here below the main characteristics of the SSUs are reported.

Level 1 SSU: Rubidium Oscillator

- Pulling range: $\pm 1*10^{-9}$
- Free-run stability: $\pm 5*10^{-11}$/month
- Hold-over stability: 24 hours, 0-45°C: $\pm 3*10^{-10}$
- PLL slaving time constant: user selectable between 100-10^5 s via RS-232C
- SSU availability: 99.989%

Level 2 SSU: Quartz Oscillator

- Pulling range: >1 ppm
- Free-run stability: $5*10^{-10}$/day
- Hold-over stability: 24 hours, -20/+70°C: $4*10^{-9}$
- PLL slaving time constant: user selectable between 100-10^5 s via RS-232C
- SSU availability: 99.989%
- Allan Deviation (ADEV) and Modified Allan Deviation (MADEV) [8][9][10][11] measurement results: see fig. 7
- Root mean square of Time Interval Error (TIErms) and Time Deviation (TDEV) [8][9][10][11] measurement results: see fig. 8

The test results reported in fig. 7, 8 were obtained in SIRTI laboratories with a test system especially designed for the purpose of clock characterization. This test system is based on a high performance time

counter, drived, via a GPIB (standard IEEE 488.2) interface, by a personal computer running the measurement software TEMPO (Timing Equipment Measurement PrOgram). TEMPO manages the real-time data acquisition and the subsequent numerical processing and evaluation of all the main clock stability quantities.

Two sequences of Time Error data were acquired by measuring the output of the SSU with the sampling period τ_0=60 ms: the first with the SSU operating in locked mode (comparing the input with the output timing signal), the second with the SSU operating in free running mode, over a measurement period, respectively, of T=5400 s and T=3600 s. The graphs shown compare the results in both the measurement configurations. More details about the measurement methodology can be found in [16].

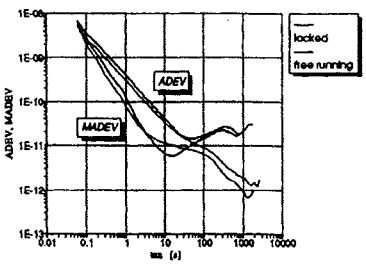


Fig. 7: ADEV and MADEV measurement results for a level 2 SSU (comparison between locked and free running configuration)

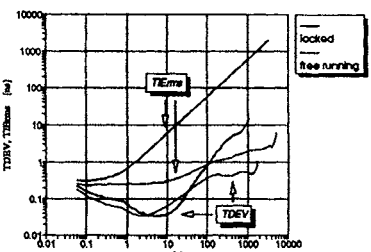


Fig. 8: TIErms and TDEV measurement results for a level 2 SSU (comparison between locked and free running configuration)

4. Conclusions

In this paper, the impact of synchronization on innovative digital networks and services was presented, with the example of the synchronization network designed and deployed for the Italian long distance operator (IRITEL). The paper pointed out the main features of this network, including the general system architecture, the management and control characteristics, the node structure and clocks performances.

References

[1] ITU-T Revised Rec. G.707 *'Synchronous Digital Hierarchy Bit Rates'*, G.708 *'Network Node Interface for the Synchronous Digital Hierarchy'*, G.709 *'Synchronous Multiplexing Structure'*, Geneva, 1993

[2] ITU-T Rec. G.702, *'Digital Hierarchy Bit Rates'*, Blue Book 1988

[3] Rob Avery, *'Strategies for Synchronization in an Evolving Telecommunications Network'*, Proc. of 6th European Frequency and Time Forum, Noordwijk (Netherlands), March 1992

[4] ITU-T Rec. 822, *'Controlled Slip Rate Objectives on an International Digital Connection'*, Blue Book 1988

[5] ITU-T Revised Rec. G.781 *'Structure of Recommendations on Equipment for the Synchronous Digital Hierarchy (SDH)'*, Rec. G.782 *'Types and General Characteristics of SDH Equipment'*, G.783 *'Characteristics of SDH Multiplexing Equipment Functional Blocks'*, Geneva, 1993

[6] S. Bregni, M. D'Agrosa, L. Vattriani, *'Jitter Testing Technique and Results at the Desynchronizer Output of SDH Equipment'*, presented at IEEE IMTC'94, Hamamatsu, Japan, May 10-12, 1994

[7] ITU-T Rec. G.803, *'Architectures of Transport Networks based on the Synchronous Digital Hierarchy'*, Geneva, 1993

[8] Draft ETS DE/TM-3017, *'Generic Requirements for Synchronisation Networks'*, Montpellier, February 1994

[9] ITU-T Rec. G.810, *'Considerationss on Timing and Synchronization Issues'*, Geneva, 1994

[10] ITU-T Rec. G.811, *'Timing Requirements at the Outputs of Primary Reference Clocks Suitable for Plesiochronous Operation of International Digital Links'*, Geneva, 1994

[11] ITU-T Rec. G.812, *'Timing Requirements at the Outputs of Slave Clocks Suitable for Plesiochronous Operation of International Digital Links'*, Geneva, 1994

[12] Peter Kartaschoff, *'Synchronization in Digital Communications Networks'*, Proc. of the IEEE, Vol. 79, No. 7, July 1991

[13] ITU-T Rec. G.703, *'Physical / Electrical Characteristics of Hierarchical Digital Interfaces'*, Geneva, 1991

[14] ITU-T Rec. G.704, *'Synchronous Frame Structures Used at Primary and Secondary Hierarchical Levels'*, Blue Book 1988

[15] ITU-T Rec. G.821, *'Error Performance of an International Digital Connection Forming Part of an Integrated Services Digital Network'*, Blue Book 1988

[16] S. Bregni, M. Carbonelli, M. D'Agrosa, D. De Seta, D. Perucchini, *'Different Behaviour of Frequency Stability Measures in Independent and Synchronized Clocks: Theoretical Analysis and Measurement Results'*, presented at IEEE SUPERCOMM/ICC'94, New Orleans, USA, May 1-5 1994

New Developments in Atomic Frequency Standards (A 1)

ABSOLUTE FREQUENCY MEASUREMENT IN THE VISIBLE AND NEAR INFRARED RANGES

D. Touahri, M. Abed, L. Hilico, O. Acef, A. Clairon and J.-J. Zondy
Laboratoire Primaire du Temps et des Fréquences
LPTF/BNM/OP, 61, av. de l' Observatoire, 75014 Paris, France

Y. Millerioux
Institut National de Métrologie
INM/BNM/CNAM, 292, rue Saint-Martin, 75141 Paris, France

R. Felder
Bureau International des Poids et Mesures
BIPM, Pavillon de Breteuil, F-92312 Sèvres Cedex, France

F. Nez, F. Biraben, L.Julien
Laboratoire Kastler Brossel
ENS/Université de Paris VI, 4 place Jussieu, 75005 Paris, France

Abstract

Future development in frequency synthesis and measurement in the visible and infrared ranges, based on some modifications of the LPTF chain towards an extended tunability, are presented. These modifications will enable to generate up to the 28th harmonic of the frequency of the CO_2 /OsO_4 secondary reference standard at 29 THz. As an immediate implementation, we project to measure the frequency of a laser diode locked on the two-photon transition of the rubidium atom at 778 .1 nm (385 .3 THz, 13th harmonic). Preliminary studies of this system enlighten its excellent metrological performance with potential short term stability and reproducibility more than one order of magnitude better than those of the HeNe/I_2 standard at 633 nm . Besides its potentiality as an accurate secondary frequency standard (on the 10^{-12} level) in the near infrared, the two photon transition is a more appropriate reference for the spectroscopic study of the 2S-nS/nD (n = 6-12) two-photon transitions of atomic hydrogen (750-820 nm), since it synthesizes nearly perfectly the 2S-8D hydrogen frequency.

1 - INTRODUCTION

The absolute frequency of the red HeNe/I_2 laser (473 THz/633 nm) has recently been measured at the Laboratoire Primaire du Temps et des Fréquences (LPTF) by comparison with the synthesized 16th harmonic of a CO_2 laser stabilized by saturated absorption on an OsO_4 line at 29.5 THz /1/. The accuracy was only limited by the poor reproducibility of the HeNe/I_2 standard. The uncertainty of this measurement ($\Delta\nu/\nu = 7.0\ 10^{-12}$) is twenty times better than that of the measurement performed at the NBS in 1983 /2/. This measurement has brought a substantial contribution to the *Mise en Pratique* of the Meter /3/, since the absolute frequency of the "i" component of the R(127)11-5 iodine line deduced from the 1983 measurement has been found 137 kHz wrong-shifted in the red. That redshift has been indirectly corroborated via the consistency of two simultaneous determinations of the Rydberg constant using different and independent frequency chains /4-5/. Since our chain did not introduce significant noise in the measurement (as it can be checked from the Allan variance of the

measurement in fig.1), its ultimate resolution is bounded to the accuracy of the CO_2 /OsO_4 secondary reference (10^{-12}). To benefit from its full resolution , a better secondary standard in the visible should be defined.

The new value of the HeNe/I_2 laser standard has been used for the first frequency measurement of an optical transition in hydrogen (two-photon transition 2S-8S/8D at 385 THz) to be performed without any interferometry, by comparison of the frequency of the Ti:Sa laser used for the atomic excitation with the difference frequency between the two optical standards at 3.39 μm (HeNe/CH_4) and 633 nm (HeNe/I_2) /6/. An unprecedented uncertainty of 2.2 parts in 10^{11} on the knowledge of the Rydberg constant has been deduced from this measurement. The contribution of the HeNe/I_2 frequency uncertainty in that final uncertainty amounts to 0.9 10^{-11} which is already less than the contribution of other uncertainties such as that of the electron-to-proton mass ratio (1.1 10^{-11}) and the uncertainty of 2S Lambshift and QED calculations (1.3 10^{-11}).

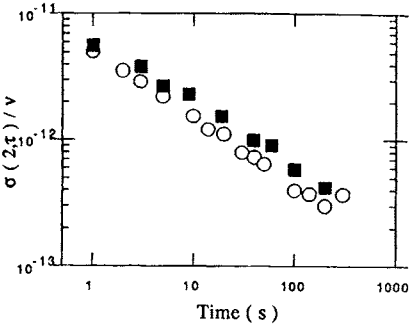


Fig. 1 : Square root of Allan variance :

○ Beatnote between INM12 and BIPM4 He-Ne/I2 lasers (3600 measurements/1s).

■ The measurement of the INM12 frequency obtained from our chain (202 meas. /1s).

Further reduction of this contribution calls for another standard in the near IR, having a better reproducibility. Owing to the coincidence between the 2S-8S/8D two-photon transition frequency and the $5S_{1/2}$-$5D_{3/2,5/2}$ two-photon transition in rubidium, the best candidate for this future standard seems to be a laser diode locked on a hyperfine component of the alkali metal transition.

2 - GaAlAs LASER DIODE STABILIZED ON A HYPERFINE COMPONENT OF THE TWO-PHOTON RESONANCE IN RUBIDIUM (λ=778.1 nm)

The hyperfine structure of $5S_{1/2}$-$5D_{3/2,5/2}$ two-photon transition in Rb (fig.2) was investigated several months ago at the ENS laboratory using a Ti:Sa laser /7/. Very interesting metrological results were obtained, among which narrow linewidths (≈ 500kHz), well resolved structure (the average neighbour -to-neighbour peak separation is 10 MHz) and high signal-to-noise ratio fluorescence signals owing to the detection in the blue on a zero-background noise level. These promising features led us to develop at the LPTF in collaboration with the ENS, INM and BIPM laboratories two devices in which a commercially available GaAlAs laser diode is frequency-stabilized to one of these hyperfine transitions.

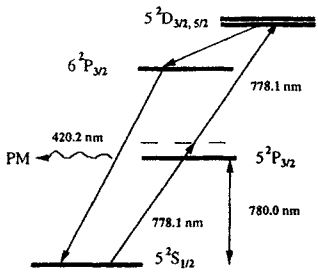


Fig.2 : Partial energy level diagram of Rb. The two-photon transition probability is enhanced by the proximity of the virtual state from the 5 P states (Δ = 1.05 THz).

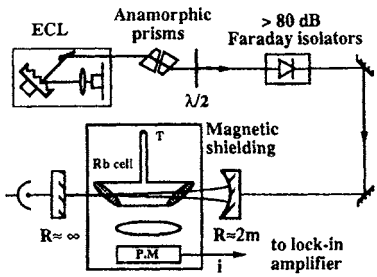


Fig. 3 : Schematic of the Rb two-photon setup. The cavity enables to better define the two counter-propagating modes and to avoid residual first order Doppler effect. The laser frequency is modulated at 70 kHz. A servo loop controls the intracavity power. The fluorescence signal is demodulated by a lock-in amplifier and the error signal fedback to the ECL current and grating servos.

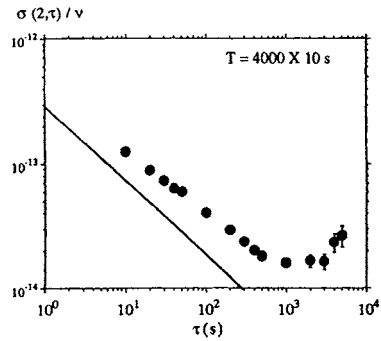


Fig.4: Square-root of the 2-sample Allan variance obtained from 4000 samples of 10s integration time. The line represents $\sigma(2,\tau)/\nu = 3.0\ 10^{-13}/\sqrt{\tau}$. The stability for τ <300s is one order of magnitude lower than that of the HeNe/I_2 standard.

The principle of such a set up is given in fig.3. The laser diode is set in an ECL (extended-cavity laser) configuration while the Rb cell is put in an external cavity to allow for the control of the power and geometry of the beam.

We obtain currently a frequency stability, in term of the Allan variance of 2 parts in 10^{14} for an integration time of 2000s (fig.4). The frequency repeatability is 200 Hz when one of the two devices is reset in the same operating conditions. The frequency difference is 1.7 kHz when the two devices are operating in the same conditions. The light shift is 3 kHz which corresponds to an intra cavity power of 30 mW and a "w" parameter of 0.5 mm /8/.

This is the most important observed spurious effect affecting the frequency of each setup, but the light shift is straightforwardly extrapolable. Further characterization of the metrological performance of this system is under way, but henceforth it is unquestionable that these are at least one order of magnitude better than those of the HeNe/I_2 laser at 633 nm. Consequently this is potentially a new candidate for the realization of the definition of the Meter. Furthermore, other criteria pledge for this candidacy. One is the spectral location of the Rb (5S-5D) transition frequency with respect to the two other available frequency standards at λ =3.392 μm (HeNe/CH_4) and at λ=0.633 μm (HeNe/I_2). The frequency sum of the Rb transition and the HeNe/CH_4 synthesizes quasi perfectly the frequency of the red HeNe/I_2 , providing a mean of intercomparison between the three systems, which constitutes a first step to the settlement of a reference frequency grid in the visible. Another important application is that it can provide a frequency reference for optical telecommunication at 1.556 μm which coresponds to half the frequency of the Rb two-photon line. Finally one may use the fact that the difference frequency between a diode laser at 1.556 μm and the Nd:YAG laser at 1.064 μm (281.6 THz) happens to nearly match the frequency of the IR HeNe/CH_4 standard to define new visible frequency standards near 532 nm, where interesting lines of the B-X transition of I_2 can be used to lock a diode-pumped frequency-doubled monolithic Nd:YAG laser /9/.

3 - EXTENDED SYNTHESIS OF VISIBLE AND NEAR IR FREQUENCIES USING A CO_2 /OsO_4 REFERENCE (29 THz)

Before adressing the specific setup for the Rb two-photon transition frequency measurement, we shall describe the modifications being currently brought to the chain described in /1/ so as to impart it the capacity to synthesize quasi-continuously any frequency from ≈0.5 μm to 1.1 μm, by generation of high order harmonics of CO_2 lasers. For instance, to synthesize the frequency of the two-photon transition in Rb (385.3 THz), we need to generate the 13th harmonic of 29.6 THz.

Fig. 5 shows the modified architecture. Conceptually the left part of the figure is identical to the chain implemented for the red HeNe/I_2 frequency measurement, except for the fact that the 4th harmonic is now optically synthesized through SHG or SFG of the CO_2 lasers (to produce 5 μm radiation) followed by a second stage of SHG in an appropriate nonlinear crystal. This all-optical scheme circumvents the cumbersome use of MIM diodes for the nonlinear mixing of the KCl:Li laser (2.5 μm) with the CO_2 lasers /1/.

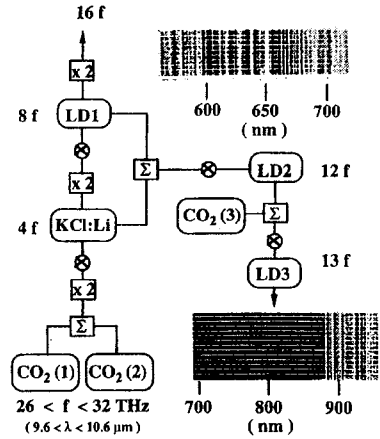


Fig. 5 : Synthesis of the 12th , 13th and 16 th harmonics of CO_2 laser frequencies. and associated spectra. Upper:16th harmonic spectrum with 2 CO_2 lasers (1 and 2) spanning the laser lines range 26-32 THz. Lower:12th and 13th harmonic spectrumwith 3 CO_2 lasers (1, 2 and 3). The resolution is < 0.1 nm (50 GHz) in the darkest part of the spectrum.

A better reliability and a higher flexibility of the chain would result from this modification. The tunability extension of the chain requires the synthesis of other harmonics than the 4th, 8th and 16th. It also takes advantage of the numerous CO_2 lasing lines belonging to the R and P branches at 10 μm. Third harmonic generation of the tunable KCl:Li transfer laser (i.e. 12th harmonic of CO_2) enables to cover the range 800-930 nm where powerful GaAlAs diode lasers are yet available. That frequency tripling is realized indirectly by summing the frequency of the KCl:Li and LD1 lasers. To this point, the frequency of the D_1 and D_2 lines of Cs atom can be measured. The range of synthesized spectrum is further extended to the visible by successive upconversion of the 3rd, 4th... LDi's transfer oscillators with additional CO_2 lasers, generating hence a 29 THz-spaced frequency comb from the near IR to the visible. The 13th to the 16th CO_2 harmonics can be synthesized in that way. Note that this path to the 16th harmonic generation is potentially richer than the direct path on the left part of the figure since the harmonic can be generated with many combinations of CO_2 lines. The tunability of that frequency comb is guaranteed by more than 300 available lasing lines in the 9.4 - 10.6 μm range, using all the C and O isotopes combinations /10/. These CO_2 lasers can be readily phaselocked to a CO_2/OsO_4 reference laser and are easily phaselocked to one another with the use of MIM diodes and submillimetric sources. These numerous lines, in conjunction with the large tunability of the KCl:Li laser (2.4-2.8 μm) enables to scan quasi-continuously the optical spectrum from 700 nm to 900 nm using the 12th and the 13th harmonics of CO_2 . The upper inset in fig.5 displays the 16th harmonic spectrum computed from only one CO_2 (1) laser using the left part of the chain. It is obvious that the synthesis via the 3rd harmonic generation of the color center

laser would provide a denser spectrum. The lower inset shows the combined 12th and 13th harmonic spectrum with the use of three CO_2 lasers. In this case, the average resolution is less than 50 GHz in the near IR, within reach of fast photodetectors. With the current development and progress of frequency difference measurement techniques on the THz scale /11-13/, most often only one CO_2 laser would be necessary to realize the synthesis.

Furthermore, other additional harmonics can be generated with such a setup by sum or difference frequency mixing of two transfer oscillators. The gap from 900 to 965 nm (11th harmonic) can be covered by down conversion of LD2. The 9th harmonic (1-1.2μm) can be synthesized by difference-mixing of LD3 and the KCl:Li laser. The 17th to 21st harmonics in the violet (the latter corresponds to the 4th subharmonic near 486 nm of the 1S-2S H transition) can as well be obtained by SFG of two of the near IR /mid IR transfer oscillators (for instance LD1 and LD3 for the 21st harmonic synthesis). The range can be extended to the blue-UV spectrum (24th to 28th harmonics) by SFG of two of the near IR diode lasers. This last scheme can alternatively be used to halve a frequency difference in the near IR using the optical frequency divider /14/. This impressive potentiality illustrates the versatility of the CO_2/KCl:Li laser-based chain for the frequency synthesis in the whole optical spectrum. The only limitation would come from the availability of the various nonlinear crystal required for the above mentioned parametric processes, bearing in mind that as little as 1 nW of synthesized light is enough for the coherent tracking and counting of every beatnote with a S/N ratio of 20 to 30 dB in a 100 kHz bandwidth /1/.

To this purpose, we have spanned the whole diversity of commercially available nonlinear crystals with a computer code evaluating their performance in terms of their conversion efficiencies (taking into account deleterious effects of birefringence walkoff under critical phasematching) and tunability (the fewest number of crystals for the largest tunability of the chain).

Typically 1-3 mW power of 5 μm radiation is needed to perform the synthesis of ≈ 1 nW light at 2.5 μm. To obtain such a conversion efficiency from cw frequency doubling of a CO_2 laser with an average output power of 1 W per line, the SHG must be performed in a low absorbing crystal and enhanced in a moderate finesse cavity (F ≈12). Silver gallium selenide ($AgGaSe_2$) has been retained from a compromise between nonlinear $\chi^{(2)}$ value and low absorption (<5%cm^{-1}) among the other existing materials, though it may exhibit some thermal lensing effects. When two CO_2 lasers operating on different lines are required for the 5 μm synthesis, a type-II twin-crystal device /15/ quasi-degenerate SFM in GaSe can satisfy the resonance condition for both lines in a single enhancement cavity . The second stage SHG is performed with zinc germanium phosphate ($ZnGeP_2$). The SHG of 5 μm radiation can be occasionally replaced by parametric amplification (2.5μm - 5μm→5μm) if the 5 μm generated from the CO_2 frequency doubling is less than expected. The KCl:Li third harmonic synthesis is temperature-tuned non critically phasematched in silver thiogallate ($AgGaS_2$) over the whole range of tunability of the color center laser, so that a single crystal can sustain the tunability of the chain. An angle-tuned KTP type-II twin-crystal device /15/ can frequency double the KCl:Li from 2.4 to 2.8 μm owing to the wavelength noncritical property of this material /16/. The upconversion of near IR lasers with CO_2 lasers is achievable either by noncritical temperature-tuned phasematching or angle-tuned (at ambient temperature) in $AgGaS_2$ up to its visible transparency cutoff at 700 nm. For further upconversion, gallium selenide (GaSe) which cutoff wavelength is 650 nm is phasematchable owing to its large birefringence. As to the mid IR/near IR SFM (e.g. 8f+12f or 8f+13f) for the green-blue synthesis, material such as KTP, $KNbO_3$, $LiIO_3$ or $LiNbO_3$ can be used, and $LiIO_3$ or BBO for the near IR/near IR SFM into the VUV spectral range above 400 nm. In the latter process, a type-II walkoff-compensated twin device may be again used to enhance the conversion efficiency and perform the SFM in a single cavity, taking advantage of the birefringence experienced by the pump waves in a type-II interaction.

4 - MEASUREMENT OF THE $5S_{1/2}$-$5D_{3/2,5/2}$ TWO-PHOTON TRANSITION FREQUENCY

Fig.6 describes the simpliest case of the general architecture of fig.5 in which a single CO_2 laser operating on the P(32) line of $^{13}C^{16}O_2$ synthesizes approximately the frequency of excitation of the $5S_{1/2}$-$5D_{5/2}$ of Rb transition (13f = 385.285 THz+Δf), where Δf=186 GHz. The CO_2 laser (ν(CO_2)=29,651,657THz) is phaselocked to the laboratory standard (CO_2 /OsO_4 operating on the R(12) line) using a MIM diode and the 6th harmonic of a klystron at 92 GHz. The rather large frequency mismatch Δf=186 GHz at the end of the chain results from the choice of using a single CO_2 laser to perform the synthesis. It will be measured with a Schottky diode and the 2nd harmonic of a Gunn diode at 93 GHz. To obtain a high S/N ratio for the beatnotes, both near IR transfer lasers are frequency-locked to a high finesse (F=20000) vacuum suspended cavity using the FM sideband phase modulation technique. A beatnote width of 500 Hz has been already obtained from two diodes at 780 nm locked simultaneously to an invar cavity of a lower finesse. This cavity length can be further servoed to the Rb reference.

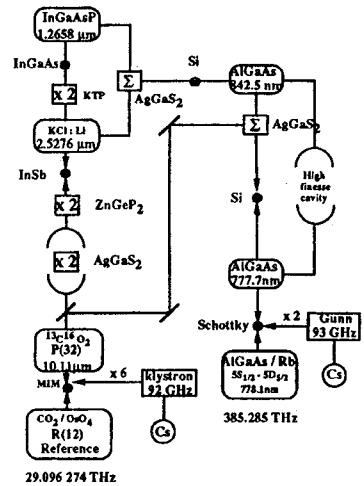


Fig. 6: Frequency chain for the synthesis of the $5S_{1/2}$-$5D_{5/2}$ two-photon transition frequency of Rubidium (=13 f [$^{13}C^{16}O_2$/ P(32)] -186 GHz).

Experimental detection of a Schottky beatnote resulting from high order harmonic mixing between two free running near IR lasers and a microwave source has been realized at the LPTF /11/. Frequency differences up to 1 THz between the two lasers have been measured with a S/N ratio exceeding 2 dB in 1 MHz analyzer bandwidth using very low microwave power. Cavity locking would certainly improve the S/N ratio. An optical frequency comb generator (OFCG) /12-13/ is also suitable for this difference frequency measurement. We intend to develop such a device at 778 nm using a resonant MgO:$LiNbO_3$ phase modulator driven at 10 GHz by a low noise DRO. This device utilizes multiple pass of the optical field inside a resonant cavity containing the phase modulator to enhance the effective phase modulation index. This enhancement results in the generation of very high order laser sidebands whose frequencies, known from the laser carrier reference frequency and the sideband order, can be compared to the frequency of the laser under test.

From the calculated output spectrum of an OFCG at 778 nm, shown in fig.7, a wide-span range up to 6-7 THz is expected if one sets the minimum sideband power for a beatnote detection to be -80 dB of the input carrier power (100 mW). The order of the generated sidebands is mainly limited by the material dispersion ($dn/d\nu = 2.38\ 10^{-15}$ Hz^{-1} at $\lambda = 778$ nm) and by the effective carrier power coupled into the cavity. The power transfer from the carrier to the sidebands is indeed interpreted by the loaded cavity as additional losses on the intracavity carrier field, causing a strong impedance mismatch between the input and cavity leakage field.

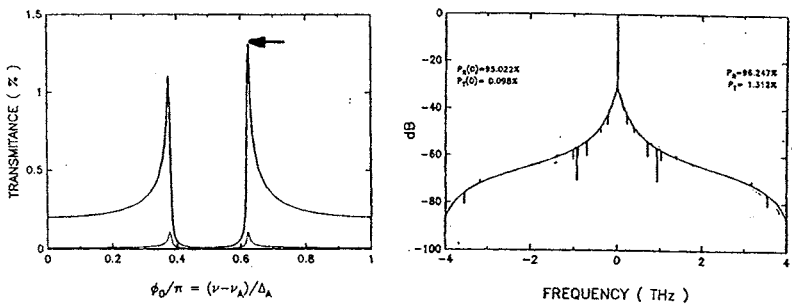


Fig. 7: Theoritical output power versus laser/cavity frequency detuning (a) and spectrum (b) of an OFCG at 778 nm using a $MgO:LiNbO_3$ phase modulator (Lc=25 mm, modulation index $\beta=0.37\ \pi$, mirror reflectivity R=0.99, intracavity loss (double pass) of 1 - exp(-2δ) =1.1% resulting in a finesse=150, ν_{RF}=10 GHz). Δ_A is the cavity free spectral range ($\Delta_A = \nu_{RF}/7$) . The spectrum in (b) corresponds to the locking point indicated by an arrow in (a).

This impedance mismatch results in a reflection of ≈98% of the incident power. By placing a second short cavity, only resonant for the carrier, in front of the modulator cavity a better power coupling should be expected.

5 - MEASUREMENT OF THE 2S-nS/nD (n=6-12) HYDROGEN FREQUENCIES

Though all these frequencies can be directly real time measured using our tunable chain (an optical fiber link is planned between the LPTF and the ENS laboratory where the hydrogen spectroscopy is performed), it is preferable for the sake of flexibility to measure them starting from the known rubidium two-photon reference. The problem is then restricted to difference frequency measurements in the near IR (fig.8).

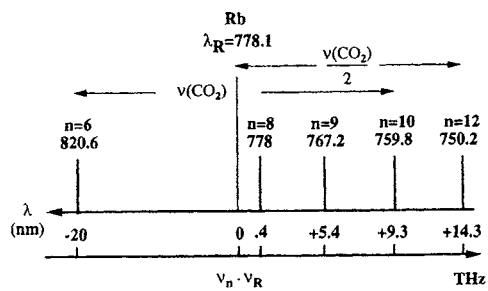


Fig. 8 : Hydrogen 2S-nS/nD two-photon transitions respective to the 5S-5D Rb lines.

The 2S-8S/8D transition ($\Delta\nu \geq 40$ GHz) can be straightforwardly measured with a wideband Si photodiode or a point contact Schottky diode. The other transitions requires the implementation of an OFCG at $\lambda = 778$ nm, capable of measuring frequency differences up to 6 or 7 THz. The measurement of ν_9 (n=9) can be directly performed with this OFCG. The other remaining frequencies, ν_6, ν_{10} and ν_{12} require the use of one stage divider combined with an upconversion process with a CO_2 laser. The easiest frequency to be measured without the use of a comb generator is the 2S-12S/12D, taking advantage that its frequency shift from the rubidium reference equals half the frequency of a CO_2 laser operating at 10.4 μm. A simple measurement setup is shown in fig.9, in which a diode laser at 808 nm is used to compare ν_{12} to both references. The same setup can be used for the measurement of ν_6 and ν_{10} , with the frequency of the diode laser around respectiveley ν_{10} or ν_6, but this time the frequency difference Δ, of the order or 5.3 THz, requires the assistance of an OFCG .

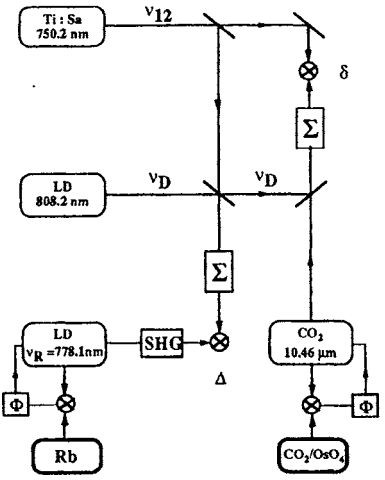


Fig. 9 : 2S-12S/12D frequency measurement setup: $\nu_{12} = \nu_R + \nu(CO_2)/2 + (\Delta-\delta)/2$

6 - CONCLUSION

We have presented the currently modified status of the LPTF chain so as to measure the absolute frequency of any radiation in the whole optical spectrum, through the generation of a tunable 29 THz-spaced frequency comb. This tunability illustrates the usefulness of a secondary reference standard at 10 μm, namely frequency stabilized CO_2 lasers. These lasers can be operated between 9.4 to 10.6 μm on numerous lines and stabilized to narrow reference rovibrational lines of molecules such as OsO_4, SF_6 or SiF_4. In spite of its heaviness, such a *classical* chain remains at the present stage of development of frequency measurement techniques, the only tool for absolute determination of optical radiation frequencies. An implementation of the modified architecture of the laboratory chain to the frequency measurement of a promising candidate for a future secondary standard - a diode laser locked to the Doppler-free two-photon resonance in atomic Rb vapour - is described. The expected accuracy is 10^{-12} , corresponding to the present accuracy of the laboratory secondary IR standard. In the future, we plan to improve this accuracy and to perform a new measurement of that reference standard from the primary time/frequency standard.

This prospective 10-fold gain in the accuracy, with respect to the red HeNe/I_2 standard, of a new near IR frequency standard should bring a consequent improvement of the definition of the length unit and in spectroscopic measurements. In particular, the uncertainty on the Rydberg constant, which is a fundamental atomic energy scale, would be henceforth limited by the atomic QED corrections, which would provide a serious test for their validity.

REFERENCES

/1/ - O. Acef, J.J. Zondy, M. Abed, D.G. Rovera, A.H. Gérard, A. Clairon, Ph. Laurent, Y. Millerioux and P. Juncar, Opt. Comm. **97** (1993) 29.

/2/ - D.A. Jennings, C.R. Pollock, F.R. Petersen, R.E. Drullinger, K.M. Evenson, J.S. Wells, J.L. Hall and H.P. Layer, Opt. Lett. **8** (1983) 136.

/3/ - T.J. Quinn, Metrologia **30** (1994) 523.

/4/ - F. Nez, M.D. Plimmer, S. Bourzeix, L. Julien, F. Biraben, R. Felder, O. Acef, J.J. Zondy, Ph. Laurent, A. Clairon, M. Abed, Y. Millerioux and P. Juncar, Phys. Rev. Lett. **69** (1992) 2326.

/5/ - T. Andreae, W. König, R. Wynands, D. Leibfried, F. Schmidt-Kaler, C. Zimmerman, D. Meschede and T.W. Hänsch, Phys. Rev. Lett. **69** (1992) 1932.

/6/ - F. Nez, M.D. Plimmer, S. Bourzeix, L. Julien, F. Biraben, R. Felder, Y. Millerioux and P. De Natale, Europhys. Lett. **24** (1993) 635.

/7/ - F. Nez, F. Biraben, R. Felder and Y. Millerioux, Opt. Comm. **102** (1993) 432.

/8/ - Y. Millerioux, D. Touahri, L. Hilico, A. Clairon, R. Felder, F. Biraben and B. de Beauvoir, to be published in Opt. Comm. (1994).

/9/ - A. Arie, S. Schiller, E.K. Gustafson and R.L. Byer, Opt. Lett. **17** (1992) 1204.

/10/ - C. Freed, L.C. Bradley and R.G. O'Donnel, IEEE J. quant. Electron.**QE-16** (1980) 1195.

/11/ - O. Acef, F. Nez and D.G. Rovera, Opt. Lett. (to be published).

/12/ - L.R. Brothers, D. Lee and N.C. Wong, Opt. Lett. **19** (1994) 245.

/13/ - M. Kourogi, K. Nakagawa and M. Ohtsu, IEEE J. Quant. Electron. **29** (1993) 2693.

/14/ - H.R. Telle, D. Meschede and T.W. Hänsch, Opt. Lett. **15** (1990) 532.

/15/ - J.J. Zondy, M. Abed and K. Salah, submitted to J. Opt. Soc. Am. **B** (1993).

/16/ - J.J. Zondy, M. Abed and A. Clairon, to be published in J. Opt. Soc. Am. **B**.

Laser Cooled Mg Atoms for an Optical Frequency Standard

V. Rieger, K. Sengstock, U. Sterr, J.H. Müller, D. Bettermann, F. Ruschewitz and W. Ertmer

Institut für Angewandte Physik der Universität Bonn
Wegelerstr. 8, 53115 Bonn, Germany
FAX: +49-228-733474

Abstract: The magnesium $3s^2(^1S_0)$-$3s3p(^3P_1)$ transition is presented as an attractive candidate for an optical frequency standard. Ultra-high resolution spectroscopic measurements have been performed on laser trapped atoms using a time-domain version of optical Ramsey spectroscopy. Observed fringes have a resolution down to 1 kHz and the spectroscopic accuracy in determining the line center has reached $2 \cdot 10^{-15}$. A description of cooling and trapping and of high resolution spectroscopy with Mg is given.

1. Introduction

For many applications both in fundamental and applied research, improvements in the current atomic frequency standards have become of great concern. Particularly for measurements at optical frequencies, the stability that can be derived from the present cesium frequency standard already shows deficits for short integration times. Thus increasing effort is being devoted to the development of optical frequency standards /1,2/. With the advent of laser cooling and trapping techniques this research has made great progress in recent years. A trapped cloud of cold neutral atoms constitutes an excellent spectroscopic sample of extremely narrow velocity width at or near zero mean velocity in combination with densities equal or larger than those obtainable in atomic beams. These factors help to overcome the influence of the 2nd order Doppler effect and the limited interaction times available at thermal energies. Several laboratories have already demonstrated that impressive results can be achieved in the microwave region /3/, and recently optical high resolution spectroscopy has been performed using laser trapped atoms /4/.

A successful integration of techniques from laser manipulation and high resolution spectroscopy requires the careful choice of a suitable atomic level scheme. Cooling and trapping is best achieved using a fast, closed transition while for spectroscopy

'forbidden' transitions to long-lived states are ideal. Different candidates have been discussed in the literature /5/, and several groups are exploring the interesting opportunities contained in the singulet and triplet states of alkaline earth elements /6-12/. The $^1S_0$ ground state is connected via a fast transition to the $^1P_1$ excited state and through a spin-forbidden transition to the metastable $^3P_1$ level. For magnesium the cooling transition is closed, whereas all heavier alkaline earth elements have low-lying ^1D-states, which limit accumulation times due to optical pumping. Except for beryllium, where little data is available, magnesium also offers the narrowest intercombination transition with a linewidth as low as 31 Hz /13/.

To explore these interesting possibilities offered by the Mg level scheme we have built a magneto-optical trap where up to $4 \cdot 10^5$ atoms can be stored at temperatures as low as 3 mK. In order to take advantage of the low velocity of the atoms and the long natural life time of the metastable state, the trapped atoms are probed *in situ* by a time domain version of the four zone optical Ramsey method /14/: A sequence of four light pulses from a highly stable dye laser is applied to the trap volume while the trapping fields are turned off. Atoms not excited to the metastable state during the spectroscopic period are recaptured as soon as the trapping fields are turned on again. Their fluorescence on the strong trapping transition is detected with a photomultiplier and provides a signal that is a 'quantum amplified' measure of the absorption on the intercombination transition. In our experiments the resolution has already reached 1.5 kHz, while the accuracy is at the Hertz level, as most types of systematic line shifts are significantly reduced for pulsed spectroscopy on trapped atoms.

The following presentation gives a review of this research and discusses in detail the results that have been achieved in the context of optical frequency standards. For a related description see ref. /15/.

2. Presentation of Experiment

2.1. The magnesium level scheme

The relevant transitions for frequency standards research within the ^{24}Mg level scheme are shown in fig. 1. The intercombination transition $3s^2(^1S_0)$-$3s3p(^3P_1)$ at 457.1 nm is a well-suited frequency reference. The 5.1 ms lifetime of the upper-level /13/ offers a resolution well beyond the kHz-level. The fast $3s^2(^1S_0)$-$3s3p(^1P_1)$ transition in the

singulet system (upper-level lifetime 2.02 ns) is closed and therefore ideal for laser cooling and trapping. One interesting detection scheme for excitation on the clock transition is found in the $3s3p(^3P_1)$-$3s5d(^3D_2)$ transition at 284.9 nm, as the $^3D_2$-state decays in a fast cascade via the $3s4s(^3S_1)$ state under the emission of 517.4 nm fluorescence.

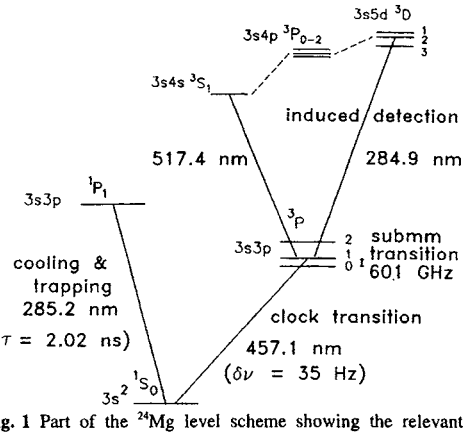


Fig. 1 Part of the ^{24}Mg level scheme showing the relevant transitions for cooling and high resolution spectroscopy.

Also of interest is the submillimeter transition $^3P_1$-$^3P_0$ within the 3s3p finestructure multiplet. It has been used by Godone and Novero /16/ for thermal beam Ramsey spectroscopy. Such research can also benefit substantially from the results of laser cooling and trapping presented here.

2.2. Magneto-optical trapping of magnesium

To prepare a dense ensemble of cold atoms we use a modified version of the well known magneto-optical trap (MOT) /17/. The trap can be filled either from a thermal or a laser-slowed atomic beam. The necessary UV-radiation is generated by frequency doubling the output of two single-mode ring dye lasers at 570 nm.

One of the intensity and frequency stabilized laser beams is guided into a UHV vacuum chamber and steered by a combination of mirrors and quarter-wave plates as indicated in fig. 2 to form six orthogonal circular polarized waves as required for magneto-optical trapping. By using a single beam for the trap and placing all optical components in the vacuum large accumulative power losses of the UV-light can be avoided. Using a trapping beam diameter of 3.4 mm we are able to store up to $4 \cdot 10^5$ atoms from a laser decelerated beam in an approximately gaussian shaped cloud with a diameter of 250 µm corresponding to a peak density of 10^{10} cm^{-3}. Due to the large velocity capture range of 50 m/s of our MOT, trapping is possible directly from a thermal atomic beam. This leads to a much simpler experimental setup, where only one UV-laser beam is needed, although the number of trapped atoms is reduced by three

orders of magnitude.

The dynamics of the trap density is generally well described by a rate equation of the form /18/

$$dn/dt = R/V - \alpha n - \beta n^2 . \quad (1)$$

Here R denotes the external loading rate, V the trap volume and α and β loss coefficients due to collisions of the trapped atoms with background molecules and other trapped atoms, respectively. In the absence of collisions between trapped atoms (β=0, or *n* small) eq. (1) yields an exponential loading or decay behavior of the trap. This exactly is observed experimentally, so that we conclude that binary collisions among Mg atoms are negligible for the present operating conditions. This is relevant for frequency standards research, as collisions between trapped atoms have already been found to lead to significant pressure shifts in high resolution spectroscopy /19/. The typical value for the decay constant α at a background pressure of 10^{-8} mbar in our apparatus is $\alpha = 2.5\ s^{-1}$.

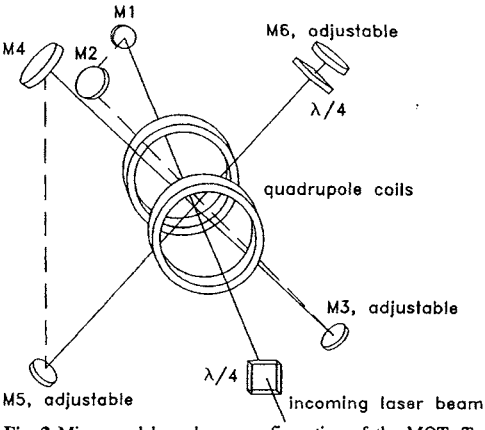


Fig. 2 Mirror and laser beam configuration of the MOT. To economize UV-light power a single beam is threaded through the trap using three adjustable mirrors (M).

2.3. Optical Ramsey Spectroscopy

The transfer of Ramsey's method of separated oscillating fields to the optical domain /20,14/ allows to use the phase information of the atomic coherence for ultra-high resolution spectroscopy. A detailed calculation of the excitation probability P as a function of the laser frequency can be performed using the theory developed by Bordé *et al.* /21/. The probability P can be described as the sum of two coherent parts (recoil doublet of the Ramsey fringes) plus an incoherent background in the form of a Doppler-profile and a saturation dip:

$$P = A_+(\Delta)\cos[2T(\Delta+\delta)+\Phi] + A_-(\Delta)\cos[2T(\Delta-\delta)+\Phi] + A_i(\Delta) \quad (2)$$

Here $A_\pm(\Delta)$ and $A_i(\Delta)$, describe amplitude factors for the low-frequency (+) and high-frequency recoil component (-) and the incoherent background (*i*) that depend slowly on

the difference $\Delta = \gamma\omega_L - \omega_0$ between the laser frequency ω_L and the atomic frequency ω_0 corrected for the relativistic second order Doppler effect by the Lorentz factor $\gamma = (1-v^2/c^2)^{-1/2}$; $\delta = \hbar k^2/2m$ denotes the recoil shift, for an absolute value k of the wave-vector of the laser beams and the atomic mass m. The phase Φ is the total phase difference of the four laser pulses taken at the interaction times with the atom. The periodicity in frequency space of the Ramsey fringes is determined by the time T between the copropagating laser pulses. In a pulsed experiment such as the one described here, it can be set with high accuracy, whereas in any beam experiment it is given through the time of flight of the atoms between the lasers beams.

The narrow-bandwidth 457.1 nm radiation is produced by a home-built stilbene 3 dye laser frequency stabilized to a highly stable reference cavity. Typically 100 mW of blue light is produced within a linewidth of less than 1 kHz, as can be estimated from spectroscopic results. Long-term frequency stability can be achieved by additionally locking the laser frequency to a Ramsey resonance obtained on a thermal atomic beam.

The complete setup of the experiment is shown in fig. 3. Acoustooptic modulators

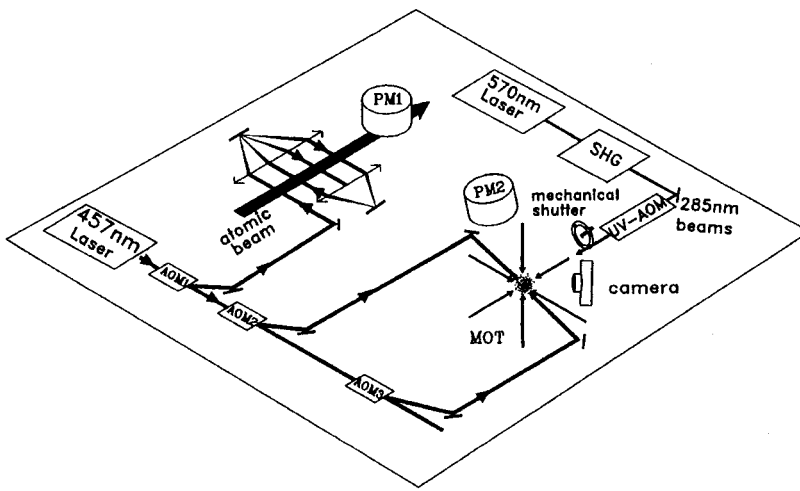


Fig. 3 Schematic view of the experimental setup for parallel spectroscopy on a thermal atomic beam and laser cooled Mg atoms in a magneto-optical trap (MOT). The laser light is switched by acoustooptic modulators (AOMs), to achieve sufficient extinction the UV beam is additionally shuttered by a mechanical chopper. Fluorescence is observed with photomultipliers (PMTs) and trap diagnostics performed with the help of a CCD-camera.

(AOMs) direct the light from the laser to a thermal beam apparatus and from two directions to the MOT. AOM1 is operated at a variable frequency to permit a frequency offset between the two spectroscopic signals. The whole experiment is time-controlled by purpose-built electronic pulse generators and interfaced to a computer.

2.4. Detection through *electron-shelving* and *atom-recycling*

Figure 4 illustrates the principle of pulsed Ramsey spectroscopy on atoms trapped in a MOT: Atoms that have accumulated in the trap are irradiated by four pulses of light at 457.1 nm after the UV trapping light has been turned off. Copropagating pulses are separated by the time T, each pulse has a width τ. The magnetic fields of the MOT are turned off simultaneously with the trapping light and a homogeneous magnetic field is turned on to select the $\Delta m=0$ component of the transition. Atoms that have been excited to the metastable $^3P_1$ state, corresponding to a 'shelving' of the electron in this level, and have sufficient radial velocity leave the trap before decaying back into the ground state. The remainder, which typically are 98% of the initially present atoms, are recaptured and thus recycled for the next spectroscopic excitation process, and their UV fluorescence is monitored by a photomultiplier tube to yield the spectroscopic signal, completing one measurement cycle. The whole experiment is repeated at a rate of up to 95 Hz.

The influence of the blue light on the trap dynamics can be understood as an additional loss mechanism and described as a constant loss rate. For each new frequency value of the 457.1 nm light a new equilibrium number N_e of trapped atoms is reached which is given by /4/

$$N_e = \frac{N_0}{1 + P\tau_0/\tau_C} \quad . \qquad (3)$$

Here N_0 and τ_0 denote the number of atoms and the lifetime, resp., of the undisturbed trap and τ_C the duration of one cycle (i.e. τ_C^{-1} = 95 Hz). Atom-recycling leads to an amplification of the excitation

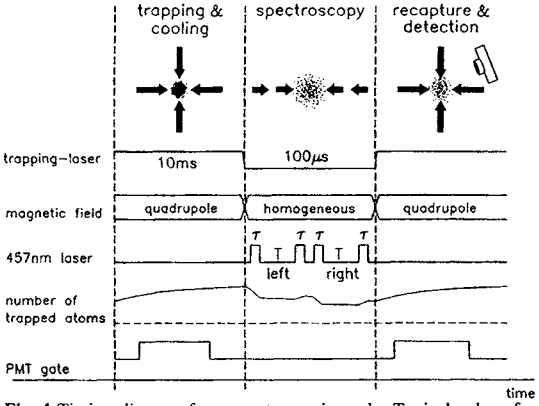


Fig. 4 Timing diagram for a spectroscopic cycle. Typical values for the duration of different sections of the cycle are indicated

probability P by the factor τ_0/τ_c which can take maximum values of 70. As a result of the combined effects of electron-shelving and atom-recycling it is possible to detect an excitation probability P = 1% with only about 1000 trapped atoms as a decrease in UV-fluorescence of 40%. with an integration time of 2 s per frequency point.

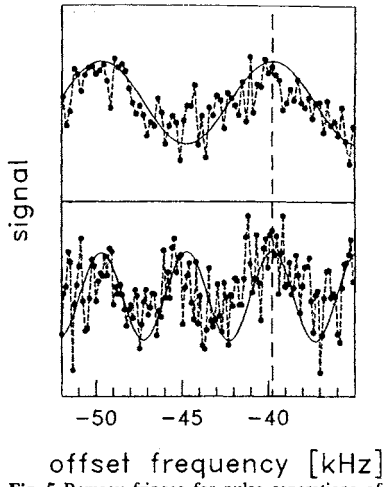


Fig. 5 Ramsey fringes for pulse separations of T=50μs (top) and T=100μs (bottom), obtained with integration times per point of 1.5s and 3s, resp.

3. Experimental Results

3.1. High resolution Ramsey fringes

Typical results for Ramsey fringes obtained on the MOT are presented in fig. 5. For clarity the incoherent background has been subtracted from the signals. The fringe periodicity can be varied over a wide range, and the highest resolution (FWHM) obtained so far is 1.3 kHz. The position of the fringe maximum is characterized by a high degree of accuracy with respect to the 'true' transition frequency as will be discussed in more detail below. This fact is exemplified by a comparison of the MOT signal with Ramsey fringes obtained on a the thermal beam apparatus as shown in fig. 6. As a result of the influence of the second order Doppler effect the thermal signal suffers from an asymmetry and shift of the fringe maximum relative to the 'true' position. In a series of such comparisons the size of second order Doppler effect for thermal atoms could be measured to (-1.5±0.4) kHz /4/, where the error results from phase uncertainties in the *thermal* signal.

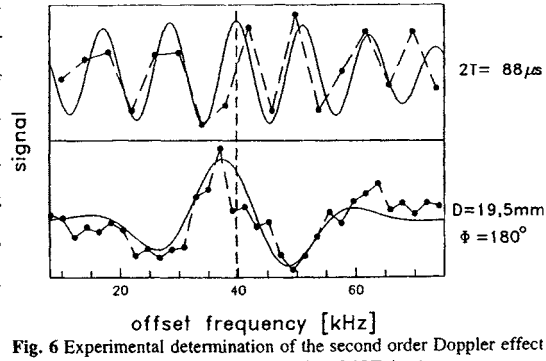


Fig. 6 Experimental determination of the second order Doppler effect for thermal atoms (bottom) vs. atoms in a MOT (top)

3.2. Frequency stability

Locking the frequency of our laser oscillator to one of the spectroscopic resonances yields a relative frequency stability $\delta\nu/\nu$ that can be determined from the resolved line Q, the signal-to-noise ratio S/N and the integration time τ according to

$$\frac{\delta\nu}{\nu} \approx \frac{1}{Q}\frac{1}{S/N}\frac{1}{\sqrt{\tau/sec}} \quad . \qquad (4)$$

Figure 7 presents examples of high resolution scans both on trapped and on thermal atoms. From the resolved linewidth and signal-to-noise ratio the following relative frequency stabilities can be inferred:

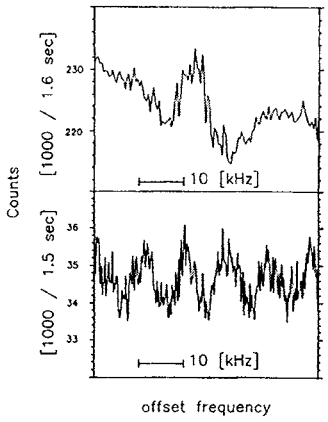


Fig. 7 Examples of high resolution fringes on a thermal atomic beam (top) and cooled atoms in a MOT (bottom)

Tab. 1 Frequency stability of a laser oscillator frequency locked to optical Ramsey fringes

	atomic beam	trap
$\delta\nu/\nu(\tau=20s)$	$3.4\cdot10^{-13}$	$8.7\cdot10^{-13}$

At present the superior resolution obtained at the trap cannot be utilized in full because of the lower signal-to-noise ratio.

3.3 Frequency accuracy

In evaluating the frequency accuracy of an oscillator locked to one of the Ramsey fringes of fig. 7, it is important to consider possible line shifts that influence the position of the spectroscopic signal. For thermal ensembles the error budget is often dominated by uncertainties as a result of the Doppler effect. On the MOT these scale with the *mean* velocity of the cloud of trapped atoms. In studies of the velocity distribution we have determined that this is less than 5 cm/s so that the associated errors are reduced by four orders of magnitude. Due to the substantial reduction of these velocity dependent errors pulsed Ramsey spectroscopy enters a new regime, where uncertainties of 1 Hz or less become significant. Table 2 summarizes the major of them and gives a comparison with the error budget of thermal beam spectroscopy /15/.

Tab. 2 Error budget for high resolution optical Ramsey spectroscopy on Mg atoms

effect	trapped atoms (T=3 mK)	thermal beam (T=700 K)
signal phase uncertainty / residual 1st order Doppler effect	1 Hz	400 Hz
uncertainty due to the 2nd order Doppler effect	< 1 mHz	300 Hz
asymmetric Lamb-dip background	< 1 Hz	50 Hz
wave front curvature	< 1 Hz	10 Hz
residual electric field	< 1 Hz	< 1 Hz
magnetic field uncertainty	< 1 Hz	< 1 Hz
Sagnac effect	5 mHz	< 1 Hz
gravitational effect	1 Hz	1 Hz
total uncertainty	1.5 Hz	500 Hz
$\Delta\nu / \nu_0$	$2 \cdot 10^{-15}$	$7 \cdot 10^{-13}$

3.5. Direct comparison of thermal beam and MOT Ramsey spectroscopy

In order to obtain an independent and direct evaluation of the difference in frequency stability between a laser stabilized to a thermal beam Ramsey resonance and one stabilized with the help of a cooled and trapped ensemble we have compared the frequency of the laser locked to the thermal beam resonance with the center of a Ramsey fringe obtained on the MOT.

The spectroscopic signal of both the thermal beam and the trap was measured simultaneously for roughly 0.8 s each at +20 kHz, 0 kHz, and - 20 kHz relative to the approximate minimum of the Ramsey resonance used for stabilization. The frequency spacing for interrogation was determined by the periodicity of both signals, which was chosen to be 80 kHz on both apparatus. After such a measurement cycle at three optical frequencies the difference in count rates at ±20 kHz was determined, and depending on the sign of this difference the offset frequency between the laser and the cavity-resonance changed by fixed steps of ±200 Hz. By taking small frequency corrections low servo gain was obtained and thus for short times the shot-noise induced frequency fluctuations

could be avoided. The center frequency of the trapped atoms Ramsey resonance relative to the reference cavity was determined from a parabolic fit to the count and frequency data at the three measurement points. By deliberately detuning the laser frequency we have found that the parabolic fit yields a good approximation to the mainly cosinusoidal dependence of the fringe shape as long as the frequency difference between the dye laser and the trapped atoms Ramsey resonance is not too large.

From the difference between the calculated center frequency of the Ramsey fringe on the MOT and the actual offset frequency we were able to compute the two-sample (Allan) variance of the frequency difference between the independent Ramsey resonances. The result in fig. 8 shows a proportionality to $\tau^{-1/2}$ for integration times of less than 300 s and a distinct deviation for longer times. The linear dependence on $\tau^{-1/2}$ can be attributed to shot-noise. Calculating the contributions from the trap and the thermal atomic beam we find the measured Allan variance in accordance with the value derived from the trap data and a factor of two above the thermal beam value. Since our measurement is only sensitive to the more instable of the two resonances the data at short times therefore mainly shows the shot-noise behavior of the trap. However, the flat region at large integration times must be attributed to long term variations in the frequency of the thermal beam Ramsey resonance. From independent measurements we can assume that pointing instabilities and power fluctuations of the laser beam have caused this behavior resulting in a long term frequency error of 500 Hz. Thus our measurement yields evidence for the superiority of cold atom optical frequency standards over their thermal counterparts.

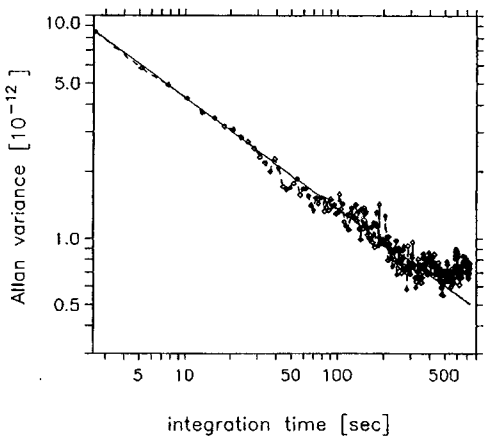


Fig. 8 Allan variance of the frequency difference between the laser stabilized to a minimum of a thermal beam Ramsey fringe and the position of a Ramsey fringe obtained on atoms in a MOT

4. Conclusion and Future Prospects

We have reported studies on ultrahigh resolution spectroscopy

on laser cooled magnesium atoms. Significant progress has been made in the integration of laser manipulation and high resolution spectroscopy techniques. We have developed a novel way to utilize laser trapping not only for the preparation of high quality spectroscopic samples but also for efficient detection of high resolution signals from few neutral particles. Ramsey fringes on trapped, slow atoms constitute a new frequency reference which shows very small systematic line shifts and has already achieved a resolution better than previous thermal beam techniques. In comparing both methods we have found evidence for long term frequency errors in thermal beam Ramsey resonances.

Several improvements should bring considerable progress in the near future. Spectroscopy on the trap is at present mainly limited by S/N, which is governed by shot-noise fluctuations in the number of atoms. Thus increasing the number of trapped atoms will be the next step. Also, phase errors of the 457.1 nm laser beams will be minimized through an additional external optical frequency stabilizer /22/ and drastically improved mechanical stability. With these improvements it seems possible to reach a stability of $\sigma \leq 10^{-14}$ within averaging times of 1 s or less.

5. Acknowledgements

We have appreciated the experimental contributions of J.v. Würzen, F. Dingler, and A. Pabst. Financial support was provided by the Deutsche Forschungsgemeinschaft within the "Schwerpunktprogramm Präzisionsspektrometrie mit laserpräparierten Atomen und Ionen".

References:

/1/ J. Helmcke, A. Morinaga, J. Ishikawa, F. Riehle: IEEE **IM 38**, 524 (1989)

/2/ S. Rolston, W.D. Phillips: Proc. IEEE **79**, 943 (1991)

/3/ J.L. Hall, M. Zhu, P. Buch: J. Opt. Soc. Am. **6**, 2194 (1989); M. Kasevich and S. Chu: Phys. Rev. Lett. **63**, 612 (1989); A. Clairon, C. Salomon, S. Guellati, W. Phillips: Europhys. Lett. **16**, 165 (1991); G. Santarelli, P. Laurent, A. Nadir, A. Clairon, S.N. Lea: these proceedings

/4/ K. Sengstock, U. Sterr, G. Hennig, D. Bettermann, J.H. Müller, W. Ertmer: Opt. Comm. **130**, 73 (1993)

/5/ W. Ertmer, R. Blatt, J.L. Hall in: *Laser-Cooled and Trapped Atoms* (Proceedings of the Workshop on Spectroscopic Applications of Slow Atomic Beams), W.D. Phillips ed., Natl. Bur. Stand. (U.S.) Spec. Publ. **653** (1983) p. 154

/6/ R.L. Barger, J.C. Bergquist, T.C. English, D.J. Glaze: Appl. Phys. Lett. **34**, 850 (1979)

/7/ J. Helmcke, D. Zevgolis, B.Ü. Yen: Appl. Phys. B **28**, 83 (1982)

/8/ A. Morinaga, F. Riehle, J. Ishikawa, J. Helmcke: Appl. Phys. B **48**, 165 (1989)

/9/ U. Sterr, K. Sengstock, J.H. Müller, D. Bettermann, W. Ertmer: Appl. Phys. B **54**, 341 (1992)

/10/ A. Witte, Th. Kisters, F. Riehle, J. Helmcke: J. Opt. Soc. Am. B **9**, 1030 (1992)

/11/ T. Kurosu, M. Morinaga, F. Shimizu: Jpn. J. Appl. Phys. **31**, L273 (1992)

/12/ N. Ito, J. Ishikawa, A. Morinaga: J. Opt. Soc. Am. B **8**, 1388 (1991)

/13/ A. Godone and C. Novero: Phys. Rev. A **45**, 1717 (1992)

/14/ Ch.J. Bordé, S. Avrillier, A. van Lerberghe, Ch. Salomon, D. Bassi, G. Scoles: J. de Physique **42**, C8-15 (1981)

/15/ K. Sengstock, U. Sterr, J.H. Müller, V. Rieger, D. Bettermann, W. Ertmer: submitted to Appl. Phys. B, special issue on high resolution spectroscopy

/16/ A. Godone, C. Novero: Metrologia **30**, 163 (1993)

/17/ E.L. Raab, M. Prentiss, A. Cable, S. Chu, D.E. Pritchard: Phys. Rev. Lett **59**, 2631 (1987)

/18/ M. Prentis, A. Cable, J.E. Bjorkholm, S. Chu, E.L. Raab, D.E. Pritchard: Opt. Lett. **13**, 452 (1988)

/19/ K. Gibble and S. Chu: Phys. Rev. Lett. **70**, 1771 (1993)

/20/ Y.V. Baklanov, V.P. Dubetski, V.P. Chebotayev: Appl. Phys. **9**, 171 (1976)

/21/ Ch.J. Bordé, Ch. Salomon, S. Avrillier, A. Van Lerberghe, Ch. Bréant, D. Bassi, G. Scoles: Phys. Rev. A **30**, 1836 (1984)

/22/ M. Zhu and J.L. Hall: J. Opt. Soc. Am. B **10**, 802 (1993)

Frequency Standard Investigations on Trapped Ytterbium Ions Using Semiconductor Laser Sources

Chr. Tamm and D. Engelke
Physikalisch-Technische Bundesanstalt
Bundesallee 100, D-38116 Braunschweig

SUMMARY

We discuss experiments on laser cooling of ytterbium ions (Yb^+) in a radiofrequency quadrupole trap (Paul trap). In an experimental system entirely based on semiconductor diode lasers, single $^{172}Yb^+$ ions are cooled to kinetic temperatures of less than 0.8 K. The system is also capable of laser cooling of $^{171}Yb^+$ without requiring an additional radiation source to avoid optical hyperfine pumping. The specific hyperfine-sublevel structure of $^{171}Yb^+$ permits an optical-clock scheme where the reference transition is at a wavelength of λ = 435 nm, suitable for using a frequency-doubled diode laser as the local oscillator.

1. INTRODUCTION

Ion traps offer unique possibilities for improved frequency standards in the microwave and optical frequency range since long interaction times between atomic systems and external radiation fields are possible in an environment with small perturbations. A potential accuracy in the range of 10^{-18} is estimated for optical frequency standards based on single laser-cooled ions in electrodynamic traps /1/.

Ions with alkali-like level system permit sensitive optical detection and efficient laser cooling. Among these ions, $^{199}Hg^+$ and $^{171}Yb^+$ play a particular role since reference transitions without linear Zeeman effect are available in level systems with relatively simple hyperfine and magnetic-sublevel structure. More easily than Hg^+, Yb^+ can be laser cooled with the use of semiconductor laser sources. The use of diode lasers offers new experimental possibilities and permits the

realization of particularly compact and reliable experimental systems. For Yb^+, several optical-clock schemes have been proposed /2,3/. Microwave spectroscopy on a single cooled $^{171}Yb^+$ ion has recently been demonstrated /4/.

In this contribution, we show experimental results demonstrating laser cooling of single $^{172}Yb^+$ ions and of small ion clusters in a Paul trap, using for optical excitation a frequency-doubled diode laser and a diode-pumped fiber laser. It is intended to apply this experimental system to the initial demonstration of a novel optical frequency-standard scheme based on the specific hyperfine-sublevel structure of $^{171}Yb^+$. The scheme can be realized with the use of a frequency-doubled semiconductor laser as the local oscillator.

2. ATOMIC SYSTEM

Fig. 1 shows a section of the energy-level scheme of $^{172}Yb^+$. Optical detection and laser cooling of trapped $^{172}Yb^+$ ions relies on the excitation of the λ = 370 nm $^2S_{1/2}$ - $^2P_{1/2}$ resonance transition whose natural full halfwidth is $\delta\nu$ = 23 MHz. From the $^2P_{1/2}$ level, the metastable $^2D_{3/2}$ level is populated by spontaneous decay.

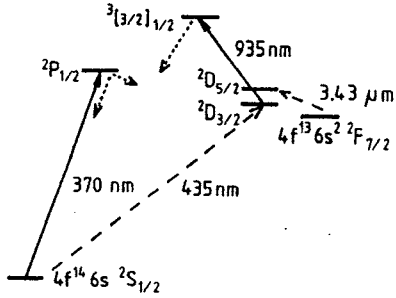


Fig. 1. Partial energy-level scheme of Yb^+. The full, dashed and dotted lines indicate transitions of the laser cooling cycle, potential clock transitions, and relevant spontaneous-decay channels, respectively.

For efficient laser cooling, ions dwelling in the $^2D_{3/2}$ level are rapidly transferred to the ground state by excitation of a transition at λ = 935 nm /5/. Excitation of ions to the $^2D_{5/2}$ level would lead to spontaneous decay to the $^2F_{7/2}$ level whose calculated lifetime is in the order of years /6/. The λ = 3.43 μm $^2F_{7/2}$ - $^2D_{5/2}$ transition, which has a calculated natural linewidth of 60 Hz, could be used as a reference transition of a frequency standard in the mid-infrared wavelength range /3/.

For $^{171}Yb^+$, continuous narrow-bandwidth optical excitation of the $^2S_{1/2}$ - $^2P_{1/2}$ resonance transition leads to optical pumping between the ground-state hyperfine levels, which strongly reduces the fluorescence scattering rate. One way to avoid hyperfine pumping is alternating optical excitation of the quasi-cycling F = 1 - F' = 0 component and of the F = 1 - F' = 1 component of the $^{171}Yb^+$ resonance transition. If a diode laser source is used for optical excitation, the required frequency switching by 14.8 GHz is not difficult to realize /7/.

Recent observations on $^{171}Yb^+$ indicate a hyperfine splitting of approximately 2.5 GHz in the $^3[3/2]_{1/2}$ state /8/. Since this splitting is much larger than the homogeneous linewidth of the $^2D_{3/2}$ - $^3[3/2]_{1/2}$ transition, the F = 1 - F = 0 hyperfine component of the transition can be excited selectively. For this case, selection rules imply that the rate of transitions to and from the $^2D_{3/2}$(F = 2) level can be made negligible even if the resonance transition is driven strongly. As a result, ions dwelling in the $^2D_{3/2}$(F = 2) level are effectively decoupled from the laser-cooling excitation cycle. Under these conditions, the electron-shelving technique /1/ can be applied to sensitively detect excitation of the $^2S_{1/2}$(F = 0) - $^2D_{3/2}$ (F = 2) electric-quadrupole transition whose natural linewidth is approximately 6 Hz /6/. The small natural linewidth of the transition suggests that it could be used as the reference transition of an optical frequency standard.

3. $^{172}Yb^+$ LASER COOLING EXPERIMENT

Fig. 2 shows the setup of the $^{172}Yb^+$ laser cooling experiment. The trap uses conical electrodes machined from molybdenum; the diameter of the ring electrode is 1.3 mm. The trap is driven at a frequency of 16 MHz and is typically operated to provide an axial pseudopotential well depth of 15 eV. For ion generation, an Yb oven and an electron source are mounted close to the trap. In order to minimize static electric fields in the confinement region, adjustable compensation voltages are applied to oven, electron source, and to one endcap electrode of the trap after loading of ions. The trap system is housed in a fused-silica envelope that is evacuated to a pressure in the range of 10^{-9} mbar.

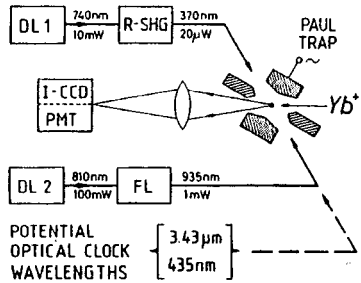


Fig. 2. Schematic of the experimental setup. DL1, DL2, diode lasers; R-SHG, resonant second-harmonic generation; FL, fiber laser; I-CCD, intensified video camera; PMT, photomultiplier.

The laser sources make use of two commercially available AlGaAs laser diodes. For excitation of the $^2S_{1/2}$ - $^2P_{1/2}$ resonance transition, the output of an optically stabilized, tunable diode laser emitting at λ = 740 nm is resonantly enhanced and frequency-doubled in a $LiIO_3$ crystal. Up to 20 μW of λ = 370 nm radiation are generated in a linewidth of approximately 200 kHz /9/; frequency switching as required for optical excitation of $^{171}Yb^+$ is readily obtained. The λ = 370 nm radiation is focused to a beam diameter of approximately 20 μm at the center of the trap. The λ = 935 nm radiation driving the $^2D_{3/2}$ - $^3[3/2]_{1/2}$ transition is generated by a Nd^{3+}-doped fiber laser pumped by a laser diode. The fiber laser is tuned and spectrally narrowed to approximately 200 MHz by frequency-selective elements placed in the open section of the laser cavity /10/.

Fig. 3 shows an image of five trapped, laser-cooled $^{172}Yb^+$ ions as obtained by spatially resolved detection of the resonance fluorescence at λ = 370 nm. The spatial arrangement of the ions corresponds to a local minimum of the effective potential resulting from the trap pseudopotential, Coulomb interaction between the ions, and an incompletely compensated static electric field. The observation of "crystals" formed by trapped Yb^+ ions complements earlier observations on other ionic species /11/.

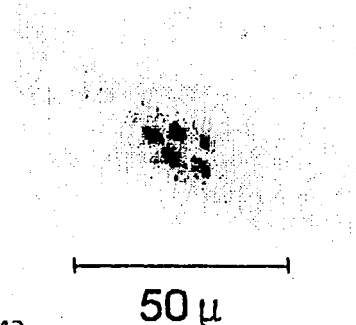


Fig. 3. Image of five trapped, laser-cooled $^{172}Yb^+$ ions. The obtained spatial resolution is determined by imaging aberrations. The ions rest in the symmetry plane of the trap which is at an oblique angle with the direction of observation. Light collection time is 1 s.

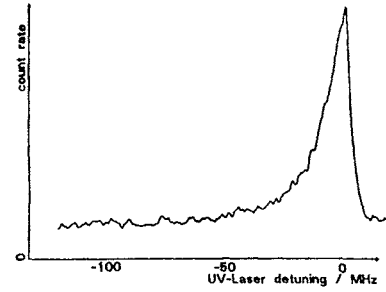


Fig. 4. Variation of fluorescence intensity with laser frequency at λ = 370 nm for the case of a single cooled $^{172}Yb^{+}$ ion. The maximum photon count rate is in the range of 10^3 s^{-1}.

For the case of a single cooled ion, Fig. 4 shows the variation of the fluorescence count rate with optical-excitation frequency. The asymmetry of the excitation spectrum is due to the onset of ion heating when the excitation frequency approaches the line center from lower frequencies. Similar excitation spectra have been observed in other experiments on trapped Yb^{+} /4,5,12,13/.

The absorption profile of Fig. 4 is well approximated by a section of a Lorentzian profile whose full halfwidth corresponds to the natural linewidth of the $^2S_{1/2}$ - $^2P_{1/2}$ resonance transition. From this one estimates that for the excitation spectrum shown in Fig. 4, the Doppler broadening due to residual ion motion is significantly smaller than the homogeneous broadening. If the Doppler width were equal to the homogeneous linewidth $\delta\nu$ of the resonance transition, the kinetic temperature T corresponding to ion motion along the propagation direction of the driving laser field would be given by $T = \lambda^2 m(\delta\nu)^2/(8k\ln 2) \sim 0.8$ K with λ being the excitation wavelength, m the atomic mass, and k the Boltzmann constant.

4. CONCLUSION

To our knowledge for the first time, we have demonstrated laser cooling of Yb^{+} ions in a trap using an experimental system entirely based on semiconductor laser sources. This system should provide an economical and reliable basis for optical high-resolution spectroscopy and frequency-standard investigations on trapped ytterbium ions.

We have proposed a novel optical frequency standard scheme which takes advantage of the hyperfine-sublevel structure of $^{171}Yb^+$. The scheme relies on the fact that for $^{171}Yb^+$ ions, the $^2D_{3/2}(F = 2)$ level can be effectively decoupled from the optical-excitation cycle of fluorescence detection and laser cooling. Thus, the $^2S_{1/2}(F = 0)$ - $^2D_{3/2}(F = 2)$ transition can be used as a reference transition of an optical frequency standard. The corresponding transition wavelength is λ = 435 nm, suitable for employing a frequency-doubled diode laser as the local oscillator.

REFERENCES

/1/ See, e.g., W.M. Itano, J.C. Bergquist, D.J. Wineland, Science 237, 612 (1987).

/2/ R. Blatt, R. Casdorff, V. Enders, W. Neuhauser, P.E. Toschek: In: "Frequency Standards and Metrology", ed. by A. DeMarchi (Springer, Berlin, Heidelberg 1989), pp.306.

/3/ A. Bauch, D. Schnier, Chr. Tamm, J. Mod. Opt. 39, 389 (1992).

/4/ V. Enders, Ph. Courteille, R. Huesmann, L.S. Ma, W. Neuhauser, R. Blatt, P.E. Toschek, Europhys. Lett. 24, 325 (1993).

/5/ A.S. Bell, P. Gill, H.A. Klein, A.P. Levick, Chr. Tamm, D. Schnier, Phys. Rev. A 44, R20 (1991).

/6/ B.C. Fawcett, M. Wilson, At. Data. Nucl. Data Tabl. 47, 241 (1991).

/7/ R.N. Watts, C.E. Wiemann, Opt. Commun. 57, 45 (1986).

/8/ A. Bauch, D. Schnier, Chr. Tamm, Proc. EFTF 1993, pp. 583.

/9/ Chr. Tamm, Appl. Phys. B 56, 295 (1993).

/10/ Chr. Tamm, D. Schnier, Opt. Commun. 87, 240 (1992).

/11/ see, e.g., F. Diedrich, E. Peik, J.M. Chen, W. Quint, H. Walther, Phys. Rev. Lett. 59, 2931 (1987).

/12/ H. A. Klein, A. S. Bell, G. P. Barwood, P. Gill, Appl. Phys. B 50, 13 (1990).

/13/ P.T.H. Fisk, M.A. Lawn, C. Coles, Appl. Phys. B 57, 287 (1993).

PRELIMINARY RESULTS IN A CESIUM ATOMIC FOUNTAIN FREQUENCY STANDARD

G. Santarelli, P. Laurent, S. N. Lea*, A. Nadir and A. Clairon.

LPTF-BNM, Observatoire de Paris, 61 avenue de l'Observatoire, F-75014 Paris, FRANCE

*Laboratoire de Spectroscopie Hertzienne de l'Ecole Normale Supérieure,

24 rue Lhomond, F-75231 Paris cedex 05, FRANCE

Abstract

An atomic frequency standard based on a fountain of cold cesium atoms offers a large improvement in both stability and accuracy over current primary cesium atomic beam frequency standards. Our prototype fountain is now in operation. We intend to present the whole set-up and some preliminary results concerning the accuracy and stability performance of this new standard.

Introduction

Neutral atom trapping and cooling has opened up a new way of realising the next generation of atomic frequency standards /1,2/. Several laboratories are involved in the construction of atomic fountains using laser cooled atoms /3,4/. The feasibility of the fountain has been demonstrated a few years ago using sodium /5/ and cesium /6/. In our atomic fountain, the atoms are first cooled or trapped in a low pressure Cs cell and are then launched upward using laser light. The atomic cloud, following a parabolic flight, passes twice through a μ-W cavity. The long time between the two interactions leads to a sub-hertz Ramsey resonance, two orders of magnitude narrower than in classic thermal beam devices. Many of the systematic effects which affect the accuracy of an atomic frequency standard depend to the linewidth and the atomic velocity distribution. With this narrow resonance we have all the traditional shifts easily evaluable at the 10^{-14} level. The stability performance of fountain standards is also very promising /7/.

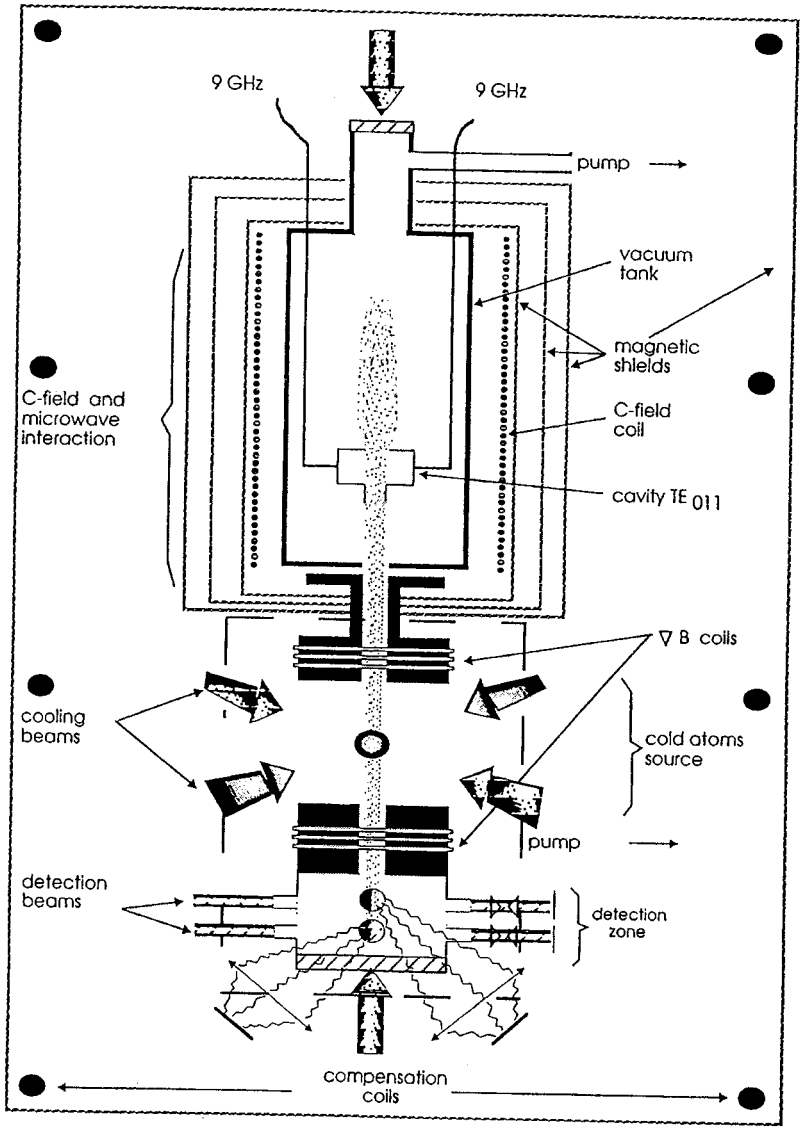


Figure 1: The L.P.T.F. atomic fountain prototype

1. The LPTF atomic fountain prototype.

The scheme of the LPTF fountain frequency standard is depicted in fig.1. We can distinguish three fundamental parts: the cold atom source, the microwave and C-field region, and the detection zone. These regions are separated by graphite tubes in order to reduce the Cs pressure in the last two regions. In the cold atom source, partially described in /8/, six independent laser beams capture and cool the cesium atoms by a pure optical molasses or a magneto-optical trap (M.O.T.), in a low pressure vapor cell ($\sim 10^{-8}$ Torr). In M.O.T. operation, the quadrupole magnetic field gradient (~8 G/cm) is generated by two anti-Helmoltz coils. We are able to cool from 10^5 to a few 10^8 atoms down to 5μK by 3-D σ^+-σ^- M.O.T. or lin⊥lin optical molasses. As will be explained later, it is possible to launch the atoms by the 1-D moving molasses technique into the μ–Wave cavity and the C-field zone. In this second part of the fountain the atoms pass once through an high-Q (~ 30000) cylindrical TE_{011} cavity situated 30cm above the cooling zone and fly ballistically in the C-field region where an highly homogenous static magnetic field of about 1.7 mG is produced by an axial solenoid and a few compensation coils. In this part of the fountain the pressure is kept low (few 10^{-10} Torr) to avoid collisions with the background gases. The third part of the fountain is the detection zone where two laser beams detect by fluorescence the populations of both F=3 and F=4 levels by the time of flight technique (T.O.F.) /6/. Three mu-metal magnetic shields surround the μW cavity and C-field region. There are compensation coils between the magnetic shields in order to reduce the rotation of the magnetic field and consequently to reduce the Majorana transitions. The whole set-up is surrounded by a fourth magnetic shield. The shielding factor of this arrangement is about 10^4 for 20-30 mG field fluctuations. To further reduce the axial magnetic field fluctuations to less than 0.1 μG, four coils controlled by an active servo system are used. The C-field region is about 70cm high and the whole set-up height is about 1.5 m.

All the laser frequencies necessary for the fountain operation are provided by four narrow linewidth (~100 kHz) diode lasers /9/. We use the F=4-F'=5 transition of the D_2 line to cool and detect the atoms. A first, grating tuned, diode laser is locked 2 MHz below this transition by saturated absorption in a cesium vapor cell. The output beam, which is also used for the detection, injection-locks two 150mW power laser diodes through an acousto-optical modulator (A.O.M.). The output of these two slave lasers is used for the horizontal beams and the upward and downward beams respectively. The frequency and the intensity of the beams are controlled via A.O.Ms and electro-mechanical shutters. All the beams are spatially filtered and expanded up to 1.5 cm in diameter and the maximum optical power is about 10 mW/cm^2 for the cooling beams and

1 mW/cm^2 for the detection. A fourth laser, frequency stabilized on the F=3-F'=4 transition is used as a 'repumper' and is superimposed an the horizontal cooling beams.

2. The fountain operation.

The fountain operation is pulsed. First we load for about 0.4s the cesium atoms by M.O.T. or by pure optical molasses. The loading time is inversely proportional to the partial cesium pressure which can be adjusted by varying the temperature of a cold finger. In this phase the laser beams are at the maximum intensity and they are red detuned by ~3Γ (Γ is the cesium natural linewidth ~ 5.3 MHz of the D_2 line). The temperature reached for this detunig is about 60μK. A variety of methods are available for launching the atoms. The total launching time is obviously limited by the beam size; the best method is which gives the necessary acceleration in the shortest time, in order to have a long post-cooling phase. The configuration that seems to work best is the 1-D moving molasses, which is obtained by switching off the horizontal beams and symmetrically switching the frequency of the upward and downward beams to create a moving frame. We are able to reach 5m/s in 200μs which means about 10^3g acceleration. This operation heats the atoms. To re-cool them we switch on the horizontal beams at the maximum intensity with 3Γ detuning to recapture the atoms in a 3-D moving molasses in 300μs. After, we reduce the intensity to about 2mW/cm^2 and the detunig is increased to ~10Γ. The final temperature is about 5μK which corresponds to an r.m.s. velocity of 2cm/s. The atomic cloud passes once through the μW TE_{011} cavity for a first $\pi/2$ pulse and flies ballistically in the C-field region. We can adjust the flight time from 0 to 700ms by changing the launching velocity. The atoms fall down and interact a second time with the μW to allow a Ramsey separated oscillatory field method. Finally the atomic populations and the vertical velocity distribution are then measured in the detection region. A first beam tuned 2 MHz below the F=4-F'=5 transition detects by fluorescence the population in the F=4 level in a standing wave which works as a Doppler molasses. The last part of the beam is a travelling wave which pushes away the atoms in the F=4 level by radiation pressure. Then the remaining atoms which are in F=3, are optically pumped to F=4 by the 'repumper' beam, which is tuned on the F=3-F'=4 transition. The atoms are then detected with the same procedure as described before. The transition probability is obtained by dividing the F=4 population by the total atom number; this normalisation procedure rejects the shot-to-shot atom number fluctuations, the equivalent noise of the detection is about 100 atoms per shot.

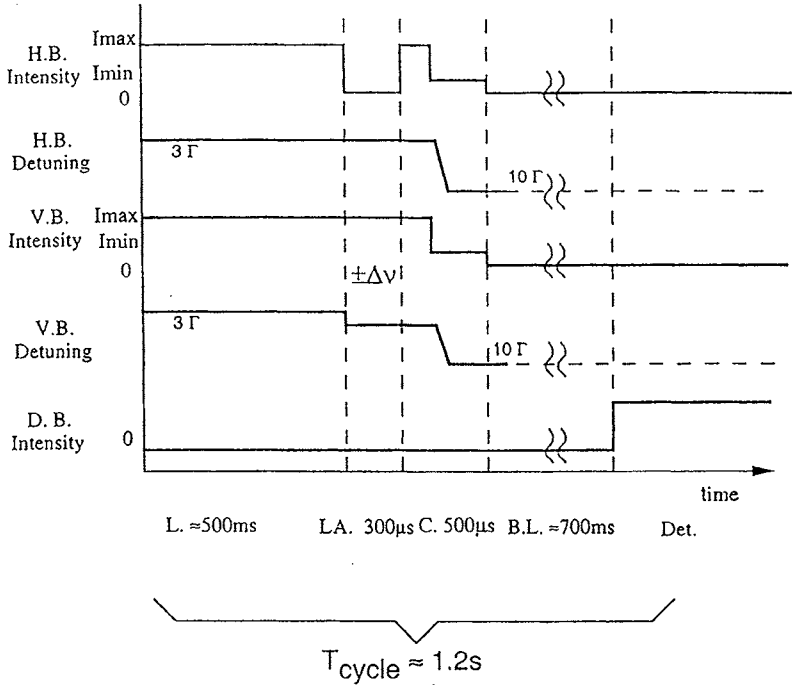


Figure 2: Fountain timing, H.B. horizontal beams, V.B. vertical beams, D.B. detection beams, L. loading phase, LA. 1-D moving molasses launching phase (symmetrical switching $\Delta\nu$ of the upward and downward beam frequencies), C. 3-D moving molasses cooling, B.L. ballistic fight, Det detection phase

Figure 2 shows the time sequence for the laser beams in our apparatus. Figure 3 shows 1Hz width Ramsey fringes, which are obtained with 500ms flight time. The central fringe depicted in fig.4 is the narrowest linewidth achievable in our system and is less than 700mHz wide (FWHM), two orders of magnitude narrower than in conventional thermal beam devices /10/. Another characteristic of the fountain as show in fig.3 is the narrow Rabi envelope (50Hz), due to the long interaction time in the cylindrical μW cavity. The large number of fringes is due to the near monokinetic velocity distribution of the atoms.

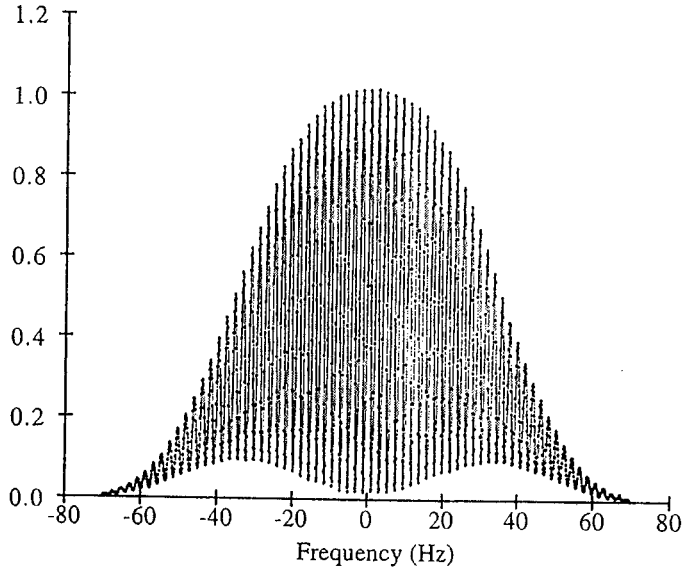


Figure 3: Ramsey fringes

In order to operate the fountain as a frequency standard, we use the Ramsey central fringe as a frequency discriminator. A μW synthesis chain multiplies a 10MHz quartz oscillator phase-locked to an H-maser up to 9.19..GHz. The multiplied signal is mixed with a commercial frequency synthesizer and the beat-note is used to phase-lock a dielectric resonator oscillator (D.R.O.). This enables the interrogating signal to be modulated. At each fountain cycle the interrogation signal switches from one side of the fringe to the other. The difference between the two transition probabilities which results is calculated by a personal computer which digitizes the signals from two detection channels, giving an error signal which is used to control numerically the frequency of the synthesizer. This digital servo system enables the frequency stability of the fountain to be measured directly against the H-maser /11/. The personal computer also generates the timing signal for the fountain cycle.

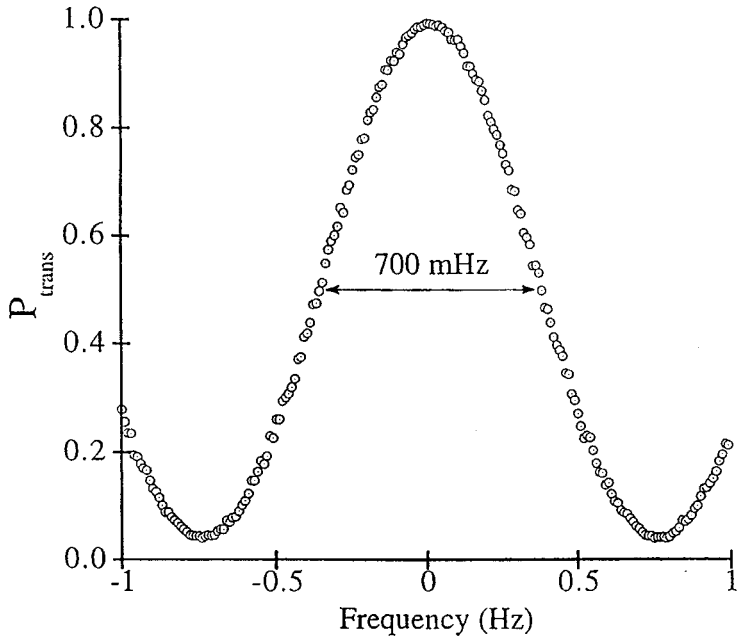


Figure 4: Central Ramsey fringe

3. Results and discussions.

The expected fractional frequency stability of the fountain as an atomic frequency standard is evaluated by the square root of the two-sample Allan variance defined as

$$\sigma_y(2,\tau) = \frac{\delta\nu}{\pi\nu_0}\left(\frac{N}{S}\right)\sqrt{\frac{Tc}{\tau}}, \qquad (1)$$

where ν_0 is the clock frequency, $\delta\nu$ the width of the central Ramsey fringe, $\frac{N}{S}$ is the noise-to-signal ratio in a single fountain cycle, Tc the time for one complete cycle and τ is the integration time. At present, when we operate with pure optical molasses with 10^5 atoms in $m_F=0$ at the detection, the shot-to-shot signal-to-noise ratio is about 300 measured on the top of the fringe, which is close to the detection noise. The expected

frequency stability with 1Hz linewidth is $10^{-13}\,\tau^{-1/2}$. The measured frequency stability is 5-6 $10^{-13}\,\tau^{-1/2}$, which is of the same order of the best optically pumped cesium frequency standard. This degradation is probably due to an aliasing phenomena consequent upon the pulsed operation of the fountain. As pointed out by J. Dick /12/, the local oscillator frequency noise is down-converted in the interrogation process, leading to a degradation of the signal-to-noise ratio at the detection. Figure 5 shows the square root of the Allan variance versus the number of fountain cycles. It is interesting to observe that in less than one hour we reach the 10^{-14} stability level, which simplifies the preliminary evaluation of the fountain frequency standard.

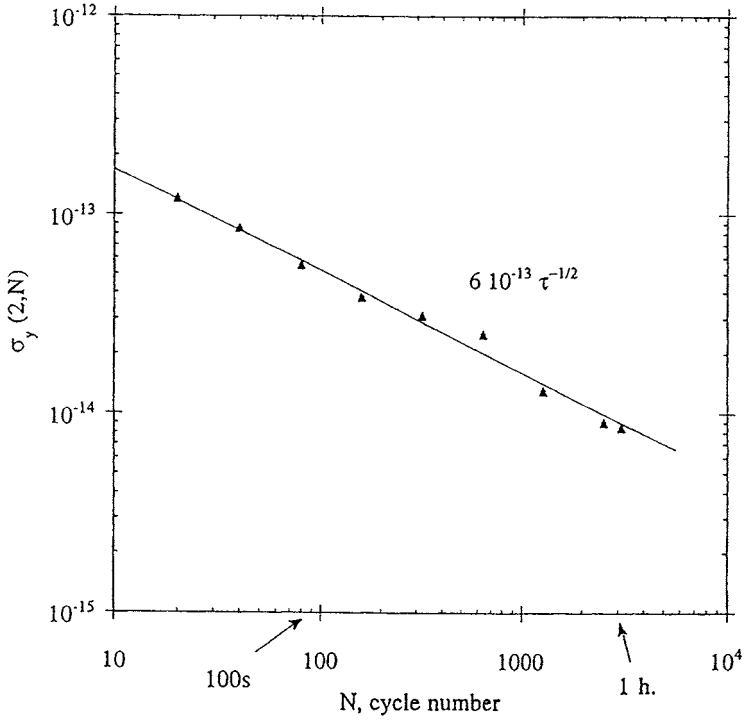


Figure 5: Fractional frequency stability

The operation of the fountain as a frequency standard requires the evaluation of the systematic effects which affect the accuracy of the standard itself. The accuracy capability of a such a Cs fountain is very attractive. Most of the effect depend on the atomic velocity and resonance width and therefore are considerably reduced in the fountain. For instance, the combined second-order Doppler effect and the gravitational red shift which is a consequence of general relativity theory, can be calculated using the following formula

$$\omega(H,v) = \omega_0 \left[1 + \left(\frac{v}{c}\right)^2 - \frac{1}{2} g H^2 \right] \qquad (2)$$

The resulting relative frequency shift is

$$\frac{\Delta\nu}{\nu_0} \cong \frac{1}{c^2 \delta\nu^2} \leq 10^{-17},$$

which is negligible.

The black body shift, which is related to the absolute temperature of the thermal radiation, is about $2\ 10^{-14}$ at 300 K and it can be calculated at 10^{-16} if we know the radiation temperature within 0.5K /14/.

The cavity related effects, which depend on the central Ramsey fringe-width are also considerably reduced. Due to the low atom number, we are far from maser oscillation and cavity pulling can be calculated as in a passive cesium tube /13/. With a cavity detuning of 100 kHz and a cavity Q-factor of 30000, we have

$$\frac{\Delta\nu}{\nu_0} = \frac{\nu_{cav} - \nu_{at}}{\nu_{at}} \left(\frac{Q_{cav}}{Q_{at}} \right)^2 \approx 10^{-16}. \qquad (3)$$

The residual first-order Doppler shift resulting from the imperfect standing wave in the μW cavity, can be evaluated as follows:

$$\frac{\Delta\nu}{\nu_0} = \frac{\langle \delta\phi \rangle \delta\nu}{\pi\nu_0}, \qquad (4)$$

where $\langle \delta\phi \rangle$ is the average distributed phase shift .

For an high-Q cylindrical TE_{011} cavity, A. De Marchi and co-workers have calculated a maximum distributed phase shift of 5 μrad for our geometry /15/. The resulting shift is of the order of 10^{-16} which is much less than in conventional Ramsey

cavities.

As a consequence of the narrow resonance line, we can considerably reduce the static magnetic C-field which is used to separate the mF transitions . Due to the long interaction time in the cylindrical μW cavity, the Rabi envelope is about 50 Hz wide and in consequence the C-field can be reduced to about 1.7 mG, yielding more than 1kHz spacing between the mF transitions. Unlike a classical thermal beam device, we can also map the magnetic field distribution by launching the atoms to different heights. At the present, with $\sigma(\bar{B}) \leq 5\ \mu G$, the uncertainty due to the C-field is

$$\frac{\Delta\nu}{\nu_0} \cong \frac{427}{10^{10}}\,\bar{B}\,\sigma(\bar{B}) \leq 510^{-16}, \qquad (5)$$

The Ramsey /16/ and Rabi shifts which are the consequence of spurious $m_F=0 \rightarrow m_F=\pm 1$ transitions and pulling from the neighbouring $m_F=\pm 1 \rightarrow m_F=\pm 1$ transitions, are, as a result of the large spacing between the mF states and the symmetry in the populations of the different Zeeman sub-levels, negligible at the 10^{-14} level but at present are not fully evaluated.

The new effect which arises in the fountain operation is the spin-exchange frequency shift due to collisions between ultra-cold atoms. Currently, this spin-exchange collisional shift is expected to be one of the limiting factors in the realisation of the potential accuracy of the atomic fountain as a frequency standard. The theory developed by Verhaar and co-worker states that this shift for s-wave scattering is proportional to the average atomic density over the flight /17/. The resulting shift for a tall atomic fountain, can approximately written as:

$$\frac{\Delta\nu}{\nu_0} \approx \frac{10^{-22}\,N}{(2\pi)^{3/2}\sigma_i^{\,2}\,\sigma_f} \qquad (6)$$

where N is the initial atom number and σ_i and σ_f are the initial and final cloud r.m.s. sizes. The measurements performed by the Stanford group /18/ estimate the collisional shift in the cases where the atoms are distributed in all mF states equally populated or where only the mF=0 state is populated, leading to two different results

$$\frac{\Delta\nu}{\nu_0} \cong 5\ 10^{-22}\,\bar{n}, \quad \text{all } m_F \qquad (7)$$

$$\frac{\Delta\nu}{\nu_0} \cong 1.5\ 10^{-21}\,\bar{n}, \quad m_F=0 \qquad (8)$$

where $\bar{n}$ is the average atomic density over the flight. Some preliminary measurements performed in a M.O.T. configuration in our laboratory with all the m_F states populated are in agreement with the published results. In our case we operate in a pure optical molasses configuration with 10^5 atoms at the detection in $m_F=0$ and 10μK temperature. The total shift obtained using (7) is less than $3\ 10^{-15}$. Further reduction of this shift can be obtained by rejecting from the fountain those atoms which contribute to the shift but not to the signal i.e. atoms in the wings of the transverse velocity distribution. We can achieve the selection of the ultra-cold atoms by using 2-D velocity-selective Raman transitions. At Stanford /19/ and in the LPTF /7/ preliminary 1-D experiments give encouraging results, but the complexity of the fountain is increased.

Conclusions.

The first operation of an atomic frequency standard based on cold cesium atoms has been achieved. Although far from being an evaluation of the atomic fountain as a frequency standard, this work shows the very attractive performance of this new device. The potential stability of $10^{-13}\ \tau^{-1/2}$ is apparently limited by the local oscillator frequency noise but we hope to overcome to this problem by the new multiplication chain being constructed in our laboratory.

Acknowledgements.

This work forms a part of the programme of scientific developments of the Bureau National de Metrologie (BNM). The assistance of A.H. Gerard , M Dequin and M. Lours is gratefully acknowledged.

References.

/1/ See for instance the special issue, J. Opt. Soc. Am. B6, 7989.

/2/ K. Gibble and S. Chu, Metrologia 1992, 29, 201-212

/3/ A. Clairon, Ph. Laurent, A. Nadir, G. Santarelli, M. Drewsen, D. Grison, B. Lounis, C. Salomon, in: Frequency Stabilized Lasers and their Applications, (SPIE proceedings Vol. 1837, 1992) 306.

/4/ A. Michaud, M. Chowdhury, K. P. Zeite, C. J. Cooper, G. Hillenbrand, A. Steane and C. Foot, Proceedings of 7th EFTF 1993, Neuchatel, 525-530.

/5/ M. Kasevich, E. Riis, S. Chu and R. De Voe, Phys. Rev. Lett., 63, 1989, 612-615.

/6/ A. Clairon, C. Salomon, S. Guellati and W. D. Phillips, Europhys. Lett., 1991, 16 , 165-170.

/7/ S.N. Lea, A. Clairon, C. Salomon, P. Laurent, B. Lounis, J. Reichel, A. Nadir and G. Santarelli. accepted for publication in Physica Scripta

/8/ A. Clairon, Ph. Laurent, A. Nadir, M. Drewsen, D. Grison, B. Lounis and C. Salomon, Proocedings of 6th EFTF 1992, Noordwijk, 27-33.

/9/ G. Santarelli, A. Clairon, S.N. Lea, G. Tino, Opt. Comm.,104, 1994, 339-344.

G.D. Rovera, G.Santarelli and A. Clairon. Rev. Sci. Instrum., 1994.

/10/ A. Bauch, B. Fischer, T. Heindorff, R. Schroder, Proceedings of 7th EFTF 1993, Neuchatel, 531-535.

/11/ G.D. Rovera, E. de Clecrq, A. Clairon, UFFC, 1994.

/12/ J. Dick, Proceedings of PTTI 1987, 133-1147.

/13/ J. Viennet, C. Audoin and M. Desaintfuscien, IEEE IM, 21, 1982, 204-209.

/14/ W.M. Itano, L. Lewis and D. Wineland, Phys. Rev. A, 45, 1982, 1233.

/15/ G. Vecchi and A. De Marchi, IEEE IM, 42, 1993, 434-438.

/16/ L. Cutler, A. De Marchi, C. Flory and R Giffard, J. Appl. Phys., 1991, 2780-2792

/17/ E. Tiesinga, B.J. Verhaar, H.C.T. Stoof and D. Van Bragt, Phys. Rev. A, 45, 1992, 2671.

/18/ K. Gibble and S. Chu, Phys. Rev. Lett., 70, 1993, 1771.

/19/ M. Kasevich and S. Chu, Phys. Rev. A, 69, 1992, 1741.

First results on the ON magneto optical trap

G. Dudle, N. Sagna, P. Thomann

Observatoire Cantonal - Rue de l'Observatoire, 58 - CH-2000 Neuchâtel - SWITZERLAND

Abstract

This paper reports on the cold atoms activity at Observatoire de Neuchâtel. It gives a description of the experimental set-up of the magneto-optical Cesium trap that is under development with the purpose of building an atomic clock. In order to characterize the cold atoms source, experimental studies have been undertaken: first results about the trap parameters and the number of cold atoms as a function of background pressure, laser detuning and intensity are presented and compared to the predictions of a simple theoretical 1D model.

Keywords: Doppler cooling and trapping, Magneto optical trap.

1. Introduction

In recent years, important progress has been made in optical trapping of neutral atoms /1/. This has opened up exciting perspectives in the field of atomic frequency standards /2/. The goal of our work is to realize an atomic clock using cold Cs atoms. In such a clock, the number of atoms that contribute to the signal strongly depends on the efficency of the capture process of the magneto-optical trap (MOT). In this paper a study of how to optimize this capture process is presented: measurements on our MOT were performed and one investigates how the number of trapped atoms varies with the background pressure, the laser detuning, intensity and diameter. To get a better understanding of this trapping process, a simple one dimensional model that considers both Doppler cooling and trapping of atoms is developed; using a simple numerical simulation this model enables to make predictions about the number of trapped atoms for any configuration of magnetic field.

2. Experimental Set-up

The optical set up used to control the cooling beams, the vacuum system containing the cesium and the coils producing the magnetic gradient are presented. Besides, a short description of the detection chain is given.

Several conditions concerning laser frequency, its linewidth and polarisation as well as the laser intensity have to be fulfilled in order to slow thermal atoms. In different steps we meet the requested conditions. The linewidth of the used light has to be smaller than the natural linewidth of the cesium D_2 transition (λ= 852nm, $\Gamma = 2\pi$ 5.3 MHz). Previous work /3/ has shown that commercially available laser diodes are too broad to be used directly and must be narrowed. By placing the diode in an extended cavity linewidths of less than 1 MHz are easily reached. The drawback of this technique is the loss of power. Only about 25% of the output of the laser diode is extracted from the extended cavity.

This low power laser beam is divided in two parts. One of them is used to lock the master laser with an tunable offset to the transition F= 4 -> F' = 5 of the D_2 line of cesium resolved in a vapor cell by saturated absorption.The realised set up allows frequency shifts of the laser light in the range between 0 to $\approx$ - 4 Γ. The second part of the tunable and spectrally narrowed output of the extended cavity laser is used to control the output of a high power (typically 40mW) laser by injection locking.

Passing through a set of anamorphic prisms and a telescope the slave output is spatially filtered and brought to the final circular diameter of up to 14 mm. In the following step the beam is divided in three parts in order to irradiate the trap in the x,y and z direction of space. A typical value of the intensity per beam is 5 mW/cm^2 which corresponds to $I/I_{sat} \approx 2$. Before penetrating the vacuum system the circular polarisation needed for the operation of the MOT /4/ is given to each beam by $\lambda/4$ retarder plates. Since all beams entering the vacuum system are parallel, two of them (y and z) are redirected inside by mirrors, the third (x) crosses the chamber without changing of direction. In order to obtain a standing waves the three beams are retroreflected by mirrors and $\lambda/4$ plates placed outside of the vacuum chamber.

The light of an additional laser is mixed to one of the cooling beams after the telescope. This laser, locked to the transition F = 3 -> F' = 4 using a Doppler-free saturated absorption cell, prevents atoms from accumulating in the state F = 3 which is decoupled from the cooling beams. This repumping beam is provided by a DBR laser with 3mW output.

Magnetic gradient: Two loops with a diameter of 2r = 50mm are placed at a distance of 2d = 43 mm. The ratio $2d/r = \sqrt{3}$ corresponds to the situation where the gradient is constant over the largest range. The coils are disposed in a horizontal plane, their common axis (z) being vertical. The zero divergence of the magnetic field and the symmetry of the geometry imply $dB_x/dx = dB_y/dy = -0.5\ dB_z/dz$ (so called "spherical quadrupole" field geometry). The maximum z component of the gradient is 2.6 G/cm.

The vacuum chamber: The vacuum chamber is made of stainless steel with several optical windows to introduce the laser beams as well as to observe the trapping region. It contains the coils for the magnetic gradient and four mirrors which redirect two of the three laser beams. The density of the cesium vapor inside the assembly is controlled by varying the temperature of a finger with an initial 1g load of cesium. By measuring the light absorption accross the chamber this density can be verified. The residual pressure, essentially formed by CO and CO_2, is of the order of $1\ 10^{-8}$mbar

The Visualisation and measurement of the cold cloud: For the visualisation as well as for the different measurements performed on the cold atoms, a CCD camera has been installed. Connected to a video analyser, the chain allows estimates of the light radiated by the trapped atoms. If the cloud of atoms is optically thin, the total power is easily deduced by taking into account the solid angle of the collection optics. The light emitted by the cloud is a function of the scattering rate which is itself dependent on the incident laser intensity and detuning. In the case that the cloud ceases to be optically thin the assumptions of the calculations yield too small a number of atoms. With that instrument studies of the number of atoms as a function of different parameters such as laser detuning and intensity or background pressure have been performed. The method used to determine the trap spring constant consists in measuring the displacement of the cloud while creating an intensity imbalance between the two counterpropagating waves of a given direction by attenuating the retroreflected beam /5/.

3. Experimental results

Figures 1-4 show experimental results on how the number of atoms depends on different parameters. Fig.1, represents the number of atoms versus the laser detuning, for a fixed magnetic gradient of 2.6 $G.cm^{-1}$. The experimental curve exihibits a pronounced maximum at a detuning of -1.8 Γ.

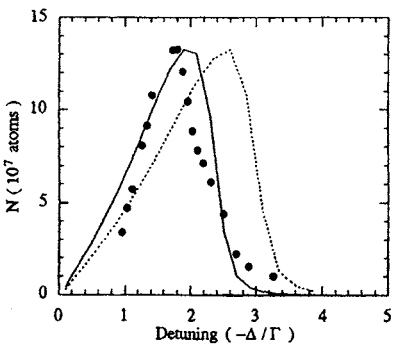


Fig. 1. The number of trapped atoms versus the laser detuning. The diameter is 1 cm, the intensity in each of the six beams is 2.8 mW/cm^2 and the magnetic field gradient is 2.6 G/cm. The dots represent the experimental results obtained. The dotted line shows the prediction of the three-level model multiplied by a normalization factor of 0.51 while the solid line is the prediction of the multi-level model (see section 4), multiplied by a factor of 1.4.

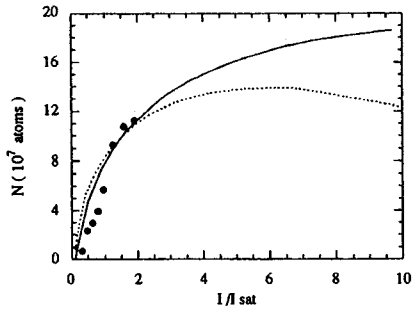


Fig. 2. The number of trapped atoms versus the intensity of each laser beam. I_{sat} = 2.2 mW.cm^{-2} is the saturation intensity and I is the average intensity over the 1/e^2 diameter of the beam. The 1/e^2 diameter of the beam is 1 cm. The detuning is kept constant at $\Delta = -1.7\,\Gamma$ and the magnetic field gradient is 2.6 G/cm. The dots represent the experimental results obtained. The dotted line shows the prediction of the three-level model multiplied by normalizing factor of 0.51 while the solid line is the prediction of the multi-level model (see section 4), multiplied by a factor of 1.1.

Fig. 2 represents the number of atoms as a fonction of the laser beams intensity, for a fixed detuning and diameter. By intensity, one means the average intensity over the 1/e^2 diameter of the beam. As expected the measured number of atoms increases with the intensity.

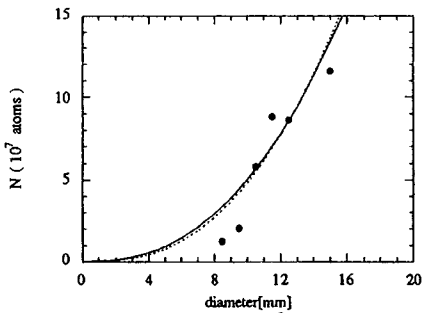


Fig. 3. The number of trapped atoms versus the 1/e^{2}diameter of each laser beam. The detuning is kept constant at $\Delta = -1.7\,\Gamma$, the magnetic field gradient is 2.6 G/cm and the average intensity of each of the six beams is 2.8 mW/cm^2. The dots represent the experimental results obtained. The dotted line shows the prediction of the three-level model multiplied by a normalization factor of 0.51 while the solid line is the prediction of the multi-level model (see section 4), multiplied by a factor of 0.6.

In Fig. 3 one shows how the number of atoms increases as the diameter of the laser beams increases; the experimental results obtained are to be taken carefully regarding the difficulties of maintaining the laser beam intensity constant while its diameter changes. Finally, Fig. 4 shows how the number of atoms depends on the residual background pressure. One sees that a residual pressure of 10^{-8} mbar is required to optimize the capture process.

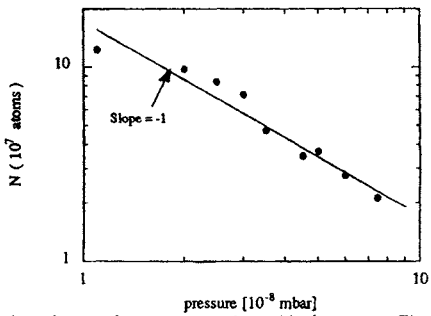


Fig. 4. The number of trapped atoms versus the residual pressure. The detuning is Δ=–1.7 Γ, the magnetic field gradient is 2.6 G/cm and the intensity of each of the six beams is 2.8 mW/cm^2. The circles represent the experimental results.

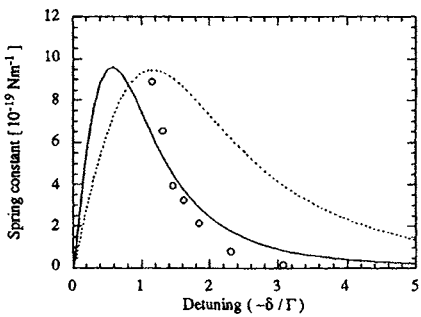


Fig. 5 The spring constant of the trap versus the laser detuning. The magnetic field gradient is 2.6 G/cm and the intensity of each of the six beams is 6 mW/cm^2. The circles represent the experimental results obtained. The dashed and the solid lines shows the spring constant deduced from the three-level model and the multi-level model. The normalization factor to the experimental results is 28 for the three -level and 35 for the multi-level one.

Fig.5 represents the spring constant of the trap as a function of the laser detuning; the results of our measurement are comparable to those of other groups /4, 6/.

4. Theoretical models

In this part, two entirely one-dimensional models are presented. They describe the capture process and give predictions about the number of trapped atoms. While analytical treatments /7, 8/ of the capture process and the number of atoms neglect the effect of the magnetic field, both models presented here use a simple numerical simulation in order to take into account the magnetic field. They are different in the type of optical transition each of them considers but apart from that they are based on the same set of assumptions:

One assumes that the atoms move on the x-axis and are subjected to the radiation pressure of two counterpropagating, circuraly polarised ($\sigma \pm$), laser beams, with a wave vector $\mathbf{k} = k\,\mathbf{e_x}$, and to the effects

of the magnetic field $\mathbf{B}(x) = b\, x\, \mathbf{e_x}$. This description considers only the magnetic field along the central axis of the beams where it is parallel to the wave vector. Although most of the atoms enter the trap off axis, one will see that for small magnetic field gradients b and small beam diameter d such that $\frac{\mu_B}{\hbar}\frac{b}{\Gamma}\frac{d}{2} << 1$, the capture process as well as the influence of the magnetic field on it may be described by considering the on-axis case. These models neglect the gaussian profile of the beams, two other pairs of laser beams and the optical saturation they induce.

As described in /7/ one assumes that the capture process is isotropic in a three-dimensional magneto-optical trap: when an atom enters the trapping volume, it sees a pair of co- and counter-propagating beams. Therefore the capture process can be described by a one-dimensional model. The equation of evolution of the number of atoms in a magneto-optical trap is given by /6/:

$$\frac{dN}{dt} = R - \frac{N}{\tau_{trap}} \tag{1}$$

where R is the trap capture rate and $\frac{1}{\tau_{trap}} = \frac{1}{\tau_{Cs}} + \frac{1}{\tau_{bg}}$ is the loss rate due to collisions between the cold cesium atoms and atoms of the background with thermal velocity (Cs and non-Cs background gases). Intratrap collisions that induce nonlinear terms in the rate equation (1) are ignored. Such a contribution is indeed negligible /9/, for the densities (10^{11} cm^{-3}) that we obtained. The trap loss rate is given by:

$$\frac{1}{\tau_{trap}} = (n_0\ \sigma_{Cs}\ \bar{v}_{Cs}) + (\sum_i n_i\ \sigma_i\ \bar{v}_i) \tag{2}$$

where n_0 is the density of the cesium background gas and $\sigma_{Cs} = 2\ 10^{-13}\text{cm}^2$ is the cross section of collision between cesium atoms /10/, $\bar{v}$ is the mean relative velocity of the colliding atoms; since the cold atoms are approximately at rest, $\bar{v}$ is the mean velocity of the background atoms /6/. The sum in equation (2) is over the different components of the residual gas. Besides the capture rate can be written as the number of atoms entering the trapping region -a sphere with diameter d- per unit time, multiplied by that fraction of the atomic velocity distribution that lies below the capture velocity v_c of the trap /4/.

$$R = \frac{4}{\pi^2}\, n_0\, S\, \frac{v_c^4}{v_{th}^3} \tag{3}$$

where $S = \pi\ d^2$ is the surface area of the trap volume, and $v_{th} = \sqrt{\frac{8}{\pi}\frac{k_B T}{m_{Cs}}}$ is the mean velocity of the background cesium atoms. The steady-state number of trapped atoms is $N_S = R\tau_{trap}$; to determine N_S, one has to compute the maximum velocity v_c with which an atom can enter the trapping zone and can still be confined in it. This is done by numerically solving the Newton equation $m\dot{v} = F(x,v)$ where F is the force on an atom induced by the combined action of the light and the magnetic field gradient. This force is supposed to be a Doppler-type one: it corresponds to a difference of radiation pressure of the two beams, taking into account both the Doppler and the Zeeman shifts. The model neglects the interferences between the two beams and the sub-Doppler processes that they induce. Such an assumption is justified because sub-Doppler cooling and trapping play a negligible part in the capture process of atoms from a thermal vapor.

The expression of the force F depends on the transitions involved in the cooling process. In the first model, one supposes that under the action of the two circularly polarized beams, the atoms are Zeeman pumped towards $m_F = \pm 4$; this assumption, together with the consideration of the Clebsch-Gordan coefficients, allow to reduce the $F = 4 \rightarrow F' = 5$ cesium transition to a $J = 0 \rightarrow 1$ transition. Since the

sublevel $\left|J = 1, m_J = 0\right\rangle$ is decoupled, the system reduces to a three-level one. The force can then be written:

$$F(x,v) = \hbar k \Gamma \beta \left[\frac{1}{1 + 4\beta + 4\left(\frac{\Delta - (kv + \mu_B \hbar^{-1} bx)}{\Gamma}\right)^2} - \frac{1}{1 + 4\beta + 4\left(\frac{\Delta + (kv + \mu_B \hbar^{-1} bx)}{\Gamma}\right)^2} \right] \tag{4}$$

where Δ is the laser detuning, $\beta = I/I_{sat}$ is the saturation parameter of each beam and $I_{sat} = 2.2\ mW/cm^2$. At the opposite, the second model (multilevel model) neglects the effect of optical pumping: all the ground state Zeeman sublevels are assumed to be equally populated. Therefore the force takes into account all the $\Delta m = \pm 1$ transitions from the F=4 ground state and is given by the expression:

$$F(x,v) = \frac{1}{9} \hbar k \Gamma \sum_{m_g=-4}^{m_g=4} \left[\frac{\beta^+_{m_g}}{1 + 4\beta^+_{m_g} + 4\ \delta^{+\ 2}_{m_g}(x,v)} - \frac{\beta^-_{m_g}}{1 + 4\beta^-_{m_g} + 4\ \delta^{-\ 2}_{m_g}(x,v)} \right] \tag{5}$$

where $\beta^{\pm}_{m_g}$ is the saturation parameter associated to the transition $\left|F = 4, m_g\right> \rightarrow \left|F' = 5, m_g \pm 1\right>$; $\delta^{\pm}_{m_g}(x,v)$ is the corresponding effective detuning, normalized to Γ and is given by:

$$\delta^{\pm}_{m_g}(x,v) = \frac{1}{\Gamma}\left(\Delta \pm kv - \frac{\mu_B bx}{\hbar}((m_g \pm 1)\ g_e - m_g\ g_g)\right) \tag{6}$$

where g_e=0.4 and g_g=0.25 are the Landé g factors of the excited and the ground levels. From expressions (5) and (6) the predictions of the two models concerning the number of trapped atoms are computed and compared to the experimental results.

5. Interpretation

In this part we discuss the predictions of the two models by comparing them to the experimental results. The three-level model predicts a number of trapped atoms that is about twice those measured. As shown in Figs 1-3 its predictions reproduce the general behaviour of the number of atoms as a function of parameters such as the laser beams detuning, intensity and diameter. Although this simple model takes into account the magnetic field, it fails, however, in predicting the optimum detuning for the capture process, the error being of the order of the cesium natural linewidth.

The multi-level model exhibits an even better agreement with the experimental data. Despite that its predictions require to be multiplied by a normalisation factor varying between 0.6 and 1.4, this model reproduces precisely (figs.1-2) how the number of trapped atoms depends on the laser beams detuning and intensity and in particular it predicts the optimum detuning within 0.2 Γ. Such an agreement justifies a posteriori the assumption of no optical pumping this model makes. This assumption matches the situation in a three-dimensional magneto-optical trap: during the capture process, the magnetic field changes in direction and magnetude and therefore tends to equalize the population of the Zeeman sublevels of the ground state by Larmor precession. This agreement also justifies the use of an entirely one-dimensional model to describe the trapping process in a MOT with small magnetic field gradient and laser beams diameter. Such a situation in which the atoms entering the trap on-axis are representative of the general capture process is in contrast to that analyzed by Lindquist /10/. Meanwhile, as shown in Fig. 3,

the predictions of the model exhibit a small difference with the experimental results concerning how the number of atoms N depends on the laser beams diameter d. The model predicts that N increases as $d^{3.4}$ whereas the experimental results show that N is proportional to $d^{2.6}$. Such a difference can be explained by the fact that the model assumes a square intensity profile of the laser beams whereas in reality they are gaussian.
As shown by expressions (6) and (7) both models take into account the power broadening due to the two counterpropagating beams. That is why the predictions of the number of atoms versus the laser intensity differ from the ones of /10/ that considers power broadening by only one beam. Our models indeed predict that as the laser beam intensity increases the number of atoms reaches a maximum and then tends to a lower value. Such a behaviour has to be confirmed by further experimental results with more powerful cooling beams. Besides one should be aware that the real physical situation corresponds to a six-beam power broadening that, according to our models, would lead to an even more pronounced maximum.
The predictions of both models about the spring constant shown in Fig. 5 are computed by linearising the forces given by expressions (6) and (7) around x=0 and v=0; this linearisation is justified because the spring constant caracterizes the effect of the trap on slow atoms located near the center of the trap /5/. The predictions of both models fail to reproduce the experimental results; they in particular require important normalization factors. This emphasizes the domain of validity of the models and their limits: as long as one considers phenomena that are related to the capture process, Doppler-type models give predictions that match experimental results. However, for parameters such as the friction coefficient or the spring constant whose magnitude can only be understood in terms of sub-Doppler mechanisms /5/, the models provide underestimations as expected.

6. Conclusion

This work gives an understanding of the role of the different parameters of a magneto-optical trap. It points out how to optimize the number of trapped atoms. Two simple one-dimensional models allowing predictions about the number of trapped atoms are presented. One shows that the most accurate predictions are given by the model that assumes no optical pumping but still takes into account all the Zeeman sublevels of the Cesium ground state. The characterization of the capture process represents the first step towards the use of an anisotropic magnetic field in order to continuously extract cold atoms from the magneto-optical trap.

7. Acknowledgement

This work was supported by the Swiss National Fund for Scientific Reseach and the Swiss Federal Office of Metrology.

8. References

/1/ J. Opt. Soc.Am B 6, 11 (1989)

/2/ G. Santarelli, P. Laurent, A. Nadir, A. Clairon, " Preliminary Results in a Cesium Atomic Fountain", Frequency Standard" These proceedings.

/3/ N. Sagna, C. Mandache, and P. Thomann, (1992), Proc 6th EFTF, Noordwijk , pp. 521/25

/4/ E.L. Raab, M. Prentiss, A. Cable, S. Chu and D.E. Pritchard , (1987), Phys. Rev. Lett, 59, pp. 2631/34

/5/ A.M. Steane, M. Chowdhury and C.J. Foot (1992) , J. Opt. Soc.Am B 9, pp. 2142/58

/6/ D.W. Sesko, T.G. Walker and C.E. Wieman (1991) , J. Opt. Soc.Am B 8, pp. 946/58

/7/ C. Monroe, W. Swann, H. Robinson and C.E. Wieman, (1990) , Phys. Rev. Lett, 65, pp. 1571

/8/ K.E. Gibble, S. Kasapi and S. Chu, (1992), Opt. Lett, 17, pp. 526-/28

/9/ D.W. Sesko, T.G. Walker, C. Monroe, A. Gallagher and C.E. Wieman, (1989), Phys. Rev. Lett, 63, pp. 961

/10/ K. Lindquist, M. Stephens and C.E. Wieman, (1992), Phys. Rev. A, 46, pp. 4082/90

SPATIAL PHASE DISTRIBUTION WITHIN CAVITIES FOR CESIUM FOUNTAIN STANDARDS

A. Khursheed and A. De Marchi,

Dip. di Elettronica, Politecnico di Torino,
I-10129 Torino, Italy.

ABSTRACT
This paper presents a numerical method for calculating the spatial phase distribution of cavities for cesium fountain studies. The method is used to asses the effect of holes in the cavity end cap and coupling of a waveguide port into the cavity side wall.

1. INTRODUCTION

In order to design the proposed fountain cavity for a cesium primary frequency standard providing an accuracy better than 10^{-15}, a number of properties of the RF field which interacts with the atoms will need to be controlled. Amongst the most significant of these parameters is the spatial distribution of the phase of the microwave H-field.

To calculate the spatial phase variation inside the fountain cavity, a computational approach is used. The merit of using a numerical approach is that it can calculate the spatial phase distribution inside cavities having a wide variety of boundary shapes, whereas analytical methods are limited to cavities having geometrically simple shapes. In this paper, a boundary perturbation numerical method is presented, which is used to take into account the effect of holes in the end caps of the cavity, and analyse asymmetric and symmetric coupling of waveguide ports into its side wall.

2. AN ANALYTICAL APPROACH

The proposed fountain cavity is based on a TE_{011} mode in a cylindrically symmetric cavity. An idealized model of this cavity, neglecting the coupling of the source at the side wall and the beam passing holes at the end caps, is shown in Figure 1. This figure shows magnetic field lines in the r-z plane. A cavity radius of a=25.5mm and length L=26.075mm are chosen to provide a nominal resonant frequency of 9.19GHz. The cesium beam going through the cavity is confined to a region of 4mm radius around the axis. The longitudinal component of the magnetic field, which is the relevant one for the atomic transition of interest, is maximum at cavity centre.

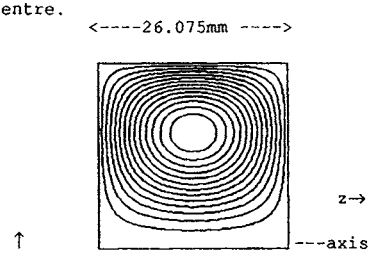


Figure 1: Idealized Fountain Cavity

The spatial phase distribution in this idealized model of the fountain cavity can be found analytically by performing first order perturbation of the fields with respect to the cavity resonant wave number k. It can be shown that the effect of losses at the metal walls of a cavity causes a shift Δk in the wave number [1], given by

$$\Delta k = (j-1)\frac{k}{2Q} \qquad (1)$$

where Q is the quality factor of the cavity for the considered mode. The longitudinal component $H_z(r,z)$ of the magnetic field can be solved for

explicitly in the cylindrical TE_{011} cavity mode with perfect metal walls, and is found to be

$$H_z(r,z) = J_0(k_c r)\cos(\beta z) \quad (2)$$

where $\beta=\pi/L$, k_c is the transverse wave number given by $k_c^2=k^2-\beta^2$, and J_0 is the zero order Bessel function.

The effect of wall losses on the above field component is then given by

$$H_z(r,z) = J_0((K_c+\Delta k_c)r)\cos((\beta+\Delta\beta)z) \quad (3)$$

where Δk_c is the change in the transverse cutoff wave number and $\Delta\beta$ is the change in the propagation factor. By using (1), Δk_c and $\Delta\beta$ can be expressed in terms of the cavity Q as

$$(j-1)\frac{k}{2Q} = \frac{k_c}{k}\Delta k_c + \frac{\beta}{k}\Delta\beta \quad (4)$$

The Q of the TE_{011} mode for this idealized cavity can be analytically derived [2], and is given by

$$Q = \frac{aZ_0k^3}{2R_s[k_c^2+2(\frac{a}{d})\beta^2]} \quad (5)$$

where R_s is the surface resistance of the metal, and Z_0 is the intrinsic vacuum impedance. By substituting for the Q in (4), and separating transverse losses from longitudinal losses, Δk_c and $\Delta\beta$ can be explicitly solved

$$\Delta k_c = (j-1)\frac{k_c}{k}\frac{R_s}{Z_0}\frac{1}{a}$$
$$\Delta\beta = (j-1)\frac{2\beta}{k}\frac{R_s}{Z_0}\frac{1}{L} \quad (6)$$

The effect of metallic wall losses on the field component $H_z(r,z)$ can then be found by expanding (3). Since the tangent of the phase is given by the ratio of the imaginary part of the field to its real part, the spatial phase distribution for the idealized fountain cavity is then given by

$$\tan(\theta(r)) = -\frac{r}{a}\frac{k_c}{k}\frac{R_s}{Z_0}\frac{J_1(k_c r)}{J_0(k_c r)} \quad (7)$$
$$\tan(\theta(z)) = -2\frac{z}{L}\frac{\beta}{k}\frac{R_s}{Z_0}\tan(\beta z)$$

This result shows that the spatial phase distribution in the cavity can be approximated by two simple one dimensional equations. The radial phase distribution is plotted in Figure 2, assuming copper metal walls. This phase variation is similar to the one already presented by Vecchi and De Marchi [3].

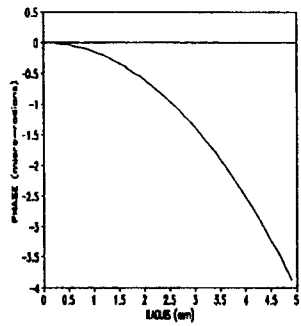


Figure 2: Radial phase variations in the idealized fountain cavity

3. A NUMERICAL APPROACH

The analytical method just described cannot be used to calculate the spatial phase distribution for the fountain cavity with holes in the end caps and waveguide ports coupled into the side wall. This is due to the difficulty of finding an analytical closed form solution for these more complicated boundary conditions. To circumvent this difficulty, numerical methods of calculating the cavity spatial distribution are considered. However, the numerical methods normally used to solve electromagnetic problems do not have the necessary precision to directly calculate the phase perturbations with the desired accuracy. In fact phase variations smaller than 100 μradians are relevant to this application. Alternative

numerical methods must be devised.

A numerical method based on a perturbation approach was found, which is suitable for this application. The fields are first calculated for perfect metal walls by numerical methods such as the finite element or finite difference method, which solve the wave equation as an eigenvalue problem. These fields will be referred to as the primary fields of the cavity.

The magnetic field vector is the sum of a primary field $\mathbf{H}_0$ and a secondary field $\mathbf{h}$, so that $\mathbf{H}=\mathbf{H}_0+\mathbf{h}$. The time-independent wave equation for the magnetic field is given by

$$\nabla^2\mathbf{H} + k^2\mathbf{H} = 0 \qquad (8)$$

The wave number k appearing in this equation can be expressed in terms of the resonant wave number k_0 and its shift Δk, given by (1). By substituting these expressions into the wave equation, and neglecting second order terms, the following equation can be derived

$$\nabla^2\mathbf{h} + k_0^2\mathbf{h} = -\frac{k_0^2}{Q}\mathbf{H}_0 \qquad (9)$$

This equation is a quasi-static inhomogeneous wave equation for the secondary field, since it has a fixed wave number and a distributed source term related to the primary field. A similar secondary field equation for the electric field can also be derived.

The boundary conditions appropriate for each secondary field equation can be found by connecting the secondary field to the primary field at the metal wall through the surface impedance and Maxwell's equations. Once the secondary field is found, the spatial phase can be directly computed from the ratio of the imaginary part of the total field to its real part.

It should be noted that the solution to the above secondary field equation is not unique. To every solution, a field proportional to the primary field can be added, and still satisfy the same boundary conditions and field equation. Any numerical solution of the secondary field equation will therefore need to prevent the primary field from being mixed in with the secondary field solution.

To calculate the effect of holes in the end caps of the fountain cavity, a cylindrically symmetric model of the cavity can be used. Figure 3(a) depicts such a model. On the other hand, to estimate the effect of coupling a waveguide port into the cavity sidewall, the constant cross-section cavity model depicted in Figure 3(b) can be used.

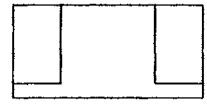


Figure 3(a): R-Z plane cross-section of fountain cavity

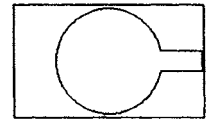


Figure 3(b): X-Y plane cross-section of fountain cavity

The electric field for the TE_{011} mode has only the azimuthal component. To obtain the secondary fields, it is therefore convenient to solve for the electric secondary field $e(r,z)$, with $u(r,z)=re(r,z)$. The secondary field equation in this case is given by

$$\frac{\partial^2 u}{\partial r^2} - \frac{1}{r}\frac{\partial u}{\partial r} + \frac{\partial^2 u}{\partial z^2} + k_0^2 u = -\frac{k_0^2}{Q}U \qquad (10)$$

where $U(r,z)$ denotes the corresponding primary field. The boundary condition at the metal wall is given by $u(r,z)=rZ_s H_{t0}$, where H_{t0} is the tangential primary magnetic field at the boundary, and Z_s is the surface impedance. The longitudinal component of the magnetic secondary field can then be derived from the function

u(r,z), through Maxwell's equations.

For the constant cross-section cavity, the secondary field equation can be directly formulated in terms of the longitudinal component of the magnetic field h(x,y) so that

$$\frac{\partial^2 h}{\partial x^2} + \frac{\partial^2 h}{\partial y^2} + k_c^2 h = -\frac{k_0^2}{Q} H_0 \quad (11)$$

The boundary conditions for h(x,y) require that its normal derivative be given by

$$\frac{\partial h}{\partial n} = \frac{k_c^2}{k} \frac{R_s}{Z_0} H_0 \quad (12)$$

The secondary fields u(r,z) and h(x,y) are defined at the nodes of a numerical mesh within a finite element or finite difference discretization of the secondary field equation. Let the vector [u] denote the values of the function u(r,z) at n mesh nodes in the r-z plane, and the vector [h] be the values of h(x,y) at m mesh nodes in the x-y plane. Let [U] and [H] be their corresponding primary field vectors. The finite element or finite difference discretization of the secondary field equations leads to the following matrix equations

$$\begin{aligned} [A][u] &= [B][U] \\ [C][h] &= [D][H] \end{aligned} \quad (13)$$

where [A] and [B] are n by n matrices and [C] and [D] are m by m matrices that arise from the discretization process and incorporation of the appropriate boundary conditions. These matrices are symmetric and have a sparse number of non-zero elements. The standard way to solve these equations is to use iterative methods such as the successive over relaxation or the conjugate gradient method.

At each iteration in the solution of the above matrix equations it was found necessary to eliminate the spurious appearance of the primary field in the secondary field solution.

4. RESULTS

The solution of the matrix equation for the secondary fields in the r-z plane was first tested against the phase distribution for the idealized fountain cavity. A finite difference program solving the secondary field equation was written. The results showed good agreement with the analytical predictions, having a maxmum error of 0.25% at r=5mm and z=10mm.

Figure 4(a) shows magnetic field lines for the primary field close to an 8mm diameter hole at the end cap in the r-z plane. The figure clearly shows that beyond a critical radius, $H_z(z)$ changes sign at some z close to the entrance of the beam passing hole. As a consequence its phase will undergo a π radian change. The critical radius for which this occurs was found to be around 65% of the hole radius.

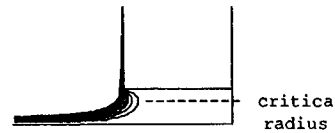


Figure 4(a): Primary magnetic field lines close to hole in cavity end caps

The secondary magnetic field lines for the fountain cavity having holes in its end caps are shown in Figure 4(b). The maximum value of the secondary field was found to be $6*10^{-5}$ of the maximum primary field. The longitudinal phase distribution on cavity axis through the holes in the end caps is compared in Figure 4(c) with the case of the idealized cavity without holes. The results show that, for a cavity with holes, the phase has a much slower variation than predicted for the idealized cavity.

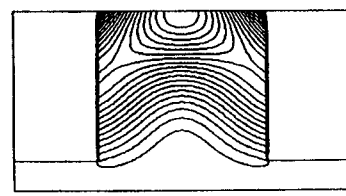


Figure 4(b): Magnetic field lines for the secondary field in the fountain cavity with holes

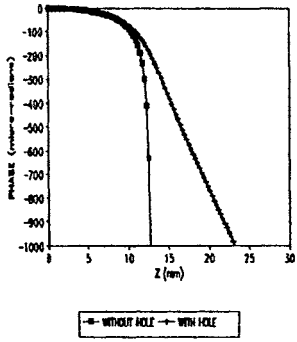


Figure 4(c):Longitudinal phase variation in fountain cavity

Figure 5(a) shows the electric field lines of the primary field in the x-y plane when a waveguide port of 10mm in height coupling into the side wall of the fountain cavity.

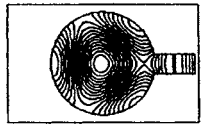


Figure 5(a): Electric field lines for the primary field with asymmetric waveguide coupling

The secondary field in this case was solved by use of the finite element method, which was found to be more flexible in applying normal derivative boundary conditions. The secondary electric field lines calculated for the primary field shown above are depicted in Figure 5(b)

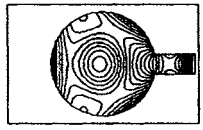


Figure 5(b): Electric field lines for the secondary field with asymmetric waveguide coupling

The longitudinal component of the magnetic secondary field was found to vary from $-3*10^{-5}$ to $3*10^{-5}$ of the maximum primary field value. In this case, the geometrical axis of the cavity is significantly different from the electromagnetic centre of the field. Figure 5(c) confirms that the phase distribution relative to the geometrical centre of the cavity is shifted to one side, and shows that there is considerable variation between vertical and horizontal phase distributions.

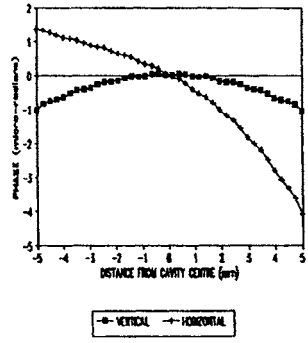


Figure 5(c):Radial phase variations in asymmetrically coupled fountain cavity

A comparison of vertical and horizontal phase distributions relative to the centre of the fountain cavity was also made for waveguide ports opened symmetrically on either side of the cavity. Figure 6(a) and 6(b) show the primary and secondary electric field lines respectively for such a layout. These diagrams demonstrate that another waveguide port can be used to correct for the field asymmetry displayed in the preceeding cavity layout. Figure 6(c) shows that although the phase variation in the horizontal direction is considerably reduced, it has a steeper variation in the vertical direction when compared to the idealized cavity without holes.

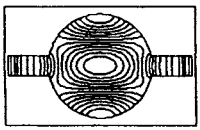


Figure 6(a): Electric field lines for the primary field with symmetric waveguide coupling

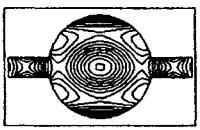


Figure 6(b): Electric field lines for the secondary field with symmetric waveguide coupling

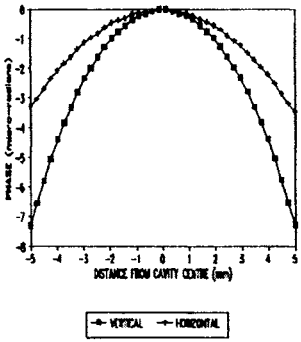


Figure 6(c):Radial phase variations in symetrically coupled fountain cavity

5. CONCLUSIONS

A numerical method has been presented which is able to calculate spatial phase variations for cavities with a wide variety of shapes, and in particular for the one proposed for use in cesium fountain standards. The method has been used to calculate the effect on the phase distribution of holes in the cavity end caps and of waveguide ports on the cavity side wall. Future work will extend the method to include the source. Generalization to three dimensional analysis is also anticipated.

References

1. R. E. Collin, Foundations for Microwave Engineering, 1966, p354

2. ibid, p327

3. G. Vecchi and A. De Marchi, Spatial phase variations in a TE_{011} microwave cavity for use in a cesium fountain primary frequency standard, IEEE Trans. Instruments and Measurements, Vol. 42, No. 2, April 93

Resonators (B 1)

NEW-DOUBLY ROTATED QUARTZ CUTS FOR THE DESIGN OF SAW RESONATORS LOW SENSITIVE TO THERMAL AND MECHANICAL PERTURBATIONS

S. Ballandras, E. Bigler, C. Bonjour, G. Martin
Laboratoire de Physique et Métrologie des Oscillateurs du C.N.R.S.
associé à l'Université de Franche-Comté-Besançon
32 avenue de l'Observatoire - 25044 Besançon Cedex - France

Summary :

In this paper, three new families of doubly rotated quartz cuts for the design of Surface Acoustic Wave (SAW) resonators are presented. These SAW quartz cuts have been theoretically found exhibiting a very low sensitivity to radial symmetrical in-plane stresses and to statical temperature effects. Their interest for the design of ultra-stable SAW resonators is based on their very low power flow angle. The theoretical approach used to point out the existence of such quartz cuts is exposed and also the first experimental results of the sensitivity of the new cuts to quasi-static temperature variations is presented.

1. Introduction :

The design of very stable oscillators with a high spectral purity mainly depends on the existence of resonators with a low sensitivity to the variations of physical parameters such as temperature or mechanical stresses. Such resonators can be built lising anisotropic properties of the propagation substrate (mostly quartz plates) coupled with particular geometries of these plates. This approach has been implemented by Eernisse /1/ who has pointed out the existence of a Bulk Acoustic Wave (BAW) quartz cut which is compensated for both temperature and mechanical stress effects when associated with a thin circular plate geometry. This cut (namely the SC-cut : $\Phi = 22°$, $\theta = 34°$) is doubly-rotated and then implies more complicated operations to be cut, which is ever a disadvantage in terms of industrial applications. Nevertheless, its properties were sufficiently

attractive to overcome these practical difficulties and the SC-cut is now an industry standard in the design of BAW resonators.

Eernisse's theoretical approach has been transposed to the case of Rayleigh surface wave propagating on a quartz substrate. Two crystalline orientation families have been theoretically found exhibiting low sensitivities to quasi-static temperature variations and radial in-plane stresses (case of thin circular plates). The first, B.K. Sinha's STC cut /2/, is defined by the cut angles $\Phi = 0°$, $\theta = 41.8°$ and the propagation direction $\Psi = 48.2°$ and the second, studied by G. Théobald *et al.* /3/, is defined by the angles ($\Phi = 0°$, $\theta = -10°$, $\Psi = 30°$) and is called TG cut (cut angles given according IEEE standard /4/). Their physical properties have been checked using delay lines built on thin disks /5/ and their sensitivity to dynamical thermoelastic variations has been experimentaly mesured /6/ and found an order of magnitude lower than the (ST,X) cut one (the (ST,X) cut is the SAW quartz cut industry standard). Unfortunately, these two singly-rotated cuts did not present a zero power flow angle which is an important criterium in the design of very high stability SAW resonators built with classical interdigital transducers.

But, since only singly-rotated SAW quartz cuts have been studied to find an equivalent of the SC-cut, the existence of SAW devices with a low sensitivity to stress and temperature effects and a power flow angle close to zero among all the doubly-rotated quartz cuts had to be verified. Then, a systematic study of all the doubly-rotated quartz cuts has been done to find new cuts satisfying to all three criteria : zero temperature effects, zero planar isotropic stress effects, zero power flow angle. New families of cuts have been found and will be called LSP1, LSP2 and LSP3 in the present paper (LSP means Low Sensitivity to Perturbation). The theoretical approach used for this study will be described in the first part of the paper. It is based on the classical harmonic model of the Rayleigh SAW propagation on piezoelectric media yielding characteristic of the non-perturbed wave, particularly the power flow angle /7/, and on the perturbation method developed by H.F. Tiersten to describe temperature and stress effect sensitivity of SAW devices /8/.

The second part is devoted to more precise calculations on the sensitivity of the new cuts to symmetrical radial in-plane stress effects. Particularly, this

sensitivity will be evaluated using an anisotropic analytical model of stress distribution on a 2D circular plate and the stress sensitivity coefficients calculated using H.F. Tiersten's perturbation method.

In the third part of the paper, the first experimental results showing the sensitivity of the LSP cuts to quasi-static temperature variations are presented. It is emphasized that a zero first order static temperature coefficient can be found experimentally for doubly-rotated SAW quartz cuts with a smooth correction of the propagation direction Ψ (2° to 3° typically). These experimental results provide good reasons to hope that the theoretically predicted characteristics of the new LSP cuts will be experimentally confirmed.

2. Investigation tools and method

2.1. Electromechanical coupling factor and power flow angle

The propagation parameters of Rayleigh surface wave on a semi-infinite piezoelectric crystal derives from a very classical linear harmonic analysis developed in many references /7,9/. The electromechanical coupling factor will be defined as follows /10/.

$$\frac{K_s^2}{2} = \frac{(V_\infty - V_0)}{V_\infty} \qquad (1)$$

where V_∞ is the velocity of the wave propagating on a free surface and V_0 the volicity of the wave propagating on a metallized surface. The power flow angle η is defined as

$$\eta = Arctg\left(\frac{W_3}{W_1}\right) \qquad (3)$$

where W_1 and W_3 are the components of the surface electro-acoustical Poynting vector in the SAW reference axes (Fig.1). Both quantities $K_s^2/2$ and η only depends on the anisotropic properties of the crystal and can be mapped as a function of the angles Φ, θ and Ψ. For practical applications, a zero

power flow angle and an electromechanical coupling factor of the same order of magnitude as for the (ST,X) cut are criteria to be fullfilled.

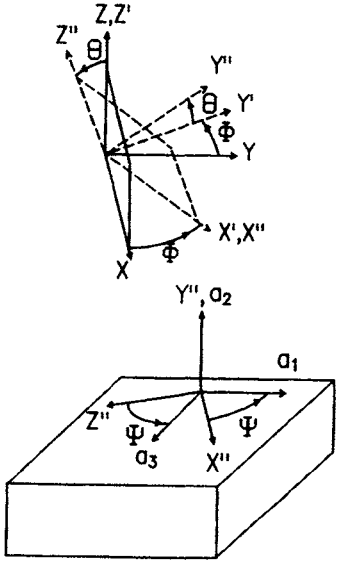


Fig. 1 : *Diagram showing the definition of the axes relatively to the surface of a semi-infinite propagation substrate*

2.2. Sensitivity to 1st order temperature effects

A very satisfying evaluation of SAW quartz cut sensitivity to first order temperature effects can be obtained following Sinha and Tiersten's procedure /11/. It is based on a perturbation method /8/ involving variational equations of small elastic fields superposed on a bias /12/. This approach allows a rigorous computation of nonlinear properties of quartz for thermoelastic effects /13/. First, the propagation substrate is assumed to expand freely in all the spatial directions during the static temperature variations. This implies that no stress arises and that the mechanical strains $\overline{S_{ij}}$ will be related to temperature T via the thermal expansion coefficients a_{ij} as follows

$$\overline{S_{ij}} = a_{ij} (T - T_o) \tag{4}$$

where T_o represents the reference temperature (generally equal to 25°C). Assuming now that no rigid rotation takes place around the fixed centre of the crystal plate ($a_1 = a_2 = a_3 = 0$) /11/ and taking into account the first derivatives of the fundamental elastic constants of quartz /13/, the relative frequency shift $\Delta F/F_o$ due to a statical thermal perturbation limited to first order effects will be expressed as follows

$$\frac{\Delta F}{F_o} = {}^{\theta}a (T - T_o) \tag{5}$$

where the sensitivity coefficient to first order statical temperature variations ${}^{\theta}a$ depends on the nonlinear anisotropic thermoelastic properties of the crystal and the non-perturbed characteristics of the Rayleigh wave propagation. As previously in the case of the power flow angle, the value of ${}^{\theta}a$ will depend on the angle Φ, θ and Ψ.

2.3. Sensitivity to mechanical stresses

The perturbation method described in the previous section can also be used to characterize the sensitivity of SAW quartz cuts to mechanical stress effects /14/. The hypothesis of no rigid rotation around the fixed centre of the crystal plate will be retained. This approach allows to relate directly the relative frequency shift $\Delta F/F_o$ with the statical stress tensor $\overline{T_{ij}}$ by defining the stress sensitivity coefficients ${}^{s}a_{ij}$ /15/. These coefficients are obtained by associating the nonlinear elastic properties of the substrate and the propagation characteristics of the Rayleigh wave as in the case of θa. The relative frequency shift due to mechanical perturbation will be expressed as follows

$$\frac{\Delta F}{F_o} = {}^{S}a_{ij} \overline{T_{ij}} \tag{6}$$

The validity of this model can be discussed in order to clearly establish its application fields. For example, if the stress field applied to the substrate invalidates the imperturbed non lossy Rayleigh wave model, equation (6)

will not allow a precise evaluation of the induced frequency shift. However, since the purpose of the calculation developed in the present paper is to determine the sensitivity of SAW quartz cuts to radical in-plane stresses due to diametrical symmetrical compression or normal hydrostatic pressure effects /14,16/ which can be considered as small perturbations, this model will provide enough precision to compare the performances of various SAW quartz cuts versus the cut angles Φ, θ and Ψ.

2.4. Criteria of investigation for the new cuts

It has been recalled in the first paragraphs that the power flow angle (and the electromechanical coupling coefficient) and the SAW sensitivity to static temperature variations do not depend on the geometry of the propagation substrate (which is assumed to expand freely in all of the spatial directions).

But the SAW sensitivity to mechanical perturbations will mainly depend on the geometry of the propagation substrate which will condition the stresses induced by the mechanical biases. Since there is no mathematical reason to find all the stress sensitivity coefficients to be zero for a given set of angles Φ, θ and Ψ, it is of interest to select a geometry well-suited to compensate the larger number of mechanical perturbations.

Previous works /2, 15, 16, 17/ have shown that in the case of thin isotropic disk submitted to double radial symmetrical in-plane compression /18/, the planar stress distribution at the center of the plate will be expressed as follows

$$\overline{T_{11}^{o}} = -\frac{2P}{\pi d}, \overline{T_{33}^{o}} = -\frac{2P}{\pi d}, \overline{T_{13}^{o}} = 0 \qquad (7)$$

where P represents the force per unit thickness of the plate, d is the diameter of the disk and the upperscript "o" denotes that the stresses are calculated in the plate axes.

The stress distribution (7) is then transformed by a tensorial rotation which gives access to the new stress field expressed in the SAW reference axes

$$\overline{T_{11}} = \frac{1}{2}\left(\overline{T^{o}_{11}} + \overline{T^{o}_{33}}\right);\ \overline{T_{33}} - \frac{1}{2}\left(\overline{T^{o}_{11}} + \overline{T^{o}_{33}}\right);\ \overline{T_{13}} = 0 \quad (8)$$

It can be noted that this stress distribution does not depend on the azimuthal angle β between the force and the SAW propagation direction. But since quartz is an anisotropic material, eq. (8) will not faithfully describe the mechanical stress state of the plate. However, this isotropic approximation has been validated for singly-rotated SAW quartz cuts /16, 17/ and has been estimated sufficient for a first investigation. Then , using eq. (8) and (6), the SAW sensitivity to double diametrical compression will be expressed as follows.

$$\frac{\Delta F}{F_o} = -\frac{2P}{\pi d}\left({}^{S}a_{11} + {}^{S}a_{33}\right) \quad (9)$$

Thus, a systematic investigation based on numerical computations has been implemented in order to check the conjunction of a zero power flow angle and a low sensitivity to first order temperature and mechanical planar stress effects for SAW quartz cuts defined by the angles Φ, θ and Ψ. The criteria of investigation are summarized below

$$\eta = 0\,,\ {}^{\theta}a = 0\,,\ {}^{S}a_{11} = -\,{}^{S}a_{33} \quad (10)$$

2.5. Computation results

A FORTRAN program has been written in order to calculate systematically the previously defined properties of SAW quartz cuts. Taking the symmetries of quartz into account /4/, the parameters of exploration by numerical computation have been fixed as follows. :

- Φ varying from $-$ 60° to + 60° with a step of 10°
- θ varying from $-$ 90° to + 90° with a step of 5°
- Ψ varying from 0° to 90° with a step of 5°

The results presented in Fig. 2 under the form of maps show for a given value of Φ the superimposition of the criteria (10) for each θ and Ψ. Using these maps, it is possible to point out three regions of cuts which satisfy simultaneously the three research criteria. Enlargements of the contour map of Fig. 2 are presented for each family ot the new cuts called LSP1 (Fig. 3a), LSP2 (Fig. 3b) and LSP3 (Fig. 3c). These enlargements allow to define more precisely the angles of the cuts and also to establish what kind of tolerances can be expected for the angles Φ, θ and Ψ. Figure 3a shows that the LSP1 cut presents a zero power flow angle with a good tolerance to cut errors on the angle θ, but the triple intersection is more sensitive to the angle Φ than the other cuts.

LSP2 cut does not exhibit a real triple intersection as it can be seen on Fig 3b, but the stress compensation and the zero power flow angle appear to be low sensitive to errors on the propagation direction Ψ (parabolic shape of the curves).

Properties of LSP3 cut illustrated in Fig 3c seem to be particularly interesting because the conjunction of the three criteria is less sensitive to angle Φ as for the other cuts. Nevertheless, it must be emphasized that this cut exhibits a lower electromechanical coupling coefficient (half of the two other cuts). LSP1 and LSP2 cuts offer a electromechanical coupling as high as (ST,X) cut's one and will certainly present at least similar electrical characteristics as the SAW industry standard. Table I and Table II summarize the value of angles Φ, θ and Ψ for the LSP cuts and some of their propagation characteristics (velocity and electromechanical coupling coefficient).

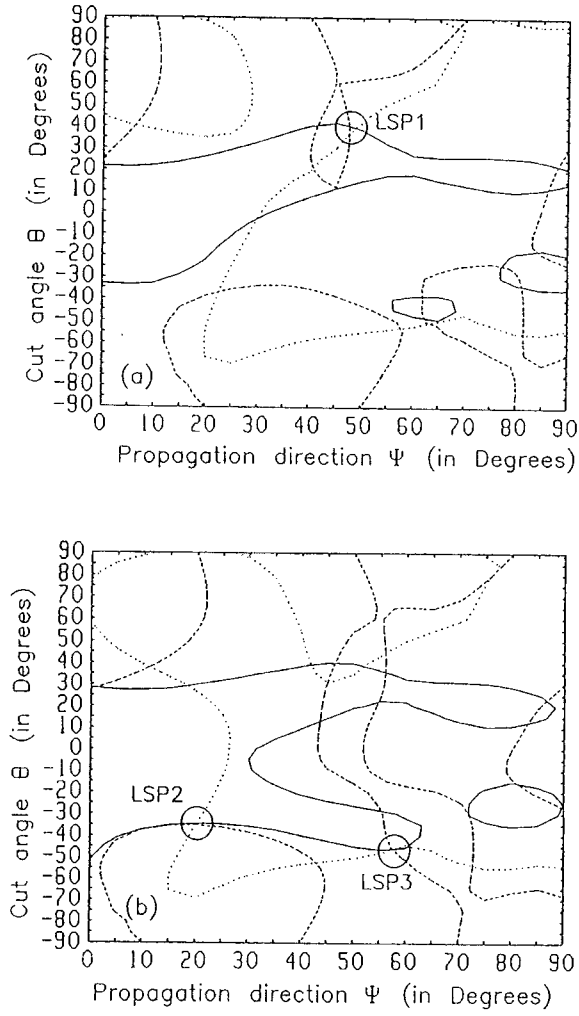


Fig. 2 : *Maps for the determination of cuts in accordance with the following criteria :*

------ zero power flow angle (η = 0°)

..... zero 1st order temperature coefficient (${}^{\theta}a$ = 0 ppm.K^{-1})

—— zero stress sensitivity coefficient (${}^{s}a_{11} = -{}^{s}a_{13}$)

(a) Φ = -10° (b) Φ = -20°

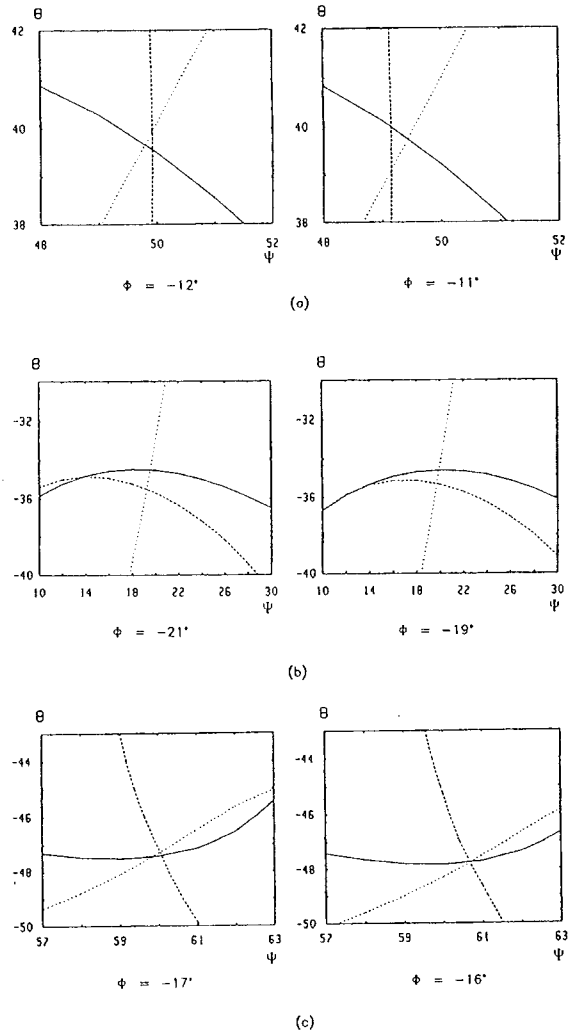


Fig. 3 : *Enlargements of the maps presented in Fig. 2 in the vicinity of the 3 LSP cut families*

----- zero power flow angle ($\eta = 0°$)
..... zero 1st order temperature coefficient (${}^{\theta}\alpha = 0$ ppm.K^{-1})
—— zero stress sensitivity coefficient (${}^{s}\alpha_{11} = -{}^{s}\alpha_{13}$)
(a) LSP1 cut (b) LSP2 cut (c) LSP3 cut

Table I : *Families of SAW quartz cuts with zero power flow angle and low sensitivity to both mechanical stresses and temperature variations*

Designation	Cristallographic cut angles according to /4/		Propagation direction
	Φ (in degrees)	θ (in degrees)	Ψ (in degrees)
LSP1	−11.7	39.7	49.5
LSP2	−20	−35.5	20
LSP3	−16	−47.7	60.8

Table II : *Electromechanical coupling factor and Rayleigh velocity of the LSP families of cut compared with (ST,X) cut characteristics*

Designation	Electro-mechanical coupling coefficient $\Delta V/V$ ($\times 10^{-2}$ %)	Velocity of the Rayleigh Wave (free surface) in $m.s^{-1}$
LSP1	5.21	3288.24
LSP2	6.85	3443.59
LSP3	2.34	3405.88
ST,X	5.80	3157.57

3. Improvement of the diametrical compression model

A simple isotropic model has been used to characterize the sensitivity of SAW quartz cuts to radial symmetrical in-plane compression. This isotropic approximation is, generally, valide for singly-rotated cuts. Then it is necessary to evaluate how good this approximation is for doubly-rotated cuts. This can be theoretically achieved using either an analytical anisotropic model of diametrical compression /18/.

The analytical anisotropic model is based on a 2D calculation of the stress field induced by a single or double symmetrical compression. The statical stress tensor is calculated for each azimuthal angle β by mean of complex integration /18/. For the case of single diametrical compression, its terms will be expressed in the plate axes as follows

$$\overline{T^{\rho}_{11}} = \frac{P}{\pi d}\left(\frac{4A}{1+\delta+\varepsilon} - \frac{A+B}{1-\varepsilon-\gamma}\right); \overline{T^{\rho}_{33}} = \frac{P}{\pi d}\left(\frac{A+B}{1-\varepsilon+\gamma}\right); \overline{T^{\rho}_{13}} = 0 \qquad (11)$$

where γ,δ and ε are combinations of compliance constants expressed in the plate axes and A and B are two coefficients calculated by complex integration. After transformation of the statical stress tensor in the SAW reference axes by tensorial rotation, it is possible to calculate the frequency shift induced by compression effects using eq. (6) which remains valid for the anisotropic model. All the calculation steps are more detailed in reference /5/.

This anisotropic model, shortly described, has been applied to the new LSP cuts. Numerical calculations have been performed for an azimuthal angle β varying from 0° to 180° by step of 5°, taking into account the symmetry of SAW propagation. Results are presented on figures 4a, b, c, and compared to the classical isotropic model used in section II of the present paper. These figures emphasize a satisfying coherence between the two models. Nevertheless, a focus on the 4-point compression shown in figures 5a, b, c, for the three LSP cuts points out an evident discrepancy between isotropic and anisotropic calculations. However, it can be remarked that for the LSP1 and LSP3 cuts, three orders of magnitude separate the 2-point compression case from the 4-point symmetric compression case, and more than 2 orders of magnitude for the LSP2 case.

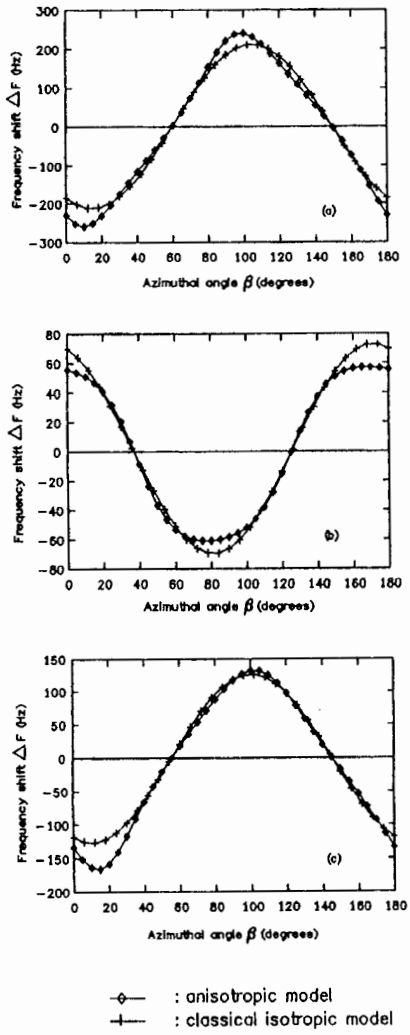


Fig. 4 : *Comparison between theoretical force-frequency effects due to single diametrical compresion applied on LSP cuts, super-imposition of isotropic stress calculation and full anisotropic stress model results*

(a) LSP1 cut (b) LSP2 cut (c) LSP3 cut

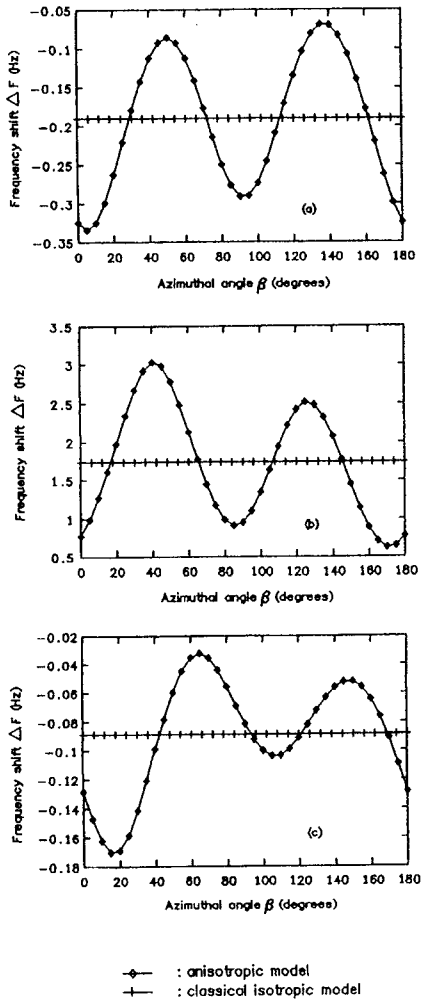


Fig. 5 : *Comparison between theoretical force-frequency effects due to double diametrical compresion applied on LSP cuts, super-imposition of isotropic stress calculation and full anisotropic stress model results*

(a) LSP1 cut (b) LSP2 cut (c) LSP3 cut

4. Experimental measurement of sensitivity to statical thermal effects

The theoretical approach exposed in the first parts of the present paper have been provided satisfying results in the case of BAW resonators /1/ and singly-rotated SAW quartz cuts /2, 3/. Even if the BAW SC-cut is a doubly-rotated cut, its realization implies only two rotations (Φ and θ) as in the case of singly-rotated SAW quartz cuts like TG or STC cuts (rotation of θ and Ψ). But the LSP cuts require three rotations (Φ, θ and Ψ). This fact may introduce some difficulties to verify experimentally the physical properties of these cuts. Particularly, the side of the crystal plate where the SAW must propagate is not indifferent because of the symmetries of quartz which impose that /4/.

$$(\Phi, \theta, \psi) = (-\Phi, \theta, -\psi) \tag{12}$$

More, if an error occurs during the rotation of the angle Φ, it will be amplified by the following rotation θ and Ψ. However, a set of LSP cut has been tested to check first the statical temperature compensation effect theoretically predicted.

The experimental measurements have been performed on SAW delay lines composed of two interdigital transducers deposited on the quartz substrates. The period of the transducer was λ = 34,4 µm (Rayleigh wavelength). The delay of the device depends strongly on the SAW velocity and is varying from 1.3 to 1.6 µs for the case of LSP cuts. The delay-lines are looped with a selective amplifier and the whole device provides a frequency in the vicinity of 100 MHz.

The measurement process involves generally two thermal cycles. The first cycle allows to evaluate the error on the propagation direction Ψ which is corrected using the map of isothermal temperature sensitivity coefficient previously used to determine the LSP cuts. A simple linear regression provides with a satisfying accuracy the correction on the angle Ψ. Then a second thermal cycle allows to measure the temperature of inversion and the second order temperature sensitivity coefficient.

The SAW quartz delay-lines are placed in a brass box which is submitted to quasi-static temperature variations from − 20 °C t_o + 100 °C in an oven controlled by a PC-type computer. The duration of the cycles is long enough (8 hours) to ensure that no dynamical temperature effects occur.

For each LSP-cut, a correction of the propagation direction has been performed to find the compensation effect. For the LSP1 cut, the value of this correction was found equal to +3°. The second thermal cycle has been applied to the cprrected delay line. The result shown in Fig. 6a points out a turnover temperature equal to 88 °C. This discrepancy between theoretical and experimental results may be explained taking into account the precision on the cut angles and on the elastic coefficients of quartz. However, using a polynomial fit method, the first and second order temperature sensitivity coefficients have been evaluated and $^{\theta}a$ has been found equal to 3.8 ppm/°C. This value of $^{\theta}a$ corresponds to an error of 1.25° on Ψ to obtain the temperature compensation effect.

The correction applied on Ψ for the LSP2 cut was equal to −3.5°. In that case, the turnover temperature has been found equal to 26.1 °C. The frequency-temperature curve is reported in Fig. 6b. The second order temperature coefficient has been evaluated and is of the same order of magnitude as the (ST,X) cut one.

For the LSP3 cut, a correction of +4° has been applied on Ψ and the frequency versus temperature dependence of the modified delay line has been measured. The result appears in Fig. 6c and shows that the correction was too strong. Indeed, the turnover temperature is equal to 75 °C and $^{\theta}a$ is equal to 2.25 ppm/°C. As in the case of the LSP1 cut, this LSP3 cut needs more precision on the angle Ψ as for the LSP2 cut which appears the most interesting cut in regard with the first experimental results. Table III summarizes experimental data for each LSP cut and also for the (ST,X) cut /5/.

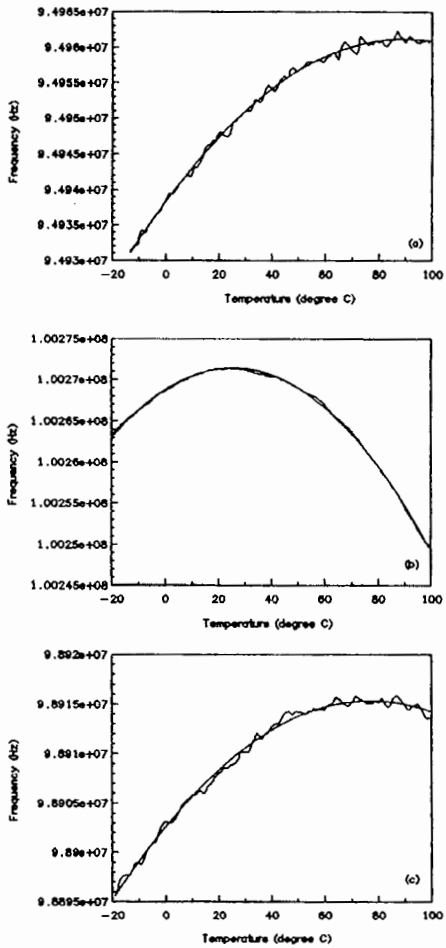


Fig. 6 : *Experimental measurements of frequency-temperature dependence for the LSP cuts*

(a) LSP1 cut : $\Phi = -10°$, $\theta = 40°$, $\Psi = 51°$
(b) LSP2 cut : $\Phi = -20°$, $\theta = -35°$, $\Psi = 16{,}5°$
(c) LSP3 cut : $\Phi = -20°$, $\theta = -45°$, $\Psi = 62°$

Table III : *Experimental characteristics of delay lines built on LSP cuts submitted to static temperature effects.*

	Fo (MHz)	Turnover temperature (°C)	First order temperature coefficient $^{\theta}\alpha$ (ppm/°C)	Second order temperature coefficient $^{\theta}\beta$ (ppb/°C²)
LSP1 cut $\Phi = -10°, \Theta = 40°, \Psi = 51°$	94.95	88	3.8	−30
LSP2 cut $\Phi = -20°, \Theta = -35°$ $\Psi = 16.5°$	100.27	26.1	≈0	−40.2
LSP3 cut $\Phi = -20°, \Theta = -45°$ $\Psi = 62°$	98.91	75	2.25	−21.6
(ST,X) cut /5/	91.3	25	≈0	−34

5. Conclusion

The existence of SAW quartz cuts low sensitive to statical temperature and mechanical in-plane stress effects and presenting simultaneously a power flow angle close to zero has been theoretically pointed out. The validity of the mechanical model selected for systematical numerical investigation has been verified and it appears to be precise enough to emphasize SAW quartz cuts stress compensated. First experimental results on LSP cut sensitivity to first order temperature effects have shown that the theoretically calculated properties of the new cuts can be practically found with almost a slight correction of the propagation direction. Even if the second order temperature sensitivity of the LSP cuts is not better than the (ST, X) cut, one can expect that if the new cuts are really stress-compensated, they will be very low sensitive to dynamical thermal perturbations, like the TG and STC cuts. Further experiments will be performed to check this hypothesis.

Acknowledgements

The authors would like to thank Pr H. Merigoux and Dr J.-F. Darces for their help in controlling the crystalline orientations of the new cuts

References

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

/2/ B.K. Sinha, "A stress and temperature compensated orientation and propagation direction for surface acoustic wave devices", IEEE Trans. Ultrason. Ferro. Freq. Cont., vol. UFFC-34,n° 1, pp. 64-74 (jan. 1987).

/3/ G. Théobald, E. Bigler, J.J. Gagnepain, "Stress-compensated cuts for surface acoustic waves", Proc. 1986 Ultrasonics Symp., IEEE cat. n° 86CH2375-4, pp. 209-212.

/4/ IEEE Standard on Piezoelectricity 176-1949, Proc. IRE, vol. 37 (1949), pp. 1378-1395.

/5/ S. Ballandras, E. Bigler, "Surface acoustic wave devices with low sensitivity to mechanical and thermoelastic stresses", J. Appl. Phys. Lett., vol. 72, n° 8, pp. 3272-3281 (october 1992).

/6/ S. Ballandras, E. Bigler, "Experimental study of SAW quartz cuts with low sensitivity to dynamical thermal effects", Elect. Lett., vol. 27, n° 15, pp. 1343-1345 (july 1991).

/7/ A.J. Slobodnik, E.D. Conway, R.T. Delmonico, "Microwave Acoustics Handbook, vol. 1A : Surface Wave Velocities", Air Force Cambridge Research Labs, 1973.

/8/ H.F. Tiersten and B.K. Sinha, "A perturbation analysis of the attenuation and dispersion of surface waves", J. Appl. Phys., vol. 49, n° 1, pp. 87-95 (jan. 1978).

/9/ K.A. Ingebrigtsen, "Surface waves in piezoelectrics", J. Appl. Phys., vol. 40, n° 7, pp. 2681-2687 (1969).

/10/ E. Dieulesaint, D. Royer, "Ondes élastiques dans les solides", Masson éd., 1974.

/11/ B.K. Sinha and H.F. Tiersten, "On the temperature dependence of the velocity of surface waves in quartz", J. Appl. Phys., vol. 51, n° 1, pp. 4659-4665 (jan. 1980).

/12/ J.C. Baumhauer and H.F. Tiersten , "Nonlinear electroelastic equations for small fields superposed on a bias", J. Acoust. Soc. Am., vol. 54, n° 4, pp. 1017-1074 (oct. 1973).

/13/ B. K. Sinha, H. F. Tiersten, "First temperature derivatives of the fundamental elastic constants of quartz", J. Appl. Phys., Vol. 50, n° 4, pp. 2732-2739 (1979).

/14/ B. K. Sinha, H.F. Tiersten, "On the influence of a flexural biasing state on the velocity of the propagation of surface waves", Wave Motion, vol. 1, pp. 37-51 (1979).

/15/ E. Bigler, G. Théobald, D. Hauden, "Stress sensitivity mapping for surface acoustic waves on quartz", IEEE Trans. Ultr. Ferro. Freq. Contr., vol. UFFC-36, n° 1, pp. 57-62 (jan. 1989).

/16/ S. Ballandras ,"Sensibilité des oscillateurs à quartz à ondes de surface aux contraintes et gradients thermiques", Thèse de doctorat, Université de Franche-Comté, Besançon, mars 1991.

/17/ B.K. Sinha, W.J. Tanski, T. Lukaszek, A. Ballato, "Influence of biasing stresses on the propagation of surface waves", J. Appl. Phys., vol. 57, n° 3, pp. 767-776, 1984.

/18/ D. Janiaud, "Modélisation de l'influence d'une accélération sur la fréquence des résonateurs à quartz", Thèse, Université de Besançon, 1978, pp. 63-67.

/19/ S. Ballandras, E. Bigler, "An analysis of QAS-like circular structures for reduced stress sensitivity in SAW devices", Proc. 47th Ann. Freq. Cont. Symp., Salt Lake City (UT), june 1993.

G- SENSITIVITY OF HIGH FREQUENCY QUARTZ CRYSTAL RESONATORS

M. ZAFRANY, T. SOUBIROUS, V. CANDELIER, P. CANZIAN, G. WAGNER

THOMSON / CEPE
44, Avenue de la glacière - BP 165
95105 Argenteuil - FRANCE

I. INTRODUCTION

Acceleration sensitivity of quartz crystal resonators is commonly attributed to the resonator design parameters such as cut angle, geometry of the blank, orientation of the mounting, number of supports /6/.

Experimental results published up to now partly supports this point of view. Concerning the blank design, no real differences are observed between the different cut angle ; Filler /1,2/ reports similar performances on AT, BT or SC cut resonators.

The sensitivity of resonators is strongly related to the radius of curvature for a plano-convex or bi-convex blank ; the better the symmetry is, the lower the sensitivity is. A ratio of 10 in performances is reported between plano-plano and plano-convex resonators for the AT cut while the SC cut appeared insensitive to this kind of asymmetry /1,3/.

The third overtone gives a better performance than the fundamental mode /1,2,3/. However, the fifth overtone use do not bring any improvements. The aspect ration D/2h (D : blank diameter, h : blank thickness) is suspected to have much more influence on the G-sensitivity than the thickness alone /1/.

A lot of results are underlining the influence of the orientation mounting of the blank. Hanson /8/ indicates a minimum

sensitivity at $\alpha = 20°$ and a maximum at $\alpha = 90°$ for a 2 point-SC resonator, while Filler /1/ do not observe any influence for a 4 point-SC cut resonator in the 0°-45° range.

A performance of $0.28.10^{-10}/g$ is reported by Lee /9/ at $\alpha = 0°$ for a 4 point-AT cut, and the orientation $\alpha = 60°$ leads to a reduction by a factor of 2 compared to $\alpha = 90°$ on AT resonators /1/. The optimum angle appears to depend on the number of supports and the cut angle.

The understanding of the G-sensitivity, largely studied in the recent years has led to two major statements /see review 10/ :

1/ the maximum symmetry must be achieved through both the mounting structure and the mass loading distribution /6/.

2/ the mounting must lead to a minimum deformation of the blank.

The Finite Element Analysis (F.E.A) is emerging as a powerful method to calculate the stress distribution of a complete structure /11,12/, and thus is becoming a very helpful tool in designing low G-sensitivity quartz crystal resonators.

The present paper is dedicated to survey the typical performances of resonators made on an industrial scale through a manufacturing process that inherently introduces scattering at different steps of fabrication. The first part recalls some typical results obtained at low frequency with an emphasis on the "QAS" design that yields the best performance in terms of absolute values and reproducibility.

The second part of the paper is dealing with recent work done on high frequency resonators with a special focus on blank design and mounting structures.

The modeling by F.E.A of this type of structure has been undertaken in the last part.

II. G-SENSITIVITY OF LOW FREQUENCY RESONATORS

The G-sensitivity of current fabrication of low frequency HC 40 resonators is given in Table 1. The four first resonators type have exactly the same mounting structure : 2 points ZZ' supported by nickel strip bounded with an epoxy silver filled adhesive.

(10^{-10}/g)		X	Y	Z	Γ
Frequency	type				
10 MHz	AT P5	2	20	32	36
5 MHz	AT P5	3	24	32	40
10 MHz	SC P3	3	22	6.5	24
5 MHz	SC P3	2	22	13	27
10 MHz	SC P3	1.6	2	2.3	3.6
	(QAS)				
10 MHz	SC P3	2	2.4	5	6
	(QAS+IP)				

Table 1

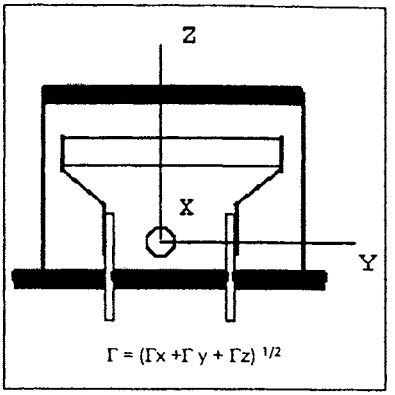


Figure 1 : Orientation axes

The measurement method described in details elsewhere /13/ has been implemented on an industrial control equipment that is controlling 10 resonators at a time with 1.10^{-10} /g reproducibility and a resolution better than 5.10^{-11}/g .

The standard deviation linked to the mentioned values(Table 1) varies typically from 1 to 3.10^{-10}/g in all directions, but may be as high as 10.10^{-10}/g for anomalous batches. Actually the standard deviation is related not only to the measurements variations but also to the variations in the manufacturing process.

The in plane and normal sensitivity analysis lead to the following comments :

a /The X sensitivity is very low and seems independent of both frequency and cut angle. In this case the deformation of the structure occurs in the strip plane

b /The Y sensitivity appears to be strongly linked to the HC 40 packaging with an average of 22.10^{-10}/g whatever the frequency or the cut angle are. The deformation involved in this direction corresponds to the flexion of the strips that induces tension/compression stress in the blank plane.

The effect of the mounting structure is then clearly seen in the QAS types of resonators ,reducing the sensitivity by a factor of 10. The Intermediate Plate (I.P) mentioned in table 1 is described in Ref. /14/ (the main advantage of this structure is to reduce the barometric sensibility, having a still good G-sensitivity).

c /The Z sensitivity shows an influence of the cut angle: the SC lots are almost 3 times better than the AT ones. The deformations involved in this direction are the bending of the strips that may induce complex stress (Torsion / compression) in the blank and the deformation of the blank itself.

The QAS case (10 and 20 MHz)

A previous study, made at CEPE, on the optimisation of the QAS design (Figure 2) has led to discard the number of mounts influence /5/. Two groups of resonators with 4 and 2 points yield the same value at 5.10^{-10}/g .

The mounting orientation of the blank exhibits a minimum sensitivity at 0°, 90°, 180° while the maximum is located at 50° and 120°.

The orientation error during the assembly step estimated at ± 5° is then inducing a variation of ± 0.5 10^{-10}/g.

Recent fabrications of QAS have the typical performances given table 2 (the first line is the average and the second is the standard deviation) :

		(10^{-10}/g)			
F	Type	X	Y	Z	Γ
10 MHz	HC 40	1.6	2	2.3	3.5
		(0.6)	(1.1)	(1.5)	(1.5)
20 MHz	HC 37	1	2.4	4.3	5
		(0.7)	(1.4)	(3.3)	(3)

Table 2

The newly designed 20 MHz QAS is partly aimed at improving the mechanical behaviour by reducing the size of the resonator.

The scaled-down has little effect on the in plane sensitivity and produces a slight increase in the normal direction.

A detailed analysis of the fabrication lots shows that two geometric defects can be introduced during the fabrication process. The first one (D1) is the misfit between the active part of the blank and the centre of the blank .The second defect (D2) is the misfit between the metal electrode centre and the centre of the active area (see figure 2).

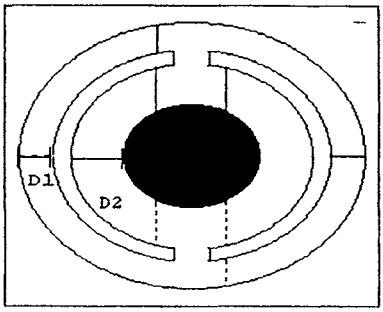


Figure 2 : QAS design

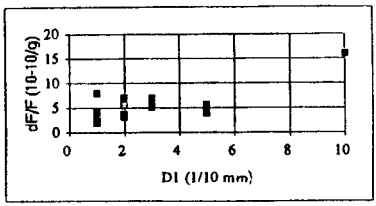


Figure 3-a

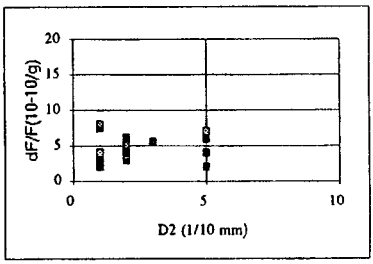


Figure 3-b

Figure 3 clearly shows that the D1 defect has no significant influence below 5/10 mm ; above the sensitivity is increased by a factor of 2 . However in the standard manufacturing process a misfit of 2/10 mm is eliminated at the ultra sonic drilling control step.

In a similar way the D2 defect doesn't seem to have any influence. It is to note that misfit above 5/10 mm are not achievable in the standard metallization process.

The QAS design offers the advantage of a $5.10^{-10}/G \pm 3.10^{-10}/g$ performance with good reproducibility from batch to batch since this structure is little affected by the manufacturing variations.

III. G-SENSITIVITY OF HIGH FREQUENCY RESONATORS

The starting point of our program on improving performances and increasing yields of high frequency resonator can be summarised by figure 4 :

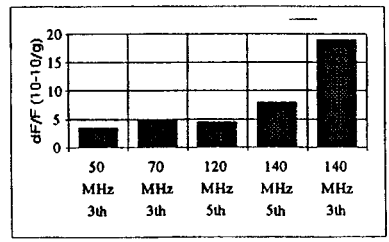


Figure 4

The standard design, namely 4 point mounts equally spaced with aluminium deposited electrodes bounded with adhesive epoxy silver filled in a Cold Weld HC35 package, doesn't give the needed performance with acceptable yields above 120 MHz.

The points reported on figure 4 are the average values computed from different batches of resonators. The standard deviation of the three first lots is $2.10^{-10}/g$ while for the two last lots it can be as high as $8.10^{-10}/g$ (we suppose, for the sake of simplicity, that the measurements obey a Gaussian law. Unfortunately, the proportion of resonators on the upper side of the distribution is not the same that the lower side ; actually the average is closer to the lower sensitivity).

A given number of design modifications were performed in order to improve the global performance described figure 4 :

1. Blank design

The improvement bring by the QAS structure - discussed in section II - is tested at frequency as high as 140 MHz.

A quartz ring surrounding the active area in the Inverted Mesa structure (I.M) is supposed to increase the stiffness of the blank comparing to the standard design (see Figure 5). In both cases the blanks were shaped through ion beam milling process.

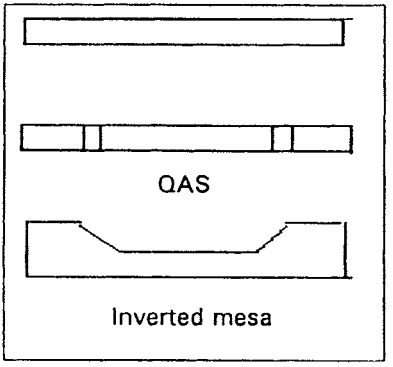


Figure 5 : Blanks design

The G-sensitivity is not improved by the I.M or the QAS structure as shown in figure 6-a. The detailed analysis of the measurements indicates that the normal sensitivity is almost the same in each case while the in plane (XY) sensitivity is doubled for the I.M and QAS structure (figure 6-b).

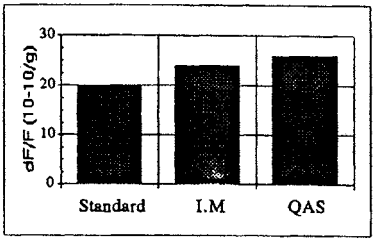


Figure 6-a

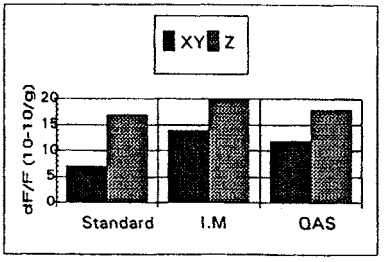


Figure 6-b

2. Electrode diameter and metal type

Two metal electrodes were deposited on the 120 MHz leading to a quality factor of $7.5.10^{12}$ for silver electrode and $6.5.10^{12}$ for the aluminium one. The measurements made on twenty resonators of each type do not confirm any quality factor effect (table 3).

	Aluminium	Silver
Average	$5.25.10^{-10}$/g	$5.45.10^{-10}$/g
std deviation	$1.8.10^{-10}$/g	$1.9.10^{-10}$/g

Table 3

The influence of electrode diameter was investigated on the 140 MHz for 1.5 mm and 2.5 mm. Table 4 doesn't evidence any improvement for the larger diameter.

	1.5 mm	2.5 mm
Average	22.10^{-10}/g	20.10^{-10}/g

Table 4

The various energy trapping conditions generated through metal type or electrode diameter are not showing any significant differences.

3. Mounting structure

Two types of mounting were realised: the direct mounting on 4 point - equally spaced and the indirect mounting on 2 points (see figure 7).

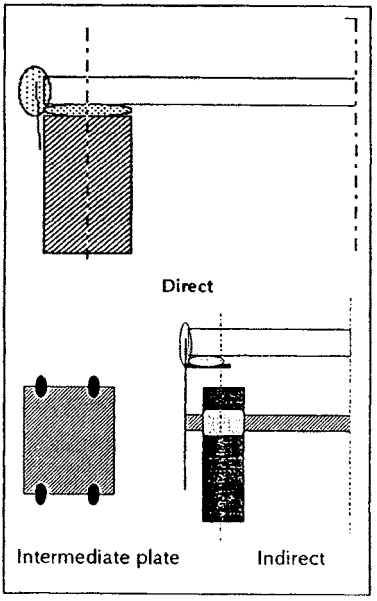


Figure 7 : Mounting structures

The indirect mounting structure was aimed at reducing the influence of the base which could induce unwanted deformation of the blank, the second reason was the use of large nickel strips that behave more like extended fixing than point fixing.

The sensitivity measured for different batches of each design is 22.10^{-10}/g and 15.10^{-10}/g respectively for the direct mounting and the indirect mounting structures. The in plane and normal analysis (figure 8) shows that the XY sensitivity is not affected by the design while the Z sensitivity is significantly reduced.

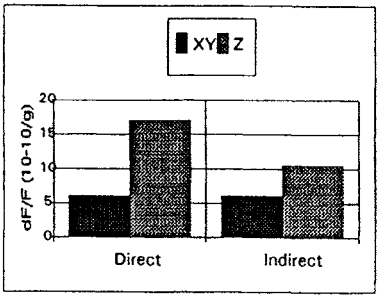


Figure 8

4. Adhesive and annealing effects

The stress induced by the different linear coefficients of expansion was investigated by two means, the first is the use of two conductive adhesives with different stiffness after curing cycle and the second is the annealing of manufactured lots at 130 °C for 8 days.

The epoxy bonding, known as soft, comparing to a cement bonding did not show any improvements : 22.10^{-10}/g in the first case and 23.10^{-10}/g in the second case.

The evolution of the G-sensitivity, after annealing, for two of the designs realised is summarised table 5 :

	X		Y		Z		Γ	
	to	+8 d	to	+8 d	to	+8 d	to	+8d
QAS	12	15	9	9	20	18	26	25
Std	10	9	7	8	16	16	22	22

Table 5

5. High frequency behaviour

The design parameters changes such as blank geometry, electrode diameter and metal, or adhesive bonding type do not seem to have much effects on the final performance.

The predominant role of the mounting is clearly seen using an indirect 2 points mounting structure. However, the performances of the 140 MHz resonator has not reaching the 50 MHz one.The main reason is the high sensibility in the normal plane measured at $10.10^{-10}/g$ in the best case. The mounting structure is not the only parameter to optimize, the blank deformation is becoming very influent. The aspect ratio already mentioned in introduction may probably have a benefic effect.
From a production point of view, the yield for a 5 $10^{-10}/g$ specification is not exceeding a few percent for the highest frequency measured here.

IV. FINITE ELEMENT ANALYSIS

The F.E.A. of the resonators already described has been undertaken in order to simulate the mechanical behaviour under various environments (Acceleration, vibrations, pressure, magnetic sensitivity,...). The eigenfrequency and the modes shape are shown figure 9 for the 20 MHz QAS blank including two mounts. This kind of analysis is now currently used to modify the mounting structure as to shift the unwanted eigenfrequency for given requirements.

The G-sensitivity simulation is realised by integrating the different stresses in the central area of the resonator for each direction of the acceleration vector. The σ_{xx} mapping under 1 g normal to the blank plane is shown figure 10.

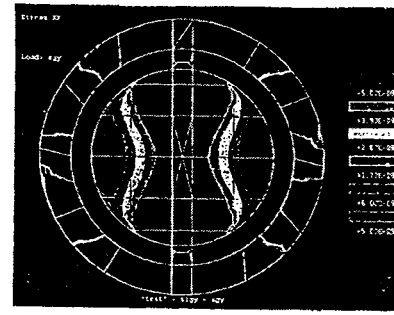


Figure 10

Using the SC coefficients given in Ref. /9/ the computed sensitivity is :

$$\frac{\Delta F}{F} = \Sigma\, k_i\, T_i$$

$\frac{\Delta F}{F}$	σ_{XX}	σ_{ZZ}	σ_{XY}	σ_{XZ}	Γ
X	2e-2	-	2e-3	-	**5e-14**
Y	-4.5e-1	1	-	2.5e-2	**3e-12**

(stress in N/m^2)

These results obtained on an "idealized" structure perfectly clamped in two points doesn't take into account the mounting influence.

Figure 11 shows the modeling of 2 and 4 points high frequency resonator under 1g perpendicular to the blank plane. The deformation in the first case is a "saddle" type while the second case is a "drumhead" type. Althought the stresses are much greater for the 2 points mount design, the computation made for a Sc cut shows a slight improvement by using a 4 points structure :

	4 points	2 points
σ xx	1.5	-4
σ yy	1.5	4.5
σ zz	-4e-2	4.5e-1
σ yz	9e-3	-2.5e-2
σ xz	1e-2	7e-2
σ xy	2e-2	3e-2
Γy	**1.2e-11**	**1.7e-11**

Table 6 : (stress in N/m²)

The extremely low sensitivity predicted by the F.E.A is also reported by different authors /11,12/.

The disagreement with measured values may arise from the following difficulties :

- The need of an accurate simulation that takes into account the details of the structure and the imperfections generated through a manufacturing process.
- The definition of the integrating area that may have complex limits /4/.

These two problems have to be resolved in order to get simulations closer to experimental results.

V. CONCLUSION

The various results obtained from 5 MHz to 140 MHz on standard structure or QAS structure with both AT and SC cut are summarised figure 12 ; the thickness e is reported in microns on the x axis :

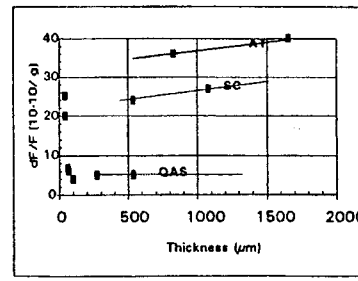


Figure 12

The better performance of the SC cut over the AT cut is clearly seen at low frequency, with a 30% factor of improvement.

The influence of the thickness for the AT cut is valid up to thicknesses above 100 µm, below this value, the sharp increase of the G-sensitivity is attributed to inadequate mounting structure. The QAS blank which produces real amelioration above 250 µm doesn't allow to reach good performances for lower thicknesses.

The low G-sensitivity can be reached through adequate mounting at low frequency; at high frequency, the mounting influence that assures a high degree of symmetry is not enough.The rigidity of the blank has to be taken into account.

Bibliography

/1/ R.L Filer, J.A. Kosinski, and J.R. Vig
Annual Frequency Control
Symposium p265-271 1983

/2/ R.L.. Filer, J.A Kosinski, and J.R. Vig
36 Annual Frequency Control
Symposium p215-219 1982

/3/ R.L Filer and J.R. Vig
35 Annual Frequency Control
Symposium p110-116 1981

/4/ B. Capelle and J. Detaint
44 Annual Frequency Control
Symposium p416-423 1990

/5/ M. Fragneau, B. Bonnotte, D.
Schottland Private communication

/6/ P.C. Lee and M.S. Tang
44 Annual Frequency Control
Symposium 1990

/7/ P.C. Lee and X. Guo
44 Annual Frequency Control
Symposium p468-473 1990

/8/ W.P. Hanson and T.R. Meeker
Annual Symposium Frequency
Control p478-487 1990

/9/ P.C.Y. Lee and M.S.H. Tang
Annual Symposium on Frequency
Control p413-415 1989

/10/ J.A. Kosinski and A. Ballato
European Frequency and Time
Forum p93-98 1993

/11/ F. Deyzac
European Frequency and Time
Forum 1993

/12/ L.D. Clayton anf E.P. Eernisse
International Frequency Control
Symposium p560-573 1993

/13/ JP. Aubry and P. Canzian
5 th European Frequency and Time
Forum 1992

/14/ JP. Aubry and JC. Craveur
1th European Frequency and Time
Forum 1987

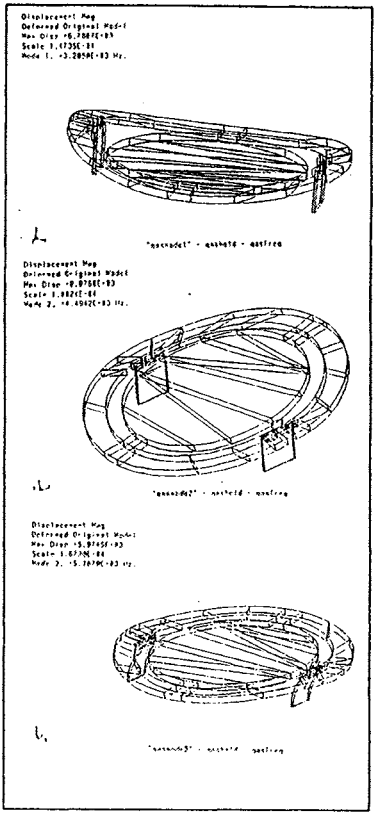


Figure 9

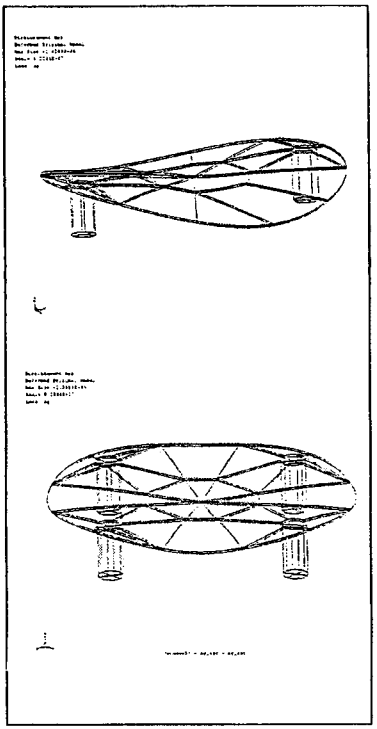


Figure 11

Measurement of Spurious Resonances of Crystal Units Using Network Analysers with Error Correction

Bernd W. Neubig
TELE QUARZ GROUP
D-74924 Neckarbischofsheim
Germany

Abstract

1. Introduction

The hybrid bridge technique for the measurement of spurios responses as standardized by IEC 283 (1968) [3] has significant limitations in accuracy and reproducibility. Significant progress has been made in the last years applying modern network analysers in connection with error correction methods to the measurement of crystal units [6].

The paper presents a systematic approach to adapt the new techniques to the measurement of spurious responses. The goal is to use the standard test set-up (such as the π-network or a s-parameter test fixture) as it is used for the determination of the crystal parameters of the main mode [5]. In connection with the published method for the measuerement of load resonance parameters[7], all crystal parameters can then be measured with one unique set-up configuration.

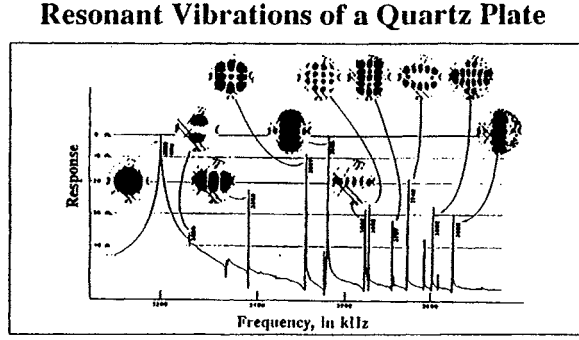
2. Classification of spurious responses

Unwanted or spurious modes have different origin and appearance. They can be

(1) anharmonic resonances of the same vibration mode (usually shear mode), whose frequency is in general above the relevant main mode (see Fig.1 [8]). These resonances are determined by the equation

$$f_{mnp} = \frac{1}{2}\sqrt{\frac{1}{\rho}}\sqrt{\frac{c_{65}\,n^2}{t^2} + \frac{c_{11}\,m^2}{l_x^{\,2}} + \frac{c_{66}\,p^2}{l_y^{\,2}}}$$

(2) unwanted modes (and overtones thereof) of other vibration modes, which may lie below or above the desired mode.



X-ray topographs ($2\bar{1}\cdot0$ plane) of various modes excited during a frequency scan of a fundamental mode, circular, AT-cut resonator. The first peak, at 3.2 MHz, is the main mode; all others are unwanted modes. Dark areas correspond to high amplitudes of displacement.

Fig. 1

Spurious modes can be

(i) isolated from the main mode, or can be

(ii) mechanically coupled to it, causing non-linear effects which depend on drive level (power dissipation) etc.

Spurious modes can lie

(I) very close to the main mode (a few kHz), such that the peak of the spur is on the slope of the main response.

(II) well apart from the main mode, and therefore can be considered as a separate peak.

Unwanted modes can cause problems

(a) in filters by distortion of the transfer characteristic in the pass band or the stop band: attenuation "dips" and phase discontiuities in the pass band, detoriation of the attenuation in the stop band

(b) in oscillators, the working frequency can "jump" to a strong unwanted mode or the frequency response can show discontinuities ("activity dips") at particular operation conditions (temperature, load capacitance/pulling voltage, drive level). VCXOs can show "dips" ins the pulling characteristic (frequency vs. pulling voltage) or in the response of the deviation vs. modulation frequency. TCXOs can show "dips" in the frequency vs. temperature characteristics.

In the following we will distinguish phenomenologically four different cases:

case A: strong spurs well "isolated" from the main mode,

case B: weak spurs well "isolated" from the main mode,

case C: strong spurs in the vicinity of the main modes,

case D: weak spurs in the vicinity of the main mode.

These cases are depicted in Fig.2.

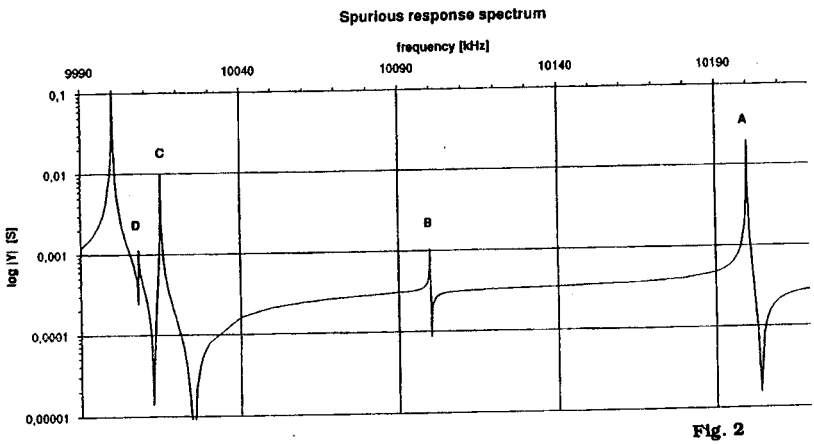


Fig. 2

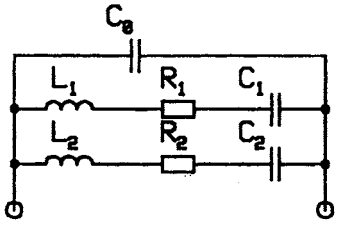
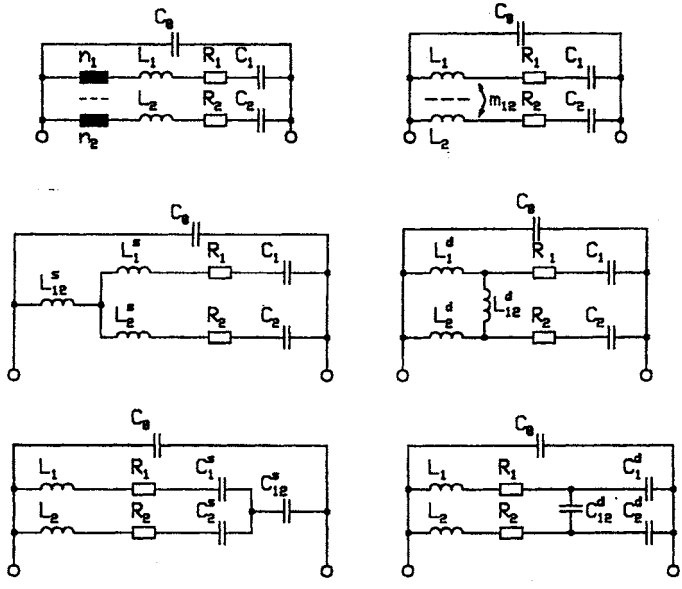


Fig. 3

3. Equivalent electrical circuits

Non-coupled resonances can be represented by seperate series resonance circuits all connected in parallel to the static capacitance C_0. In the following we use the symbols C_0 for static capacitance, L_1,C_1,R_1 for the motional parameters of the first (main) mode, and L_i, C_i, R_i with i = 2, 3, 4,... for the i-th (spurious) mode (see Fig.3).
Resonances, which are mechanically coupled, must be represented electrically by a coupling circuit such as a transformer or an equivalent coupler e.g. three capacitors or inductors in delta- or star configuration, where one or two of them may have negative element values. Possible circuits are summarized in Fig.4.

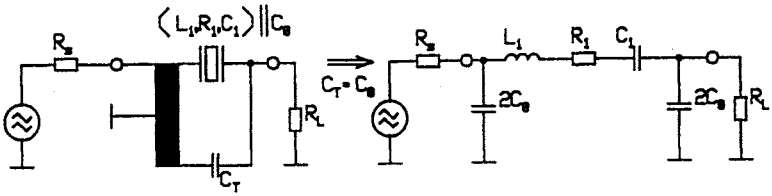
Fig. 4



4. Classical measurement methods

The classical measurement technique for spurious modes is the hybrid-coil method, described by Horton and Smythe [1] and Priebe [2], which was standardized in IEC 283 [3].
It uses a differential transformer with a variable capacitor C_c in the second branch to "compensate" the static capacitance C_0 (see Fig.5).

Fig. 5



Due to its simplicity in practical application (once, the hybrid transformer is made ...), this method is very popular. However it has significant disadvantages:

(1) Even if C_0 is compensated (C_C=C_0), it has not disappeared: It appears with twice its value in parallel to the source and load termination resistors. Therefore the frequency of minimum attenuation is not exactly equal to the series resonance frequency, and the resonance resistance computed by a formula neglecting C_0 - which is the usual practice - is wrong.

(2) Spurious resonances close to the main response can not be characterized properly. The "dB" or "differential dB" values cannot be directly converted to resistance values or to the motional parameters.

(3) Several problems are related to the hybrid transformer: It must have a flat frequency response, low winding capacitance, low stray inductance, perfect coupling and symmetry over a wide frequency range. Because this critical part is not commercially available, it suffers from poor reproducibility.

(4) The manual measurement is very time-consuming, as it is necessary to sweep in the vicinity of each resonance very slowly to achieve sufficient accuracy.

The disadvantages in item (3) can be overcome, if the hybrid transformer is replaced by a commercially available 180° hybrid coupler embedded in a 50 Ω environment, as proposed in [4]. This however does not eliminate the above mentioned problems (1),(2) and (4).

Stronger responses can also be measured by the conventional π-network method, however the measurement error increases, the weaker the spurs are and the closer to the main mode they are.

5. Alternative measurement methods

The measurement technique to be proposed is based on a test set-up with network analyser or equivalent, the π-network according to IEC 444 and the error correction technique described in [5] and standardized in IEC 444-5. The same method can also be used in conjunction with a s-parameter test system and adequate crystal test fixture.

Step 1: Calibration with short, open and reference resistor in the frequency range of interest

Step 2: Measurement of the static capacitance according to IEC 444-5

Step 3: Determination of the resonance frequency f_r (resp. series resonance frequency f_s, or load resonance frequency f_l) and the equivalent parameters R_1, C_1(and L_1, Q_1) of the main mode with one of the error-corrected methods described in IEC 444-5. If spurious modes are very close to the main mode (i.e. closer than $f_s/1000$), the values derived herewith may include non-negletable errors, but can be used as starting values for the iterative procedure described later.

Step 4: Fast sweep (approximately 1 sec) through the frequency range of interest upwards (increasing frequency) and downwards. A minimum of 500 data points is recommended. Substract the error terms, compute $Y_{up}(f)$ and $Y_{down}(f)$. Substract the admittance $j\omega C_0$ from both sets.

Step 5: Inspect the data and select computationally the peaks with a given minimum (trigger) value Y_{min}. Average the related peak frequencies of both sweeps. This yields an excellent guess for the resonance frequencies of the spurs. The sensitivity of the peak selection can be improved by mathematical differentiation of the Y(f) values. Alternatively the network analyser's internal "peak search" function can be used if implemented.

Step 6: Zoom each spurious response individually and measure Y(f) with higher frequency resolution.

The following steps depend on the type of spur to be considered (see classification in clause 2).

case A (strong spur well "isolated" from the main mode)

Zoomed data points can be selected as in IEC 444-5. Fastest approach is the iterative procedure as desribed in [5]. From a minimum of two or three test frequencies the i-th spurious resonance can be characterized by f_i, R_i, C_i (and L_i,Q_i).

case B (weak spur well "isolated" from the main mode)

Method as in case A. To improve sensitivity, the following modifications could be considered:

* averaging of several sweeps
* reduced RF bandwidth and video bandwidth (slower speed !)
* use of frequency selective, low noise network analyser. The technique using direct measurement of amplidue and phase is in this respect advantageous over the s-parameter systems.
* increase impedance R_T "seen" by the crystal in the test fixture. The s-parameter fixture with its R_T = 100Ω is advantageous over the IEC 444 π-network (R_T = 25Ω).

case C (strong spur in the vicinity of the main mode)

Data are zoomed with high frequency resolution in an interval covering the 3dB-bandwidth of the main mode and of the considered spur (its bandwidth is assumed to be approximately equal to that of the main mode - as is its Q in most cases).The parameters of both modes are then extracted by a parameter-fitting algorithm.

For uncoupled modes the equivalent circuit diagram of Fig.3 is used, i.e. three variables per resonance have to be determined (L_i, C_i, R_i or f_i, C_i, R_i). For coupled modes one circuit of Fig.4 has to be used. For N resonances 3 + 4•(N-1) elements have to be determined.

Rather good starting values for the elements can be used which speeds up the parameter search and reduces the risk of ill-conditioned iterations to wrong local minima.

These are:

f_1, R_1,C_1 (and therefore L_1 and Q_1) of the main mode from step 3

f_2 from the location of the peak maximum of Y(f) (after removal of $j\omega C_o$)

R_2 from the amplitude of the peak maximum, derived from

$$R_2 \approx \frac{1}{|Y(f_{peak})|}$$

C_2 from the assumption that the Q_2-value of the spur is approximately the same as Q_1 of the main mode

$$C_2 \approx \frac{1}{\omega_2 Q_2 R_2}$$

In the case of coupled modes, the starting value of element which determines the coupling can be arbitrarily set to zero.
The iteration procedure minimizes the error function

$$E = \Sigma w_i \, |Y_i - Y_i^M|^2 = \Sigma w_i \, ((G_i - G_i^M)^2 + (B_i - B_i^M)^2)$$

in which $Y_i = G_i + j\,B_i$ denotes the theoretical values from circuit analysis computed at the measurement frequency f_i, and $Y_i^M = G_i^M + j\,B_i^M$ represents the values derived from the measurement at f_i. The weighting factors w_i can be selected such, that the data points close to the resonances get the greatest significance, e.g. $w_i = |Y_i^M|$.
Several searching algorithms for minimization of E are applicable. One of them (Newton method) is described e.g in [9]. Another approach is the method of steepest descend, e.g. the Simplex algorithm. The evaluation of the different methods with respect to simplicity and stability is currently under progress.

case D (weak spur in the vicinity of the main mode)
The method is identical to case C, except that the starting value of R_2 hass to be derived differently, because the amplitude of the spurious peak is too errraneous, as it is located on the slope of the main resonance curve.
A more reliable starting value for R_2 is gained as follows:
The approximate admittance of the main mode

$$Y_m(f_i) = \frac{1}{R_1 + j\frac{1}{\omega C_1}\left(\frac{\omega^2}{\omega_1^2} - 1\right)}$$

is computed at the measurement frequencies f_i by using the starting values of f_1, C_1, R_1 . It is then substracted from the measured admittance values Y_i^M in the vicinity of the main mode and the weak spur. From the peak of this residual Y_{spur} the starting value of R2 can be computed as

$$R_2 = \frac{1}{|\max(Y_{spur})|}$$

It should be noted, that the achievable accuracy for the parameters of very weak spurs is reduced. However, the method proposed herein is superior to the conventional methods used so far.

6. Conclusions
The paper describes a concept for the measurement of spurious resonances, which allows the use of network analysers in conjunction with error correction techniques. The test fixture can be the π-network or the s-parameter test fixture as described in IEC 444-5. The proposal includes the determination and characterization of weak spurs,which may lie close to a strong mode. It considers also the mechanical coupling of modes. The next steps are the optimization of the parameter fitting algorithm. The experimental verification and comporative measurements have to follow. These results will be reported later.

7. Acknowledgements
The author thanks Friedrich Krug (University of Erlangen), Eberhard Seydel (KVG), and Dr. Frieder Sohre (Philips Doetinchem) for the valuable discussions during the preparation of this work.

8. References

[1] Horton,W.H. and Smythe,R.C.: The hybrid-coil bridge method of measuring unwanted modes of vibration in quartz crystals; Proc. 17th Annual Symposium on Frequency Control (ASFC) 1963, pp.316-324

[2] Priebe,F.K.: Attenuation and resistance measurements of unwanted modes of quartz crystals; Proc. 19th ASFC (1965), pp.487-508

[3] IEC publication 283 (1968): Methods for the measurement of frequency and equivalent resistance of unwanted resonances of filter crystal units

[4] IEC document 49(Germany)109 (1985): Draft - Proposal for extension of the π-network method for the measurement of main and unwanted resonances of crystal units up to 500MHz.

[5] Neubig, Bernd W.: Measurement of Quartz Crystal Units up to 500 MHz and Above by the Use of a Pi-Network with Error Correction; Proc. 11th Quartz Devices Conference, Kansas City (1989), Vol.1, p.60ff.

[6] IEC publication 444-5 (1993): Methods for the measurement of quartz crystal devices for the determination of equivalent electrical parameters using automatic network analyser techniques and error correction

[7] Neubig, Bernd W.: PI-Network Measurement of the Load Resonance Frequency without Load Capacitor;Proc. 4th European Frequency and Time Forum (EFTF) (1990), pp.481-486

[8] Vig, John R.: Quartz crystal resonators and oscillators - a tutorial; U.S. Army Electronics Technology and Devices Laboratory (1990)

[9] Peach,R.C., Morris,S.E.: A system for precision parameter measurements on quartz crystal resonators and bipoles; Proc. 39th ASFC (1985), pp.527-534

A THEORETICAL STUDY OF SURFACE ACOUSTIC WAVES IN LAYERED QUARTZ HOMEOTYPE STRUCTURES

W. WALLNÖFER and P.W. KREMPL
Physics Division, AVL List Ges.m.b.H., Kleiststraße 48, A-8020 Graz, Austria

Abstract

The surface acoustic wave properties of a quartz-type substrate (quartz, $AlPO_4$, $GaPO_4$), covered with a thin layer of a homeotypic material with the same orientation, were calculated. The SAW properties and dispersion characteristics vary strongly with the layer thickness and the combination of materials. Calculated velocities, coupling constants and temperature coefficients of acoustic waves are presented. For $GaPO_4$ on quartz and $AlPO_4$ on quartz, temperature compensated orientations were found for which a higher coupling is calculated than attainable with SAW devices with a single one of the studied materials. The predicted properties of these layered substrates, especially the second temperature coefficients, compare favourably with other temperature compensated materials.

1. Introduction

The quartz homeotypes are materials which have drawn considerable interest because of their promising piezoelectric and elastic properties. They have the trigonal structure of α quartz (SiO_2), and the silicon atoms are substituted by germanium or alternatingly by atoms with 3 and 5 valence electrons (B, Al, Ga, or Fe; P or As). From this group, $AlPO_4$ (berlinite) and $GaPO_4$ have been subject of considerable crystal growth effort, and their elastic and piezoelectric properties have been measured./1,2,3,4/.

The lattice mismatches among quartz and its two most studied homeotypes are in a range where good heteroepitaxy seems achievable. Thus, the lack of large seeds motivated research on liquid phase epitaxy. Philippot et al. /5/ grew $GaPO_4$ on $AlPO_4$ Z seeds. Krempl et al. /6/ found a method to grow $GaPO_4$ as well as $AlPO_4$ epitaxially on quartz seeds.

Besides its importance for the growth of bulk crystals, this development might also have a more direct effect on the field of surface acoustic wave devices. Since SAWs are bound to the uppermost parts of the substrate, a thin epitaxial layer suffices to change their properties. We undertook a theoretical study of the SAW properties of quartz-homeotypic substrates covered by a thin epitaxial layer of a different quartz-homeotypic material, to predict the properties of devices as shown in Fig. 1.

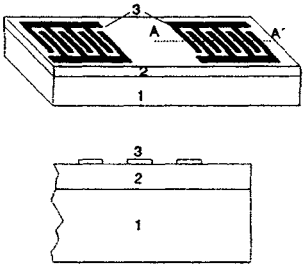


Fig. 1. Schematic drawing of a layered SAW device: 1 = substrate, 2 = epitaxial layer, 3 = interdigital transducer.

2. Calculation

For the calculation, we assumed an infinite homogeneous half-space of one material, bounded by a plane interface to a plane-parallel layer of another material. Both orientations were assumed to be equal, which corresponds to the observed way of epitaxy. At the interface, the materials are attached firmly to each other, i.e. the displacement field u_i and the components of the stress tensor acting on the interface plane T_{3i} are assumed to be continuous at the interface. The same is valid for the electrical potential φ and, since the interface is not metallic, for the component D_3. The thermal expansion parallel to the interface also must be continuous; it was assumed to be equal to that of the halfspace.

With these assumptions, a computer program was developed to calculate the properties of the acoustic waves with propagation parallel to the surface /7/. The inhomogeneous solutions of the elastoelectric equations in the quasistatic approximation were calculated separately for half-space and layer for a given wavelength. The coefficients of these solutions were determined by the boudary conditions at the interface and the surface, except the effective permittivity at the surface which was used to find the wavelength of the layer wave. Calculation time could be shortened significantly by employing the boundary condition transfer matrix algorithm by Adler /8/.

3. Results

In layered structures, the SAW properties depend on the ratio of the wavelength to the layer thickness and are thus subject to dispersion. Thus, phase velocity $v_p = \omega/k$ and group velocity $v_g = d\omega/dk$ are different. The behaviour of the phase velocities is shown schematically in Fig. 2 for the case that the acoustic waves are clearly slower or faster in one of the materials.

If the acoustic velocities in the layer are clearly higher, the layer wave velocity increases with increasing layer thickness. When it reaches the lowest bulk velocity of the substrate, the wave radiates energy into the substrate, which leads to high propagation losses. The usable range is small and in a region with very high dispersion; thus we did not study this kind of structure further.

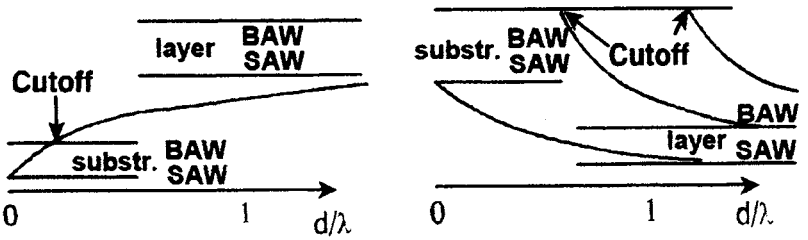


Fig. 2. Schematic dependence of the phase velocities of layer acoustic waves on the thickness/wavelength ratio when the acoustic velocities of the layer material are clearly higher (left) or slower (right) than those of the substrate material.

The acoustic velocities increase from $GaPO_4$ to $AlPO_4$ to quartz. Thus, the remaining combinations are $AlPO_4$ on quartz, $GaPO_4$ on quartz and $GaPO_4$ on

$AlPO_4$. Fortunately, these are also the combinations which make more technical and economical sense.

In these cases, the phase velocity decreases with increasing layer thickness and approximates asymptotically the SAW velocity of the layer material. In addition, other modes become bounded by the layer on crossing the substrate bulk velocity and decrease asymptotically to the slowest layer bulk velocity. In the range up to the appearance of the first additional layer-bounded mode only one mode propagates without radiative losses; this single-mode range is preferable for most applications. The dispersion decreases steadily with the thickness-to-wavelength ratio. In the single mode range it goes down to $\frac{\lambda}{v}\frac{dv}{d\lambda} \approx 2..8\%$. If other modes are tolerable, it can be made arbitrarily small.

The calculations were restricted to the two groups of orientations for which materials with point symmetry 32 show no beam steering. In the x-boule orientations with Euler angles λ=0° and θ=0°, where the propagation is along the twofold rotation axis x, the displacements are in the sagittal plane (Rayleigh-like waves), while in the x-cylinder orientations (λ=0° and θ=90°) the only displacement component of the so-called Love waves is perpendicular to the sagittal plane and parallel to the x axis. The angle μ is variable in both cases. Thus the layer wave properties for each material combination and orientation type were calculated as a function of μ and the thickness-to-wavelength ratio d/λ.

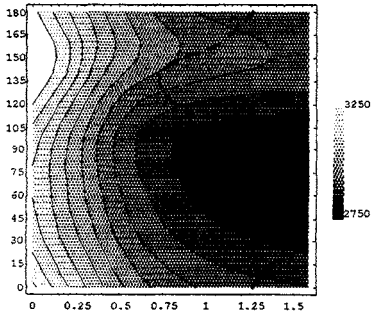


Fig. 3. Phase velocity of acoustic layer waves in m/s, in $AlPO_4$-on-quartz structures with x-boule orientation as a function of d/λ and μ, with the boundary of the single-mode range (dashed).

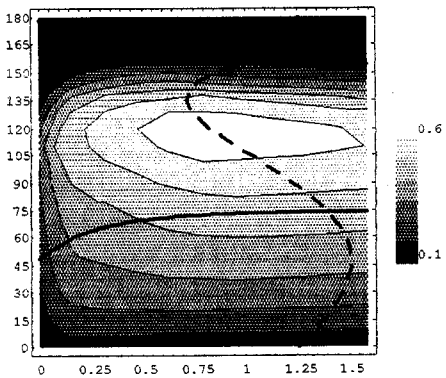


Fig. 4. Coupling factor k^2 of acoustic layer waves in %, in $GaPO_4$-on-quartz structures with x-boule orientation, with the boundary of the single-mode range (dashed) and the locus of temperature compensation of delay (thick line).

In Figs. 3 and 4 the phase velocity and the coupling constant k^2 of x-boule orientations of $AlPO_4$-on-quartz structures are shown. The single mode range is at the left of the dashed line. The thick line in the k^2 plot denotes the temperature compensated structures; it links the ST cuts of quartz and $AlPO_4$. With $d \approx \lambda$, properties similar to bulk $AlPO_4$ without additional modes are attainable. (A second branch of temperature compensated structures is in the dark low-coupling area at the top of the plot). For some angles μ, the coupling is higher for $d/\lambda = 0.5..1.5$ than for bulk $AlPO_4$. For the x-boule orientations of the other two material combinations, the behaviour is similar, but the properties are somewhat less favourable.

However, these findings are overshadowed by those in the x-cylinder orientations. The coupling factors of the layer Love waves for $GaPO_4$ on quartz are shown for all values of μ in Fig. 5 and for $\mu = 30°$ in Fig. 6, where also the very low coupling factors of the corresponding Bleustein-Gulyaev waves in the bulk materials are indicated. The function $k^2(d/\lambda)$ shows a very strong and pronounced maximum near $d/\lambda = 0.3$ for all μ values in all 3 combinations, which amounts to 1.0% for $GaPO_4$ on quartz, 1.3% for $GaPO_4$ on $AlPO_4$ and 1.4% for $AlPO_4$ on quartz; this is more than twice the maximum value for single crystal SAWs in one of the concerned materials. For large d/λ, k^2 decreases as $1/(d/\lambda)$.

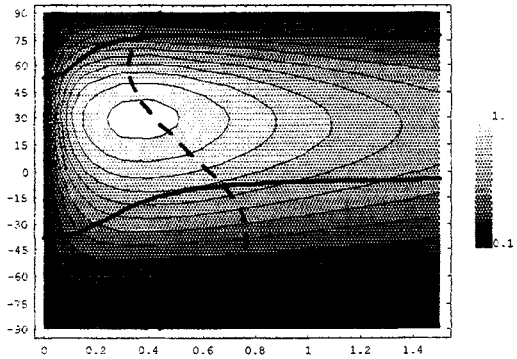


Fig. 5. Coupling factor k^2 of Love waves in %, in $GaPO_4$-on-quartz structures as a function of d/λ and μ, with boundary of the single-mode range (dashed) and locus of temperature compensation of delay (thick line).

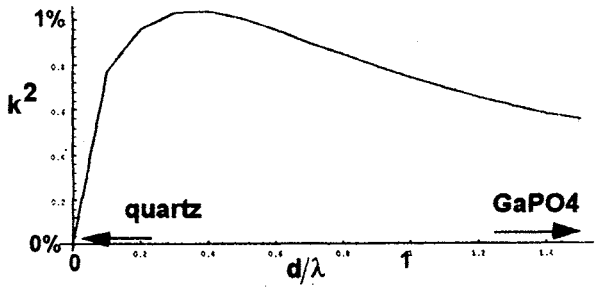


Fig. 6. Coupling factor k^2 of Love waves in %, in $GaPO_4$-on-quartz structures with x-cylinder orientation and μ = 30°, with the coupling factors of the single crystals of the same orientations.

A possible explanation of this behaviour is that the energy of the wave is "trapped" in the layer which acts as a wave guide; the amplitude decays strongly in the substrate material, while in the layer its depth dependence is similar to a quarter wave with a node little below the interface. Since this form remains constant at larger d/λ, the ratio of the total energy to the squares of the mechanical and electrical amplitudes at the surface is proportional to d/λ. However, a certain layer thickness is necessary for an effective trapping. This effect is clearly stronger for transverse waves, which in single crystals are only weakly bound electrically to the surface and extend far into the substrate, than for Rayleigh waves.

4. Conclusions

Table 1. Comparison of temperature compensated SAW orientations; measured and calculated values (most advantageous items in bold letters)

Material	Quartz /9/	$AlPO_4$ /1/	$GaPO_4$ /3,10/	$Li_2B_4O_7$ /11/	$GaPO_4$ on quartz	$AlPO_4$ on quartz	$GaPO_4$ on $AlPO_4$
Temp. comp. orientation	ST	ST	ST	X,Z	X-cylinder	X-cylinder	X-cylinder
SAW velocity, m/s	3158	2741	2335	3542	2600-2800	3000-4000	2600-2800
Coupling k^2, %	0.11	0.48	0.29	**1.0**	**0.7**	**1.0**	**0.7**
2.TC of freq., $10^{-9}/K^2$	-32	-220	**8**	-230	**10..20**	-30..-15	20..30
Dispersion	**no**	**no**	**no**	**no**	yes	yes	yes
Max. temp., °C	<573	<588	**930**		<573	<573	<588

The properties of the best temperature compensated orientations for these Love wave structures are compared with other temperature compensated SAW cuts in Table 1. The velocities can be varied over a certain range due to the dispersion. The coupling factors of the layered structures are distinctly higher than in the single crystal quartz homeotypes and similar to those in lithium tetraborate; compared to the latter they have the advantage of their chemical resistivity and insolubility in water. The calculated second temperature coefficients of frequency are remarkably good.

To conclude, the described layered structures, according to our calculations, seem to have some clear advantages. The combination of high coupling and very good temperature stability is unique among the presently known materials. Thus, these structures could have the potential to fill the gap between high coupling materials such as $LiNbO_3$ and $LiTaO_3$ and the quartz ST cut. The dispersion might be used for dispersive delay lines for chirp filters. For the structures with quartz as the lower material, large wafers at low cost are readily

available as substrates. The epitaxial layer growth process, once developed, would be of short duration and allow an efficient production.

However, there are also some problems. The dispersion, which is an inherent property of such layered structures, makes them unsuitable for some SAW applications. Also, an industrial liquid phase epitaxy process yielding an adequate quality of the layer and the interface would have to be developed. An exact thickness control would be required in the growth process and/or the following processing of the coated wafers.

The weighing of these advantages and problems will determine whether this theoretically interesting effect will also meet the practical interest of the SAW community.

References

/1/ Z.P. Chang & G.R. Barsch, IEEE Trans. on Sonics and Ultrasonics SU-23 (1976) 127

/2/ D.S. Bailey, J.C. Andle, D.L. Lee, W. Soluch, J.F. Vetelino, & B.H.T. Chai, Proc. 1983 IEEE Ultrasonics Symposium, 335-340

/3/ W. Wallnöfer, J. Stadler, & P.W. Krempl, Proc. 7th European Frequency and Time Forum, Neuchatel 1993, p. 653

/4/ W. Wallnöfer & P.W. Krempl, to be published in Phys. Rev. B (1994)

/5/ E. Philippot, A. Ibanez, A. Goiffon, B. Capelle, A. Zarka, J. Schwartzel, & J. Détaint, Proc. 6th European Frequency and Time Forum, Noordwijk 1992, p.383

/6/ P.W. Krempl, G. Voborsky, U. Posch, & W. Wallnöfer, Austrian patent pending

/7/ G.W. Farnell & E.L. Adler, "Elastic Wave Propagation in Thin Layers", in W.P. Mason (Hrsg.), "Physical Acoustics", Vol. 9, Academic Press, New York 1972

/8/ E.L. Adler, IEEE Trans. Ultrason. Ferroelec. Freq. Contr. 37 (6) (1990) 485

/9/ B.J. James, Proc. 42nd Annual Frequency Control Symposium (1988), 146

/10/ P.W. Krempl, J. Stadler, W. Wallnöfer, W. Ellmeyer & R. Selic, Proc. Fifth European Frequency and Time Forum, Besancon 1991, p.143-147

/11/ B. Lewis, N.M. Shorrocks & R.W. Whatmore, Proc. 1982 IEEE Ultrasonics Symposium, 389-393

CRYOGENIC SAPPHIRE DIELECTRIC RESONATOR MICROWAVE OSCILLATOR WITH IMPROVED STABILITY

D.G. Blair, M.E. Costa, A.N. Luiten, A.G. Mann, M.E. Tobar
Dept. of Physics, University of Western Australia

ABSTRACT

Microwave oscillators of exceptional short term stability have been realized from cryogenic sapphire resonators with loaded Q factors in excess of 10^9 at X-band and 6 K by employing a synthesis of loop oscillator operation with active Pound frequency stabilization. These oscillators have exhibited a short term fractional frequency stability of 3 to 4 x 10^{-15} for integration times from 0.3 to 100 seconds and 1 x 10^{-14} at 1000 seconds. Long term fractional frequency drifts are within a few parts in 10^{13} per day. Substantially improved performance can be expected with simple improvements in instrumentation. Sapphire oscillators are simple, reliable and robust and should be ideal flywheel oscillators for the new generation of atomic standards. This paper describes resonators with Q-factors up to 8 x 10^9 and oscillators based on them. A new method of characterisation using the time evolution of relative phase is described, and oscillator performance is presented.

1. INTRODUCTION

This paper describes recent results in the development and evolution of cryogenic sapphire dielectric resonator oscillators. Such oscillators are needed in a range of areas including space telemetry, radar and radar mapping, VLBI radio astronomy, spectroscopy and precision experiments in general relativity and gravitation, including gravitational radiation detection by pulsar timing and planetary-baseline doppler tracking of spacecraft. In the last mentioned case frequency standards with stability of order 10^{-15} to 10^{-16} for long integration times (τ >1000 s) will be necessary to set new limits on low frequency gravitational radiation. The hydrogen maser is presently the most stable frequency standard for medium integration times, $100 < \tau < 30{,}000$ s. The new generation of microwave atomic frequency standards now being developed shows promise of exceeding the stability of the best hydrogen masers. Of these perhaps the best known is the Jet Propulsion Laboratory's linear ion trap standard /1/, based on the 40.5 GHz

transition of Hg^+ which has achieved a fractional frequency stability of 7×10^{-14} $\tau^{-0.5}$. However phase noise in the local oscillator used to interrogate these microwave resonances also impinges on the long term stability /1, 2, 3/. Thus, as these new generation sources approach 10^{-14} $\tau^{-0.5}$ stability, they will require "flywheel" oscillators at least an order of magnitude more stable than present state-of-the-art quartz oscillators, which have a stability σ_y ($1<\tau<100$ sec) $\approx 1 \times 10^{-13}$ /4/.

The present status of microwave oscillators is summarized in Figure 1 below. Short term stability of order 10^{-14} can at present be provided only by cryogenic oscillators with resonator quality factors in excess of 10^8. However, if optical frequency standards could be referenced down to radio frequencies, such stability would be available at room temperature using frequency-stabilized Nd:YAG lasers /5/.

The cryogenic hydrogen maser (CHM) operating below 1K should in principle be 2 to 3 orders of magnitude more stable than a room temperature one, owing to lower thermal noise, a reduced H-H spin-exchange relaxation rate and the resulting higher signal power. Spin-exchange effects may, however, be a severe limitation /6/. Stability measurements have so far been thwarted by the lack of another oscillator of comparable stability /7, 8/.

Using a superconducting niobium cavity with a loaded Q-factor of over 10^{10}, Stein and Turneaure achieved a short term stability of $2\text{-}6 \times 10^{-16}$ in the late 1970s /9, 10/, in a superconducting cavity stabilised oscillator (SCSO). Despite this performance it is recognised that SCSOs suffer from stability limitations set by mechanical deformation of the cavity due to vibration and tilt which, combined with the relative difficulty in obtaining ultra-high Q cavities, have restricted their practical development.

Cryogenic oscillators built around mechanically more robust mono-crystalline sapphire dielectric resonators (SDR's) have demonstrated fractional frequency stabilities of 10^{-14} or better over 1 to 1000 s. Of these the best two are JPL's Superconducting Cavity Maser Oscillator (SCMO) /1, 11/ and the University of Western Australia (UWA) Sapphire Oscillator /12, 13, 14/. Both of these

oscillators employ a superconductor-shielded cylindrical sapphire resonator. In contrast to JPL's lead-plated resonator, the UWA sapphire resonator operates in a high order "whispering gallery" mode and is surrounded by a niobium cavity at some distance from the dielectric, resulting in much smaller fields at the superconducting walls (see Figure 2).

In the sapphire oscillator at the University of Western Australia the sapphire resonator forms the frequency determining element in a loop (or feedback) oscillator powered by GaAsFET amplifiers. A Pound-type IF phase stabilization technique locks the loop oscillator frequency to the centre of resonance. In contrast, JPL's SCMO resonator is powered at 2.7 GHz by a ruby maser pumped at 13 GHz. An important property of both these oscillators is the existence of a maximum of mode frequency as a function of temperature which allows fractional frequency stabilities of 10^{-15} or better with quite modest temperature regulation. The special features of the UWA sapphire oscillator are its high power (~1mW), convenient temperature of operation (6K) and simplicity of configuration. These are distinct advantages compared with the SCMO and CHM which operate well below 4.2K, require magnetic shielding and have very low output powers (1nW and 1pW respectively) which considerably restrict their phase noise performance.

Essential to achieving the highest stability in sapphire oscillators are [a] highest possible resonator Q-factors, [b] a robust simple oscillator configuration and [c] an effective means of characterisation. We report new results in all these areas below. Section 2 reports the highest observed Q-factors in dielectric resonators, Section 3 describes an improved oscillator configuration, and Section 4 describes a new method of combined phase noise and Allan variance determination, which gave rise to the sapphire oscillator data in Figure 1. Oscillator performance results are presented in Section 5.

2. HIGH Q RESONATORS

In a cylindrical dielectric resonator the modes may be designated H-type (quasi-TE) or E-type (quasi-TM). Each mode may be described by three mode numbers

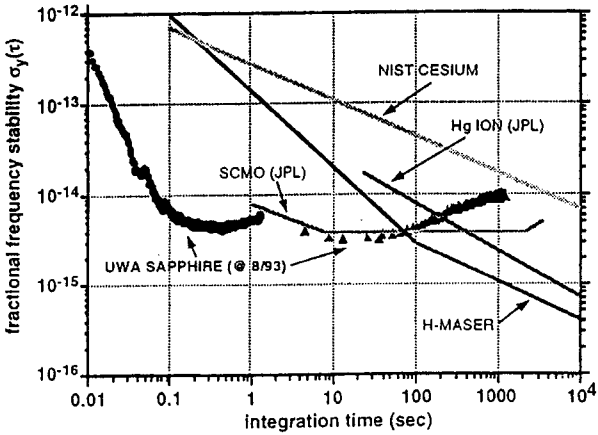


Figure 1

Measured fractional frequency stability of the UWA Sapphire Oscillator compared with the present performance of various other ultra-high stability oscillators.

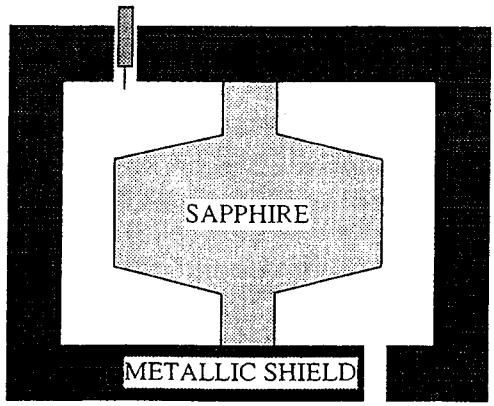


Figure 2

Cross-sectional view of the sapphire dielectric resonator shown with a single electric field probe.

m, n and p characterizing the field variations in the azimuthal, radial and axial directions respectively. We have focused on the "whispering-gallery" modes which exhibit the highest field confinement to the dielectric: namely the E_{m11} ($TM_{m1\delta}$) and H_{m11} ($TE_{m1\delta}$) modes having m azimuthal, one radial and one axial field variations. The resonator is in the form of an optically polished spindle, 50mm in diameter and 30mm long, with tapered end surfaces and integral 10mm support posts (see Figure 2). It was fabricated from high purity HEMEX grade sapphire provided by Crystal Systems Inc. (1991) with the crystal c-axis oriented along the geometric axis (to within 1 degree). To achieve the high quality factors reported here the sapphire is ultrasonically cleaned in a bath of nitric and hydrofluoric acid. The sapphire is mounted inside a cylindrical metallic shield (inside diameter 8 cm and height 5 cm). Coupling to the microwave field is via small probes which protrude through holes in the end wall of the shield. An electric field probe, suitable for E modes, is shown in Figure 2. Once assembled the resonator is placed in a small vacuum pot which is then evacuated and permanently sealed. This pot is suspended and thermally isolated in an evacuated chamber submersed in a liquid helium bath. The temperature is monitored with a carbon-glass resistor.

Q measurements were made by displaying the reflected power as a function of time as a signal generator was swept across resonance. The generator comprised a high resolution RF synthesizer whose frequency-swept output was mixed with a microwave synthesizer. Both synthesizers were locked to a reference crystal oscillator with a fractional frequency drift of 10^{-10} per hour. Using a very slow sweep (~1-10 Hz s^{-1}) we obtained the conventional bandwidth determination of Q which is accurate at lower Qs but near 10^{10} is corrupted by the frequency instability of the reference. For improved accuracy we have used a sapphire oscillator as a reference, allowing the Lorentzian response of the cavity to be measured to mHz resolution.

The unloaded electrical quality factor of a dielectric resonator is given by

$$Q_0^{-1} = F_c (\tan\delta + \chi'') + R_S \Gamma^{-1} \qquad (1)$$

where F_c is the confinement factor, representing the ratio of the mode's energy stored in the dielectric to the total stored energy, tan δ is the loss tangent of the

dielectric and χ'', the imaginary part of the ac susceptibility, represents losses due to paramagnetic impurities in the dielectric. The third term represents losses at the walls: R_S is the surface resistance and Γ is a geometry factor describing the magnetic field at the walls. For the modes considered here, $E_{18\,1\,1}$ (12.7 GHz) in a copper shield and $H_{14\,1\,1}$ (11.9 GHz) in a superconducting niobium shield below 6 K, the wall losses are negligible.

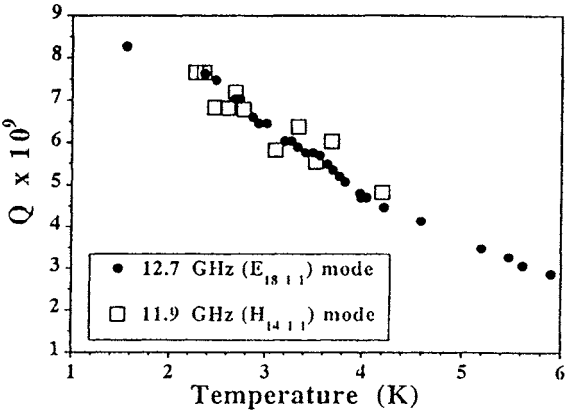


Figure 3

Unloaded Q as a function of temperature for $E_{18\,1\,1}$ and $H_{14\,1\,1}$ modes.

The Q as a function of temperature at 12.7 GHz ($E_{18\,1\,1}$) and 11.9 GHz ($H_{14\,1\,1}$) is shown in Figure 3 and is essentially mode independent. The Q factor shows a steady improvement on cooling: from 2.8 x 10^9 at 6 K, to 4.5 x 10^9 at 4.2 K, to 8.3 x 10^9 at 1.55 K. Previous measurements in similar size resonators fabricated from Union Carbide sapphire (manufactured in 1989) with chromium and iron impurities at the 1 ppm level have shown maximum Qs of only 1.5 x 10^9 at 4.2 K /17, 19/ which is similar to the results of other workers: 1.3 x 10^9 at 4.2 K /15/ 1.7 x 10^9 at 1.5 K /16/ and 1 x 10^9 (corrected) at 4.2 K /18/.

Braginsky /15/ has shown that below 50 K the dielectric loss is strongly dependent on the level of crystal imperfection, as determined by the rate of crystal growth. The Crystal Systems HEMEX sample reported here exhibits chromium and iron impurity levels of the order of parts per billion /17/ as determined by measurement of the mode frequency shift (χ') as the electron spin resonances (ESRs) at 11.45 GHz (Cr^{3+}) and 12.05 GHz (Fe^{3+}) are saturated /19/. This demonstrates that paramagnetic losses from the Cr^{3+} and Fe^{3+} ESRs, which follow a T^{-1} Curie law, are clearly negligible. We therefore speculate that the record Q may be attributable to the lower defect density expected from the Heat-Exchange Method of crystal growth as compared to the conventional Czochralski process employed by Union Carbide. It is thought that further material improvement is possible /20/, which is likely to yield even lower values of tan δ at liquid helium temperatures. We have also measured the temperature dependence of mode frequency in the range 1.6-20 K and find significant departures from the Curie law, especially below 4.2 K. Further details of this aspect of our work will be published elsewhere.

3. HIGH STABILITY OSCILLATORS

We now describe the simple robust oscillator circuit which we have used to create an ultra high stability oscillator using the sapphire resonators described above. The oscillator utilises a microwave amplifier in a loop oscillator configuration and three servos: a Pound type frequency servo, a resonator power servo, and a temperature control servo.

The improved loop oscillator circuit, shown in Figure 4, is essentially the configuration previously described /13, 14/ with the addition of a power level servo. In transmission the resonator is the primary frequency determining element of the free-running loop oscillator while in reflection it is the dispersive element in an active, Pound-type frequency stabilisation scheme /21/. A bandpass filter ($Q \sim 10^3$) is necessary to select the chosen SDR mode. The circuit losses are compensated by the gain of the GaAs FET amplifier (A) which runs in slight saturation. The loop phase is servo controlled to lock the oscillation frequency to

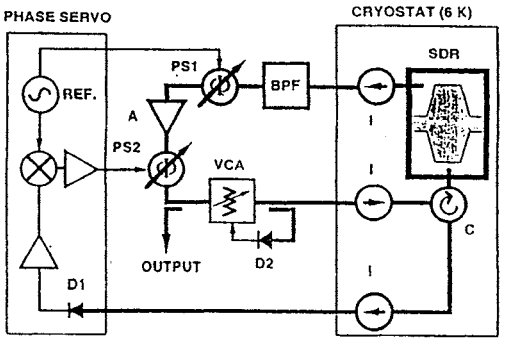


Figure 4

The loop oscillator circuit with power regulation.

the centre of resonance. The carrier power incident on the SDR is phase modulated at a rate (1.4 MHz) much larger than the resonator bandwidth. PM-to-AM conversion upon reflection provides the discriminator signal which is recovered by synchronous detection with the 1.4 MHz reference. The SDR input coupling is set near unity for maximum carrier suppression to permit square law operation of the detector diode (D1) and thereby maximize discriminator sensitivity. After amplification and filtering the discriminator signal is fed back to a second varactor diode phase shifter (PS2) to control the frequency of the loop oscillator. The transfer function of the phase stabilization circuit is essentially that of a single integration and has a unity gain crossover point at several kHz.

All of the loop oscillator components are at room temperature except for the circulator (C) and isolators (I) that are placed near the resonator (typically 3 per transmission line) to minimise frequency pulling effects. The microwave power dissipated in the resonator is typically 1mW and is partially regulated by the power servo comprising the tunnel diode detector (D2) and voltage controlled attenuator (VCA). This servo fully compensates the variation of insertion loss of PS2 with the error voltage and amplifier power fluctuations but not the attenuation increase in the transmission lines with falling liquid helium level.

4. OSCILLATOR CHARACTERISATION BY TIME EVOLUTION OF RELATIVE PHASE

A common technique for measuring oscillator stability is to study the beat waveform between two similar oscillators. A frequency counter is used to determine the fractional frequency stability by measuring the frequency fluctuations in the beat waveform, while the phase noise spectrum is determined by spectral analysis of that waveform.

Here we describe an improved method for evaluating oscillator stability on short time scales by measuring the *time evolution of relative phase* (referred to hereafter as TERP) between a pair of oscillators. The method involves digitising and recording the beat waveform between the two oscillators. The advantage that this technique offers over other methods, for example /22/, is that the precise form, rather than estimates, of both the fractional frequency stability and the phase noise spectral density can be determined from the same measurements. It therefore provides a more thorough way of describing and characterising oscillator performance.

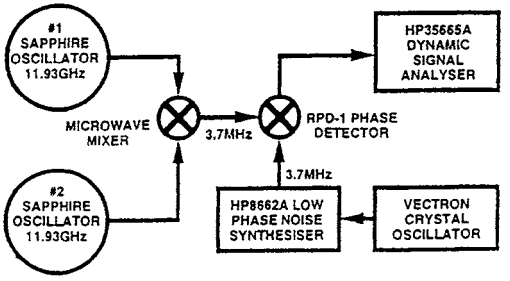


Figure 5

Hardware arrangement for recording the phase noise on a beat waveform from two ultra-stable cryogenic sapphire oscillators. The HP8662A is a low phase noise synthesiser which is used to offset the 3.7 MHz difference frequency without contaminating the waveformwith unwanted phase noise.

The TERP method relies on the usual phase detector technique in which a high-sensitivity mixer is used to compare two oscillator signals. If the signals have the same frequency and are in phase quadrature the waveform, hereafter denoted V(t), at the output of the mixer is proportional to the phase difference between the signals. For those cases in which it is not possible to tune the oscillators to the same frequency a low phase noise frequency synthesiser can be used to offset the difference frequency and trim the phase to produce the signal V(t). See Figure 5 for an example of the apparatus.

The unique feature of the TERP method is that the waveform V(t) is digitised, recorded and transferred to a computer for off-line processing. The measured values of V(t) are converted to the phase $\varphi(t)$ by scaling with a factor that can be easily determined from the apparatus. The fractional frequency stability and phase noise spectrum of the oscillators can then be directly computed. The phase noise spectral density is deduced by application of a conventional Fast Fourier Transform algorithm while the fractional frequency stability, in the form of the Allan standard deviation, is determined at integration time τ from the formula,

$$\sigma_y(\tau) = \frac{\sqrt{\langle[\varphi(t+2\tau) - 2\varphi(t+\tau) + \varphi(t)]^2\rangle}}{\sqrt{8}\pi\nu_o\tau}, \qquad (2)$$

where the symbols $\langle\ \rangle$ denote average value, $\varphi(t)$ is the phase in radians at time t, τ is the time interval between the measured values of $\varphi(t)$ and ν_o is the fundamental frequency of the oscillators. It follows that σ_y can also be determined at a multiple of the raw sampling interval by integrating the measured values of $\varphi(t)$ to yield values at the required time intervals.

The arrangement of the measurement system is similar to that usually used for phase noise spectral analysis. However, as long as the oscillators have high intrinsic stability there is no need for a phase-locked loop. A small residual frequency offset can be removed from the measurements in the off-line processing by subtracting a straight-line trend.

An important consideration is the maximum tolerable residual frequency. Only when the frequency of the beat waveform is precisely 0Hz does that waveform contain the true form of the oscillator phase noise. If there is a small residual offset frequency the phase noise sidebands in the oscillator signals are not correctly demodulated. This can be understood by considering in the frequency domain the process of mixing the oscillator signals. When the carriers are translated to 0Hz the lower sideband is aliased and superposed on the upper sideband. If there is a small residual frequency the lower and upper sidebands will be offset from each other resulting in distorted phase noise. The Fourier transformation of measurements that are collected over a time T produces a space spectrum with a frequency resolution of $\frac{1}{T}$. To avoid the effect of the distortion caused by a residual frequency offset the following condition must be satisfied,

$$\frac{1}{T} >> \Delta\nu \,, \qquad (3)$$

where $\Delta\nu$ is the residual frequency offset.

5. OSCILLATOR PERFORMANCE

Using the above methods we have measured the performance of two identical oscillators of the type described in Section 3.

The relative stability of two ultra-stable 11.9GHz cryogenic sapphire oscillators /13/ was measured as an example of the use of the TERP methodology. The 3.7MHz difference frequency between the two oscillators was offset in a high-figure-of-merit RPD-1 Mini-Circuits phase detector using a HP8662A synthesised oscillator phase-locked to a low phase noise CO246 Vectron crystal oscillator (Figure 5). The beat waveform was measured at 1ms intervals in a HP35665A signal analyser and 20480 values were recorded on magnetic disk for later analysis in a Sun SPARCstation 2.

After the data were scaled to radian units the Allan standard deviation was computed for a range of integration times using the equation (2) and the phase noise power spectrum was computed with the SPCTRM FFT algorithm /23/. The

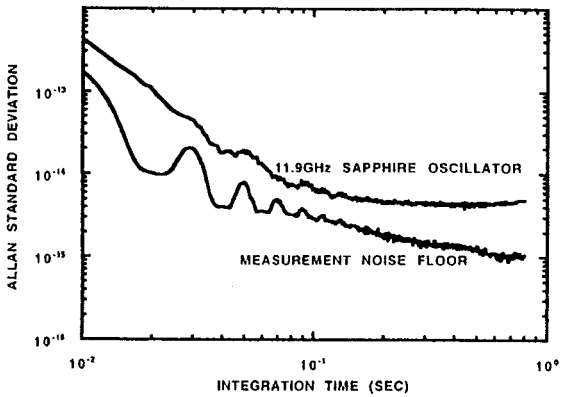


Figure 6

Allan standard deviation curves of a 11.9GHz cryogenic sapphire oscillator and of the equivalent stability of the measurement system.

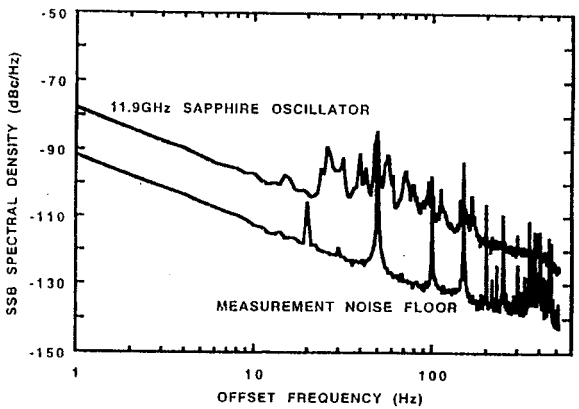


Figure 7

Single-sideband spectral density of the phase noise of a 11.9GHz cryogenic sapphire oscillator and of the measurement system.

Allan standard deviation values are presented in Figure 6 for integration times from 10 ms to 0.8s. By convention the phase noise spectrum was scaled to yield single-sideband spectral density and is shown in Figure 7. Assuming both oscillators contributed equally, the measures were also scaled by $\frac{1}{\sqrt{2}}$ and $\frac{1}{2}$, respectively, to find the nominal stability of a single oscillator.

The noise floor of these measurements is also shown in Figures 6 and 7. The noise floor curves were derived from TERP measurements between a HP5065A Rubidium frequency standard and the HP8662A phase-locked to the CO246 Vectron crystal oscillator.

In the power spectrum the rapid roll-off above 400Hz is evidence of the anti-aliasing in the HP35665A signal analyser. The numerous bright lines are attributed to a variety of sources including the microwave signal chain, the 50Hz mains power supply and its harmonics and possibly the servo circuits in the sapphire oscillators. The excess noise in the range 20Hz to 200Hz is attributed to the effect of seismic noise on the vibration-sensitive components of the sapphire oscillators and the microwave signal chain. In another experment /24/ a lower phase noise of -127 dBc/Hz at 100Hz and -140 dBc/Hz at 1 KHz has been measured at the expense of D.C. stability. The stabilisation circuitry adds phase noise above 10Hz and cancels it below. This lower noise was achieved in a free running oscillator that was vibrationally isolated against seismic noise.

Allowing for the bright line structure the underlying phase noise is divided into two spectral index regimes by a breakpoint near 35Hz. At the lower frequencies the index is about -2 corresponding to white frequency noise. In the higher range up to 400Hz, the spectral index is -1 corresponding to a flicker phase noise process.

The Allan standard deviation of the oscillators is near 4×10^{-15} for integration times from about 0.2s to 0.8s. This is similar to that determined from independent stability measurements at much longer integration times using a frequency counter. The composite oscillator stability, combining conventional long integration time data and this data was given in Figure 1. The small bump

around 2 s corresponds to the 0.2 Hz beat between the two 1.4 MHz phase servo reference oscillators. The origin of our apparent flicker frequency floor, ~3-4 x 10^{-15}, is uncertain. It is only a factor of 1.5 higher than the estimated noise floor of the power servo, but could also be due to intrinsic resonator flicker frequency noise. Ground vibrations of order 10^{-6} g might also be significant since the phase noise below 1 Hz is partially correlated with accelerometers mounted on the cryostat and the acceleration sensitivity in one resonator appears to be degraded from a few times 10^{-10} g^{-1} to a few times 10^{-9} g^{-1}. Beyond 100 seconds the power servo reduces some of the long term drift, but the performance begins to degrade as approximately 3 x 10^{-16} $\tau^{1/2}$. The long term stability measured against the rubidium reference is at the limit of the rubidium, about 10^{-12} per day. The relative drift of the two oscillators over 1 day is a few times 10^{-13}, which is an order of magnitude improvement over the oscillators based on 3 cm resonators /14/. The long term drift is dominated by a high sensitivity to room temperature and is about 10^{-13} K^{-1}, when the temperature sensitivity of D2 is eliminated by placing it at the resonator.

6. CONCLUSION

We have shown that sapphire oscillators are simple, robust, and give excellent short to medium term frequency stabilty and phase noise. They show very low sensitivity to magnetic fields, and ~10^{-10} per g sensitivity to vibration and acceleration. They have already been used in VLBI radio astronomy and to demonstrate extremely narrow linewidths in a trapped Yb ion source. Further improvements can be anticipated.

7. ACKNOWLEDGEMENETS

This work was supported by the Australian Research Council. We thank Mr. J. Searls for presenting this paper.

8. REFERENCES

1. Prestage J.D., Tjoelker R.L., Wang R.T., Dick G.J. and Maleki L.,"Hg^+ Trapped Ion Standard with the Superconducting Cavity Maser Oscillator", IEEE Trans. Instrum. Meas. IM-42 April 1993, pp.200-205.

2. Audion C., Candelier V. and Dimarcq N., "A Limit to the Frequency Stability of Passive Frequency Standards due to an Intermodulation Effect", IEEE Trans.Instrum. Meas., IM-40, 1991, pp. 121-125.

3. Szekely C., Walls F. W., Lowe John P., Drullinger R. E. and Novick, A., "Reducing the Effect of Local Oscillator Phase Noise on the Frequency Stability of Passive Frequency Standards", 1993 IEEE International Frequency Control Symposium, pp. 81-86.

4. Norton J.R. and Beeson R.J.,"Tactical BVA Quartz Resonator Performance", IEEE International Frequency Control Symposium, 1993, pp. 609-613.

5. Sampras N.M., Gustafson E.K. and Byer R.L. "Long Term Stability of Two Diode-Laser-Pumped Nonplanar Ring Lasers Independently Stabilized to Two Fabry-Perot Interferometers", Optics Letters, 18, 1993, pp. 947-949.

6. Hayden M.E., Hurlimann M.D.,and Hardy W.N.,"Atomic Hydrogen Spin-Exchange Collisions in a Cryogenic Maser", IEEE Trans. Instrum. Meas. IM-42 April 1993, pp. 314-319.

7. Hurlimann M.D., Hardy W.N., Hayden M.E., and Cline R W ,"Performance of the UBC Cryogenic Hydrogen Maser", in "Frequency Standards and Metrology" ed.: A De Marchi, Springer-Verlag Berlin, 1989, pp. 420-424.

8. Walsworth R.L., Mattison E.M.and Vessot R.F.C.,"Recent Investigations with the Harvard-Smithsonian Cryogenic Hydrogen Maser", IEEE International Frequency Control Symposium, 1993, p. 129.

9. Stein S.R.and Turneaure J.P.,"Supeconducting Cavity Stabilized Oscillators with Improved Stability" Proc. IEEE, 63, Aug 1975, p. 1249 .

10. Turneaure J.P. and Stein S.R.,"Development of the Superconducting Cavity Stabilized Oscillator" in "Near Zero: New Frontiers of Physics" ed.: J.D.Fairbank et al, New York, W.H.Freeman, 1988, pp. 414-430.

11. Dick, G. J. and Wang, R. T. : "Ultra-Stable Performance of the Superconducting Cavity Maser", IEEE Trans.Instrum. Meas., IM-40, 1991, pp 174-177.

12. Luiten A.N., Mann A.G. and Blair D.G.,"Cryogenic Sapphire Microwave Resonator-Oscillator with Exceptional Stability", Electronics Letters, 30, accepted (1994).

13. Luiten A.N., Mann A.G., Giles A.J. and Blair D.G., "Ultra-Stable Sapphire Resonator-Oscillator", IEEE Trans.Instrum. Meas., IM-42 , 1993, pp 439-443.

14. Giles A J , Jones S K , Blair D G and Buckingham M J , "A High Stability Microwave Oscillator Based on a Sapphire Loaded Superconducting Cavity", Invited Paper, Proc. 43rd Annu. Freq. Control Symp., Denver 1989, 1990, pp. 89-93.

15. Braginsky, V. B, Ilchenko, V. S, and Bagdassarov, Kh. S.: "Experimental observation of fundamental microwave absorption in high quality dielectric crystals", *Phys Lett A* , **120**, 1987, pp. 300-305.

16. Dick, G. J., and Saunders, J.: "Measurement and analysis of a microwave oscillator stabilized by a sapphire dielectric ring resonator for ultra-low noise", *IEEE Trans.on Ultrason Ferroelec Freq Control*, **37**, 1990, pp. 339-345.

17. Mann, A. G., Luiten, A. N., Blair, D. G., and Buckingham, M. J.: "Ultra-stable cryogenic sapphire dielectric microwave resonators". Proc. IEEE 46th Annual. Frequency Control Symposium, Hershey PA, 1992, pp. 167-1718.

18. Jones, S. K., and Blair, D. G.: "High-quality factors for a sapphire-loaded superconducting cavity resonator", *Electron Lett.* , **23**, 1987, pp. 817-818.

19. Mann, A. G., Giles, A. J., Blair, D. G., and Buckingham, M. J.: "Ultra-stable cryogenic sapphire dielectric microwave resonators - mode frequency temperature compensation by residual paramagnetic impurities", *J. of Phys. D - Applied Physics*, **25**, 1992, pp. 1105-1109.

20. Felt, D. M., Crystal Systems Inc., 27 Congress Street, Salem, MA 01970, U.S..A., private communication.

21. Stein, S. R. and Turneaure, J. P. : "The Development of the Superconducting Cavity Stabilized Oscillator", Proc. 27th Annual Frequency Control Symposium, 1973, p 414.

22. Rutman, J., Sauvage, G.: 'Measurement of Frequency Stability in Time and Frequency Domains Via Filtering of Phase Noise', IEEE Trans. Instr. Meas., IM-23, 1974, p.515.

23. Press, W.H., Teukolsky, S.A., Vetterling, W.T. and Flannery, B.P., "Numerical Recipes" (Cambridge University Press, Cambridge, 1992).

24. Tobar, M.E. and Blair D.G., "Phase Noise Analysis of the Sapphire Loaded Superconducting Niobium Cavity Oscilator", IEEE Trans. MTT, Vol. 42, No. 2, 1994.

HIGH-Q T.E. STABILISED SAPPHIRE MICROWAVE RESONATORS AND LOW NOISE OSCILLATORS

#J.H. Searls, #S. Edwards, *E.N. Ivanov, *M.E. Tobar, *R. Woode

Poseidon Scientific Instruments Pty Ltd
14 Mouat St. Fremantle, WA, 6160, Australia Fax (61 9 335 4650)

* Dept. of Physics, University of Western Australia
Nedlands, WA, 6009, Australia Fax (61 9 380 1014)

1. SUMMARY

This paper reports the performance of X-Band, low noise, High-Q, Sapphire Loaded Cavity (SLC) resonators with unloaded Q factors near $2x10^5$. Measured oscillator phase noise performance of -104 dBc/Hz, 1 kHz from a 9.000 GHz carrier for a simple loop oscillator topology is consistent with modelled performance.

Resonator Q, residual noise and environmental performance is reported.

Predictions of performance for oscillators which use noise reduction techniques such as described by Galani and Bianchini [1], Walls [2] or McPherson [3] indicate that oscillators based upon such Sapphire Loaded Cavity architecture will be capable of providing the lowest phase noise microwave sources yet reported in non-cryogenic systems.

2. INTRODUCTION

Research at the University of Western Australia (UWA) into the properties of microwave resonators based upon Sapphire dielectric material began in 1980 in support of UWA's Gravitational Radiation Laboratory's Gravity Wave Antenna detection scheme, where a microwave signal source with unprecedented low phase noise was required. This work has concentrated on oscillators incorporating Sapphire Loaded Cavity resonators operating in quasi-Whispering Gallery modes with superconducting walls. Q factors as high as 8×10^9 and oscillators with fractional frequency stabilities of 3×10^{-15} at 100 seconds [4] have been reported.

Past work by two of the authors: Ivanov [5] and Tobar indicated that at elevated temperatures such resonators could provide Q factors at X-Band of 2×10^5, which is sufficiently high to exceed the performance of other resonator technologies as reviewed by Driscoll [6].

3. RESONATOR

The performance of several resonators is reported. Two were designed to operate 10.000,000 GHz with unloaded Q's of 200,000, and a third was designed to operate at 9.000 GHz with a loaded Q of 98,000 to match the Q factor of an Interferometric Sapphire Transducer [7] under development at UWA. Other data includes resonators operating at 6.8 and 10.2 GHz.

The resonators are configured as shown in Figure 1 and are formed by a cylindrical piece of high-purity single crystal sapphire. This cylinder is mounted in a metallic cavity. The resonator centre frequency and Q factor are determined almost completely by the dimensions of the sapphire [5].

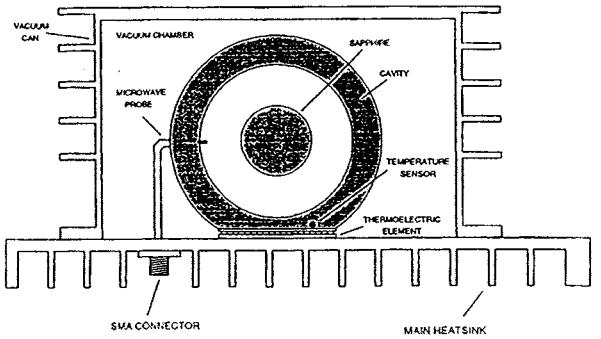


Figure 1. Schematic High-Q Microwave Resonator Configuration.

This configuration forms a complex set of boundary conditions resulting in a system which has a very large number of resonances of varying frequency and Q factor. The desired mode of operation is determined by numerical solution, and unwanted resonances are suppressed by modifying the cavity walls to cause such modes to be radiative [8].

The resonant cavity and sapphire are cooled and temperature controlled via a servo which includes a Lock-In Amplifier, PID controller and Thermo-Electric (TE) module. Temperature stability is better than 1 mK per hour (including ambient effects).

All resonators have been configured as 2 port devices. Resonator port coupling can be set over a range $0.05<\beta<1$ at the time of assembly but is fixed thereafter.

4. RESONATOR PERFORMANCE

4.1 Resonator Q factor

The mode structure employed is quasi-Whispering Gallery, of moderate azimuthal number, and resonator Q factor is determined primarily by the

dielectric losses of the sapphire. The cavities in the 10.000 GHz resonators are cooled to near 0°C where their Q factors are in excess of 200,000.

The resonator Q factor is strongly temperature dependent and the Qf product is constant to within the accuracy of the measurement system shown in Fig. 2.

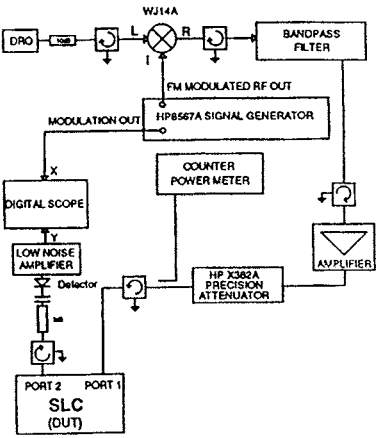


Figure 2.

Schematic of the measurement system used to measure resonant frequency, loaded Q-factor and port coupling at various temperatures.

Measurement repeatability of this system is 2% amplitude, 100 Hz centre frequency and 1% bandwidth.

Table 1 summarises resonator Q characteristics at various frequencies with constant cavity temperature.

f (GHz)	Q_o	T (°C)	$Q_o f$ x 10^{15}	Q_s	$Q_s f$ x 10^{15}
6.8	243,500	25.0	1.67	265,415	1.82
9.0	187,000	25.0	1.68	185,130	1.67
10.0	176,000	25.0	1.76	181,280	1.81
10.2	172,000	25.5	1.76	175,440	1.79
		Mean	1.72		1.77
		Dev	2.8%		4.0%

Table 1. Q factors of various Resonators and associated Q of Sapphire Dielectric

Where:

$$Q_o = \text{Unloaded Resonator } Q = Q_L \times (1+\beta_1+\beta_2) \quad (1)$$

and β_1 and β_2 are the port couplings.

Q_s is calculated from:

$$\tan\delta_{(Resonator)} = \tan\delta_{(Sapphire)} + \tan\delta_{(Cavity)} \quad (2)$$

as follows:

$$Q_S = Q_O / \{1 - Q_O \cdot \tan\delta_{(Cavity)}\} \quad (3)$$

and $Q_O.\tan\delta_{(Cavity)} \sim 5\%$ because the majority of the microwave energy is contained in the sapphire. Wall losses are estimated from consideration of the field strength at the cavity walls as described in [5].

These results indicate the loss tangent of sapphire is 5.52×10^{-6} at 10 GHz and 298 Kelvin. This result is lower than that reported by Braginsky [9] by 10% and is consistent for all resonators, even though the sapphire used in the resonators is from various sources and is grown by each manufacturer using a different method. The Qf product appears to be constant from 6.8 GHz to 10.2 GHz. Both these results are consistent with the theory describing the resonator as developed by Ivanov [5] provided the value of the loss tangent of high purity sapphire is taken to be the value stated here.

4.2 Resonant Frequency Temperature Characteristics

The resonators reported herein have a temperature coefficient of frequency of approximately -72ppm per °C, hence the use of the integral temperature controller. The main source of temperature dependence of frequency is thought to be due to the relatively strong temperature dependence of the dielectric constant of the sapphire. Changes in sapphire diameter due to thermal expansion also effect the temperature dependence of the resonant frequency. For the mode of operation used, not enough experimental data has been taken to allow the separation of these effects yet.

Figure 3 below shows the measured frequency vs temperature characteristics of a 10 GHz SLC resonator.

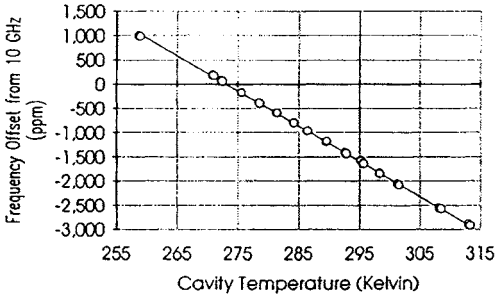


Figure 3. Measured frequency vs temperature data for a 10 GHz SLC resonator (in ppm near 273K).

The data indicates that the temperature dependence over this temperature range is linear. We assume a straight line fit to the data of the form:

$$f_o(T) = \left[f_o(T_o) \times (1 + TC \times (T - T_o)) \right]$$

$$\text{where} \quad f_o(T_o) = 10.0\ GHz\ , \quad TC = -7.25 \times 10^{-5}\ ^\circ C^{-1}\ , \quad T_o = 273\ Kelvin \tag{4}$$

At 10 GHz the frequency-temperature coefficient is -72 kHz/°C for a cavity that is not temperature controlled .

4.3 Frequency Drift Effects

The drift of the resonant frequency over time is determined solely by the ability of the temperature servo system to stabilise the frequency of the sapphire. There are three main factors which have an effect on the servo system's ability to do this: i) changes in ambient temperature, ii) changes in microwave power, and iii) aging effects in the electronics.

To reduce the effect of changes in ambient on the temperature of the sapphire, the cavity is thermally insulated in a vacuum using cryogenic techniques.

Figure 4 shows the medium term frequency stability against time.

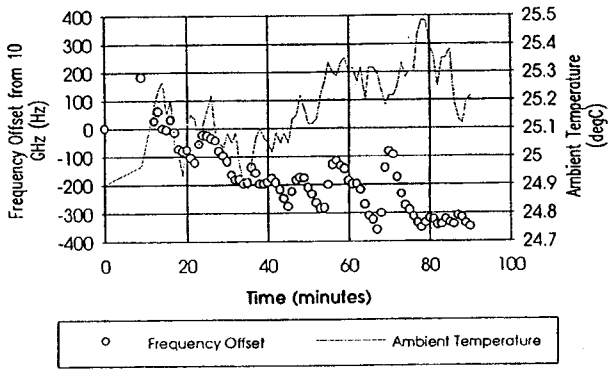


Figure 4. Measured medium term frequency stability vs time in minutes of a free running loop oscillator using a 10 GHz SLC resonator as a bandpass element. Ambient temperature changes are shown on the overlay.

Presently we consider that the long term frequency stability of the resonator is limited by the aging rate of the sense thermistor. Figure 5 shows the frequency stability of a 10 GHz resonator over a period of 4.3 days. This figure includes the settling period of the resonator after turn-on. The long-term drift is -3.5×10^{-7} per day, or -1.5×10^{-8} per hour.

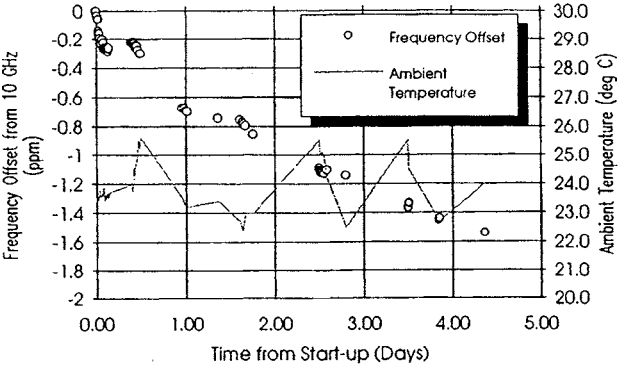


Figure 5. Measured long term frequency stability vs time in days of a free running loop oscillator using a 10 GHz SLC resonator as a bandpass element. Ambient temperature changes are shown on the overlay.

4.4 Q factor Temperature Characteristics

Braginsky et al [9] have reported on the significant decrease in loss tangent of sapphire as a function of temperature at cryogenic temperatures. The reported dependence was of the form $Q_o \propto T^{-4.75}$ for temperatures below 200 K.

Our data is measured near 273 K and over a temperature range of 48 °C. We assume a temperature dependence of Q of the form:

$$Q_o(T) = Q_o(T_o) \times \left(\frac{T}{T_o}\right)^{g(T)} \qquad (5)$$

Where T_0 = 273 K, $Q_o(T_o)$ = 215,000, and g(T) = -2.8.

We assume that the exponent g(T) is a function of temperature since the value we obtain is quite different from that of Braginsky's. This is quite plausible as the temperature coefficient of expansion of most materials varies in a similar manner.

Because the resonant frequency changes linearly with temperature, a better guide for predicting the Q at a particular temperature is the Qf product. Figure 6 shows measured data of the Qf product as a function of temperature. The line of best fit is of the form:

$$Qf(T) = Q_o(T) \times f_o(T)$$
$$= \left[Q_o(T_o) \times \left(\frac{T}{T_o} \right)^{\varepsilon(T)} \right] \times \left[f_o(T_o) \times (1 + TC \times (T - T_o)) \right] \qquad (6)$$

where the values of $f_O(T_O)$ and the temperature coefficient TC are given above.

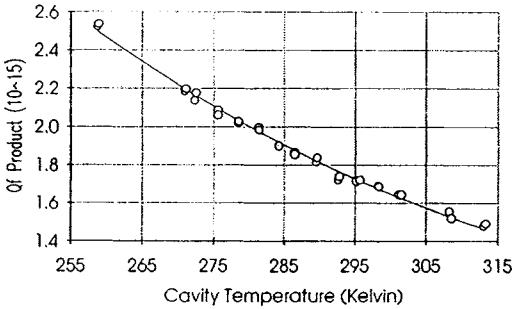


Figure 6. Measured Qf product vs Temperature for a 10 GHz SLC Resonator. The curve of best fit is described by equation 3 above.

4.5 Resonator Residual Noise

Residual phase noise measurements of the resonators were performed using a phase sensitive bridge circuit as shown in Fig 7. This type of phase noise measurement is limited by the noise of the driving oscillator and the mixer phase noise, provided the source noise is sufficiently low. Otherwise the dispersive characteristic of the resonator will decorrelate with the result that the phase detector will measure predominantly the oscillator noise.

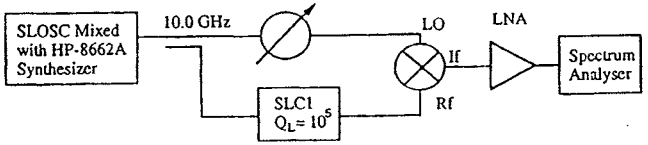


Figure 7. Residual Noise Measurement System

The signal source used is a very low noise cryogenic based Sapphire Loaded Superconducting Cavity (SLOSC) oscillator developed for the UWA Gravitational Radiation Detection Antenna. It has very low phase noise performance of -140 dBc, 1 kHz from a 9.7 GHz carrier. This oscillator is described in [4]. The SLOSC signal is mixed with an HP 8662A low noise synthesiser to the required frequency for measuring resonator residual noise. This technique has been previously reported [8].

A typical resonator residual noise measurement is shown is Fig 8. In each case where we have attempted to measure the SLC resonator residual noise the measurement has been limited by the measurement system, predominantly by the noise floor of the mixer used as the phase sensitive detector (in each case a WJ14A). This result has important implications to oscillator design, where the SLC resonator will have lower noise than any of the active devices used in the oscillator design and therefore will, until superior active devices become available, not limit oscillator performance.

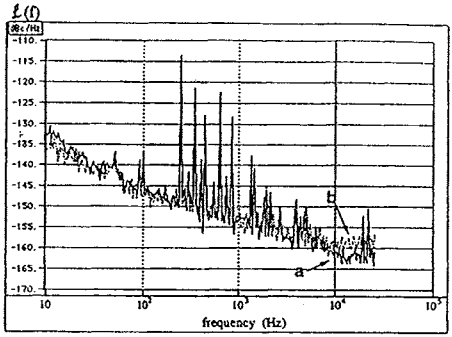


Figure 8. Residual Noise Measurement of a 10 GHz SLC resonator. The measurement noise floor is also shown.

4.6 Resonator g-Sensitivity

A measurement system similar to that shown in Figure 2 was used to detect static "g" effects. The 9.00 GHz resonator was connected to the system and rotated though 180° in each axis providing a net change in "g" force of 2g in each axis.

No variation in the resonator centre frequency was detected to the level of 100 Hz, being the resolution of the measurement system. This indicates an upper limit to the resonator g-sensitivity of the order of 10^{-8} per g. Tilt measurements on cryogenic resonators at UWA [4], indicate resonator g-sensitivity below 10^{-9} per g for similarly constructed resonators.

Phase shifts and changes to resonator centre frequency which are associated with mechanical resonances in the resonator structure have not been able to be

measured. A more sophisticated and accurate measurement of the resonator g-sensitivity is currently being investigated by one of the authors, Woode.

5. OSCILLATORS BASED UPON SLC RESONATORS

As described in the introduction, the main motivation for the development of this technology is to provide resonators with improved performance for microwave signal processing systems, in particular microwave oscillators and signal generation without the requirement for cryogenic cooling systems.

The theory of oscillators is well developed and the authors use the model presented by Leeson [10] for the analysis of oscillator noise contributions and optimisation.

Lesson's model may be generalised to account for all the major noise sources in the oscillator to become:

$$S_{\phi Osc}(f_m) = S_{\phi Comp}\left(1+\frac{1}{f_m^{\,2}}\left(\frac{f_o}{2Q_L}\right)^2\right) \tag{7}$$

with

$$\mathcal{L}(f_m) = \frac{1}{2}S_\phi(f_m) \tag{8}$$

where:

fo	=	resonator centre frequency
f_m	=	frequency offset from carrier
$S_{\phi Comp}(fm)$	=	the Sum of the spectral density of phase fluctuations in the oscillator active components
$S_{\phi Osc}(f_m)$	=	the spectral density of phase fluctuations in the Oscillator
$\mathcal{L}_\phi(fm)$	=	the ratio of noise in a 1 Hz bandwidth at an offset fm (provided the total phase excursion is small)

In this simple model the noise contributions of all the oscillator components such as amplifiers, varactor phase shifters or residual noise of the resonator are embodied by $S_{\phi Comp}$. In the case of X-Band oscillators the major source of 1/f noise is usually the GaAs amplifier sustaining stage, and simple loop oscillator performance may be predicted by (7).

Figure 9 shows the schematic diagram for a simple 9.000 GHz loop oscillator and the phase noise measurement system. Figure 10 shows the measured phase noise performance of the amplifier and Fig. 11 shows the measured loop phase noise and that predicted by this model with only the amplifier noise contribution considered in the equation. The noise performance of the other loop components has been previously measured and is well below that of the amplifier in close to the carrier.

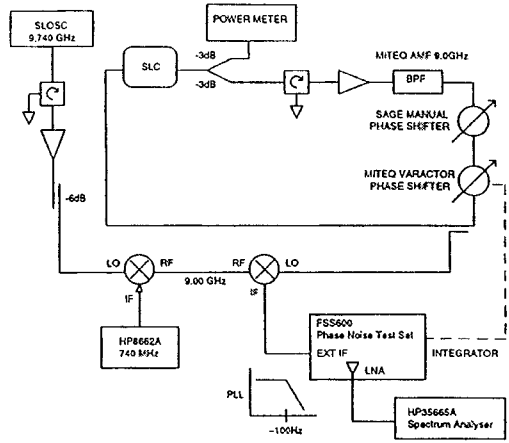


Figure 9. SLC-Based Loop Oscillator
Phase Noise Measurement system.

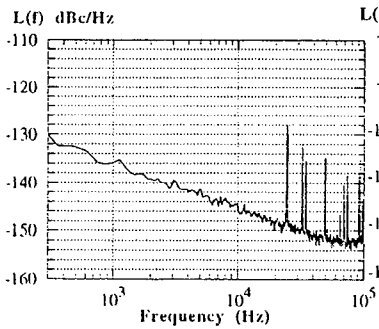


Figure 10. Measured Phase Noise of an Avantek AFT-12663-10F amplifier. No power dependence of phase noise was observed.

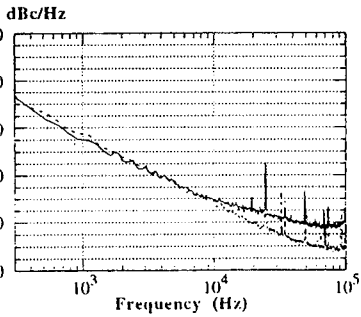


Figure 11. Measured Phase Noise of the 9 GHz Loop Oscillator and that predicted by using the amplifier noise in equation (7)

5.1 Oscillator Noise Reduction Schemes

The noise performance of the simple loop oscillator as measured above is in excess of other technologies (such as multiplied quartz or SAWs) at these frequencies and offsets. However the full potential of the high Q factors provided by the SLC resonator technology will not be realised until either lower

flicker noise gain devices become available or noise reduction schemes such as those employed by Galani [1], Walls [2], or McPherson [3] are used.

Essentially, in each of these schemes the oscillator noise is detected by using a discriminator and a phase sensitive detector (a Double Balanced Mixer). The output of the phase detector is then amplified by a low noise amplifier and this error signal is fed back into the oscillator to reduce the system phase noise.

With the availability of very low noise Operational and Low Frequency radio amplifiers the predominant source of noise in these systems becomes the flicker noise of the mixer. Figure 8 indicates the level and form of typical mixer noise.

A correctly configured oscillator topology and servo design can reduce the term $S_{\phi Comp}(f_m)$ in equation (7) to that level associated with the Double Balanced Mixer phase detector, resulting considerable improvement in oscillator phase noise.

Essential to this type of noise reduction scheme is a frequency discriminator with high Q factor and residual noise better than or equal to the noise of the phase detector. The resonators described herein satisfy both criteria.

To estimate the oscillator performance which may be obtained we assume:

1) Resonator unloaded Q factor = 200,000, centre frequency = 10 GHz;
2) A phase detector noise level equal to that shown in Figure 8;
3) Noise reduction amplifier noise is less than the phase detector noise;
4) Equation (7) can be used to predict oscillator performance with $S_{\phi Comp}(f_m)$ being equal to only that noise of the phase detector;
5) Servo Loop Gain is optimised to reduce noise at offsets between 100 Hz and 10kHz.

We obtain the following predicted oscillator performance shown in Figure 12. Currently, work on such an oscillator is in progress by the authors.

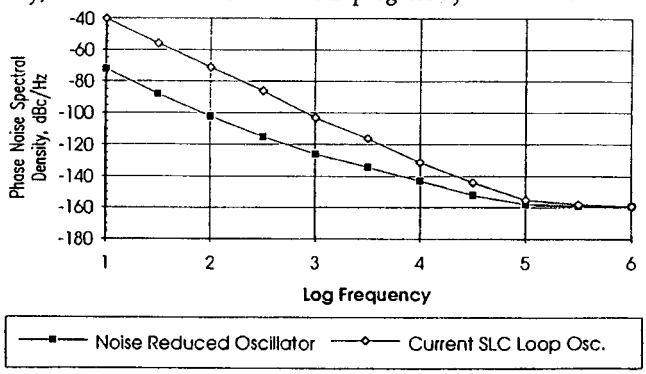


Figure 12. Predicted Phase Noise of a 10 GHz Oscillator using an SLC resonator and a DBM noise reduction scheme.

6. CONCLUSIONS

Several Sapphire Loaded Cavity resonators have been built. These resonators can be constructed to a specified operating frequency with an accuracy of better than 5ppm. Resonator Q factor exceeds 200,000 at temperatures near ambient, and resonator Qf product is constant at constant operating temperature.

This resonator topology has a temperature coefficient of -70 ppm/°C which is reduced by temperature control to realise a long term centre frequency stability of approximately 2×10^{-8} per hour.

Resonator residual noise is at a level where oscillator performance will be limited primarily by active components.

Oscillators incorporating SLC based resonators and noise reduction schemes have the potential to provide signal sources with previously unequalled short term noise.

ACKNOWLEDGMENTS

Portions of this work were supported by the Australian Research Council.

The authors would like to acknowledge the efforts of Cameron McNeilage for his assistance with the construction of the resonators and data acquisition, Dr AG Mann for his technical contributions and Dirk Ulex Precision Engineering for the precision machining of the resonators and housings.

REFERENCES

[1] Z. Galani, M.J. Bianchini, R.C. Waterman Jr., R.W. Laton, R. Dibiase, J.B. Cole, *IEEE Trans. Microwave Theory Tech.*, MTT-32, 1984, pp 1556-1565

[2] F. W. Walls, C.M. Felton, T.D. Martin, "High Spectral Purity X-Band Soucre", Proc 44th Symposium on Frequency Control, pp542-547 (1990)

[3] H. McPherson, "An X-Band Frequency Agile Source with Extremely Low Phase Noise for Doppler Radar", Proc RADAR 92, Brighton, England

[4] D.G. Blair, M.E. Costa, A.N. Luiten, A.G. Mann, M.E. Tobar, "Cryogenic Sapphire Dielectric Resonator Microwave Oscillator With Improved Stability" Proc European Frequency and Time Forum 1994 (to be published)

[5] E.N. Ivanov, D.G. Blair, V.I. Kalinichev, "Approximate Approach to Design of Shielded Dielectric Disk Resonators", *IEEE Trans. Microwave Theory Tech.* vol. 41, no. 4, pp.632-638 (1993)

[6] M.M. Driscoll, "Low Noise Microwave Signal Generation: Resonator/Oscillator Comparisons", Proc 43th Symposium on Frequency Control, (1989)

[7] H. Peng, D.G. Blair, "Test of An Interferometric Sapphire Transducer with The Super Attenuator of Virgo Graviational Wave Antenna" (to be published.

[8] M.E. Tobar, A.J. Giles, S.J. Edwards, J.H. Searls, "High-Q TE Stabilised Sapphire Microwave Resonators For Low Noise Applications." Proc 47th Symposium on Frequency Control, pp749-756, (1993)

[9] V.B. Braginsky, V.S. Ilchenko, and K.S. Bagdassarov, "Experimental observation of fundamental microwave absorption in high-quality dielectric crystals." *Physics Letters A*, Vol 120, no. 6, pp, 300-305

10 D.B. Leeson, "A simple model of feed back oscillator noise spectrum," in Proc. IEEE, vol, 54, No. 2, pp 329-330, Feb 1966.

Telecommunications and Teletiming (A 2)

Telecommunications Network Synchronization Using Improved SDH Pointer Processing Techniques

by Ralph Urbansky, Philips Kommunikations Industrie AG

Abstract

The SDH as the new international transmission standard is not fully compatible with the existing equipment with respect to timing aspects. It requires synchronization and cannot provide synchronization links for the clock signal distribution between digital exchanges.

In this paper the application of synchronizer/desynchronizer concepts known from PDH (Plesiochronous Digital Hierarchy) will be discussed. This concept obviates the need to synchronize the SDH network. The application of the adaptive threshold modulation scheme for waiting time jitter reduction provides timing-transparent synchronization links. Measurement results from field trial equipment demonstrate the feasibility of this approach.

1 Introduction

The Synchronous Digital Hierarchy (SDH) is accepted as the new emerging transmission standard, which will replace the existing Plesiochronous Digital Hierarchy (PDH). It is characterized by homogeneous bit rates and data formats in conjunction with a flat multiplexing structure. The SDH frame format and the nominally synchronous operation offer switching capabilities for the multiplexed signals by timeslot interchanging techniques. This feature provides for flexible bandwidth allocation within the transmission network.

In currently deployed SDH equipment the time switching facility requires synchronization for any SDH network element. Furthermore, current implementations do not provide timing transparency for network synchronization purposes. In SDH standardization bodies therefore, an alternative timing distribution scheme is being proposed. However, this new timing distribution concept is not fully compatible with the existing network synchronization philosophy which is based upon the slip performance of digital exchanges.

This paper demonstrates, that this lack of timing transparency is not a consequence of the SDH frame format and the pointer-based phase alignment technique but, rather, that it results exclusively from current pointer processor implementations. It will be demonstrated, that an advanced pointer justification algorithm provides for a timing transfer performance comparable to that achieved in PDH equipment. In Section 2 background information of timing aspects in PDH and SDH equipment will be given. In Sections 4 and 5, the SDH pointer processor and the synchronizer will be discussed, respectively. In Section 6 theoretical results of different synchronizer schemes will be analyzed, whereas in Section 7 measurement results obtained with field trial equipment will be presented.

2 Background

PDH transmission links are used in the telecommunications network to connect digital switches. For this application, point-to-point configurations are predominant. The line bit rate was derived from free running oscillators at relative frequency tolerances ranging from 15 ppm to 50 ppm (parts per million). Bit rate adaptation is performed using the well known bit stuffing (justification) technique. Based on waiting time jitter analysis [1], positive stuffing was most commonly employed. The stuff ratio is defined as the number of stuff bits per stuff opportunities and allows the calculation of low frequency jitter components. In PDH systems, the line bit rates including tolerances and the frame repetition rate, i. e. the stuffing frame format, are chosen with respect to waiting time jitter. Under network synchronization aspects two basic features are covered:
- The PDH transmission network elements need no synchronization as a result of the bit justification technique.
- The PDH can be used as a synchronization network for digital exchanges thanks to the timing transparency.

In contrast to this straightforward network application, SDH was specified in order to introduce new services in the transmission network by means of increased flexibility. This is achieved by a byte oriented frame structure, which allows the identification of the time slots of each individual channel. The slip buffering technique generally used for phase alignment purposes in digital switches would create unacceptable transmission delays. Therefore, the so-called pointer was introduced. This indicates the phase difference as an offset

value between the transport frame (STM-N, Synchronous Transport Module, level N) and the payload frame (VC-n, Virtual Container, level n). A constant offset compensates for constant transmission delays in a fully synchronized network. The network wander is defined as the phase difference between data and clock signal provided via different transmission links. It is caused by varying temperature-dependent propagation delays or by phase uncertainties in clocks and is limited to 18 µs as per ITU-T Recommendation G.823/824. These phase variations are accommodated by adjustments of the pointer values, so-called pointer justification events (PJE). The pointer value can be incremented or decremented. Therefore, the pointer is used to perform +/0/- justification for bit rate adaptation in plesiochronous operation.

However, in current pointer processor implementations the network elements require synchronization to keep the propagation of PJEs within the network under control. In an ideally synchronized network all pointers, which are generated by a single free running network element clock (holdover), will cancel at the pointer processor of the next node. Due to phase detector errors in the pointer processor decision loop and due to clock noise, some of these pointers will feed through. The desynchronizer as the gateway element between SDH and PDH has to accommodate the large phase hits associated with PJEs superposing the PDH signal. It consists of a very low bandwidth PLL (< 1 Hz) which can be considered as a low pass filter smoothing the phase hits. Therefore, the buffer size of the desynchronizer is the bottleneck for accumulations of PJEs in the same direction within its time constant. Although PJE accumulations are no problem within the SDH network, they have to be limited to avoid buffer overflows/underflows at the desynchronizer. As no other mechanisms are implemented in current pointer processors to keep the PJE accumulation and propagation in the network under control, standardization bodies propose the synchronous operation of the SDH.

Therefore, in terms of network synchronization, two major differences arise on the introduction of the SDH:

- Due to the current pointer processor implementation the SDH network elements need synchronization to prevent PJE accumulation.
- The SDH cannot provide 2-Mbit/s timing transparent synchronization links for network synchronization due to the current pointer processor implementation.

3 STM-N Timing

In the context of network synchronization a distinction is to be made between
- the provision of a synchronization signal and
- the provision of a synchronization link (third party timing).

In standardization bodies mainly the first aspect is being covered with the focus on the synchronization of SDH network elements with respect to the pointer activity. It is assumed, that the synchronization concept also meets the requirements of digital exchanges. Therefore, it is proposed to use the SDH as a synchronization network for the public switched telephone network (PSTN). The PDH, however, provided synchronization links between corresponding switches and their clocks only.

No fully PDH compatible solution is currently available for the provision of synchronization links in SDH networks, due to the pointer processor implementation. Therefore, the problem of third party timing is dealt with only by relaxing the synchronization signal wander performance from 18 µs to 40 µs for certain network configurations.

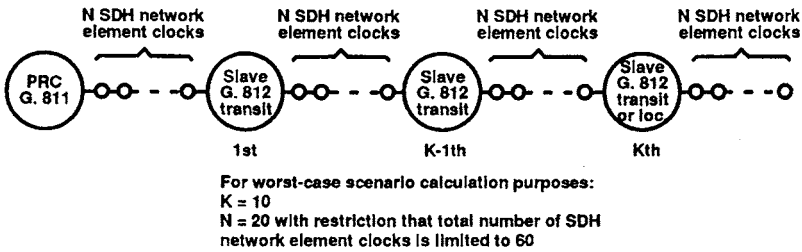


Fig. 1 Synchronization network reference configuration

In **Fig. 1** the timing reference configuration as per ITU-T Recommendation G.803 is reproduced. The signal of the primary reference clock (PRC) is distributed via the SDH line signal (STM-N). Two types of clocks are used in this chain:
- synchronization supply units (SSU)
- SDH equipment clocks (SEC).

SSUs may be part of switches or SDH equipment with a holdover performance and clock noise as per ITU-T Rec. G.812 and a PLL bandwidth in the range of several milliHertz. SECs are intended to be cheaper due to relaxed holdover characteristics and higher PLL bandwidth (1..100 Hz). Therefore, the holdover period of an SEC should be limited to a short period of time.

The general master slave clock distribution scheme consists of a hierarchical structured performance level (stratum), which is defined mainly by the holdover characteristic. According to this concept, a clock of a certain stratum can be

clocked only by a clock of equal or higher stratum. From **Fig. 1** this clock distribution rule is obviously violated, because the SSUs are referenced to an SEC. Therefore, the so-called timing marker has been standardized as an indicator for the highest stratum reference clock in the chain. Unfortunately, currently deployed SSUs are referenced via 2-Mbit/s or 2-MHz signals without timing marker (timing marker chain is interrupted by SSUs). The problem of signalling the SSU status (holdover or traceable to PRC) is still under discussion.

The problem of creating timing loops in a ring configuration can be solved by introducing a guard time, which allows the propagation and settling of a new timing marker before switching the clock to a different reference input. This prevents the clock gateway element in a ring from switching its reference signal to its own looped signal, when it has lost the external reference. Another problem in ring configurations is due to the fact that in case of a synchronization signal failure within the ring, all network elements have to switch to the other reference. The holdover time requirements in conjunction with the timing marker relaying and guard time considerations are still under discussion.

A repeater-based timing concept in SDH network elements is called "Through Timing", i. e. in Add-Drop-Multiplexers the timing content of the STM-N signal is passed through in both directions. In ring configurations no extra timing marker is required, because only the network element adjacent to the line failure has to loop back its timing signal. No timing reconfiguration is necessary in the other network elements. The Through Timing concept is a straightforward solution, but it can be applied only in small ring configurations. However, it is not standardized by ITU-T Recommendations.

4 The SDH Pointer Processor

The pointer processor performs two main tasks. It adapts the payload frame rate to the transport frame rate by inserting positive or negative stuff bytes and corresponding adjustments of the pointer value, which represents the phase offset. Furthermore, it provides a mechanism which absorbs incoming PJEs and prevents a direct feedthrough.

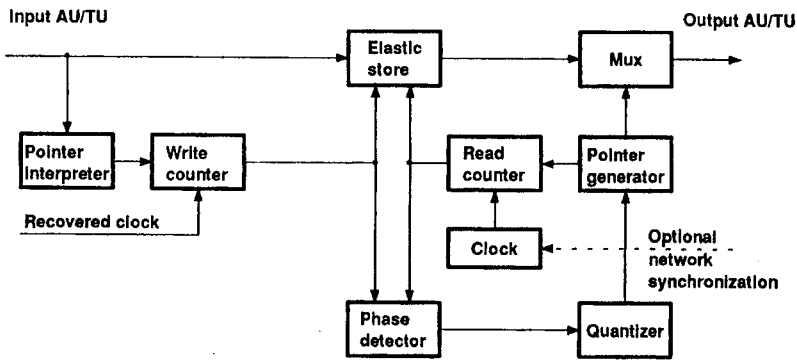


Fig. 2 Block diagram of the classical pointer processor

4.1 The Classical Pointer Processor

This general functionality is realized by first recovering the payload frame phase by pointer interpretation as shown in **Fig. 2**. The data is buffered in an elastic store. The buffer fill is a measure of the phase offset between the output transport frame defined by the local clock and the payload frame. A threshold hysteresis is applied as a nonlinear filter function, which accommodates phase variations of the payload frame or the local clock phase before a PJE is generated. This prevents a direct feedthrough of incoming PJEs. If the thresholds are crossed, a pointer justification request is forwarded to the pointer generator. The write and read counters operate as address counters for the elastic store thereby generating the frame structures. The phase detector errors associated with frame structure effects are not in the scope of this paper. An ideal phase detector is assumed, which monitors the buffer fill resulting from PJEs as well as phase variations of the line input signal with respect to the local clock phase scaled down to the corresponding payload frame rate. The

quantizer input step size corresponds to the hysteresis threshold spacing. Applying this concept, only this parameter is available for accommodating the conflicting requirements of PJE feedthrough and timing transparency.

4.2 The Generalized Pointer Processor

The two main tasks of the pointer processor are shown in **Fig. 3** as separate functional blocks subdivided in the desynchronizer block and the synchronizer block [2]. Obviously, the functionality of these blocks corresponds to those in PDH systems.

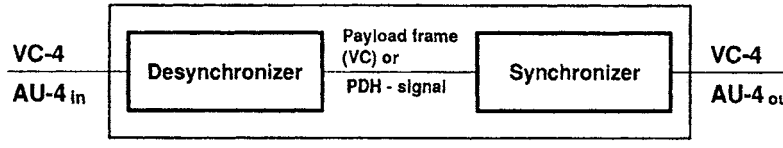


Fig. 3 Functional model of a pointer processor

The desynchronizer recovers the incoming signal phase by extracting fixed and variable stuff bits associated with pointer or bit justifications. The resulting data gaps are smoothed by the low pass function of the desynchronizer PLL. This linear filter function prevents the feedthrough of incoming phase hits to the synchronizer output. The probability of PJE accumulations depends on the relationship between the spectral content of the phase quantization error, generated by the synchronizer of the preceding pointer processor, and the PLL-filter cut-off frequency.

The synchronizer adapts the desynchronizer signal to the transport frame rate given by the local clock. No additional threshold spacing is necessary for the absorption of incoming PJEs. The timing transparency is therefore improved.

Fig. 4 shows the block diagram as in **Fig. 2**, but extended by a filter representing the desynchronizer function. The simple quantizer is replaced by a general decision circuit. This allows the optimization of the synchronizer portion with respect to the jitter characteristics associated with the timing transfer performance.

The functional distinction of a pointer processor between desynchronizer and synchronizer allows for a separate optimization of parameters and algorithms with respect to otherwise conflicting requirements. The design of the desynchronizer part is an straightforward issue of implementing a digital filter with respect to the spectral phase error characteristic of the synchronizer. The remainder of the paper will focus on the synchronizer part.

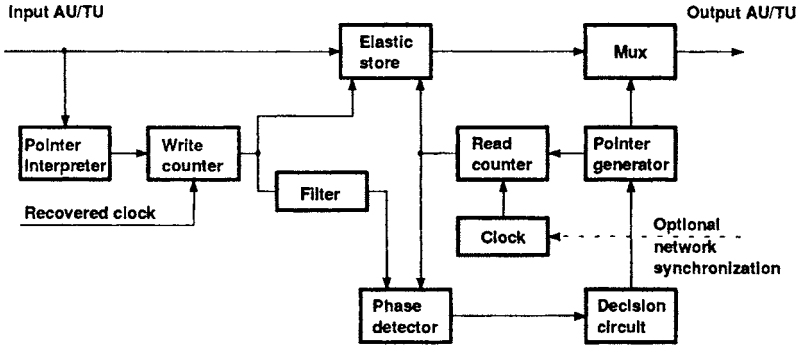


Fig. 4 Desynchronizer / Synchronizer model

5 The Synchronizer Model

The model of the synchronizer in **Fig. 5** as a part of a pointer processor represents the phase transfer function in conjunction with the phase noise resulting from the phase quantization. The phase detector is modelled as an ideal device without quantization effects and nonlinearities associated with the frame structure. The sampling rate corresponds to the maximum frequency of PJEs (4 times the frame period).

The input signal is given by the stuff ratio, which is equivalent to the input frequency. Write and read address counters are modelled by integrators, which are represented by unit delay blocks marked with a "T" in a feedback loop. The phase difference is calculated by the phase detector as the input for the decision circuit. The decision circuit generates the pointer justification requests. The delay in the synchronizer feedback loop corresponds to the pointer interpretation delay in the next network element representing the delay between (low order) pointer interpretation and the reaction on the associated address

counter. At the next node, only the write address counter is shown, which subsequently feeds the low pass filter representing the desynchronizer function. A comparison between different decision circuits will be made on the basis of the phase error of the synchronizer as the difference between the write phases of each network element and the subsequent network element.

5.1 The Decision Circuit

The decision circuit processes the phase difference as a measure of the buffer fill at the elastic store and generates pointer justification requests. It determines the spectral characteristic of the synchronizer phase error.

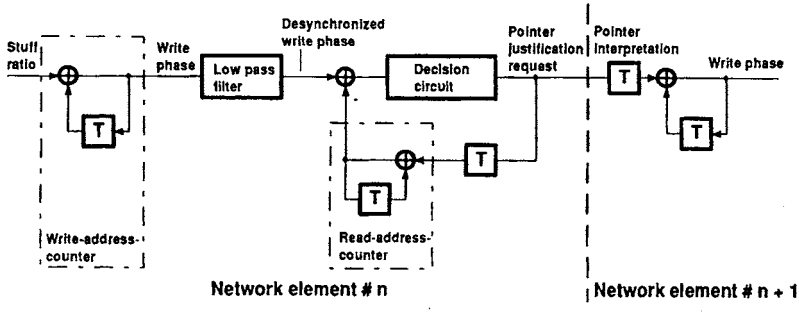


Fig. 5 Simulation model of the synchronizer

A quantizer as a decision circuit is shown in **Fig. 6a**. It provides a three-level output signal for +/0/- pointer justification requests. As a result of the inclusion of the low pass filter, no additional input hysteresis for pointer absorption is required. Therefore, the optimum threshold spacing is given by the pointer step size of the corresponding level (AU-3/4, TU-11/12/2/3). If these thresholds are crossed by the phase difference, a pointer justification request is generated. As known from waiting time jitter analyses [2],very low frequency jitter components are generated due to the small (nominally zero) stuff ratio.

A known method for waiting time jitter reduction is called threshold modulation [3]. The superposition of a modulation signal to the phase difference before quantization as shown in **Fig. 6b** shifts the jitter spectrum to higher frequencies. The addition of a frame-synchronous sawtooth signal of the amplitude of 1 UI

(Unit Interval) may be considered as an increase of the stuff ratio by the inverse of the sawtooth period (in frames). The sawtooth can be subdivided in the sum of a linear phase ramp and a staircase function of opposite sign. The jitter spectrum also consists of two parts: one part resulting from the sum of the input stuff ratio and the stuff ratio associated with the phase ramp, and the other part resulting from the staircase. Therefore, very low frequency jitter components are avoided. The sawtooth period in conjunction with a nominally zero input stuff ratio determines the remaining waiting time jitter (low frequency sawtooth desirable) as well as the lowest jitter spectrum components due to the modulation signal (high sawtooth frequency desirable). This contradiction can be relaxed by alternative modulation signals with improved spectral characteristics [4].

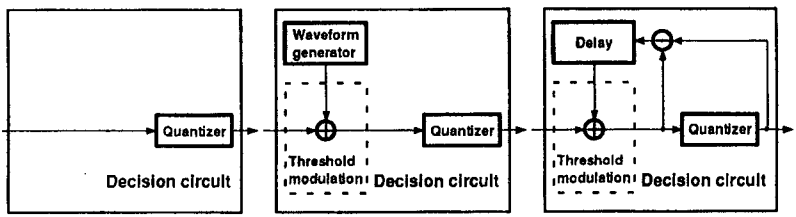


Fig. 6 a
Quantizer

Fig. 6 b
Threshold modulation

Fig. 6 c
Adaptive threshold mod.

The contradiction may be resolved by adapting the modulation signal with respect to the jitter spectrum. In the decision circuit shown in **Fig. 6c** the modulation signal is derived from the delayed quantization error. This decision circuit forms a first order Sigma-Delta Modulator. The integrator function of the read address counter in the outer feedback loop of the synchronizer creates a second order Sigma-Delta Modulator. An exact mathematical analysis has been carried out for non-overloaded quantizers and certain input signals [5, 6]. Two main results can be summarized:

- the quantization error sequence has a white noise spectrum
- the quantization error is statistically independent of the input signal.

As a result of the white noise spectrum of the quantization error, the output jitter due to PJEs corresponds to a time discrete second order differentiation. At the next network element this signal is integrated by the write counter. Therefore, a high pass filtered spectrum of the white noise quantization process is applied at the filter representing the desynchronizer.

The statistical independence of input signal from quantization error is an important property under network jitter and PJE accumulation aspects, provided a suitable input low pass filtering is carried out to meet the DC input condition. Simulation results have been presented [7], which indicate uncritical jitter propagation characteristics even in networks of up to 100 network elements.

6 The Synchronizer Phase Error

In the following diagrams the synchronizer phase error at VC-12 level (corresponding to 2 Mbit/s) is shown. It is defined as the difference between the write phases of pointer processors of each network element and the subsequent network element. The low pass filter representing the desynchronizer is not included to simplify the interpretation of the resulting phase error and PJE sequences. The stuff ratio in the examples is 1/200.

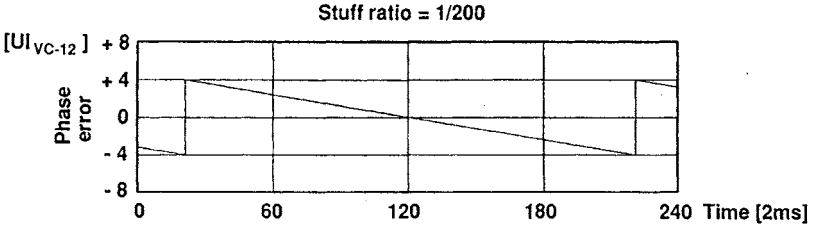


Fig. 7a Phase error

In **Fig. 7a** the phase error associated with the decision circuit of **Fig. 6a** is shown. Resulting from the integration of the input frequency offset, a falling phase ramp is obtained. If the quantizer threshold is exceeded, a phase step of (1 $UI_{VC\text{-}12} \hat{=} 446ns$) associated with a PJE compensating the phase error is generated. In the case of an irrational input stuff ratio, no exact periodicity of the sawtooth-like phase error can be observed. For general rational input stuff ratios, the phase error is composed of the superposition of sawtooth signals with amplitudes corresponding to the numerator and denominator, respectively. The stuff ratio 1/200 in **Fig. 7a** results in a sawtooth period of 200 stuff frames. Decreasing the absolute value of the stuff ratio leads to an arbitrary low fundamental sawtooth frequency. The corresponding jitter components cannot be reduced by the low pass filter.

In **Fig. 7d** the stuff ratio of 1/200 is applied again using the same decision circuit (**Fig. 6c**). The falling phase error slope resulting from the stuff ratio can also be observed here. The mean short term phase error of 4 UI_{VC-12} results in

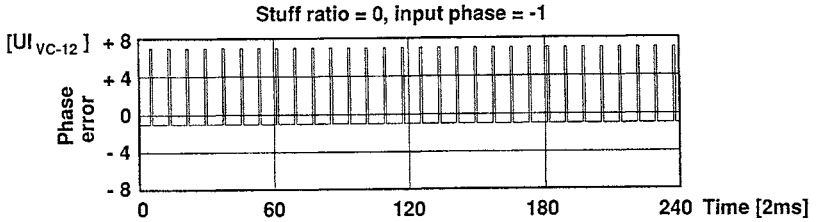


Fig. 7c Phase error, Adaptive threshold modulation

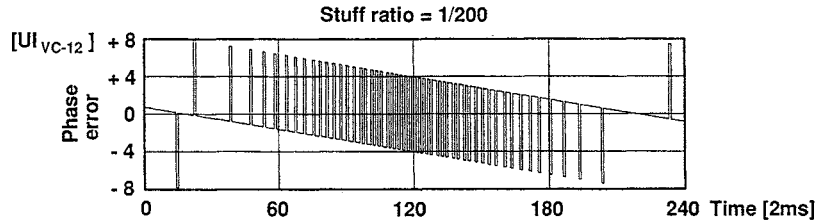


Fig. 7d Phase error, Adaptive threshold modulation

subsequent positive and negative PJEs with highest frequency jitter components compensating the phase error, whereas a phase error near zero requires relatively infrequent PJEs. As a PJE is generally immediately followed by a PJE of opposite polarity, the corresponding spectrum consists mainly of high frequency components, which can easily be reduced by the low pass filter representing the desynchronizer. Although the stuff ratio of 1/200 is applied, no short periodicity of the phase error signal can be observed. This confirms the white noise characteristic of the quantization error.

7 Measurement Results

Measurement results of the timing transfer characteristic of SDH pointer processors of 140-Mbit/s signals have recently been published [2]. The phase error of the classical pointer processor of 12 bytes associated with the stuff

In **Fig. 7b** the phase error associated with the decision circuit of **Fig. 6b** is shown. Due to the application of the same stuff ratio, basically the same falling slope is obtained. A sawtooth modulation signal is employed with the amplitude of a pointer step size (8 UI_{VC-12}) and a period of 8 stuff frames. A large number of PJEs is generated as a consequence of the threshold modulation. Negative PJEs are observed in a fixed sequence (pattern) resulting from the 8-frame modulation period. Positive PJEs are inserted to compensate the current short term phase error value. Therefore, the threshold modulation scheme employing a sawtooth signal can be considered as a pulse width modulation of the phase error. The stuff ratio associated with the phase ramp of the sawtooth results in a jitter component, which cannot be further reduced. In **Fig. 7b** this sawtooth stuff ratio can be interpreted as an 8-step quantization of the phase difference and subsequent pulse width modulation. This confirms that the application of threshold modulation requires optimization due to the contradictory requirements relating to resolution and lowest jitter frequency component.

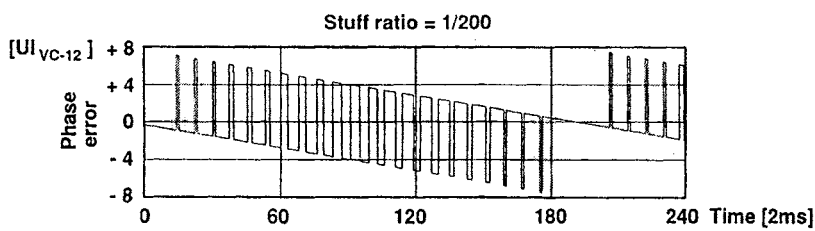


Fig. 7b Phase error, Threshold modulation

In **Fig. 7c** the phase error associated with the decision circuit of **Fig. 6c** employing adaptive threshold modulation is shown. In this example the nominal input frequency (stuff ratio zero) is applied in conjunction with a constant phase error of -1 UI_{VC-12}. This illustrates the relationship to the threshold modulation. A similar pulse width modulated phase error is observed in association with a sawtooth-shaped quantization error, which is applied as the modulation signal. The modulation signal is adapted employing the adaptive threshold modulation scheme, thus providing the highest possible pulse width modulation frequency for the encoding of the (slowly varying) input signal.

threshold hysteresis spacing was reduced by the application of the adaptive threshold modulation to values of about 2 UI corresponding to an improvement by a factor of about 50. The residual phase error emanates from the limited resolution of the phase detector.

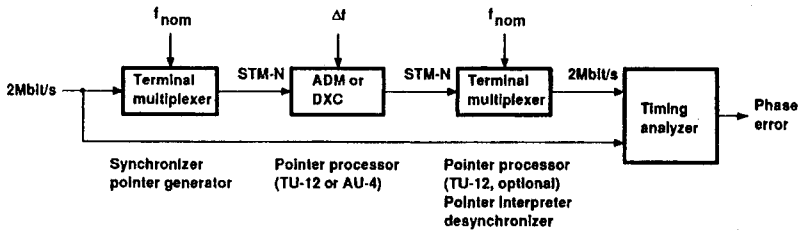


Fig. 8 Test set - up

In **Fig. 8** the measurement configuration of the equipment for the analyses of the timing transfer characteristic of 2-Mbit/s signals is shown. The network consists of three elements of the PHASE (Philips Advanced SDH Equipment) system family:

- a terminal multiplexer comprising the synchronizer, which is fed by the 2-Mbit/s test signal
- an Add-Drop multiplexer (ADM) comprising the pointer processor (selectable between high and low order)
- a terminal multiplexer configured as a desynchronizer with an optional low order pointer processor.

The phase difference between the test signal and the desynchronizer output is recorded and plotted. A constant frequency offset is applied at the intermediate ADM, which allows a comparison with the phase error results of **Fig. 7**. The main difference with respect to the block diagram in **Fig. 5** is the inclusion of the desynchronizer filter function.

The phase error scaled in UI of 2 Mbits/s over time is shown in **Fig. 9**. In this configuration the pointer processor of the intermediate ADM is switched to low order mode (TU-12, 2 Mbit/s level) without jitter reduction. (ITU-T mode). The pointer processor at the ouput terminal multiplexer is switched off. The frequency offset of the pointer processor is 0.2 ppm. The amplitude of the resulting sawtooth-like phase error is only slightly reduced by the filter function

of the desynchronizer and is about 7 UI. Applying a frequency offset of opposite polarity would result in the inverted phase error function. The peak-to-peak value between positive and negative offsets corresponds to the threshold hysteresis scaled down from the VC-12 to the 2-Mbit/s bitrate and is about 14 UI (7 µs).

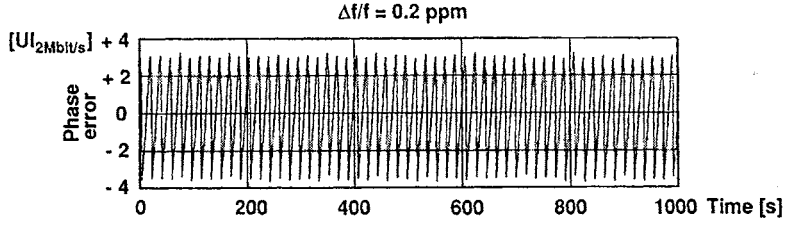


Fig. 9 Phase error, ITU - T mode

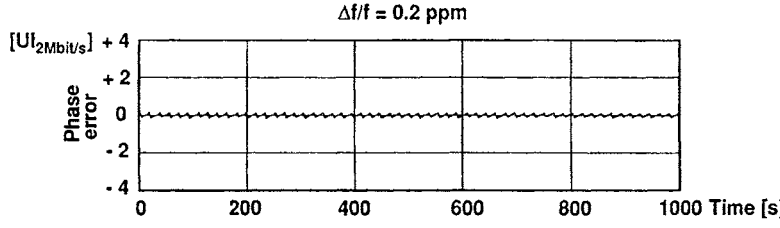


Fig.10 Phase error, Adaptive threshold modulation

Switching the pointer processor to adaptive threshold modulation would result in a residual phase error without phase hits, which is caused by implementation specific imperfections and frame structure effects. To demonstrate the frame structure effects, the network configuration associated with the phase error plotted in **Fig. 10** consists of two pointer processors switched to adaptive threshold modulation mode. The first pointer processor performs AU-4 level payload frame alignment and is operated with a clock frequency offset of 0.2 ppm resulting in a constant PJE rate. The second pointer processor is configured in TU-12 mode. It is located at the input stage of the output terminal multiplexer and is operated with the nominal clock frequency. Therefore, the AU-4 PJEs of the ADM are not absorbed, but rather recalculated into TU-12

PJEs, encoding and propagating all phase detector imperfections including frame structure effects. The signal is subsequently fed into the desynchronizer stage of the terminal multiplexer.

The residual phase error represented in **Fig. 10** results in a sawtooth characteristic. The amplitude is about 0.3 UI and corresponds to a similar improvement by a factor of about 50 as for 140-Mbit/s signals. The phase detectors employed in the pointer processors are based on an averaging algorithm, which calculates the mean phase value over an integer multiple of the frame period in order to suppress all regular frame structure effects. However, the TU-12 pointer processor is fed by a payload frame (VC-4), which is moved within the transport frame (STM-1/AU-4) to compensate for all frequency offsets applied at the ADM. The VC-12 bytes within a VC-4 are, therefore, shifted over the overhead gaps of the STM-1 frame resulting in the small sawtooth-shaped phase error.

8 Conclusion

The deployment of SDH networks based on network elements employing the classical pointer processor causes timing related compatibility problems with existing equipment. Concerning these problems a distinction can be made between the jitter at the PDH desynchronizer due to PJE accumulation, and the network synchronization aspect. In current standards, the probability of PJE accumulation is expected to be sufficiently reduced by the improvement of the clock quality. In the context of network synchronization some regulatory issues and synchronization network restrauration concepts are still under study.

This paper shows, that suitable input filtering (as a desynchronizer function within a pointer processor) prevents PJE accumulation. Network synchronization via synchronization links can be achieved by the application of the adaptive threshold modulation algorithm.

The measurement results confirm the theoretical results obtained on the basis of simplified models. The expected implementation-specific imperfections such as phase detector resolution and residual frame structure effects have been demonstrated.

This paper demonstrates the feasibility of operating the SDH without synchronization, an approach which will lead to a smooth transition from the PDH to the SDH and a substantial reduction in costs.

9 References

[1] D. L. Duttweiler: "Waiting Time Jitter", BSTJ., vol. 51, no. 1, January 1972.

[2] M. Klein, R. Urbansky: "Network Synchronization-A Challenge for SDH/SONET", IEEE Communications Mag., September 1993.

[3] G. L. Pierobon, R. P. Valussi: "Jitter Analysis of a Double Modulated Threshold Pulse Stuffing Synchronizer", IEEE Trans. on Commun., vol. 39, no. 4, April 1991.

[4] R. G. Kusyk, T. E. Moore, W. A. Krzymien: "Spectral Analysis of Waiting Time Jitter in the Presence of Stuff Threshold Modulation", Electronics Letters, vol. 26, no. 8, April 1990.

[5] P. W. Wong, R. M. Gray: "Two-Stage Sigma-Delta Modulation", IEEE Trans. on ASSP, vol. 38, no. 11 November 1990.

[6] N. He, F. Kuhlmann, A. Buzo: "Multiloop Sigma-Delta Modulation", IEEE Trans on Information Theory, vol. 38, no. 3, May 1992.

[7] R. Urbansky: "Pointer Processing for Synchronization Signals in SDH Equipment", ETSI-STC-TM3, The Hague, Temporary Document 2/39, April 1991.

Synchronization Strategy and Phase Noise Performance in Networks including SDH Transmission Equipment

Michael Wolf, Alcatel SEL

The synchronization performance of the existing master/slave approach in the telecommunication networks can be improved by introduction of a synchronization quality management on this structure using the synchronization status messages available in SDH transmission systems. The current synchronization reference links on 2 Mbit/s are subject to network configuration actions, which potentially disturb the synchronization architecture. The SDH, however, will provide a physically stable network on N * 155 Mbit/s for the synchronization links improving the reliability of these links. These improvements will overcompensate for the additional phase wander introduced on the payload by the pointer adjustments performed in SDH for phase adaptation of payload entities.

1 Introduction

The digital switching and multiplexing techniques for 64 kbit/s (Basic Rate) channels are based on synchronous operation. The evolving systems on the Synchronous Digital Hierarchy (SDH) are defined for synchronous operation at the 155 MHz level. So the network synchronization performance becomes an essential parameter for the quality of the end–to–end connections through the network.

2 Synchronous switching and multiplexing

31 Basic Rate channels are multiplexed synchronously byte by byte to form a Primary Rate signal. One byte with a fixed bit pattern is added as a frame alignment byte. A peculiar Basic Rate channel is identified by the phase of a specific byte relative to the frame alignment.

A frame aligner is used to synchronize an incoming signal to an internal system clock by delaying the signal until it matches with the internal system frame. This delay buffer can absorb some phase wander of the incoming signal. However, when the buffer gets to overflow a controlled frame slip is performed deleting just one frame. Thus synchronization is maintained on cost of the integrity of the signal.

In SDH a payload channel is identified by an adjustable phase offset relative to the frame alignment. This phase offset is given by a pointer value in the frame overhead. This technique allows for payload resynchronization to a new frame just by allocating a new pointer value without delaying the signal significantly. Any phase wander of the incoming signal which exceeds the pointer processor buffer size is

accommodated by adapting the pointer value accordingly. This does not affect the signal integrity. It adds only a certain amount of phase instability to the signal, which is a quantization phase noise caused by the step wise adaptation of phases.

3 Synchronization architecture and performance

The network of the switches is operating synchronously at 2 Mbit/s. The switches are interconnected by 2 Mbit/s connections which are provided by the transmission network. These connections are assumed to transmit the timing transparently to provide synchronization reference links between the switches.

In an SDH transmission network the 2 Mbit/s connections between two switches is provided by means of multiplexing and routing by crossconnects. The physical signal path needs not to follow a straight line connecting both interface points, it may be routed on some path through the network, which is determined by the availability of traffic capacity (network configuration, protection switching). In order to provide high phase stability of the transported payloads all elements of the transmission network should operate at the same frequency, i. e. they need "synchronization", or more exactly "syntonization".

There is a difference of the effect of synchronization impairments in switching networks and in transmission networks. In switching networks the synchronization impairments cause signal degradations by introduction of frame slips which affects the quality of service on the end–to–end connections. In transmission networks the synchronization impairments cause jitter and wander on the transported payloads.

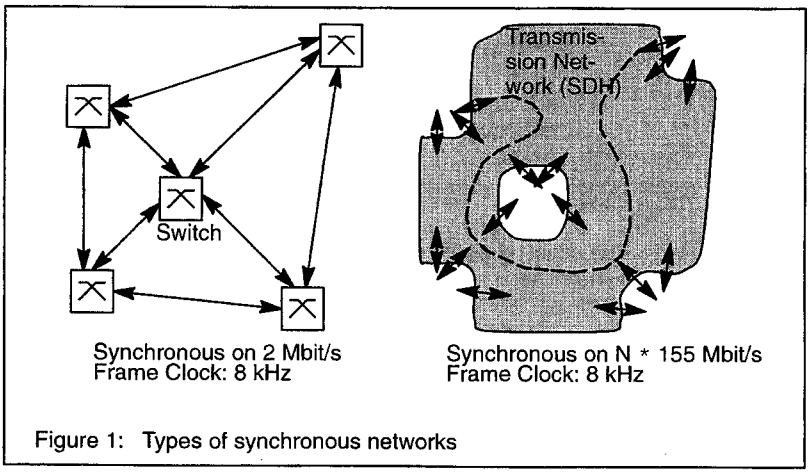


Figure 1: Types of synchronous networks

3.1 Synchronization architecture

The network nodes of the transmission network are located almost at the sites of the switches. This means that the synchronization architecture of the transmission network would look very much like that of the switching network. Both synchronization architectures can be operated separately or they can be combined. In the following it is outlined that such a combination would improve the synchronization performance and by this the quality of service on end–to–end connections.

3.2 Synchronization failures

In the switching network there is only the possibility of selecting synchronization reference inputs according to a predefined priority. This means, that an input is selected as a timing reference as long as there is a valid input signal available within some frequency tolerance. According to the hierarchical structure a large network part can be synchronized to a reference signal which has some frequency offset.

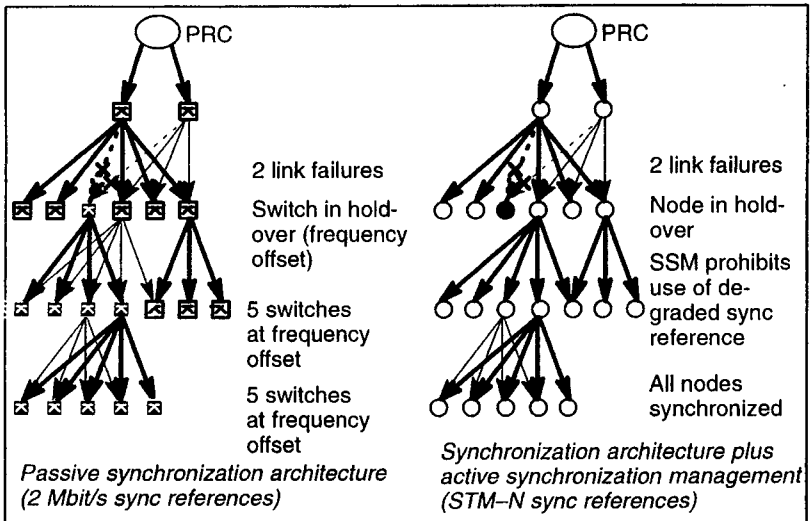


Figure 2: The effect of synchronization link failures on the switching network and on the SDH network.

In the transmission network there will be an indication of clock tolerance within the STM–N reference input signal. This is called the Synchronization Status Message (SSM) and gives the frequency tolerance of the clock source the timing reference is traceable to. This enables the network elements to select a reference according to its frequency accuracy. This flexible reference selection should be combined with a hierarchical synchronization architecture. Therefore no overall network in-

telligence is necessary for avoiding synchronization loops when selecting synchronization references.

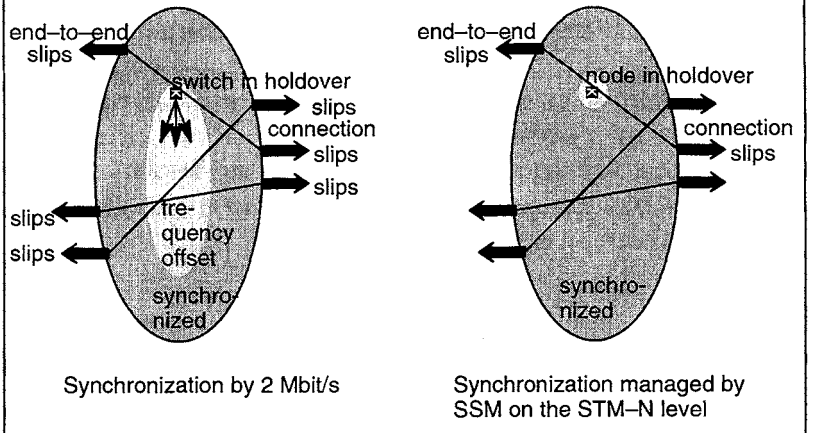


Figure 3: Improvement of frame slip performance by synchronization quality management.

3.3 Network slip performance

Now it becomes advantageous to use the improved synchronization performance of the transmission network also for the switching network. By this the probability to meet an end–to–end connection through the network which performs frame slips is decreased very much. So, the overall slip performance of the network is being improved.

3.4 Standards

There are three clock standards:

- The CCITT Rec G.811 specifies the performance of the Primary reference Clock in terms of long term phase stability of better than 10^{-11} and a moderate short term phase stability of 3 µs.
- The CCITT Rec G.812 specifies a slave clock for use in switches and major network nodes. In ETSI it is called Synchronization Supply Unit (SSU). It is characterized by low bandwidth filtering, a tight holdover performance and a moderate time interval error specification of 1 µs.
- The CCITT draft Rec G.81s specifies the Synchronous Equipment Clock (SEC) for provision of clocks for the SDH equipment. It has no specific band-

width requirement, a moderate holdover specification and a very tight time interval error specification of 40 ns.

There is a standard giving the basic principles of network synchronization architecture. It is included in the CCITT Rec G.803 which is a generic network architecture specification for SDH networks. The three items specified are:

- To have a single node clock at a network node which supplies all equipment of that node with a filtered clock or with a clock generated in the holdover mode in the case of synchronization link failures.
- To have a hierarchical master/slave synchronization architecture.
- To limit the single synchronization chains by 10 cascaded SSUs with intermediate subchains of SECs limited to a number of 20.

3.5 Realization aspects

A derived clock reference of 2 MHz can be provided by all types of SDH equipment. However, to realize the redundant structure of synchronization links it is preferable to select the derived synchronization references at that equipment, which terminates many physical links. This would be a crossconnect. At that point all the Synchronization Status Messages come together and the best timing reference can be selected easily. Therefore it is useful to synchronize a line multiplexer in a synchronization link directly to the incoming line signal and not to include it in the intra node synchronization distribution scheme.

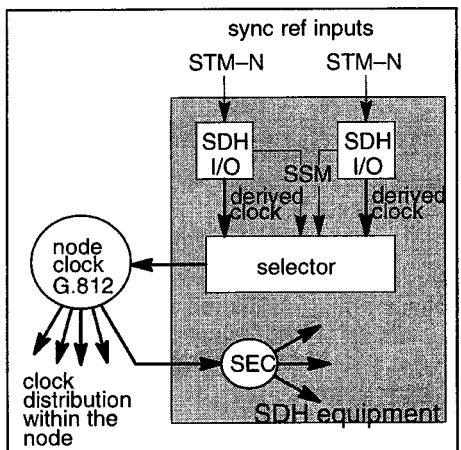


Figure 4: Clock selection and filtering example.

The node clock (G.812) may either be included in the crossconnect equipment (SSU and SEC functionality are merged) or it can be a stand alone device. It needs not to be a dedicated clock device, it can also be the clock filter of a switch. However, in that case the intra node synchronization distribution cannot be done by clock interfaces, as G.803 requires, because the switch may have no or probably not enough clock output interfaces. The filtered clock will be transferred to the crossconnect by a 2 Mbit/s interface first and then, if required, the distribution to other SDH equipment would be done by STM–N interfaces.

3.6 Conclusion on synchronization improvements

The hierarchical master/slave approach together with the synchronization quality management using the SSM in the transmission network has a potential to improve the synchronization performance of the switching network. The synchronization network consists of physically stable reference connections which are not subject to network configuration management which would potentially disturb the synchronization network. The improvement to the switching network is a reduced probability to find a connection through the network which has a low slip performance.

4. Payload phase stability at synchronization failures

Under normal network conditions the payload phase stability is quite high. It is degraded somewhat by the accumulation of jitter and wander on transmission links. However, when some network elements loose all their synchronization reference inputs and get into the holdover mode, a frequency drift is introduced on the payload and by this the phase stability is reduced. The type of phase noise added to the payload depends very much on the type of network element because of the different techniques for synchronizing the incoming payload to the internal system clock.

4.1 Switch in holdover mode

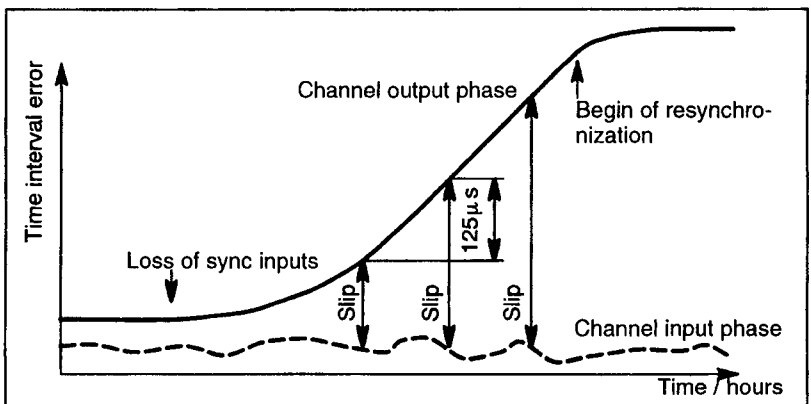


Figure 5: Output channel phase slope when a switch is in holdover

By the operation principle of the frame aligner the switch resynchronizes all payload signals entirely. If a switch is at the holdover mode it is generating a phase slope on the output signals according to the slope of its clock system. The input signals will perform some frame slips when being synchronized to the system clock.

4.2 SDH network element in holdover mode

The system clock of an SDH equipment in holdover will generate a phase slope as well. The STM–N output signal, i. e. the frame envelope for the payload, is generated by this clock. However, the phase of the payload will not follow this slope although it is embedded in the STM–N envelope. The payload is floating within this frame. This floating is not continuous, but rather quantized by the phase steps performed by pointer adjustments. When the system enters the holdover mode the output payload will experience a small phase shift which is according to the hysteresis of the pointer processor. A minimum value of this hysteresis is specified in CCITT Rec G.783.

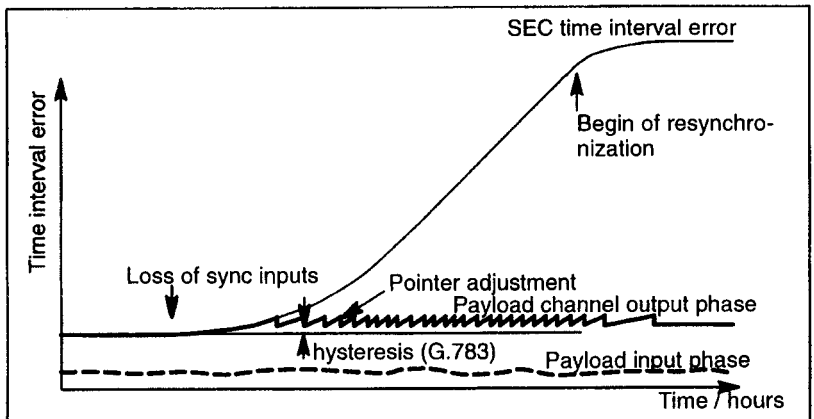


Figure6: Payload output phase characteristic for an SDH network element in holdover

4.3 Pointer processing performance and payload wander effects

In a connection through an SDH network there are always three types of SDH equipment involved.

The payload enters the SDH network at the "mapping node", the equipment at which the payload is mapped on the SDH payload entities such as virtual containers (VC). In the mapping process the bit rates are adapted by stuffing bits. The VC is positioned into the STM–N frame by setting the appropriate pointer value.

The payload passes some "intermediate nodes" at which it is routed through the network by routing the VC the payload is mapped to. At this node the pointer values are adapted as described above.

The payload gets out of the SDH network at the "demapping node". At this node pointer processing may be performed similar to the "intermediate node". At the

output port of this node the stuffing bits are removed from the VC and the remaining bits, i. e. the payload, are buffered and retimed by the recovered payload clock.

When an "intermediate node" goes into the holdover mode it will generate pointer adjustments. However, these will be absorbed by the hysteresis of the subsequent node, which is assumed to be synchronized. There will be no effect on the output phase of the payload.

When a "demapping node" goes into the holdover mode it will generate pointer adjustments. The associated phase instability is transferred to the retiming buffer and filtered by the payload clock recovery. The clock filter has to attenuate this phase noise in order to get tolerable output jitter values. This is achieved by a suitable clock design.

When a "mapping node" goes into the holdover mode the virtual container will be generated with the same frequency offset as the STM-N envelope. This causes the gradual filling of the pointer processor buffer of the next intermediate node which is assumed to be synchronized. When the threshold is reached, pointer adjustment are performed. These cause the filling of the pointer processor buffer of the next node, and so on. The filling of all these buffers in the intermediate nodes causes the output payload to perform a phase ramp, which is the negative image of the phase ramp at the mapping node, which is in holdover. When the mapping node gets synchronized again, the clock phase noise will cause sporadic pointer adjustments at the intermediate nodes, which tend to move the intermediate buffer fill towards their mid value. This causes the payload phase offset, which was accumulated by the phase ramp effect, to decay during the subsequent synchronized period.

4.4 Resulting network slip performance

The maximum wander is generated in SDH when the mapping node gets into holdover. In the core network the mapping nodes are at the locations of the switches. So, if the node is in holdover, not only the SDH equipment will be in holdover, but also the switch. The phase ramp effect of SDH will even compensate part of the wander generated by the switch. The pointer adjustments are generating only a small ripple on the phase slope of the 2 Mbit/s signal. Therefore the wander generated by SDH will not contribute to the overall slip rate.

The access network is different. The mapping nodes may get input signals, which are generated at some place different from that node. So, the wander generated by SDH in holdover would be added. However, the structure of the access network does not allow for synchronization loss with data traffic still available (a cable failure causes the failing of both, synchronization and traffic). Therefore the wander generated by SDH in holdover will not add to the overall slip performance.

4.5 Synchronization reference signals

STM–N signals

The STM–N line signals are the preferred carriers for synchronization references.

2 Mbit/s signals

In some environments 2 Mbit/s signals are required as synchronization references, e. g. when there is no clock interface available for synchronization. If the 2 Mbit/s signal is being transported by SDH systems the phase step associated with a single pointer adjustment may cause trouble in some slave clocks.

However, synchronizing on 2 Mbit/s signals will be necessary mainly in the access network area. There the SDH network structures are such that the STM–N line signal, which is used for synchronization transport, and the payload container are transported on the same physical medium. In such cases no pointer adjustments can occur and the resulting payload phase stability is sufficient for synchronizing a slave clock.

If a 2 Mbit/s signal is first multiplexed by a PDH multiplexer before mapping on SDH, the pointer adjustments performed on such higher rate payloads adds less phase instability so that the transmitted and demultiplexed 2 Mbit/s signal has enough phase stability for use as a synchronization reference.

Finally, if a 2 Mbit/s synchronization reference signal is required and the transport by SDH cannot guarantee enough phase stability, the payload can be retimed with a stable clock derived from the SDH equipment.

5 Conclusion

The combination of a hierarchical synchronization architecture and a synchronization quality management using the synchronization status message, which is encoded in the section overhead of the STM–N signals, will provide improved network synchronization performance. Therefore, preferably the STM–N signals should be used as synchronization references. As a further advantage these signals are not subject to network configuration management. This leads to a stable, hierarchical synchronization network.

The pointer processor standard G.783 provides a payload phase stability performance which is compliant to the network requirements for the provision of a low slip rate on end–to–end connections at 64 kbit/s.

For specific applications and for network evolution from PDH to SDH 2 Mbit/s signals, which have been transported by SDH, can also be used for synchronization if suitable network configurations or the retiming function are applied.

Filtering and Frequency Control for the next Generations of Mobile Communication Systems.

J. Détaint, J.Schwartzel
*FRANCE TELECOM CNET/PAB/Laboratoire de Bagneux
196, Av. H.Ravera BP 107, 92225 Bagneux France.

Abstract: A very large development of the mobile radiocommunications is expected to occur in the 15 next years. In the introduction, the most important among the present and future systems are listed and their main characteristics are briefly indicated. The case of the Universal Mobile Telecommunication System (UMTS) presently studied in the European community to enter service in the beginning of the 2000's is particularly discussed. The equipments of the future numerical radiocommunication systems will include a large number of piezo-electric devices for filtering and frequency generation. New characteristics (higher frequencies, larger bandwidth, very reduced dimensions, etc...) will be required for these devices so that new developments will be necessary to find solutions to these requirements. It is believed, that the corresponding evolutions will be very significant of the main trends in piezo-electricity for the next ten years.

This paper review the UHF and VHF filtering technologies and the frequency generation techniques the most likely to be used in the future radiomobile systems. For each kind of devices (filters for either UHF or an intermediate frequency and frequency sources), we discuss the most probable requirements, the present state of the different technologies, the new advances required to develop highly miniaturized devices for the mobile equipments or high performances sub-assembly for the base stations. In most cases, it appears that the availability of new piezo-electric materials and the obtainment of better characteristics for the existing materials will be essential factors in the achievement of the required characteristics for these devices. The possibilities of integrating piezo-electric resonators or filters with the micro-electronics contributing to the same or to related functions is also discussed together with the interest of developing new thin film piezo-electric materials permitting this integration.

1.INTRODUCTION:

In the last 10 years, the market of voice mobile communications for professional or private use was principally covered by the analogy cellular systems serving essentially car radiotelephones. The need to use more efficiently the scarce frequency resource, those of an increasing need of communication from an increasing number of persons has led, in the recent years, to the definition and the implementation of new cellular systems using advanced access techniques, digitized voice transmission, numerical modulation and error correcting coding. In western Europe, this has led to the normalisation of an unique system, the GSM (Global System for Mobile communication) which has enter service in 1992. To increase further the spectral efficiency while responding to the new demand of personal miniaturized handheld terminals following the subscriber in all his displacements, the concept of micro-cellular systems was introduced. In Europe, the Digital Cellular System at 1800MHz (DCS1800) using essentially the same technical solution as GSM, was normalized; it will enter service in the next years. In the same time, the need of specialized systems for communications directly between airplanes and earth stations or between airplanes or ships

and earth via satellites has appeared, and several are being implemented (TFTS, AMSS). The home cordless telephones are also evolving towards digital technology (CT2-CAI, DECT) while converging with other private and public personal short distance radiocommunication needs. Public point to point systems (Pointel, BiBop in France), are addressing the market of pedestrian communications in the centre of cities.

The study of a future system that will cover all the aspects of mobile and personal communications at the beginning of the next century was initiated by the European Community four years ago. This system named Universal Mobile Telecommunication System (UTMS), will include nearly all the services designed in the previous systems, plus a variety of new ones including broadband services. On figure 1 is given a representation of the coverage of the UMTS due to A.Charbonnier et al. /1/.

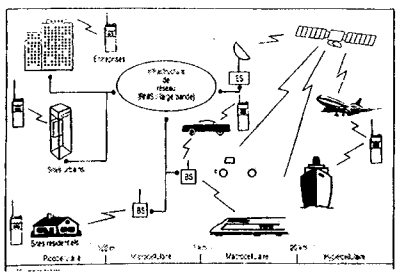


Figure 1: Coverage of the Universal Mobile Telecommunication System /1/.

The personal terminals for this system will have to be extremely miniaturized (pocket size) and low cost. The existence of a wide range of types of terminals is expected. A large number of "stations" (interfaces between the "air" and the fixed networks), will also have to meet similar requirements for the size and the cost (particularly for pico or nano cells etc...). The studies made in the RACE projects have identified two possible implementations of the radio-interface differing principally by the access technique (TDMA or CDMA) that are being evaluated.

In both cases, hypothesis concerning the design parameters of the radio-interface (baud rates, type of modulation, voice coding etc) were made. The consequences of the probable choices on the requirements for the filters and the frequency generation devices will be discussed in the next paragraphs.

The Word Administrative Radio Conference has in 1992 allocated frequency bands to the Future Public Land Mobile Telecommunication System; the UMTS is presently to be considered as the European project of implementation of the FPLMTS. In table 1 and figure 2, the main characteristics and the frequency allocations of the recent and future systems are indicated, more details can be found in references /1/&/2/.

Table 1: Principal recent or future public radiocommunication systems.

GSM (GLOBAL SYSTEM for MOBILE communication) 890-915 & 935-960MHz TDMA/FDMA, GMSK, channel 200kHz=8*22.8kbit/s	European 1992
DCS 1800 (Digital Cordless System at 1800 MHz) 1710-1785 & 1805-1880MHz same parameters as GSM	European 1994?
DECT (Digital European Cordless Telecommunications) 1880-1900 TDMA/FDMA, GMSK, channel 1728kHz=12*96kbit/s	European 1993
UMTS (Universal Mobile Telecommunication System) 1885-2025 & 2110-2200 TDMA or CDMA Under study (European Community RACE projects)	European projects opening 2002 ?

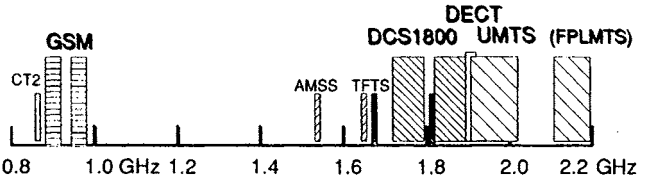


Figure 2: Frequency allocations for present and future digital radiocommunication systems.

2./ Filtering and frequency generation in the Radio sub-system of radiocommunication equipments.

The typical architecture of the radio sub-system of present mobiles or base stations is represented on figure 3. We have chosen to represent the most usual options encountered in recent equipments (direct modulation and two intermediate frequencies in the receiver), other options are possible principally on these two later points. The antenna is connected to filters constituting most often a diplexer (otherwise the filters must be switched alternatively). These UHF filters select the two sub-bands corresponding to the base to mobile and to the mobile to base traffic.

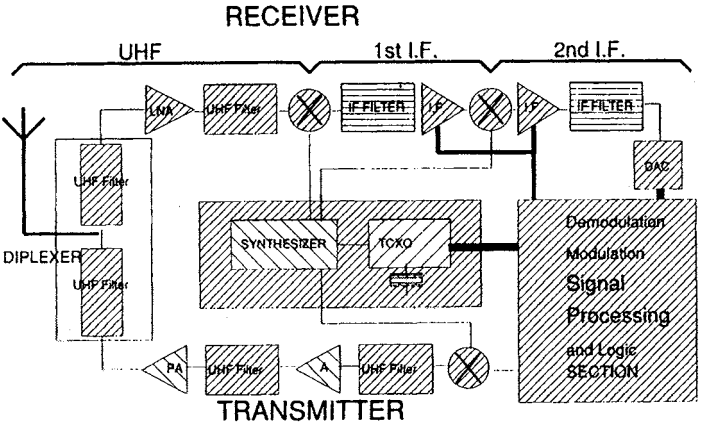


Figure 3: Typical architecture of the radio sub-system.

The receiver section includes then a low noise amplifier making a first amplification of the signal while minimizing the noise figure of the reception chain. Generally a post amplification filter is used after the LNA to complement the UHF selectivity and to remove the maximum of out-of band intermodulation products before the frequency transposition. After this one, an intermediate frequency filter selects the useful frequency "channel" which is shared by several mobiles (TDMA or CDMA). Eventually a second intermediate frequency with filtering is used to further complement the selectivity of the first one, and to limit the spectrum of the signal before the analog to digital conversion. In the figure, the case of a direct 2nd IF to digital conversion is considered. The reference frequency for the synthesizer is generated by a crystal oscillator which is usually of the VCXO type in the mobiles and a high stability OCXO in the base stations. The transmitter section includes a post modulation UHF filter and generally a second filter after

the amplification preceding the power amplifier. In many instances, the UHF voltage controlled oscillators included in the synthesizer make use of resonators of the same type as those existing in the filters, their high Q factors allows to obtain interesting phase noise characteristics.

No fundamental modifications of the basic principles of the UHF and IF architecture of the equipments are very likely to occur for the next generations. But, the use of techniques, such as space diversity or variable baud rate transmissions, that may increase notably the complexity of the radio subsystem by duplicating several functions, is very probable. Most probably, the radio sub-assembly will have to display a high flexibility (several configurations digitally chosen). For the filters, this can mean variable centre frequency or variable bandwidth numerically allocated.

As far as the market will grow, it will become more and more important to obtain enhanced performances (noise factor, intermodulation, phase noise etc..) of both the receiver and transmitter, since the possible number of subscribers within the frequency bands available is also dependant on several of these performances. The need to offer very small (pocket size) terminals having a much increased complexity will lead to research a important reduction of the volume of the different functions. This, in turn, will require important evolutions of the characteristics and technologies of the components.

All the active components of the radio sub-system will, most probably, be integrated in a very reduced number of integrated circuits (I.Cs). The characteristics of the filters and of the frequency generation devices are important factors in the overall performances; their number will most probably increase greatly in UMTS equipments. As a consequence of the integration of the electronics, they will become a major factor of the dimension of the radio sub-assembly. So that the challenge will be to improve their performances (insertion losses, selectivity, phase noise,...) while decreasing drastically their dimensions. Moreover, due to the use of higher ultra high frequencies and to the increase of the baud rates, the characteristics required for the filters and the frequency sources, will often be, quite different of those of existing devices, so that new solutions are to be found.

One can expect that the evolutions of the components required for the next generation of systems will also have noticeable impacts on those required for the future generations of equipments of the existing systems. In the next paragraph, we will review the most probable solutions and evolutions for the filters and the frequency generation devices.

3. UHF FILTERS:

The main functions of the UHF filters is to select the frequency bands in which the equipment can transmit or receive while attenuating the unwanted signals situated out of these bands.

3.1 Antenna filters, diplexers for the UMTS.

When transmission and reception occur on separate frequency bands the two UHF antenna filters are most often connected as a diplexer, even if there is no simultaneous 2 ways transmission. This result of the fact that fast T/R switching is either difficult, costly or degrade somehow the performances, whereas using a diplexer does not add practically any additional complexity and cost. To constitute a diplexer, the filters must meet several requirements concerning essentially the input impedance, and the attenuation in the pass band of the other filter. For the UMTS, the T-R frequency separation could be of the order of 200MHz (or lower), while the bandwidth required could be of the order of 100MHz, so that a T/R rejection of above 80dB will be achieved using two 4 or 5 pole filters, this imply that the ultimate out of band rejection has to be maintained below this quantity, which is a quite demanding performance at 2GHz. Reduced insertion losses in the pass-band will be searched for, since this quantity directly impacts the noise factor of the receiver and energy saving in the transmitter. A value near one dB seems to be a realistic objective. As mentioned previously, the size of the filters has to be very reduced from the present state of the art (#0.4cm3/pole). These filters have to transmit or, to be submitted, to a noticeable power (depending of the type of equipment).

3.2 Other UHF filters.

The specifications for the other filters will most probably be quite similar, with probably somehow less stringent ultimate out-of-band rejections (60dB?), and insertion losses (2dB?). Again, very small

dimensions will be required, but due the functions of this filters (see diagram of figure 3), the best solution can be to find a mean to co-integrate them with the corresponding electronics on silicon and gallium arsenide I.Cs.

3.3 UHF Filters Technologies.

Several types of filters exist for applications near two GHz, few can present simultaneously reduced insertion losses and small dimensions. Many works were, in the recent years, devoted to improve the performances of the 800-900 MHz UHF filters for the first generation of digital radiotelephones (GSM in Europe). The development of filters, presenting, at 2 GHz, better characteristics than the preceding ones, will require, at least, the same efforts.

A tentative comparison of the different technologies is made on figure 4, in the case of filters centered at 2 GHz, having a bandwidth equal to 100 MHz, and a response equivalent to that of a 4 poles resonator filter. Four technologies allows reduced dimensions and insertion losses: **the piezo-electric thin film resonators (TFR) filters, the surface acoustic waves (SAW) filters, the magnetostatic waves (MSW) filters and the coaxial dielectric resonator (DR) filters** using quarter (or half wave) resonators. In figure 4, we have indicated the present state of the art for these filters and our assumptions about what could be their characteristics at the service opening of the UMTS. Due to the very reduced insertion losses provided, the technology, using "full size" dielectric resonators enclosed in metal cavities, can also be of interest for some applications in base station equipments.

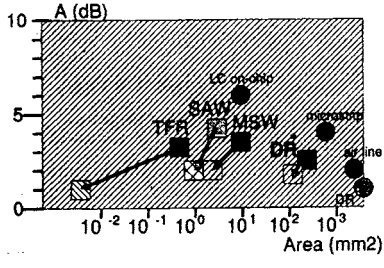


Figure 4: UHF filters at 2GHz

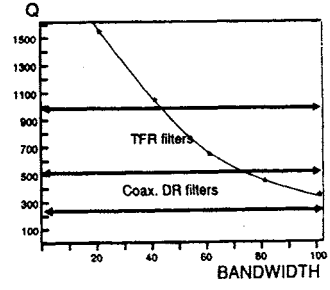


Figure 5: Q factor required for a 2 dB insertion loss (4 poles filter at 2 GHz)

An important remark to be made about figure 4, is the fact that, except for the latter one, ***no type of filters approaches now, even from far, the desirable insertion losses.***

Now, the most common UHF filters, for mobile applications (800-900MHz), are the coaxial dielectric resonators filters. The present values of the Q factors of the resonators (300-400 for the quarter wave ones) lead to insertion losses above 2 dB for a bandwidth of 100MHz at 2GHz. This limitation result much more from the geometry of the resonators and from the dissipation in the silver metallizations than from the dielectric material used. The use of larger resonators (quarter or half wave) and the advent of room temperature superconductors are among the possible solutions to improve their quality. Much higher Q factors (800) were observed at 1.8 GHz for the thin film resonators. On figure 5, we have represented, as a function of the bandwidth, the Q factors necessary for the resonators to obtain a 2dB insertion loss (4 poles filters at 2GHZ with no losses in the other elements of the filters).

3.4 Thin film resonator filters.

As indicated in figure 4, this promising technique permits to obtain extremely reduced dimensions for the filters. The filter are constituted of half wave resonators made with a piezo-electric thin film electroded with two metal films, and of coupling elements (inductance and/or capacitance). The piezo-electric thin films can be either mono-crystalline ones or constituted of textured highly oriented films. Interesting insertion losses (1.1dB at 1GHz and 3.5 at 1.8 GHz /3/ were already demonstrated together with important out of band rejections.

The resonators are obtained using various thin film deposition methods, on different substrates; the feasibility of their integration with silicon or gallium arsenide circuit has already been demonstrated /4//5/. It was also demonstrated that they can withstands power levels in excess to 0.4 Watt. This value can be increased either by the choice of the material or by using different filter architectures.

The resonators have extremely reduced lateral dimensions (of the order of 5x5 to 20x20µm at 2 GHz)/6/, their thickness is governed by the velocity of the acoustic waves, and is of the order of 1 to 2 microns at this frequency. The material mostly used now in such filters are zinc oxide films or aluminium nitride films. The most important properties of these materials are listed in table 2, where we have also indicated the properties of the most promising new thin film materials now being studied. In this table we can observe that the ZnO and AlN have medium coupling coefficients. This leads, for a 5% fractional bandwidth, (100MHz at 2GHz), to the need to use inductances in the filter to compensate the static capacitance of the resonators. This, in turn contribute to have much greater insertion losses and surface, than what could be obtained without these inductances. With materials having a much higher coupling coefficient, these inductances can be suppressed. The foreseen improvements of the characteristics of thin film resonators filters, from the present state of the art, indicated on figure 4, takes this fact into account.

Table 2: Properties of materials for applications to thin films resonators filters.

Material		velocity	k %	tfc ppm	losses
ZnO	L	6370	28	-59	high
	S	2830	28	<<0	
AlN	L	11400	25	-26	low
	S	6330	21	-26	
Li2B4O7	L	#7000	25	#0 ?	low
	S	#4000	25		
ferro-electrics	L	- -	25-60	#0 ?	low to very low
LiTaO3 X cut *	S	4140	45	#0	very low
LGS and Analogs		- -	15-30	#0 ?	very low

* properties for bulk TaLio3

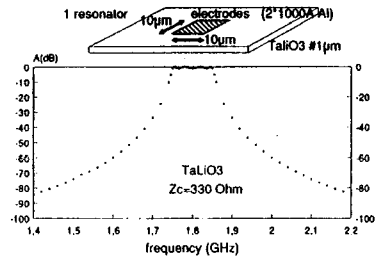


Figure 6: Computed response of a thin TFR filter having a bandwidth of 100MHz.

Other major advantages of the new materials are relative to their much better temperature coefficients (10.pmm/°C is probably the maximal acceptable value), and to their much reduced acoustic losses. Most of them can be deposited using a new process, (the sol-gel technique /7/) giving highly oriented films and presenting many advantages such as a better compatibility with I.C. integration, and a lower cost than when using vacuum deposition methods. To find the possibilities of the high coupling materials, we have computed the electrical responses of resonators made using a thin layer of lithium tantalate and aluminium electrodes. The computed resonance-antiresonance separations are very large (about 200MHz) for this type of resonators (more than 3 times the values observed for zinc oxide or aluminium nitride). On figure 7, we have represented the computed response of a four poles filter made using 4 of this thin film resonators. The equivalent scheme of the resonators were computed with the same model of composite resonators as that used to compute the response. We have used the data of table 2 and the dimensions

given in figure 8. The filter was synthetized to have a 100MHz bandwidth with a Jaumann structure, it was considered that the resonators could be adjusted to the required frequencies by variation of the thickness of one electrode.

On the whole, this filter technology appears to be extremely promising for the applications to UHF filtering in the mobile communication systems, particularly with the use of high coupling, thermally stable materials. Further research work is required concerning the materials, the modelization of thin film resonators, and the filter architectures best adapted to integration and to the different requirements encountered in the radio sub-system.

3.5 UHF SAW filters.

Important progress of the performances of low losses UHF surface acoustic waves filters were recently observed for 800-900MHz applications (A#2dB). They have mostly resulted of important research efforts made in Europe, USA and Japan, in the last five years. This effort has led to the adoption of new filter architectures (group type filters, new resonator filters, etc...) and to important progress in the conception methods /8//9/. At 900MHz, the use of SAW filters for post-modulation and pre-mixing filtering is becoming the usual solution. At UHF, the SAW filters are naturally small (a very few mm^2 at 900MHz), they can present interesting out of band rejections (40-50dB) that can probably be further improved. It was demonstrated that they can withstand an appreciable level of power (1 Watt) /9/. These filters are obtained using simple (1 mask level), batch processing techniques on large wafers of piezoelectric materials and hence are low cost devices. Lithium tantalate is the most commonly used material, it presents very low acoustic losses but presents a non negligible temperature coefficient.

Here again, there is presently no two GHz filters with sufficiently low losses, the best results being of the order of 3.5dB near 1.8GHz and of 6 dB at 2.5 GHz. Several points will be more critical at 2 GHz than at 900MHz in the development of low loss SAW filters; among them are the acoustic dissipation in the piezo-electric material and the ohmic losses in the electrodes. They will probably induce the choice of new solutions. For the materials, beside lithium tantalate, two recent ones, possessing zero temperature coefficient orientations, may be very interesting due either to a large acoustic velocity (lithium tetraborate (LBO))/10/ and/or to very reduced acoustic losses (lanthanum silico-gallate (LGS))/11/ [see table 3]. However, in this very important field, the advent of a low loss, very highly piezo-electric material, like lithium niobate, but being thermally compensated, remains very desirable. There is a reasonable hope that such a material can be found among the analogs of LGS. Here also, the advent of room temperature superconductors, may be very helpful for the electrodes.

A half micron technology will be required for the elaboration of the 2 GHz filters. Since the relative bandwidth will be much larger than those presently considered at 900 MHz, new studies concerning the filter architecture and design are also required to obtain filter with the appropriate width and losses.

4. INTERMEDIATE FREQUENCY FILTERS.

The channel width is 200kHz for the GSM and the DCS 1800; the values of the intermediate frequencies (I.F.) are generally situated in the range 70 to 175MHz for the GSM, they will probably be in the range 150 to 350MHz for the DCS 1800. The possible technologies for GSM and DCS 1800 IF filters are quite numerous, but due to the small value of the relative bandwidth, all use piezo-electric solution (either bulk wave or surface waves). these technologies and their domains of application are represented in figure 8.

For the UMTS, in a TDMA scheme, one could envisage DECT or GSM type channels (respectively of the order of 1000 and 200kHz) that could have to be implemented simultaneously at least in certain equipments. In a CDMA scheme, multiple values like 1, 5, 10, 20MHz (commutable in most cases) are envisaged.

Due to the frequency distance between the two sub-bands, the first intermediate frequency of the equipments will probably be situated in the range 190 to 350 MHz. These values are quite high for the smallest bandwidths (BW) cited. The existence of such very small value of the relative BW (of the order of 1 thousandth), will constrain to the use of materials having very low temperature coefficients. This results of the fact that the frequency shift of the response curve admitted under the effect of temperature have most often to be specified as a very small fraction of the corresponding bandwidth (to minimize the losses of the spectral resource). Equally, this small relative bandwidth imply the use of a low loss material to obtain a good amplitude response (insersion losses and shape at the corner of the pass-band).

To optimize the size and cost of the equipment, a very desirable feature, in the TDMA hypothesis, will be to obtain two commutable bandwidths from the same circuit in re-utilizing most of the components. In the case of the CDMA hypothesis, the same question appears to be more difficult since the same filter technology may not be optimal for so different relative bandwidths; in this case, still more innovative solutions are to be found.

For the amplitude response of the filters, low insertion losses are desirable (typically 4 to 8dB) to permit to have a gain in preceding stages (LNA and mixer) giving the best compromize between the noise factor and the level of interferences generated by intermodulation distorsion when high level unwanted signals are present. It is also very interesting to place nearly all the selectivity required for the channel filtering in the first I.F. filter since this situation is the most favourable to minimize the interferences produced by intermodulation in the I.F. amplifier by signals situated near the channel. So that, to increase the receiver performances in a system that is planed to have a much greater "density" of mobiles, it may be desirable to research filters more selective than those presently used in most of the first generation of digital systems. The possibility to increase largely the selectivity exists and is already employed in the stations and also in the mobiles of some systems. The phase response has to be kept very nearly linear (constant group delay) on the parts of the bandwidth containing most of the energy of the useful signal.

Two main technologies of filters can be used in this frequency range: the SAW filter and the Bulk Acoustic Wave (BAW) filters. Traditionally, the first one are considered as more convenient when large bandwidths and linear phase responses are required while the second are viewed as being most adapted for the cases of small relative bandwidth and when reduced insertion losses are searched for. These views are now to be largely reconsidered, since the evolutions of both techniques and the advent of new materials have largely extend their domains of applications (towards higher frequencies and larger bandwidth for BAW filters and towards the small relative bandwidth and the low insertion losses for the SAW filters). Moreover, for low losses I.F. filters, a convergence of the type of response obtained from both type of filter is presently observed. So that, for applications at the horizon of 2002, it will probably be wise to consider in depth all the constrains of the application and all the possibilities and potentialities of the two technologies. The mixing of both technologies in the same equipment is even of a probable occurrence in some hypothesis.

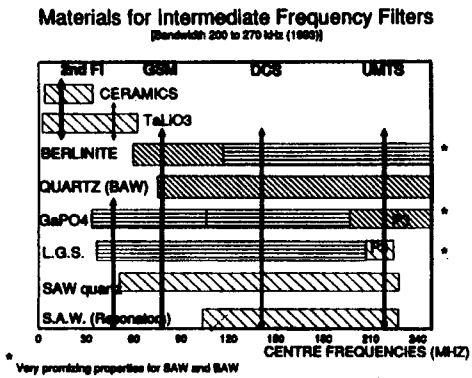


Figure 7: IF filters technologies for GSM, DCS 1800 and the smaller bandwidth of UMTS.

4.1 SAW filters.

The most usual solution for the mobiles and portables of the GSM is to use SAW devices for I.F. filtering. For this application many works were, in the five last years, devoted to enhance their characteristics (Shape factors, insertions losses, out of band rejections, dimensions), and several satisfactory compromises (with respects to the present specifications) were found. The researches are still going on and new devices will appear in the next generations of GSM equipments. As an example of this fact, figure 8a

and 8b indicate the principle and the frequency response of a filter using the new principle, recently found, of partial internal reflexions which constitute one of the most advanced solution found for the GSM SAW filters /12/.

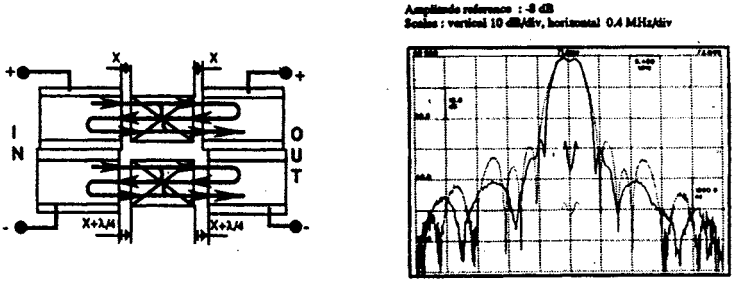


Figure 8: Principle (8a) and frequency response (8b) of a filter based upon a new principle.

On the same figure 8b, the response of an already very good but more conventional and larger low loss filter using single phase unidirectional transducers is represented (dashed line). Recently the trend to use filters of the resonator type or intermediate between conventional transversal filters and resonator filters has appeared which may lead to very interesting solutions at the highest intermediate frequencies. Nearly all SAW IF filters for the GSM are made using ST cut quartz.

For the UTMS, the different values of the relative BW, of the centre frequencies and mostly, the requirement of reducing greatly the dimensions of the filters while finding multi-bandwidth filters, will conduct to research further evolutions. The use of new materials may offer important new possibilities for the obtainment of better solutions. Particularly, for the larger bandwidth, it will be very advantageous to use a material presenting simultaneously a larger coupling coefficient than ST quartz and a good thermal behaviour. Moreover, to obtain smaller devices, it can be extremely useful to consider the material giving low acoustic velocity. In all cases, the high value of the centre frequency will favour the most thermally stable materials. On table 3, we have indicated the properties of the materials that can be interesting to obtain the different types of filters.

In this table, their is several new and much probably extremely interesting materials for which all the SAW propagation properties are presently not accurately determined. Although, it is possible to notice that the quartz like materials (quartz, $AlPO_4$, $GaPO_4$) have from far the best thermal behaviour.

Among them, berlinite ($AlPO_4$) appears to be extremely interesting for VHF SAW filters having relative bandwidth up to about 5% due to the large value of its coupling coefficient /13/ and to a smaller value of the acoustic velocity. The existence of crystalline orientations leading to better temperature coefficients than for quartz seems possible.

The SAW properties of $La_3Ga_5SiO_{14}$ (LGS), which is also a material belonging to the crystal class of quartz, are not presently available but the bulk acoustic wave properties /14/ indicate the high probability of large coupling coefficients, reduced velocities and a good thermal behaviour. Several analogs of LGS are presently being studied, it is already known that they will present much larger coupling coefficients /11/ and similar velocities than LGS. These materials may be very interesting for the larger bandwidth.

Table 3: Materials for SAW filters

MATERIAL	k^2 (%)	Relative shift for 100°C	Thermal Shift (kHz) at 225MHz	Max.BandW. at 225MHz
Quartz ** (ST cut)	0.14 dv/v 0.058*	90 10^{-6}	20	***
$AlPO_4$ (ST cut)	dv/v=0.245*	very small ?	< 20 ?	***
$GaPO_4$ (ST cut*)	dv/v=0.146	small 125 10^{-6}	28	***
$La_3Ga_5SiO_{14}$ & analogs	0.5? to much greater? **	quite small? 100-200 10^{-6}?	?	***
$Li_2B_4O_7$	>0.8 ?. **	.50-1.0 10^{-3}	112-225.	***
$TaliO_3$	1.3 -1.5 **	2.400 10^{-3}	540.	*** >15MHz
$LiNbO_3$ (coupe Y)	4,8 -5.57 **	7.2-9.4 10^{-3}	1900.	*** >25MHz

* from Wallnöfer et al. /13/, **Large wafers available

***max. bandwidth and insertion losses are not independent, they are also very dependant on the type of design used.

On the whole it appears that, in any hypothesis, further works are required either on filter architecture, filter design methods and on new materials. These latter are probably the key points to find much better solutions and, it seems important to determine accurately their constants and temperature coefficients, to compute also accurately their SAW properties and to further develop those presenting the most interesting properties for the applications. Researches about still more advanced solutions, for example, those permitting a co-integration with semiconductors by the obtainment of low losses films of high coupling, thermally stable piezoelectric materials on Si, GaAs, etc..., have probably to be started now.

4.3 BAW I.F. filters:

BAW filter were traditionally used for IF filtering in analogic radiocommunication equipments making use of FDMA, since this application require extremely selective filters at a quite low frequency (filters at 21.4 or 45 MHz having typically a 8 to 10 poles polynomial response giving very high shape factors and ultimate out of band rejection under 100dB).

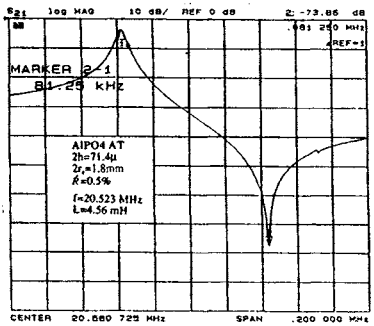


Figure 9: Response of an $AlPO_4$ resonator.

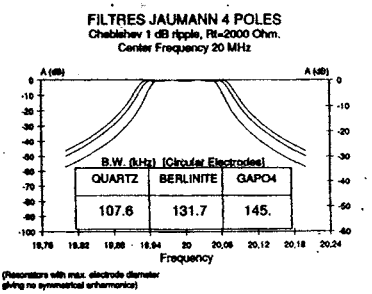
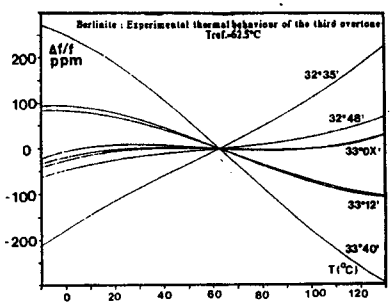


Figure 10: Four pole filters using the analogs of quartz.



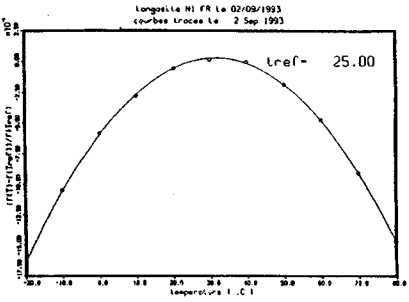


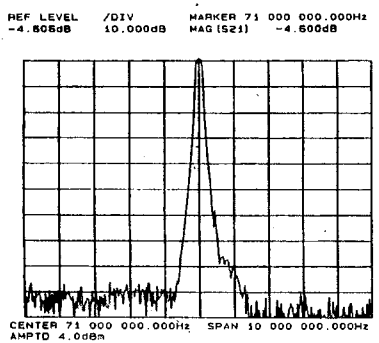
Figure 11: Thermal behaviour of $AlPO_4$

Figure 12: Thermal behaviour of LGS.

Technological progress (ion beam etching, direct to frequency plating, etc..) have, in the ten past years, extended the domain of applications of bulk wave filters to the very high frequencies and then to the ultra high frequencies /15/. More recently progress in synthesis have allowed to obtain phase responses compatible with digital transmission, very similar to what is presently obtained with low losses SAW filters. In the same time, the advent of several new materials have greatly increased the possibilities to obtain much larger bandwidth while keeping the advantage of a high thermal stability /14//16//17/.

The new possibilities bring by the high coupling analogs of quartz are illustrated in figure 9 which represents the electrical response of a AT $AlPO_4$ resonator (please note the much larger resonance anti-resonance distance than in quartz at the same frequency).

Another illustration is given in figure 10 where the responses achievable under certain hypothesis /18 / for 4 pole filters using quartz, $AlPO_4$ and $GaPO_4$ are compared. The thermal stability of two of the recent materials is illustrated on figure 12 (AT cut of berlinite ($AlPO_4$)) and on figure 11 (Y rotated compensated cut of LGS).



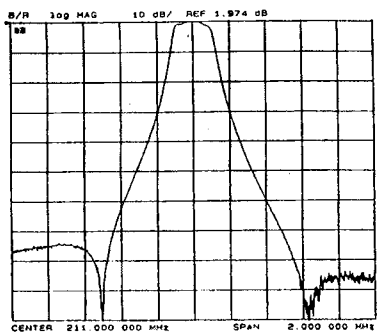


Figure 13: Response of an industrial 71MHz bulk wave filter /20/.

Figure 14: Response of an industrial 211MHz bulk wave filter /20/.

Several types of high performances BAW filters are presently used in the GSM base stations. Bulk wave filters are also used in the mobiles of several non European digital systems. The high performances possible with bulk wave filters are illustrated on figure 13 and 14 which represent the responses of industrial filters for GSM and DCS 1800 applications (Courtesy of Thomson-CEPE). We can notice the reduced insertion losses (4,6dB for the 71MHz filter of figure 13 and 3.55dB for the 211MHz filter of figure 14), the excellent shape factors, and the very large out of band rejections. A recent, and important progress in the field of VHF bulk wave devices, is the discovery of very efficient chemical methods permitting the obtainment at low cost of large quantities of VHF plates /19/. Such processes exist presently for quartz and berlinite, they may be extended to gallium phosphate and to LGS without many difficulties.

The interest of bulk wave filters for the next generations of radiocommunication equipments and particularly for the UTMS rely principally on the following facts:

-The VHF I.F. filters are very small (#.5 to 1 mm2/pole at 200MHz).
-Very selective responses can be obtained.
-BAW filters have extremely good thermal behaviours.
-Collective technologies are possible (see §4.2.1)

The largest relative bandwidth achievable with bulk wave filters using the temperature compensated material (lithium tantalate) having the highest coupling coefficient is of about 7.5%, but the same material cannot, now-a-day give very satisfactory filters for relative BW below about .5%. The most important properties of the presently known materials that can be of interest for the UMTS I.F. filters are listed in table 4.

Table 4: Properties of materials for BAW filter applications.

MATERIAL	k (%)	$\Delta f/f$ -25° to +75°C	Thermal Shift (kHz) at 225MHz	Max. practical bandW. at 225MHz*
Quartz** (AT cut)	8.	12.10^{-6}.	2,7	500.kHz
$AlPO_4$ (AT cut)	11	20.10^{-6}	4.5	1100.kHz
$GaPO_4$ (AT cut)	16	$<40.10^{-6}$	<9.0	2300.kHz ov.3 OK 200kHz
$La_3Ga_5SiO_{14}$** & analogs (**)	15 to 30 ?	100 ? to 150.10^{-6}	22.5?-33.75	1900.kHz ov.3 OK 200kHz
$Li_2B_4O_7$**	23	300 ? to 600.10^{-6}	too much for TDMA	4800.kHz ov.3 OK 200kHz
TaliO3** (X cut)	45	600.10^{-6}	too much for TDMA	>16.MHz

*The values indicated $4k^2/\pi^2$

are some how conservative to consider stray capacitances, see also /18/. **Large wafers available.

On table 4 we can notice again that the analogs of quartz present the most favourable temperature stability with a third order behaviour for quartz and berlinite, and most probably for gallium phosphate /21/. Again, the importance of the thermal behaviour that will play a very important role in the choice of material, is to be emphasized. This results, of the fact that there is a case where a small bandwidth will be used. This, in turn, will imply that the admissible frequency shift of the response of filters due to thermal effects has to be very small (see the begining of § 4).

Most of the materials listed in table 4 can be used to obtain filters with a bandwidth near 200kHz, but if the fundamental mode is to be used, the natural choice is quartz. If, as possible, economic reasons render the choice of the third overtone more interesting, then $GaPO_4$ and LGS are a good choice. Three materials appear to be very interesting for the case of a near 1000kHz BW ($AlPO_4$, $GaPO_4$, and LGS). In the case of commutable bandwidth (200/1000kHz), extra stray capacitances are to be taken into account so that the two materials with the largest coupling coefficients seem the best adapted ($GaPO_4$ et LGS). For bandwidth situated between 2 and 15 MHz the high coupling analogs of LGS and lithium tantalate seem to be the most interesting (CDMA hypothesis).

Several new developments are required to obtain BAW IF filters with optimized design, a part of

them must concern the design of the filters (geometry, sensitivity to dispersion, commutable bandwidth etc..) an other important part must be devoted to the development of materials (GaPO4, LGS and analogs). Other developments must also be made on the technological aspects.

4.4 New technologies for bulk wave filters.

Except for some rare cases concerning essentially miniaturized low frequency resonators, the bulk wave devices are presently elaborated individually on small plates. The very small size of VHF devices, the new possibilities offered by the recent polishing equipments that allow to obtain thin large wafer with a high degree of parallelism and the recent advances in chemical etching permit now to consider the obtainment of VHF BAW filters using batch processing techniques. This will constitute a very important new advance in the bulk wave technology and will lead to a large reduction of the cost of such devices. New developments will be required to set up such a technology concerning several process aspects, the design of the devices and the materials (large crystals are necessary to obtain wafers, a good crystalline quality is also required for chemical etching /22/). On figure 15, an example of a half of a filter made using the present advanced technology is represented. It concerns a small plate obtained by chemical etching containing 2 integrated resonators at 70MHz (half of a filter) /23/. At higher frequencies or with larger plates, the whole filter can be integrated on the same plate. Figure 16 is a sketch of what could be the tomorrow VHF filter technology. It may be observed that this technology remains quite simple and that wafers of 4" diameter already exist for several materials.

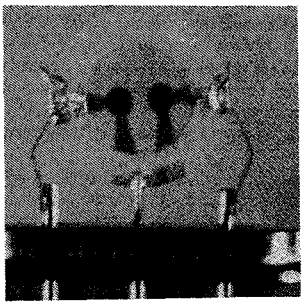


Figure 15: Filter made using the most advanced present technology.

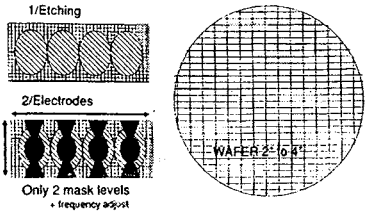


Figure 16: A sketch of what could be the tomorrow technology

Recently, an other advance was proposed that consists in bonding quartz chips upon silicon chips surface, this method can very probably be extended to other piezo-electric materials and to other semiconductors and may have an interest for UHF or IF filters.

5 FREQUENCY SOURCES.

The frequency control in the present mobiles for GSM (local synthesized oscillators and timing for the digital section) is presently made from a reference oscillator having a stability of the order of some 10^{-6}. This crystal oscillator is most often a TCXO using an analog compensation of the temperature variations.

The increase of the frequencies and almost of the bauds rates will much probably require the use of a much higher stability (particularly in the case of CDMA with large chips rates). Two choices can be made: the first is to obtain this stability in the mobile using a more stable frequency source. The other choice is to synchronise the mobile on the stations by some method. In this case however, due to the multipath propagation and to the Doppler effect, a non negligible time of observation and a lot of computations will be required (depending of the needed stability) to achieve the frequency stability. Moreover in

this case, the mobile oscillator has still to have an excellent short (or medium) term stability to interpolate between two observations. This frequency comparison can be made either in the mobiles or in the stations but requires always an exchange of data. In any cases, the phase noise of the reference oscillator of the mobiles has to be as reduced as possible.

The probable advent of low cost digital TCXO or of some other low cost method to increase greatly the frequency (and timing) stability in the mobiles may change in an important manner the conditions of this choice. However, in any cases there is a large interest to reduce the frequency aging of the resonators and to increase their Q factors (phase noise) by using an improved material (quartz).

The oscillators used in base stations are very high performances OCXO whose characteristics are mainly determined by the needs of the interface with the fixed network. They use quartz resonators with state of the art performances (aging, Q factors etc..), for which the quality of the material is determinant. New progress in the crystalline quality and purity of the material will be very interesting.

In nearly all cases the VCO of the frequency synthesizers use resonators (LC, DR,...). It will be very interesting to use high Q factor miniature resonators similar to that discussed previously (TFR, SAW) for the filters and, if possible, to integrate them to obtain very small circuits for this function while obtaining better phase noise characteristics (resulting of the better Q factor).

6. SUMMARY.

The performances of the radio interface of mobile systems equipments are very dependant on those of the filters and frequency sources. New characteristics and an important advance in the performances of the filters and oscillators will be required for the next generation of systems particularly for the UMTS. Due to the small volume required for the terminals, to a most probable large increase of their complexity, a reduction of the volume of the different functions of the radio-sub system is mandatory. For the electronics this will be achieved by a further integration. For the filters and frequency generation devices which will most probably be more numerous than in the equipment of the present digital systems, new solutions are to be found. Otherwise they will become a limiting factor in the miniaturization of the equipments.

In the review made of the different technologies for the different filters of the radio sub-system, it has appeared that several new technologies can give devices having simultaneously the required performances and very small dimensions. The case of the UHF thin film resonators filters which are co-integrable with semiconductors in mixed I.Cs is the most significative example. It was also observed that technologies commonly used at lower frequencies and for other bandwidth can be transposed for the new requirements, but with non negligible efforts in all cases.

In most case also, it was observed that these requirements concerning mainly the use of higher frequencies and larger bandwidth, will need the use of new piezo-electric materials having a larger coupling coefficient than quartz, a high thermal stability and very low high frequency acoustic losses. For the TFR filters the availability of thin piezo-electric films presenting similar properties is essential.

The discussion has revealed that most of the new piezo-electric materials presently studied ($AlPO_4$, $GaPO_4$, LGS and its analogs, $Li_2B_4O_7$, etc..) may be of interest for at least one application in the future radio-communication equipments. For the most recent materials, the developments must be pursued up to the obtainment of large crystals of high crystalline quality. Further developments in the perfection of quartz synthetic crystals, permitting to improve the long term stabilities of the oscillators and to enhance their phase noise properties are also very desirable.

The recent advances in the very high frequency BAW filters technology, give new openings to obtain very small devices with high performances. They give also, the very desirable possibility of further evolutions towards the low cost batch production methods already employed for SAW filters.

We believe that most of the trends observed for the devices destined to applications in the equipments of the next generation of mobile communication systems are very significative of the futures evolutions in piezo-electricity.

Acknowledgments: The authors wish to thank MM. N.Ruelle and M.Bon of CNET for helpful contributions in this study. They acknowledge fruitful discussions with many colleagues at CNET on the components for Radio-Mobile Systems. They thank MM. Arnaud, Buisson, Capelle, Cochet-Muchy, Marechal, Michel, Philippot, Solal and Zarka for their contributions to the material science aspects of this study.

References:

/1/ Charbonnier A., Hubert J., Attal J.,Cohen M., Evoi C.; Commutation & Transmission n° spécial "Communication avec les mobiles" 1993 p.109-121.

/2/ Margat C., Attal J., Damblin J.L., Desplanque J., Fino B., Lompech M.N.; Commutation & Transmission n° spécial "Communication avec les mobiles" 1993 p.7-20.

/3/ Anemogiannis K.,Berek S., Zottl H.; Proceedings of the 1990 IEEE Ultrasonics Symposium p.129-133

/4/ Burns S.G., R.J.Weber R.J., Braymen S.D., Proceedings of the 45th Annual Symposium on Frequency Control 1991 p.207-211.

/5/ Cushman D., Lau K.F. Garber E.M., Mai K.A., Oki A.K, Kobayashi K.W. Proceedings of the 1990 IEEE Ultrasonics Symposium p.519-524.

/6/ Vale C., Rosenbaum J., Horwitz S., Krishnaswamy S., Moore R.; Proceedings of the 44th Annual Symposium on Frequency Control 1991 p.332-336

/7/ J.M. Haussonne France Telecom CNET Lannion, Personal communication

/8/ Ruppel C.C.W, Dill R., Fisherauer A., Fisherauer G., Gawlik W., Müller F., Reindl L., Ruile W., Scholl G., Schrop I., Wagner K.Ch.; IEEE trans. Ultrason. Ferroel.Freq.Control 1993 V.40, n°5 p. 438-450.

/9/ Hikita M., Shibagaki N., Tabuchi T., Akagi T.; Electronics letters 1993 v.29 n°7 p.631-632.

/10/ Adachi M., Shiosaki T., Kobayashi H., Ohnishi O., Kawabata A.; Proceedings of the 1985 IEEE Ultrasonics Symposium p.228-232

/11/ Sakharov S.A., Larionov I.M., Medvedev A.V.; Proceedings of the 1992 IEEE Frequency Control Symposium June 1992 p.713-723.

/12/ Solal M., Hodé J.M.; Proceedings of the 7th. European Frequency and Time Forum Neuchâtel 16-18 march 1993 p.499-503

/13/W.Wallnöfer, J.Stadler, P.Krempl; Proc.7th. European Frequency and Time Forum Neuchatel 16-18 march 1993 p.653-657.

/14/ Silvestrova I.M., Pizaresky Yu.V., Bezdelkin V.V., Senyushenkov P.A.; Proceedings of the 1993 IEEE Frequency Control Symposium p.351-352

/15/ Brauge J., Fragneau M., Aubry J.P. Proceedings of the 39th Annual Symposium on Frequency Control 1985 p.504-513.

/16/ Detaint J., Carru H., Schwarzel J., Joly C., Capelle B., Zarka A.,Zheng Y., Proceedings of the 45th Annual Symposium on Frequency Control 1991 p.166-180.

/17/ Detaint J., Lançon R. Proceeding of the 30th Annual Frequency Control Symposium 1976 p.132-140

/18/ Detaint J., Schwartzel J., Joly J., Capelle B., Zarka B., Zheng Y., Toudic Y., Philippot E.; Proceedings of the 6th European Frequency and Time Forum Noordwijk 1992 p.223-227.

/19/ E.Philippot, A.Ibanez, A.Goiffon, B.Capelle, A.Zarka, J.Schwartzel, J.Detaint; Proceedings of the 6th European Frequency and Time Forum Noordwijk 1992 p.383-388.

/20/ Filters constructed by Thomson-CEPE Argenteuil France.

/21/ Deleuze A., Goiffon A. Ibanez A., Philippot E.; Proceedings of the 1993 IEEE Frequency Control Symposium Salt Lake City p.381-389

/22/ A.Zarka A., Capelle B., Zheng Y., Schwartzel J., Détaint J., Buisson X., Toudic Y. Proceedings of the 1993 IEEE Frequency Control Symposium p.696-706.

/23/ Détaint J., Schwartzel J., A.Zarka A., Capelle B., Joly C. Proceedings of the 7th. European Frequency and Time Forum Neuchâtel 16-18 march 1993 p.561-566

Compact Low Cost Low Ageing Crystal Oscillator for Telecommunication Applications and Production Statistics

by

Laurent Odiet and Kanak K. Tuladhar
Oscilloquartz S.A
2000 Neuchâtel, Switzerland

1) Introduction

This paper describes a new work carried out to realise a high performance compact oven-controlled-crystal oscillator having an uncompromised ageing characteristic, as per ITU-T recommendations, at a lower cost for the ever demanding telecommunication systems, such as PDH, SONET/SDH and satellite applications. To realise such an ambitious goal, a number of parameters have to be optimised. Of these, three most important parameters are :

1. a good choice of resonator has to be made from the host of available low ageing high-precision resonators, but without cost penalty in terms of manufacturability and yield.

2. an utilisation of optimum oscillator circuit, having a stable and optimised load-capacitor, which do not degrade ageing, but allows to have enough frequency pulling for the life time of the oscillator.

3. an utilisation of simple double oven but without Dewar flask, for reducing the size of the oscillator. Oven control circuit must be optimum and efficient for the good thermal behavior.

Detailled production statistics will be presented to show that :

a) Ageing better than 2E-11/day
b) Thermal behavior of 2E-10 from 10 to 60°C
c) Frequency pulling of 3.5E-7 from 0 to 8 V

have been achieved during routine production. Other parameters, such as Allan variance, Phase noise, oven power consumption will also be given.

Finally it will also be demonstrated that oscillators having similar characteristics can also be realised at a number of frequencies multiple and sub-multiple of 2.048 and 5MHz using 5.120 and 5 MHz oscillators respectively.

2) State-of-the-art crystal oscillator survey :

Table 1 makes a comparison between the capabilities of the 8600 BVA resonator based and conventional overtone mode AT or SC-cut resonator based oscillators, and we show there is a requirement in the middle. The main aim of this paper is to demonstrate that such an oscillator could be developed at a reduced cost and having smaller volume relative to the conventional BVA-oscillator.

3) Resonator selection :

Table 2 makes the comparison between three types of crystals resonators available on the market. This table clearly demonstrates that the conventional overtone mode AT & SC cut resonators will have difficulties in meeting ageing and frequency pulling requirements

simultaneously. In terms of low cost requirement, BVA 5th overtone AT-cut resonator is our preferred choice.

4) Oscillator block diagram

Figure 1 is the block diagram of the new oscillator.
The theoretical study of frequency deviation around the turn-over temperature of the BVA resonator shows that a temperature deviation better than 0.1 degree over the ambiant temperature range is required. Therefore, only a double oven will allow us to achieve the required thermal behavior. A thermal compensating element in the internal oven, cascaded with a serial regulator permit to generate a very stable internal voltage, which reduces parasitic pulling, indispensable for a good frequency stability. This type of architecture is well known by the oscillator designer, these consist :

- the oscillator cell, including the AGC in the feedback loop, allows to reduce number of components and to have good noise characteristic. The quartz resonator in the counter-reaction of the amplifier (emitter of a transistor) facilitates production adjustments. This method is not good for fundamental crystal, having a resistance value lower than 10 Ohms and a large C1 value, because this would not give stability required and detailed discussion can be found in the literature. The mode selector, with low Q factor to reduce parasitic influences in the loop, permits to be sure that the circuit oscillates at the right overtone mode of the quartz resonator.

- to minimize power losses and effects of hot points, the special heaters, with internal current transformer, are directly connected to the 12V supply. Without this precaution, the implication is a great sensitivity to voltage variations. Therefore, the oven control must measure the supply voltage to compensate its perturbations and assure a good dynamic regulation of the temperature.

A high-impedance input buffer amplifier separates the oscillator cell to the external environnement to minimize effects due to load variations. An optionnal multiplier permits to generate higher frequency. We also have included an oven alarm based on the temperature measurement of the two ovens. During the warm-up, because of the BVA thermal time constant, the frequency will reach the specification 1 hour after the alarm signal. It is also useful in normal function to detect a failure of the thermal regulation. The output cascode amplifier permits to adjust the output signal level to 0 dBm with 50 Ohms load.

5) Construction details:

The above considerations are not sufficient to achieve the goal, economical reasons are important. The optimisation process includes :
- the locations of the heater elements to minimize thermal gradient through the quartz.
- the optimised locations of the internal thermal measurement probes.
- the great thermal inertia and conductance of the internal oven to smooth thermal perturbations.
- the high thermal conductance of the external oven because of its large surface area and smaller relative heating section.
- the good insulation between the two ovens and the case. The best insulator is vacuum but for the low to moderate cost and smaller size one cannot allow Dewar flask as in the conventional BVA oscillator. Therefore we use the simpler and less expensive element : The air...

The figure 2 gives a good idea of assembly / construction details.

6) Measured Results and Production Statistics :

6.1) Thermal stability

During the developement phase, the first goal was to achieve the desired thermal behavior. The first prototypes have shown a tolerance better than 2E-10 over the required temperature range. Figure 3 represent a typical thermic behavior over 10 to 62 °C. Compiled maximum frequency deviations for the first 50 production pieces are given in figure 4. Somewhat larger dispersion achieved on 1st production batch is due to non

efficient turn over point adjustment and the gain of the thermal regulation. Having optimized these shortfalls, the results for the second fifty production pieces are given in figure 5. We see that the majority lay in 2.5E-10. Only a few pieces (5%) were outside.

6.2) Frequency pulling :
The pulling capability permits to adjust the frequency and two typical pulling curves having good linearity are given in Figure 6. Histogram of pulling sensitivity for 100 pieces is given in figure 7. This shows that majority have pulling sensitivity of 0.86E-7 / Volt. Figure 8 shows the dispersion of fractional frequency at center of the pulling voltage (i.e at 4V). The majority are in the range +/- 0.2E-8 / V (+/- 2.2 % of nominal value). The method of pulling adjustment is sufficient for production.

6.3) Ageing characteristic :
After the pulling range adjustment, the typical ageing characteristic is measured and a typical curve is shown in figure 9. The measurements were taken every 6 hours and we can also see the fluctuations due to environnement perturbation (supply variations, ambiant temperature) . A linear FIT is made for the last 20 points to determine the ageing at the moment of delivery. This shows ageing rate $\leq$ 1.5E-11 / day. Figure 10 shows the dispersion of ageing (on 50 pieces) after 7 days of measurement. 80 % of the pieces are better than 5E-11/day, and 70% of the pieces are better than 2E-11 / day. After studying the typical curve, we can say that it is only a matter of time to achieve the goal of 2E-11/day . In fact 90% are well within specification after 3 weeks, rest come in to specification after 4 to 5 weeks. With an ageing of 5E-11/day coupled with a pulling of 3.5E-7, the 8696 can be slaved for 19 years... which is more than adequate for most applications.

6.4) Phase noise :
Although currently for telecommunication applications, the phase noise is not specified, but with the coming of SDH, phase-noise can be an important parameter. Typical measured phase noise plot is given in figure 11. Phase noise near the carrier, $1/f^3$, is primarily influenced by the Q factor of the quartz resonator, and the 1/f noise depends on the active components of the electronics, and the white noise is dependent on the drive level to crystal resonator. Based on these measurements, we can calculate (at the intersection of the two slopes $1/f^3$ and 1/f) a loaded Q factor, which is in fact greater than 500'000.

6.5) Allan variance:
Based on NIST Technical note 1337, the measured phase noise data has been used to compute Allan variance. The computed results are given in figure 12. We subsequently measured the Allan variance, and the measured results are given in figure 13. The discrepancy of about 2 times between the computed and measured results could be due to environnement effects, dead time zone, linearity of the phase detector etc...

6.6) Other parameters :
a) The spurs level is better than -70 dBc
b) The measured harmonic content is less than -35 dBc.
c) Measured frequency sensitivity versus power supply variations (+/- 10%) is around 2E-10 peak to peak.
d) The initial switch-on power is less than 11W during first 20 minutes. After that, the current decreases and the steady state value is around 380 mA (at 25 °C ambiant temperature), hence steady state power consumption is about 4.5W.

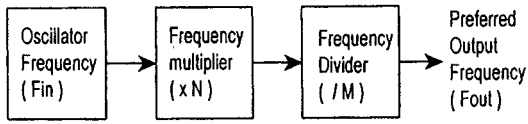
6.7) Time interval error in hold-over mode :
In telecommunication hierarchy, the frequency generated by a stratum level one (Cesium) is distributed to stratum level two, which distributes its frequency to level three and so on. The oscillators of level two are slaved in a digital PLL with primary frequency as reference. When the loop stops to receive the reference, a digital to analog converter maintain the voltage on the pulling input of the oscillator but the phase deviates as described by

figure 14. In this calculation, frequency deviation due to thermal effects has been included. In the G812 recommendation, which gives the characterisation of a level two frequency source, the ITU-T states that this effect is under discussion... But the TIE (for a time greater than 100 secondes) must be lower than 87.5 micro-secondes for 1 day. Therefore, the 8696 is very good for stratum level two frequency source as the computed TIE for the worst case is around 22µs

7) Other frequency generation :

7.1) Simple approach :

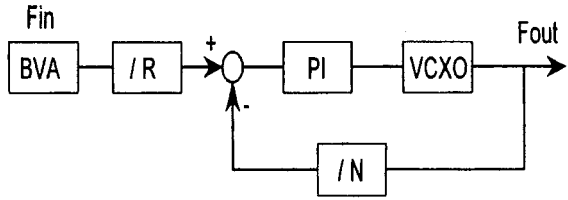
Frequency multipliers and dividers built in the oscillator extend the basic BVA frequencies of 5 and 5.12 MHz at the output of the oscillator to telecom frequencies :



One interesting value for Fout, 2.048 MHz, one of the preffered telecommunication frequencies is given below.

Fin	N (built in doubler)	M	Fout
5.12 MHz	2	5	2.048 MHz

7.2) General approach : (Proportional - Integral PLL)



8) Conclusion :

The results obtained allow us to say that :

- The oscillator having stringent requirements in terms of inproved specifications has been realised in a smaller volume and at a lower cost.
- Currently these oscillators are routinely produced in the production environnement.
- This oscillator is ideal for level two synchronisation network.

9) References :

1. Oscilloquartz Data sheets, 8600, 8665, 8696.
2. Norton J. , BVA-Type Quartz Oscillator for Spacecraft, Proceedings of the 45th Annual Symposium on Frequency Control, May 1991, pp 426 - 430.
3. Norton J, Performances of Ultrastable Quartz Oscillators using BVA resonators, These Proceedings (EFTF 1994).
4. Graf. E.P, and Peler U.R, BVA Quartz Crystal Resonator and Oscillator Production Statistics, Proceedings of the 37th Annual Symposium on Frequency Control, June 1983, pp 492 - 500.

TABLE 1 :

STATE OF THE ART CRYSTAL OSCILLATORS AT 5 MHz

	BVA	NEW (OUR AIM)	CONVENTIONAL AT OR SC CUT
AGEING [/ DAY]	0.2 to $2*10^{-11}$	$2*10^{-11}$	1 to $2*10^{-10}$
FREQUENCY PULLING [ppm]	± 0.12	± 0.3	± 0.3
THERMIC pk to pk (10 TO 60 °C)	2 to $5*10^{-10}$	$5*10^{-10}$	$20*10^{-10}$
PHASE NOISE [dBC / Hz] at 1 Hz offset 10 Hz offset	-130 to -120 -145 to -130	< -100 < -120	-95 -120
ALLAN VARIANCE 1 TO 10 s	1 to $5*10^{-13}$	$2*10^{-12}$	$5*10^{-11}$
SIZE [cc]	900 to 1900	300	65 to 140
WEIGHT [g]	850 to 1760	< 450	100 to 200
COST	MODERATE TO HIGH	MODERATE	MODERATE TO LOW

TABLE 2 :

RESONATOR SELECTION FOR NEW OSCILLATOR

	OVERTONE MODE		BVA	
	AT OR SC		5 TH OVERTONE AT	3 RD OVERTONE SC
AGEING	YES	NO	YES	
PULLING	NO	YES	YES	
FREQUENCY ADJUSTING TECHNIQUES	MATURED		MATURED	
VOLUME PRODUCTION	YES		YES	
COST	LOW		REASONABLE	HIGH
SIZE	ONLY TOTAL SIZE OF THE OSCILLATOR IS OF INTEREST			

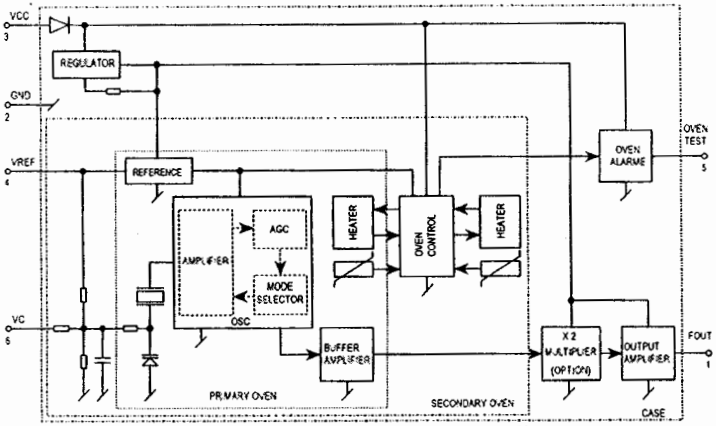


Fig 1: DETAILED DOUBLE OVEN OSCILLATOR BLOCK DIAGRAM

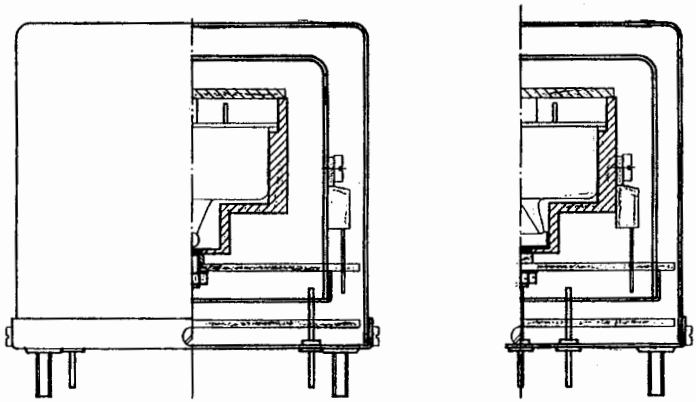


Fig 2: CROSS SECTION OF OSCILLATOR ASSEMBLY

Figure 3 :

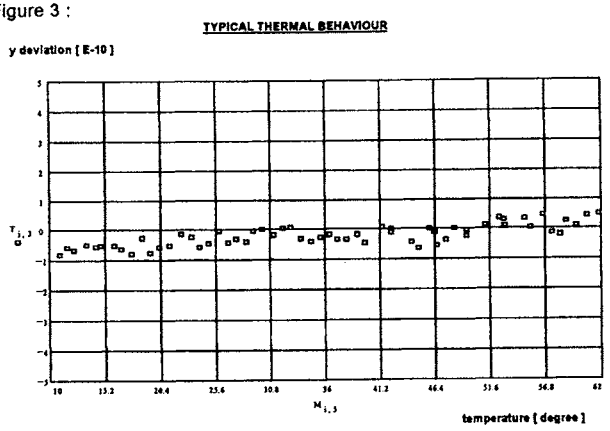


Figure 4 :

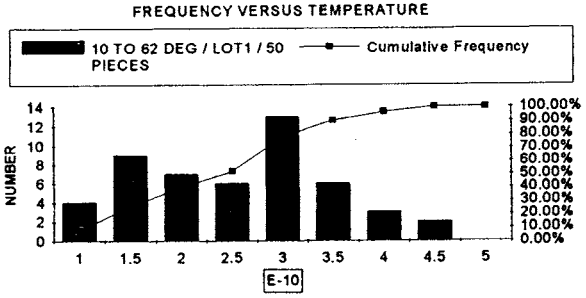


Figure 5 :

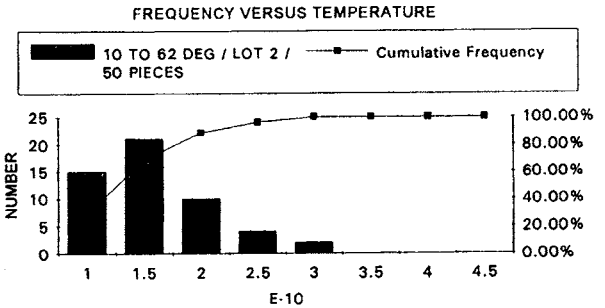


Figure 6 :

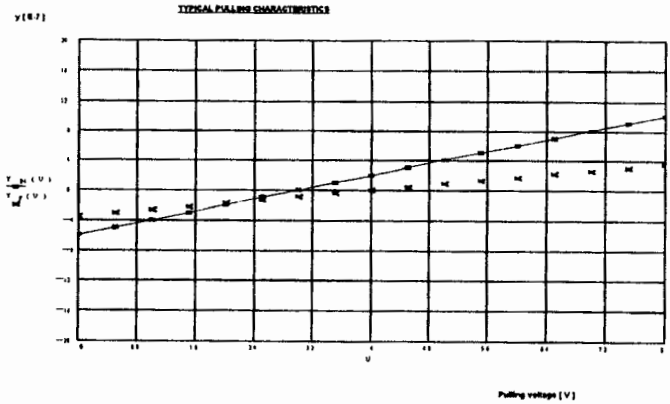


Figure 7 :

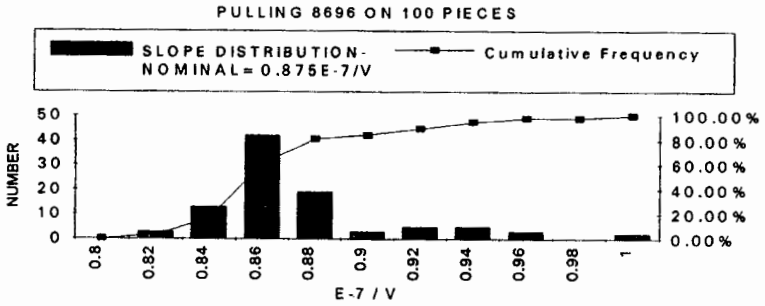


Figure 8 :

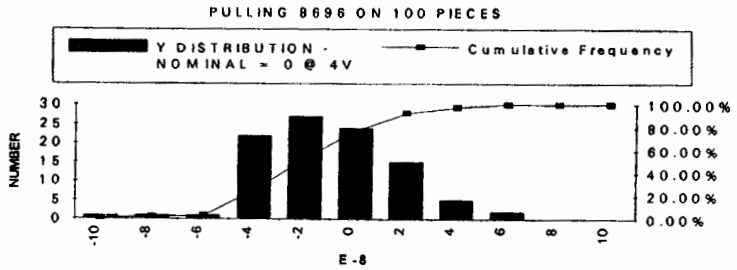


Figure 9 :

TYPICAL AGEING CHARACTERISTIC

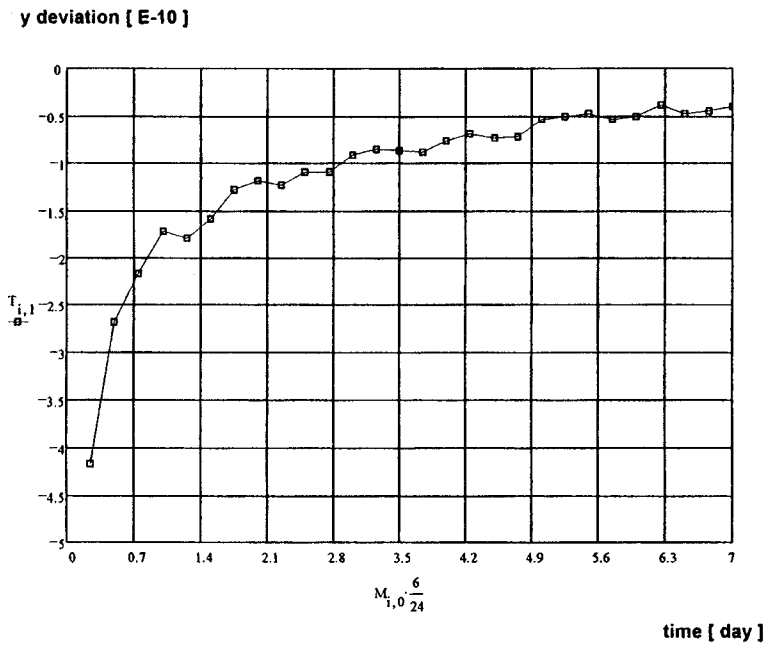


Figure 10 :

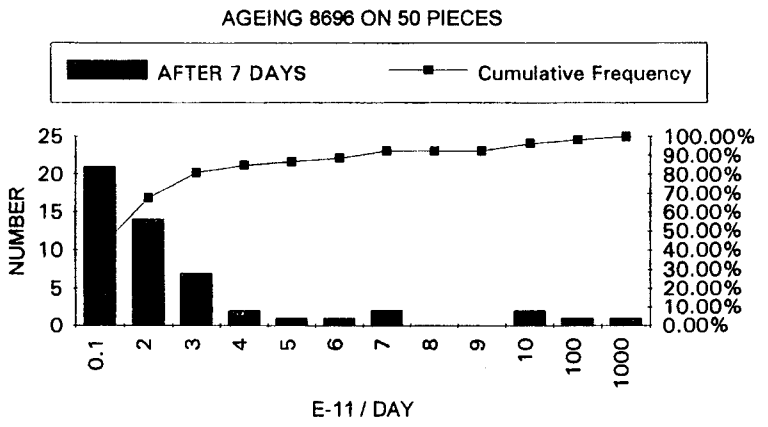
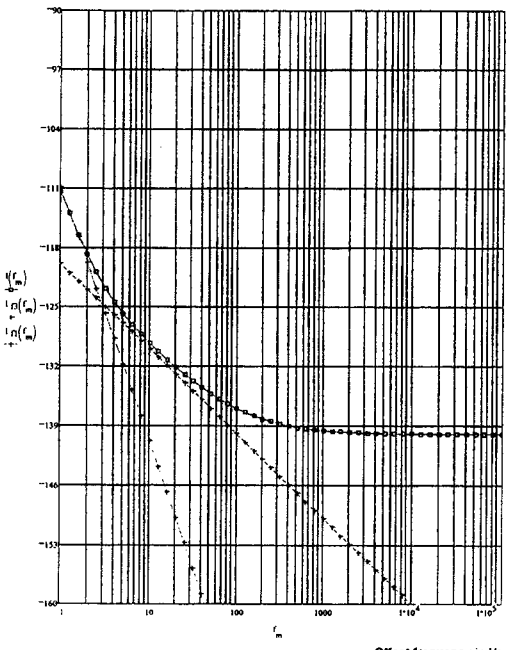


Figure 11 :

TYPICAL PHASE NOISE at 10MHz

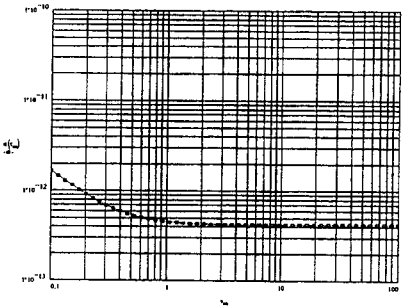
£ in dBc



Offset frequency in Hz

Figure 12 :

SQUARE ROOT OF ALLAN DEVIATION OF FRACTIONAL FREQUENCY FLUCTUATION :

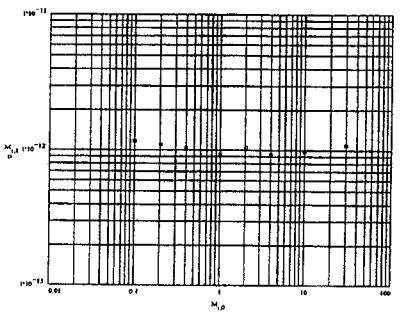


observation time [s]

Figure 13 :

MEASURED SIGMA - TAU

Sigma Tau



observation time [s]

Figure 14 :

TIE (Time Interval Error) Hold-over mode for oscillator 8696

Definitions :

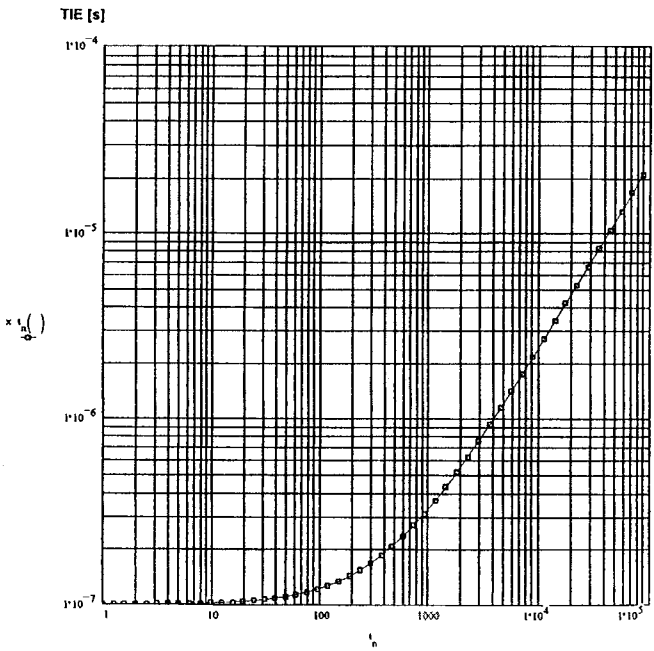
$y(t)$	fractional frequency error
$x(t)$	phase-time deviation (TIE)
x_0	synchronisation (Setting time, eg. UTC) error (1E-7)
y_θ	fractional frequency error due to temperature (2E-10)
σ_t	combined Allan variance deviation of reference frequency- and local oscillator frequency instabilities (1E-11).
y_0	initial fractional frequency error, syntonisation error, eg. DAC resolution. (2E-11)
D	linear frequency drift per second of local oscillator (ageing 2E-11 / day)

Basic Equations :

$$x(t) := x_0 + \int_0^t y(t)\,dt \qquad y(t) := y_0 + y(\theta) + D \cdot t + \sigma_t \qquad D := \frac{2 \cdot 10^{-11}}{86400}$$

$T := 86400$ observation time (elapsed time) = 1 day

$$x(t) := x_0 + \frac{D}{2} \cdot t^2 + y_0 \cdot t + y_\theta \cdot t + \sigma_t \cdot t$$



Materials (B 2)

NON-DESTRUCTIVE STUDY OF Al-hole CONCENTRATION AND FREQUENCY CHANGE OF QUARTZ RESONATORS (Part I)

K. YAMNI*, J.J. BOY**, M.A. PASQUALI**, P. ZECCHINI*

*LCCM, UFR Sciences - Route de Gray - 25030 BESANÇON Cedex - FRANCE
**LCEP / ENSMM - Route de Gray - 25030 BESANÇON Cedex - FRANCE

ABSTRACT

Ionizing radiations and sweeping are responsible to resonance frequency changes in quartz crystal oscillators. It is well-known that the main defects produced are Al-hole and Al-OH centers. They have an influence on the resonance frequency and on the different electrical parameters of quartz resonators (such as motional resistance, Q-factor, frequency temperature characteristic, ...) according to their respective concentrations in α-quartz.

In this study, we propose to follow the evolutions of these Al-hole and Al-OH centers after γ-irradiation. The added ESR (Electron Spin Resonance) measurements can permit us to determine the Al-hole concentration in different quartz material (synthetic and natural, as-grown or not) whose the impurity concentrations are measured by Inductively Coupled Plasma emission spectrometry.

In using crossed comparisons between results issued from Infra-Red, Induction Coupled Plasma emission, laser absorption at 632.8 nm and Electron Spin Resonance techniques, we intent to find the precursor defect and the way in which it is responsible of the frequency changes.

Finally, the aim is to remove these defects by technics as sweeping or pre-irradiation.

Keywords : Infra-Red, ICP and ESR spectrometries, Laser Absorption measurements, Natural and Synthetic Quartz crystal.

1. INTRODUCTION

Quartz crystal oscillators used as time-base for space applications are usually submitted to ionizing radiations. For instance, a communication satellite set in a geostationary orbit is submitted to the effects of the solar flare protons and the high-energy electrons captured by the geomagnetic field. So, during the ten last years, many authors have investigated the effect of these radiations on each component of the oscillator. Even if, sometimes, the oscillator output drops to zero for times ranging from several seconds to hours after exposure to pulse irradiations (1), ionizing radiations exposure on space oscillators (i. e. oscillators whichfor the components are especially studied for radiations exposures) can be translated by **transient and permanent shifts** in the resonance frequency due to quartz crystal resonator.

1.1. Frequency changes :

If a transient frequency shift can be explained by energy deposition or consequent temperature rise of the crystal resonator, the steady-state frequency offset is a result of **change in the elastic stiffness constants** of the quartz material due to creation of defects by radiations exposure (2).

Indeed, it is well-known that the main defects produced (as Al-OH and/or Al-hole centers - see below) have, at least, an influence on the resonance frequency and, consequently, on the different electrical parameters of the quartz crystal resonator.

So, while the oscillators were being tested for frequency shifts, studies of the radiation-induced defects in the parent quartz material can be carried out. For that, we have used the following techniques :

* **I.C.P. spectrometry** to quantify the main chemical "bulk" impurities,
* **Infra-Red spectrometry** to exhibit some absorption bands generally attributed to OH^- ions.
* **Laser absorption at 632.8 nm** to analyse chiefly the Al-hole centers.

* **Electron Spin Resonance** (so called **Electron Paramagnetic Resonance** : EPR) to qualify and to tentatively quantify the paramagnetic defects created by some treatments as sweeping or ionizing radiations exposures.

1.2. Main paramagnetic defects in α-quartz :

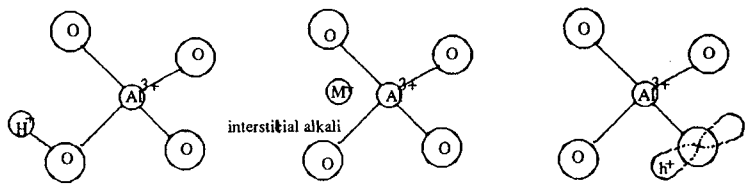
A considerable number of **impurities** in natural quartz specimens of different origins have already been reported. For instance, theirs relationships with the color centers created by natural irradiation have been summurized by HOLDEN (3). Later, FUTAGAMI (4) made some investigations on X rays-irradiated natural quartz and established the correlation with smoky color centers.

The primary techniques employed were chiefly the optical absorptions spectrometries and, since these earlier works, the smoky color has been associated with the **Al impurity content** of quartz.

The Aluminum hole centers :

* So, most of these defects are related to aluminum impurities, whichfor the most prominent paramagnetic center $[AlO_4]^0$, formed by X or γ-irradiation at room temperature (RT), has been explained as containing an aluminum 3^+ ion substituted to a silicon 4^+ ion, the **ionizing radiation having removed an electron from a neighboring oxygen** (5).

This hole center was first characterized by GRIFFITH et al. (6), and a consistent model (presented in Fig. 1) has been purposed by O'BRIEN (7). It is a **paramagnetic** defect associated with the smoky color of quartz crystal. Numerous *Electron Paramagnetic Resonance* spectra of this "Al-hole center" have been published (8).



OH- molecule

hole trapped in non-bonding oxygen p orbital

Fig. 1 : different defect centers diagrams from Halliburton (9)

* Generally, in as-grown non irradiated quartz, Al ion takes place of the Silicon tetravalent ion from the crysallographic lattice and its charge defect is compensated by an interstitial alcali or hydrogen ion. It can be represented by $[AlO_4/M^+]^0$ where M = Na, Li, K or H. These defects are **diamagnetic** and cannot be detected by EPR techniques. But irradiation at low temperature (below to 100 K) can create a hole (= removal of an electron from...) on the oxygen ion nearest neighbor to Al to give $[AlO_4/M^+]^+$. After warming, or increasing irradiation dose at sufficient high temperature, the **ion M^+ can diffuse away**, yielding $[AlO_4]^0$. During formation of this last hole center from $[AlO_4/M^+]^+$, the hole shifts from one of the "short-bonded" oxygens to one of the "long-bonded" oxygens.

This Aluminum in quartz is a source both of electrons and of cations (8).

* A **double-hole center** $[AlO_4]^+$ has also been reported by NUTTALL & al. (10).

The hydrogen hole centers :

NUTTALL and WEIL have also reported in (5) two aluminum-free species, called I and II, which appear similar to the single-hole centers $[AlO_4]^0$. In this analogy, for the species II, **three hydrogen ions** are thought to substitute for silicon (and so in place of the silicon ion). In species I, a **fourth hydrogen ion is present (acting as compensator M^+ before irradiation)**.

These two hydrogen trapped-hole centers can be represented by the following diagram (Fig. 2) :

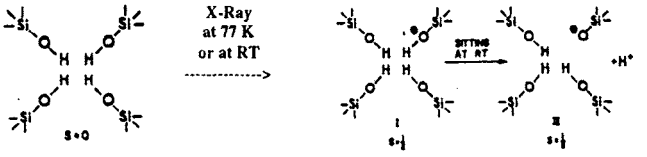


Fig. 2 : Two hydrogen trapped-hole centers

and the propose models are $[H_4O_4]^+$ for the first one and $[H_3O_4]^0$ for the second one ($I = 1/2$).

As for the previous hole center, in both I and II species, the hole is localized at a "short-bonded" oxygen. And finally, NUTTALL and WEIL suggest the existence of the diamagnetic precursor $[H_4O_4]^0$.

Other "hole centers" :

J.A. WEIL, in a review about *Electron Spin Spectrometry in crystalline quartz*, has pointed others **paramagnetic defects** associated to chemical impurities as Copper or Silver, Germanium, Phosphorus, Titanium and Iron. He notes that EPR of Fe^{3+} in α-quartz has recieved considerable attention because various varieties of quartz containing iron are known (amethyst, brown cultured quartz). Furthermore, Iron gets other oxydation states than Fe^{3+} and can be change valence under irradiation.

2. EXPERIMENTAL PROCEDURES

Our main techniques used are infrared and 632.8 nm laser absorption spectrometries. They have been employed to characterize changes in aluminum compensation in irradiated and swept cultured and natural quartz.

Though very difficult to read and to interpret, the experimental spectra issued from EPR techniques are the more complete to characterize the paramagnetic defects for qualitative and quantitative determinations. Finally, the 632.8 nm laser absorption measurement associated to EPR spectrometry can be used as indirect method to determine Al-hole concentrations.

2.1. Inductive Coupled Plasma spectrometry :

As usually, the concentration of trace elements are measured by ICP techniques in order to detect very low concentrations as in synthetic quartz crystal (about few ppb for alcali ions, for example). To determine usual chemical impurities (Al, Fe, Na, Li, K, ...), we use a PERKIN-ELMER 6500 spectrometer whichfor the emission rays released by the atom excited at very high temperature (between 5000 and 10000 K) are identified by computer. The concentrations are determined by comparison with a standard solution (12). The sample, the standard and the blank solutions have been prepared on the same conditions to improve the results. So, when it is possible, the given values correspond to the average of three values calculated with three emission rays. The detection limit is about 0.001 ppm.

2.2. Infrared spectrometry :

Infrared transmissions were measured with a NICOLET 750 Fourier spectrometer and the sample is usually cooled up to the Nitrogen liquid temperature (77 K).

The purpose of these observations is to verify the absorption peaks modifications after different treatments as irradiation or sweeping. As pointed by several authors (13) (14), as-grown OH impurities exhibit strong peaks at 3348, 3396, 3438 and 3581 cm^{-1} called s_1, s_2, s_3 and s_4. Generally, after irradiation occurs a reduction of these bands and others, related to Al-OH defects, are observed (e_1 and e_2 respectively at 3306 and 3366 cm^{-1}).

2.3. The 632.8 nm laser absorption :

This technique consists in measuring the 632.8 nm laser radiation transmitted through the two parallel polished faces of the quartz sample (11). The laser beam, emitted by a 2mW He-Ne source, is 0.8 mm width. The transmitted intensity is measured by a silicon photodiode detector and stored on the hard

disk of a HP9836 computer. The sample is translated in front of the laser beam with two micrometric motion units controlled by computer. The sample is analysed plane by plane as shown in Fig. 3. Usually, the step between two successive data points is 300 µm, and between two planes 2000 µm. The absorption values are calculated from the following relationship :

$$\sigma_L = \frac{1}{d} . \log_{10} \frac{I_0}{I_n}$$

where :

d is the thikness of the sample (generally 0.5 cm), I_0 and I_n are the measured intensities outside the sample and at the point number n.

The mean value of σ_L ($\overline{\sigma_L}$), for one plane, is given by :

$$\overline{\sigma_L} = \frac{1}{d} . \frac{1}{N} . \sum_{n=0}^{N} \log_{10} \frac{I_0}{I_n}$$

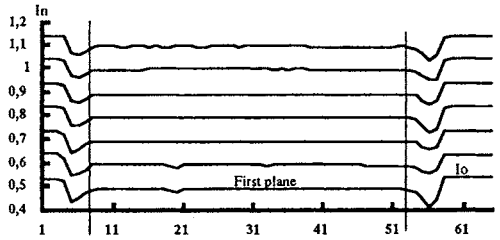


Fig. 3 : Example of laser absorption spectra (I_n versus Point number for 7 planes)

NB : For a easier reading of this 2D-representation, we have plotted, for each plane, In + p/10, p being the plane number (0 for the first one).

2.4. Electron Spin Resonance spectrometry :

This technique has been first employed to detect **paramagnetic centers of aluminum.** The EPR spectroscopy offers a most valuable tool for **quantitative** structural measurements (with very high resolutions) of the electronic and geometric configurations of paramagnetic centers existing or created in single crystals. It is applicable only when species having at least one unpaired electron are present.

In general, for non-cubic paramagnetic centers, the EPR line positions depend on the relative orientations of the externally applied magnetic field B (Zeeman). Our EPR experiments are done using a constant microwave frequency $\nu = 9.244$ GHz, and scanning the externally applied magnetic field around 3300 Gauss (or 330 mTesla). The very small sample (which is a cube whose the volume is less than 30 mm^3, its faces having been cut along the crystallographic axes) has been put in a cell cooled at the temperature of the Helium liquid (to almost of our experiments).

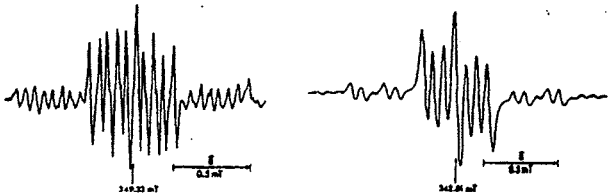


Fig. 4 : EPR spectra of $[H_4O_4]^+$ EPR spectra of $[H_3O_4]^0$ with B // c at 9.966 GHz (a) and 9.897 GHz (b) (from NUTTALL & al.)

A typical EPR spectra, such as the one shown in Fig. 4, exhibits hyperfine structure (HFS) which translate the relaxation of the spin of an unpaired electron between two different energy levels created by the presence of a magnetic field.

As we can see on the ESR spectra presented above, several hole centers are usually present in the same sample.

2.5. Sweeping and irradiation :

In LCEP laboratory, our sweeping process is operational from several years and the conditions are very classical (about 1 KV/cm for the voltage, 500 °C for the temperature). The current is monitored and the sample is put in a dry Nitrogen atmosphere.

As for the irradiation exposures, they are realised under a radioactive Cesium source whichfor the γ-ray energy is two times lower than this of a ^{60}Co source.

3. EXPERIMENTAL RESULTS AND DISCUSSION

Almost of the studied samples are Y-cut 5 mm thick plates. They come from synthetic or natural quartz which have been submitted to different treatments as sweeping and/or irradiation (see table 1).

SAMPLES		recieved	1st treat.	2d treat.	3rd treat.
A		No treat.	1,4 Mrad		
B	1	No treat.			
	2		Sweeping		
	3		200 Krad	400 Krad	1,4 Mrad
	4		1 Mrad		
C		Vac. Swept	1 Mrad		
D	1	Irradiated	Annealing		
	2		Id.	Sweeping	1 Mrad
E		Irradiated			

Table 1 : List of the analysed samples

The A sample is a synthetic quartz from SAWYER Company.

B and C samples come from the SICN company, only C sample having been swept under vaccum. α_{3500} are respectively 0.021 and 0.023 cm^{-1}.

The D sample is issued from Brazilian natural quartz .

At least, the E sample came from MOTOTROLA Company and has been irradiated with ^{60}Co source before any analysis.

The chemical analyses presented below have been performed on few grams of quartz crystal just before preparing the sample.

Samples	A				B	C	D
Elements	-X zone	+X zone	Z_g zone	Z_d zone			
B	0.08	0.065	0.025	0.001	0.10	< 0.01	0.15
Mn	0.93	0.06	0.07	0.08	0.01	< 0.01	< 0.01
Fe	1.19	0.21	0.71	0.20	1.49	0.36	1.27
Mg	0.22	0.13	0.32	0.24	0.06	0.06	0.19
Ti	0.13	0.005	0.26	0.27	< 0.01	< 0.01	0.45
Al	**2.51**	**2.92**	**1.93**	**1.57**	**1.95**	**0.72**	**37.9**
Ca	-	-	-	-	0.84	0.52	2.15
Na	2.203	3.36	0.001	0.001	1.37	0.33	1.40
Li	**0.29**	**0.61**	0.08	0.08	0.29	0.04	**1.62**
K	0.001	0.001	0.001	0.001	0.22	0.28	1.43

Table 2 : Trace elements

Before and after each treatment, we have performed infrared and absorption laser measurements to follow the evolution of the OH^- bands and the eventual apparitions of the trapped-holes. Furthermore, we have made some EPR spectra to confirm the creation of these paramagnetic defects.

3.1. Study of B samples :

Four identical Y-cut samples are extracted from a pure Z-zone of a synthetic quartz block. The two parallel faces are carefully grinded and polished to perform, first, a IR absorption coefficient measurement at 3500 cm^{-1}. We have obtained $\alpha = 0.0212\ cm^{-1}$, the quartz crystal belonging so to the A grade of the HQ quality.

The figure 5 shows IR spectra (performed at low temperature) for **B3 sample.** On these two spectra, obtained before and after γ-irradiation (up 1.4 Mrad), we can see the strong reduction of the s bands (related to free OH^- concentration) and the apparition of e_2, which is related to $Al\text{-}OH^-$ defects.

The two intermediate spectra are not presented here, but the figure 6 indicates the height of each absorption peak. Each height is calculated relatively to the absorption at 3300 cm^{-1} (relative to Si-O vibration and independant of any treatment).

Furthermore, we have to note that the sample stays clear after the last stage of irradiation, even though we see, in figure 6, a saturation effect (particularly on e_2) which can be observed from 400 Krad.

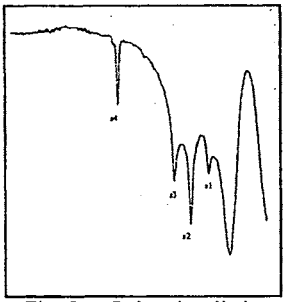


Fig. 5a : Before irradiation

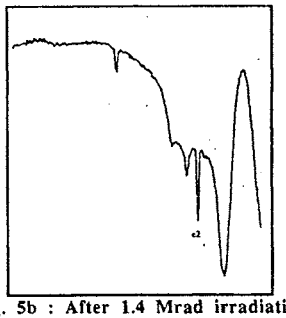


Fig. 5b : After 1.4 Mrad irradiation

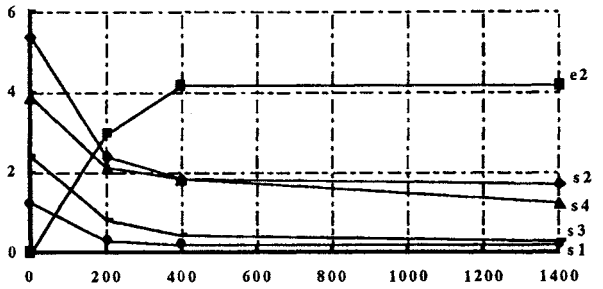


Fig. 6 : Height of s and e bands versus irradiation dose (in Krad)

In the same time, we have measured the laser absorption as described above and, as indicated in figure 7. We see that the absorption integral $\overline{\sigma_L}$ calculated for each plane increases with the irradiation dose in the same way that e_2. The values in brackets represent the mean absorption integrals for the 7 planes of measurements which increase in logarithmic manner.

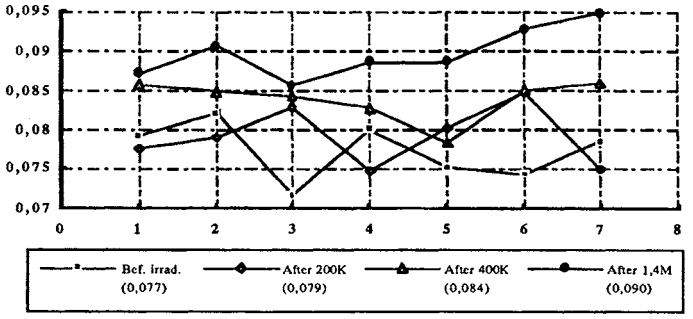


Fig. 7 : $\overline{\sigma_L}$ for each observed plane in B3 sample.

The B2 sample has been only swept during 15 days by our own means, in atmospheric pressure under dry Nitrogen. The absorption integrals obtained after sweeping have been compared at those obtained after irradiation (see table 3). The mean value of $\overline{\sigma_L}$ is about the same on B3 after 1.4 Mrad.

Sample	plane 1	plane 2	Plane 3	Plane 4	Plane 5	Plane 6	Plane 7	M. val.
B sample	0,0794	0,0821	0,0716	0,0802	0,0753	0,0744	0,0786	0,077
B2 after sweeping	0,0854	0,0890	0,0870	0,1011	0,0871	0,0968	0,0929	**0,091**
B3 after 1st treat.	0,0777	0,0791	0,0829	0,0749	0,0804	0,0847	0,0751	0,079
B3 after 2d treat.	0,0858	0,0849	0,0844	0,0828	0,0784	0,0852	0,0860	0,084
B3 after 3rd treat.	0,0873	0,0905	0,0857	0,0887	0,0887	0,0927	0,0948	**0,090**

Table 3 : $\overline{\sigma_L}$ for B2 and B3 (last column : mean values)

The IR spectrum of B2 shows a strong absorption at 3366 cm^{-1} comparable with that obtained for the sample B3 presented in figure 5b. So, **a 15-days sweeping and a 1.4 Mrad irradiation give approximatively the same concentration in Al-OH^- and Al-holes (deduced from the previous table).**

3.2. Comments about A sample :

With the A sample (Y-cut extracted from the entire as-grown block), the different growth zones (--X, +X and Z) have been observed before and after irradiation. The following figure (Fig. 8) shows the 3 studied areas more and less colored by irradiation. As for the Z zones, we can not see the dark areas colored after irradiation. The figure 9 summarizes the result obtained by laser absorption (and deduced from figure 10) which shows that the integral value $\overline{\sigma_L}$ increases with the strong darkness of the +X-zone. By the same manner, Al-hole defects are created in the -X-zone, but in a weaker proportion.

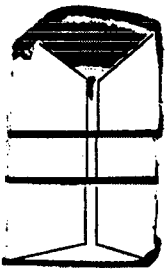


Fig. 8 : Part I (-X zone) slightly colored - Part II (pure Z-zone) and Part III (+X, Z-zone)

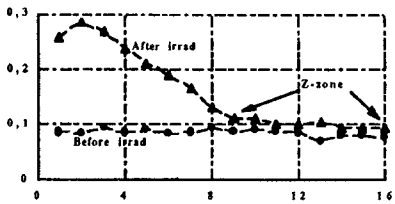


Fig. 9 : A sample (part III) : Integral absorption versus plane number

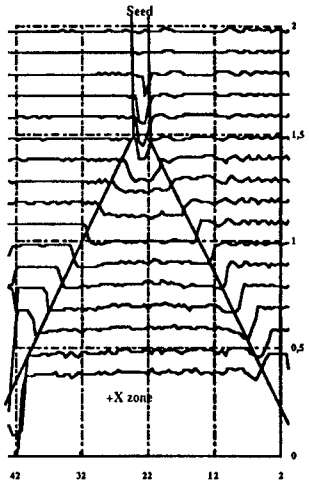


Fig. 10 : laser absorption diagram on Part III

3.3. Natural quartz sample (D) :

The chemical analysis of the D sample, issued from Brazilian natural quartz block, shows the highest concentrations of Aluminium and Lithium, nevertheless not very high for a natural quartz crystal. So, we think that a darkness can appear after irradiation, as for almost natural quartz samples. But, before irradiation, we have performed a 15-days sweeping under Nitrogen atmosphere and obtained colored zones lenghtened along Z-axis. The laser absorptions, plotted in Fig. 11b, show strong and inhomogeneous absorptions due to these banded colored zones. We suppose that their origins are due to irregular impurities distribution in the material.

After 1 Mrad-irradiation, the laser absorption plots present the same irregularities with an higher mean level of absorption (Fig. 11c).

Treatements	D1 As-recieved	D1 After annealing	D2 After sawing	D2 After sweeping	D2 After irradiation
mean $\overline{\sigma_L}$ (9 plans)	0,120	0,051	0,063	0,106	0,166

Table 4 : $\overline{\sigma_L}$ on natural quartz after sweeping and irradiation

The previous table contains informations about laser absorption measurements performed after three kinds of treatments as annealing (at 350 °C during 1h45mn), sweeping and 1 Mrad-irradiation under a ^{137}Cs source. Between the second and the third measurement, we have only cut the sample along Y crystallographic plane, D2 becoming a 5 mm thick Y-cut to be able to compare with the other analysed samples. If these two measurements are, nevertheless, different, we think that it is due principally to the partial restoration after annealing (which is confirmed by EPR spectrum) and, in a certain degree, to the polishing quality of the blank surfaces. Indeed, the figure 11a obtained after annealing shows very regular curves which are translated by a very low $\overline{\sigma_L}$, better than this of the previous A and B samples.

After sweeping, this table presents a strong mean absorption value revealing a high concentration in Al-hole defects. This will be confirm by our EPR measurements.

Finally, we have observed that the 1 Mrad-irradiation continues to darken the sample (see the 5th measurement of the table 4 and the figure 11c).

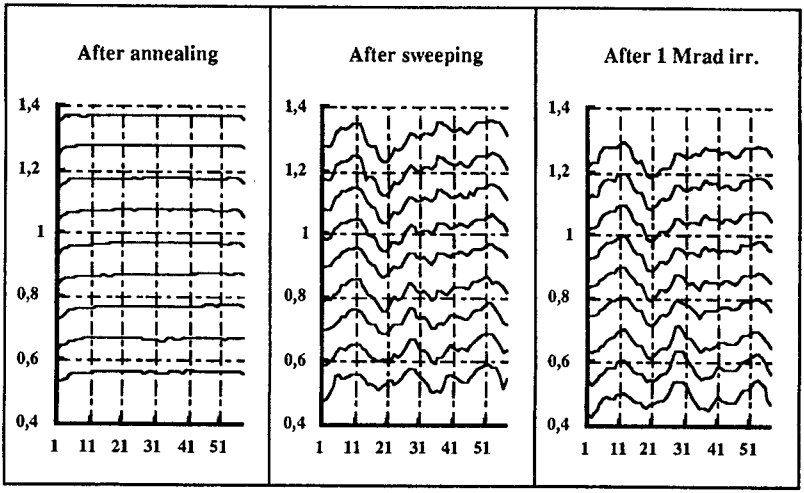


Fig. 11a, 11b, 11c : laser absorption on natural quartz (In versus point number)

On D2 sample, we have also realised room temperature spectra in the InfraRed, Near InfraRed, Ultra-Violet and Visible domains with a Lambda 9 Perkin-Elmer spectrometer. The observations are made, like usually, with the incident beam aligned with the Y-axis. The three spectra presented in figure 12 show a large difference between the first one before annealing and the second one after this treatment. So, an irradiated quartz sample absorbs largely in the full visible domain. As for the absorption at 632.8 nm due to Al-hole defects, the resolution of the spectrometer is not enough low to prefer this measurement method to the laser absorption technique.

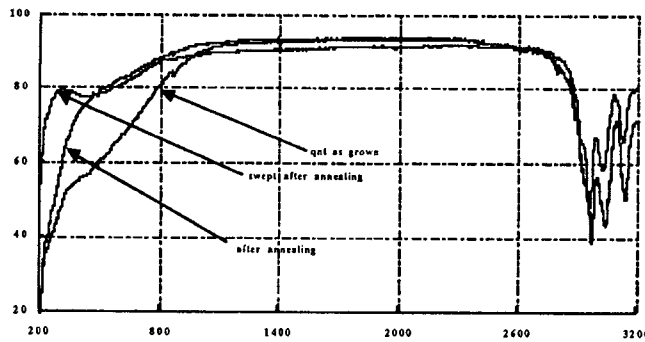


Fig. 12 : spectra on Natural quartz sample (Trasmission in % vs λ in nm)

No strong absorption is detected around 2000 nanometers, which is attributed to the frequency vibration of the H_2O molecule.

3.4. Electron Paramagnetic Resonance spectra :

Our first ESR spectra have been performed on different very small samples of quartz crystal issued from synthetic one after sweeping under vacuum or irradiation exposure and from irradiated natural quartz.

The figure 13 shows a ESR spectrum at 77 K of a vacuum swept sample (same origin than C). On the left part of the spectrum, the paramagnetic centers created can be interpreted as Al-hole, but they are not so easily to read than in spectra of irradiated quartz.

Other synthetic sample issued from B block, which was air swept by us does not show any hyperfine structure, but after γ-irradiation at room temperature, the characteristic HFS of Al-hole defect is clearly recognize as indicated in Fig. 14a and Fig. 14b, issued from (8). Though this spectra was obtained at an other temperature (20 K) and with other apparatus, we recognize the same spectrum than this presented in Fig. 16.

The Aluminium nuclei has non-zero nuclear spins (^{27}Al, I=5/2, 100% abundance), we might expect 2I+1=6 equally spaced lines. We see that the actual spectrum is more complex then that, due to the nuclear Zeeman and to the other terms which can appear in the Hamiltonian.

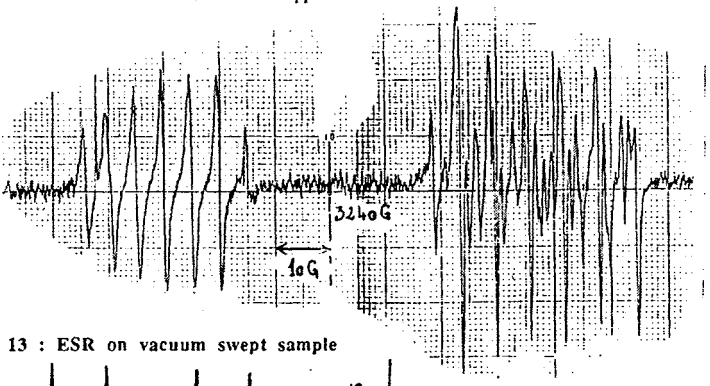


Fig. 13 : ESR on vacuum swept sample

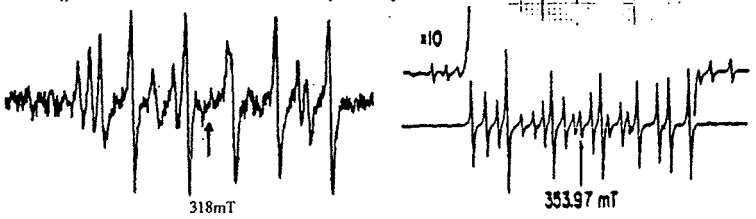


Fig. 14a (T=20K), 14b : HFS of Al-hole defect (a: ν=9.248 GHz, b: ν=9.998 GHz)

The previous spectrum (Fig. 14a) is only part of the entire spectrum which is presented in the two figures 15a and 15b. Furthermore, we show the orientation effect of the magnetic field relative to c-axis direction of the sample.

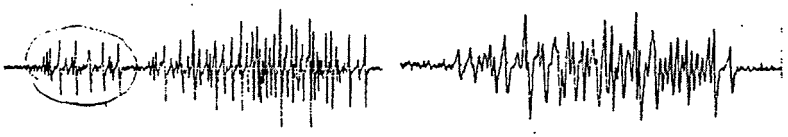


Fig. 15 : $\vec{B}$ // $\vec{c}$ and $\vec{B}$ perpendicular to $\vec{c}$

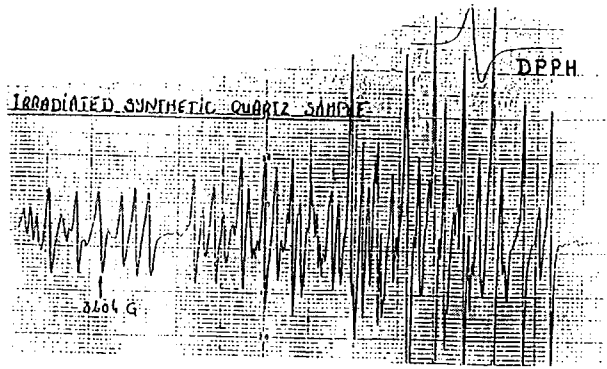


Fig. 16 : ESR of irradiated synthetic quartz (E sample at 77 K)

The first results deduced from ESR spectrometry can be summarized in the following items :

1. there are no paramagnetic defects in as-grown synthetic sample. A sweeping performed under air atmospheric give the same result.
2. only the spectra of irradiated samples contain Hyperfine Structures clearly identified as Al-hole. The other peaks can not be identify due to the complexity of the structure. Nevertheless, we know that these hyperfine structure can not be due to Iron because its centers appear at magnetic field values two times lower. We have verify that they do not exist in our samples.
3. the more interesting ESR spectra have been obtained when the magnetic field and the crystallographic axis c (so called optical axis Z) are parallel. Nevertheless, added works will be realised in turning the c-axis in the magnetic field to calculate the Hamiltonian operator of the unknown and complex HFS appearing in the right part of the spectra.
4. though the specctrum of the vacuum swept sample and this of irradiated one appear different, the complementary analyses permit us to believe that the precursor is, in all cases, the **Aluminum** center.

4. CONCLUSION

The aim of this work was to characterise the defects created by sweeping and irradiation processes. For that, we have used several techniques as Laser Absorption, IR, ESR and ICP spectrometries. Furthermore, we have proposed a dosage method of these defect centers by a non-destructive technique of measurements which can permit also to obtain the Al-hole distribution mapping in the entire analysed sample.

We have used the IR and laser absorption to monitor the Al-OH and Al-hole changements in material before and after each process. The EPR techiniques was employed to confirm the previous results obtained with 632.8 nm laser absorption measurements, the "bulk" concentration of impurities being determined by ICP. It was useful for correlate the impurity concentrations with the behavior of material during and after sweeping and irradiation. The main impurities are the **Aluminum**, of course, the **Lithium and, at least, the Sodium.** Indeed, it appears that the intensity of the color centers can be closely related to the concentration of the Lithium as indicated in the following table :

Samples		Color intensity	ICP	
			Na	Li
A (**)	-X	slightly brown	2.203	0.290
	+X	brown	3.360	0.610
	Z	colorless	< 0.010	0.08
B (*)		colorless	1.370	0.290
B (**)		colorless	1.370	0.290
C (*)		slightly dark	0.330	0.04
D (**)		brown	1.62	1.40

(*): swept (**): irradiated

For another irradiated natural quartz sample (not presented here), we have observed an important coloration after 1 Mrad irradiation. As for the previous ones, the ICP analysis has reaveled a very high concentration in Lithium (6.8 ppm) even though the Sodium concentration is only of 0.4 ppm.

Concerning the Al defect centers, we can conclude that the Al-OH and Al-hole formation in material depends on the original precursor defects in the sample and also on the environmental conditions in each process, such as hydrogen supplies during sweeping or temperature of the sample during irradiation.

Only the spectra of irradiated samples contain Hyperfine Structures clearly identified as Al-hole. But, up to now, we can not identify the right of the EPR spectra due to the complexity of the structure. We will continue to realise other EPR spectra on well-known quartz samples in relation with other techniques to point out the main defect responsible of the frequency drift in ultra-stable resonator. In the same time, we have manufactured BVA resonators in the previous characterised material, whose the results will be published in next publication (Part II).

Indeed, in the time-frequency domain, the goal is to obtain resonators quasi-insensitive to the ionizing irradiation.

ACKNOWLEDGEMENTS

The authors would like to thank Professor J.G. Théobald for helpful discussions on the experimental investigations on ESR measurements; M. P. Mériandeau from "Institut de Recherche sur la Catalyse" laboratory (Lyon University) and M. Clergeaux from "Acoustique et Optique de la Matière Condensée" laboratory (P. et M. Curie University - Paris VI) for performing the ESR measurements.

The samples irradiations are realised by Dr Lamy from C.R.T.S. of Besançon ; we thank him too.

REFERENCES

1. J. C. KING, H. H. SANDER : "Rapid annealing of frequency change in crystal resonators following pulsed X-irradiation" - 26th A.F.C.S. (1972), pp. 23, 32

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

3. E.F. HOLDEN : "The cause of colour in smoky quartz and amethyst" - Amer. Min. V(10), (1925), pp. 203, 252.

4. T. FUTAGAMI : " - Proc. Phys. Math. Soc. Jap. V(20) (1938), pp. 458.

5. R.H.D.NUTTALL, J.A. WEIL : "Two hydrogenic trapped-hole species in α-quartz" - Sol. State Commun., Vol. 33, (1980), pp. 99, 102.

6. J.H.E. GRIFFITHS, J. OWEN, I.M. WARD : "Magnetic resonance in irradiated diamond and quartz" - The Phys. Soc. London, (1955) - pp. 81, 87.

7. M.C.M. O'BRIEN : "The structure of colour centers in smoky quartz" - Proc. Roy. Soc., A231, (1955), pp. 404, 414.

8. J.A. WEIL : "A review of electron spin spectrometry and its application to the study of paramagnetic defects in crystalline quartz" - Phys. Chem. Minerals (1984), pp. 149, 165

9. L.E. HALLIBURTON, N. KOUMVAKALIS, M.E. MARKES, J.J. MARTIN : "Radiation effects in crystalline SiO_2 : The role of Aluminum" - J. Appl. Phys. 52 (5) (1980), pp. 3565, 3574.

10. R.H.D. NUTTALL, J.A. WEIL, R.F.C. CLARIDGE : " - Sol. State Commun., 19, (1976), pp. 141,

11. P. ZECCHINI, H.G. LIPSON : "Defect distribution mapping in quartz" - 43th Annual Freq. Control Symp. (1989), pp. 490, 496

12. B. VIARD, J. DAUMET, M. KHAYAR, H. MERIGOUX : "Trace elements analysis in quartz" - Proc. 3rd European Frequency and Time Forum. (1989), pp. 609, 616

13. J.J. BOY, P. ZECCHINI, B. VIARD : "Synthetic quartz material characterisations. Connexions with the quality of AT-cut resonator" - Proc. 5th European Frequency and Time Forum. (1991), pp. 93, 98

14. KATS : "Hydrogen in α-quartz" - Philips Res. Report (1962), pp. 201, 279

15. J.C. BRICE : "Crystal for quartz resonators" - Rev. Mod. Phys. 57 (1) (1985), pp. 105, 146.

ORIGIN AND MEASUREMENT OF QUARTZ RESONATOR MAGNETIC SENSITIVITY

R. Brendel, B. Cretin, A. El Habti

Laboratoire de Physique et Métrologie des Oscillateurs du C.N.R.S.
associé à l'Université de Franche-Comté-Besançon
32 avenue de l'Observatoire - 25044 Besançon Cedex - France

Abstract

The magnetic sensitivity of quartz crystal resonators is a consequence of the ferromagnetic properties of the metal used as support for the vibrating plate. Various magneto-mechanic interactions can contribute to the overall sensitivity, the most important of them is shown to be magnetostriction: when submitted to a magnetic field, most of the ferromagnetic materials undergo an expansion or contraction wich induces a stress in the sustained plate and then a change in the crystal resonant frequency.

The striking likeness between nickel magnetostriction curve and the frequency-vs-magnetic field curve of resonators is at the origin of this assumption.

Other experiments have been performed to check on this assumption, in particular an experimental setup using a heterodyne interferometric laser probe with enhanced sensitivity has been used to investigate *in situ* the magnetostriction of the plate supports.

Also, a comparison between magnetic behaviour of identical resonators mounted with different materials definitely proves the responsibility of the supports in the magnetic sensitivity of resonators and gives interesting information on its reduction.

1. Introduction

As the oscillator performance increases, new applications such as precision orbit determination need more and more accurate devices. It results that some perturbation sources ignored to date as having in most cases negligible influence, now emerge and need attention. Thus, for a few years, it has been observed that starting or stopping on-board magnetic torquers may lead to slight frequency shifts in an oscillator closed by it.

The order of magnitude of this effect (10^{-11} / gauss) is unacceptable in high precision orbitographic measurement (DORIS mission). Various papers /1/, /2/, /3/, have shown that almost all components of the oscillator circuitry may have a more or less important effect on the magnetic sensitivity. At the same time some solutions to reduce this sensitivity have been proposed, for

example by changing some component technologies /4/ or by shielding the oscillator with high magnetic permeability material /5/. Nevertheless these improvements are either not sufficient to meet the required specifications or not satisfactory because of the resulting volume and weight increase. A large part of the oscillator magnetic sensitivity finds its origin in the resonator itself. Previous works have demonstrated that it is not due to the quartz material itself but rather to the magnetic properties of the metallic springs used to support the plate /1/, /6/. Then, a reduction of the magnetic sensitivity goes through a better understanding of the magneto-mechanic interaction between springs and plate and a better knowledge of the magnetic properties of the material.

2. Resonator magnetic sensitivity measurement setup

The experimental setup used in magnetic sensitivity measurement described a few years ago /6/ has been slightly modified so as to meet the recommendations of the IEEE Standard /7/, /8/. The resonator is now kept in a fixed direction so as to avoid errors due to the earth's magnetic field or to gravitational effects. Only the magnetic field strength used is larger than the recommended value so as to compensate the reduction of the signal to noise ratio due to the experimental setup. In fact, the oscillation sustaining circuit is kept far from the resonator to insure that it is the only component submitted to the magnetic field. The other experimental measurement procedures and signal processing techniques are described in ref. /6/.

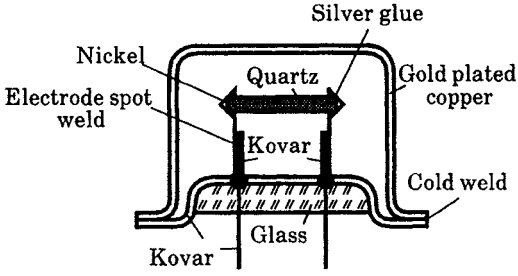


Fig. 1. *Internal assembly of a resonator.*

Most of the resonators used in the experiments are regular units as shown in Fig. 1. The quartz disc is cramped by two nickel springs soldered on two kovar supports. This latter material is used because it has the same thermal expansion coefficient as the insulating glass through which pass the supports. The electrical connection between quartz electrodes and springs is provided with a conducting silver glue. The resonator is maintained under vacuum by a gold plated copper cap cold welded on a kovar seat.

3. Possible magneto elastic interactions /9/

Various mechanisms can be considered to explain how a magnetic field induces a mechanical action in a ferromagnetic material.

3.1. Compass effect

The first effect is explained by the fact that ferromagnetic materials are constituted by small magnetic domains initially randomly oriented. When submitted to a magnetic field these domains tend to align along the same direction corresponding to the easiest magnetization direction which is usually the rolling direction of the material. As the domains reach the saturation state, their magnetization vectors oriented along the same direction are subject to a torque tending to align them with the external magnetic field just like a compass needle. This situation is sketched in figure 2.

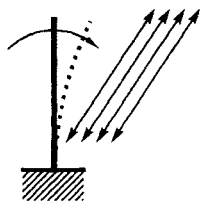


Fig. 2. *Compass effect.*

Since the supports of the quartz plate are clamped on the seat, the assembly could bend as shown in figure 3 and presents a maximum effect when the magnetic field lies in the plane of the supports.

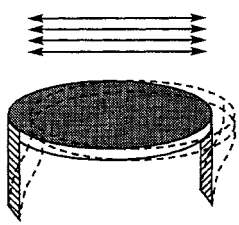


Fig. 3. *Bending of the assembly.*

It should be noted that a change in the magnetic field sense also changes the sense of the magnetization vector so that the resulting torque doesn't change.

3.2. Mutual repulsion

Two parallel ferromagnetic supports close together submitted to a magnetic field act as two small magnets of same polarity. As a consequence they tend to repel each other and to induce in the quartz plate they support a diametrical tension (see Fig. 4).

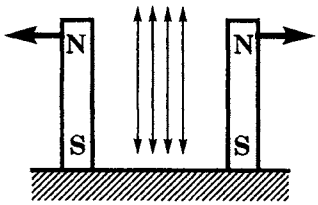


Fig. 4. *Repulsion forces.*

As in the previous case, a change in the applied magnetic field sense also changes both magnet polarities so that the resulting effect is always a mutual repulsion.

3.3. Change of elastic modulus

This effect also called *ΔE effect*, consists of a change in the Young's modulus of ferromagnetic materials when submitted to a magnetic field. As a result, the pressure exerted by a ferromagnetic spring on a sustained disc is modified by an external magnetic field (see Fig. 5).

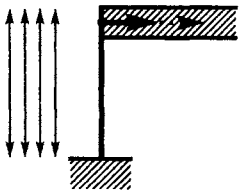


Fig. 5. *ΔE effect.*

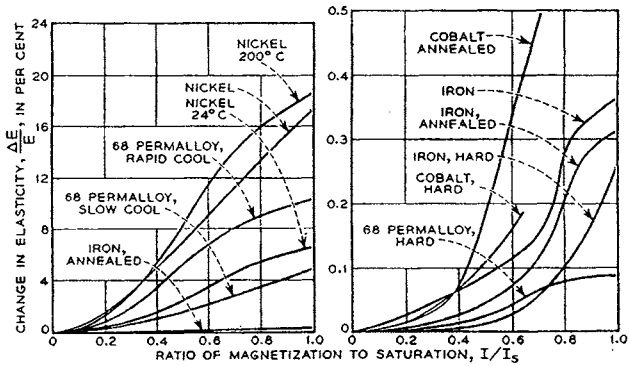


Fig. 6. *Change in Young's modulus with magnetization (after /9/).*

Fig. 6 shows the variation of Young's modulus for various ferromagnetic material /9/. The sign of this variation doesn't depend on the sense of the magnetic field.

3.4. Magnetostriction

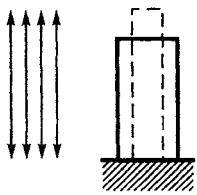


Fig. 7. *Magnetostriction.*

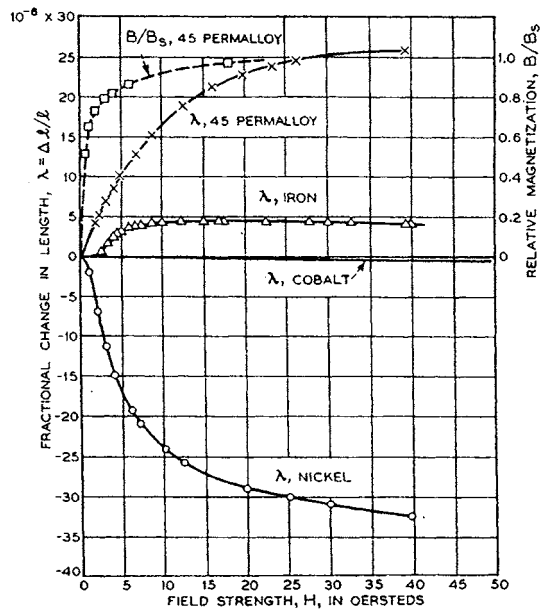


Fig. 8. *Magnetostriction of some ferromagnetic materials /9/.*

The last, and probably the most important, magneto-mechanic effect is the magnetostriction. That is a change in the dimensions of a ferromagnetic material submitted to a magnetic field. For example, as shown in Fig. 7, a rod of nickel submitted to a magnetic field undergoes a contraction in length and an expansion in the other directions so that the volume is kept approximately constant. In many cases, ferromagnetic materials may present an expansion in length rather than a contraction. In all cases, the deformation (contraction or expansion) only depends on the kind of material but not on the sense of the

applied magnetic field. Fig. 8 shows the magnetostriction coefficient of some ferromagnetic materials /9/.

The kovar used in various parts of the resonator packaging (see Fig. 1) is also a ferromagnetic material the magnetostriction of which has been measured at the *Laboratoire de Magnétisme Louis Neel* in Grenoble. The result is shown in figure 9 /10/.

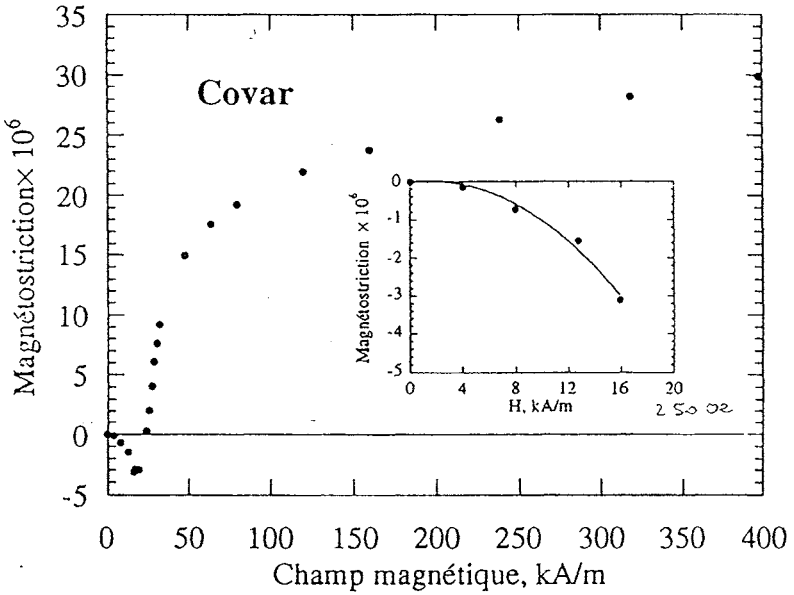


Fig. 9. *Magnetostriction of kovar /10/.*

It is easy to understand that the magnetostriction of the springs and/or the supports can induce stress in the disc they sustain.

3.5. Hysteresis

Most of the magnetic effects described above don't depend on the sense of the applied magnetic field and present a maximal effect when the field is directed along the largest dimension of the material. Also, most of these magnetic effects present a more or less pronounced hysteresis depending on the kind of material. For example, Fig. 10 shows the hysteresis of magnetostriction in nickel /9/.

Except for the ΔE effect, an external magnetic field gives rise to a mechanical displacement of the ferromagnetic springs and supports which, in turn, induces stress in the quartz plate and then a change in the resonant frequency.

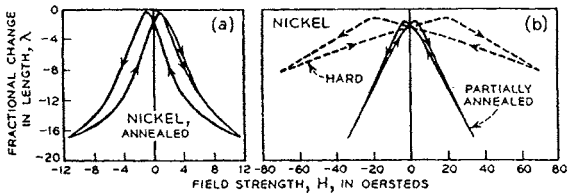


Fig. 10. *Hysteresis of the magnetostriction in nickel (after /9/).*

4. Measurements of magneto-mechanic displacements

In almost all magnetic sensitivity measurements performed on resonators, the maximal effect is observed when the magnetic field is directed along the axis of the resonator (see Fig. 11, ref. /6/ and Figs 16-19 thereafter).

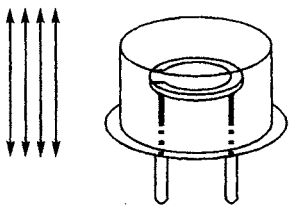


Fig. 11. *Most sensitive direction of the magnetic effect.*

This fact implies that the "*compass effect*" is probably not responsible for the observed sensitivity. So as to tell one of the remaining effect from the others, a set of experiments have been performed using a heterodyne interferometric laser probe with enhanced sensitivity enabling dynamic displacement measurements as low as a few 10^{-5} Å·Hz$^{-½}$ /11/. Fig. 12 shows a simplified bloc diagram of the experimental setup.

A ferromagnetic sample of the same material as used for the resonator springs is clamped at one end on a fixed support, a small mirror glued at the free end is used to reflect the laser beam of an interferometric probe. The sample is submitted to a 1 kHz sinusoidal magnetic field produced with an excitation coil fed by a frequency generator through a power amplifier. In addition to this dynamic magnetic field of amplitude 10 gauss a DC magnetic component in the range ± 30 gauss can be superposed.

This arrangement enables to observe the small displacements of the free end of the sample by using a lock-in amplifier and a signal processing unit. The resolution of the displacement measurement increases with the excitation frequency ,due to optical noise, while the dynamic field magnitude decreases. A 1 kHz excitation frequency has proved to be a good compromise. Figs. 13 and 14 show the dynamical displacement at the end of the ferromagnetic sample in different conditions as a function of the DC magnetic component.

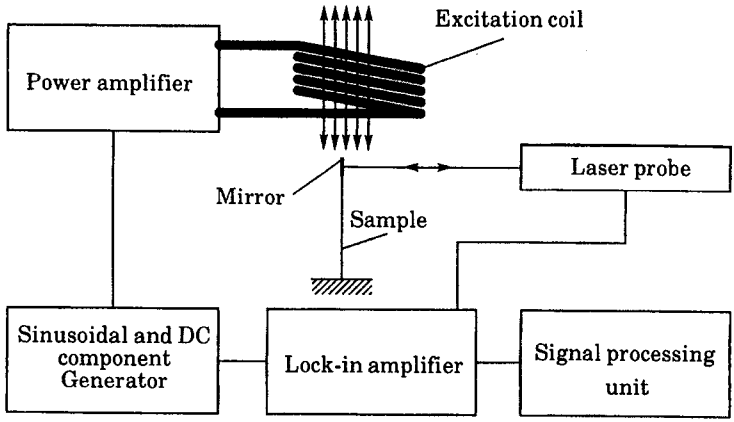


Fig. 12. *Experimental setup for the measurement of magneto-mechanic displacements.*

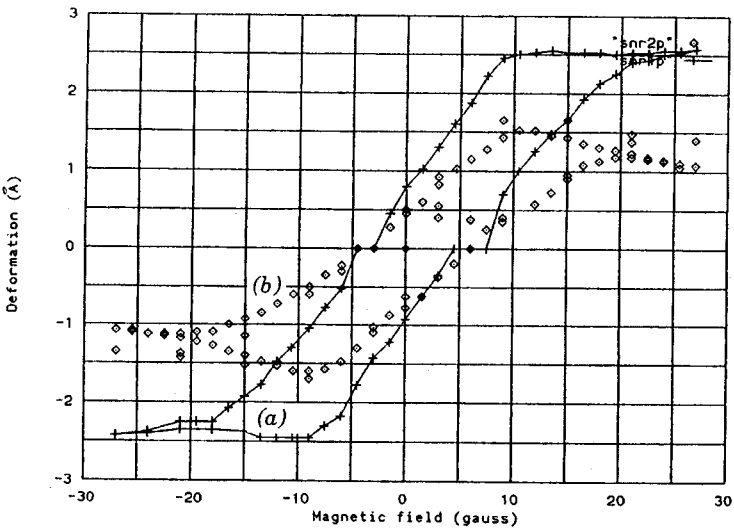


Fig. 13. *Dynamical displacement of the free end of the sample vs. DC magnetic field component with (a) or without (b) the other support present.*

Because of the particular detection system, the curves obtained represent the deformation due to the small dynamical magnetic field around a DC biasing field so these curves are not directly comparable with the magnetostriction curves shown in figure 10, rather they represent their derivative.

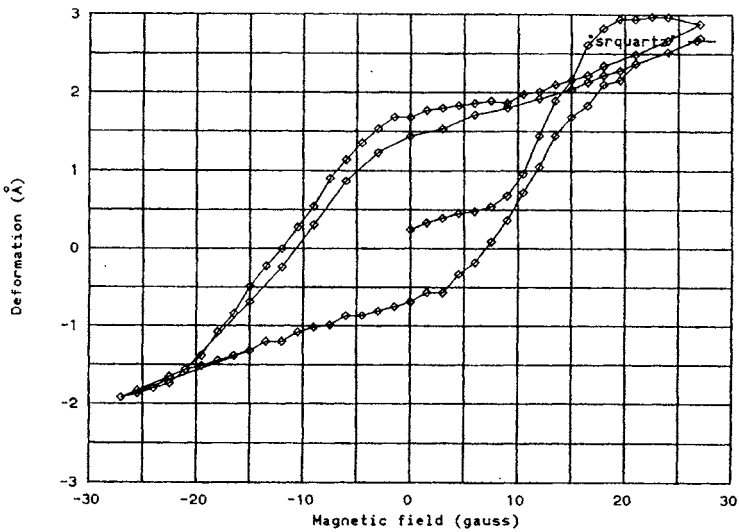


Fig. 14. *Dynamical displacement of the end of the sample vs.DC magnetic field component when the quartz disc is in place.*

Figure 13 shows two results, curve (a) is obtained by measuring the dynamical displacement of one spring in the presence of the other one when the quartz plate has been removed (see Fig. 15 a). Curve (b) is obtained in the same conditions but after the other spring has been removed (Fig. 15 b). Figure 14 shows the result obtained when the whole assembly - springs and quartz - is submitted to the magnetic field (Fig. 15 c).

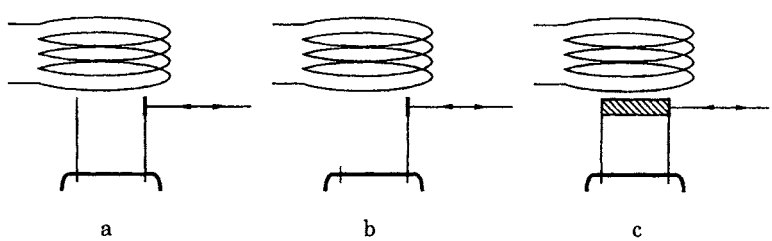


Fig. 15. *Experimental conditions for figures 13 and 14.*

From these experiments it can be deduced that the frequency shift of the quartz plate is due to the displacement induced by the magnetic field on the springs and supports, the most probable origin being their magnetostriction (ΔE effect doesn't produce any displacement) nevertheless, the difference

between curves (a) and (b) in Fig. 13 shows that the mutual repulsion effect cannot be neglected. We will show in the next section that other experimental results also incline towards the assumption of magnetostriction.

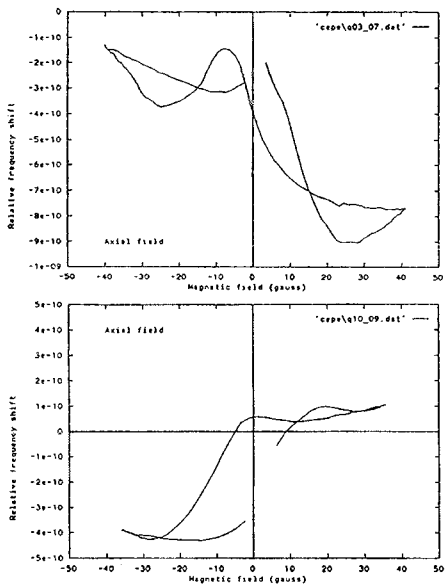
5. Magnetic sensitivity of resonators, results

A large number of experiments have been performed on resonators of various cuts, technologies and mounting. Some results have already been published /6/ and only the last results will be briefly presented here.

5.1. Influence of the resonator technology

Several lots of resonators have been measured so as to investigate for the general characters and possible dispersion of their magnetic sensitivity. All results presented here have been obtained with 10 MHz third overtone SC cut quartz resonators using nickel springs in T2111 box enclosure submitted to a 1000 s-period sinusoidal magnetic field as described in ref. /6/ and sec. 2. Although experiments have been performed with different magnetic field orientations, only the results obtained in the most sensitive direction (along the resonator axis) will be reported here.

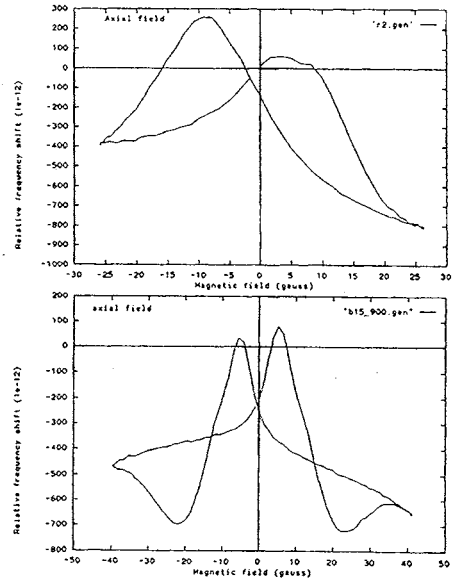
Figs. 16 a and b show the results obtained with two QHS resonators from the same batch, Fig. 17 a and b are obtained with two other resonators of same technology from another batch. The most surprising fact when looking at these results is the great variety of "*magnetic signature*" while the overall amplitude keeps approximately the same order of magnitude. Fig 18 a and b



(a)

Fig. 16. *10 MHz SC cut resonator QHS technology (same batch).*

(b)



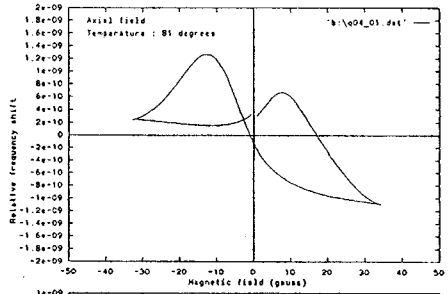
(a)

(b)

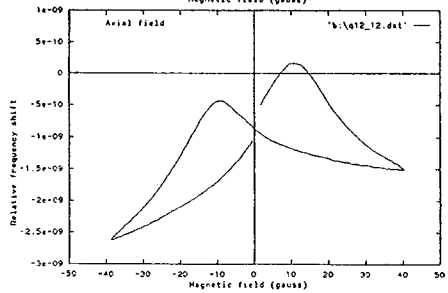
Fig. 17. *10 MHz SC cut resonators, different batch but same technology as in Fig. 16.*

are obtained with two QAS from the same batch. Their magnetic signatures are fairly similar but much larger than those of QHS (the scale in Fig. 18 is four times larger than the scale of Figs. 16 and 17). The similarity between the magnetostriction expansion coefficient (Fig. 10) and the curves shown in Fig. 18 is one of the most convincing features on account for explaining the magnetic sensitivity of resonators by the magnetostriction of the springs.

Another experiment has been performed to compare the results obtained with two QAS resonators. One of them (Fig. 19 a) is a regular one, in the other one the resonator is mounted on an intermediate plate (Fig. 19 b). The barometric sensitivity of the first resonator has been measured to be 25 times larger than the second, while its overall magnetic sensitivity is 6 times larger. Also these figures need some comments: if the magneto-elastic mechanism responsible for the magnetic sensitivity of resonator is indeed the magnetostriction of the springs, the frequency shift-vs-magnetic field curves would present the same symmetry as the magnetostriction curve (Fig. 10), so the magnetic sensitivity curve should be an even function of the magnetic field variable. This is approximately true in a number of figures presented in this section but this is not always the case (see for example Figs. 16 b and 19 a). It is highly probable that the dissymmetry of the magnetic sensitivity curves reveals in fact a dissymmetry in the mechanical mounting of the quartz plate. As a consequence, the stress induced in the plate by the magnetostriction of the springs is no longer symmetrical with respect to the plate axis and the resulting frequency shift losses the initial symmetry of the phenomenon. Hence the high barometric sensitivity of the resonator used for Fig. 19 a could

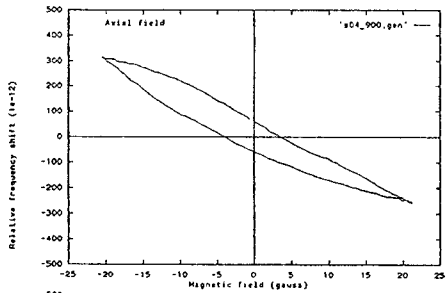


(a)

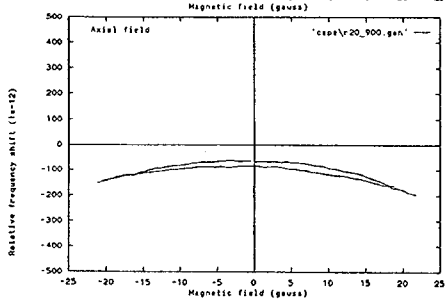


(b)

Fig. 18. *10 MHz SC cut resonators QAS technology (same batch).*



(a)



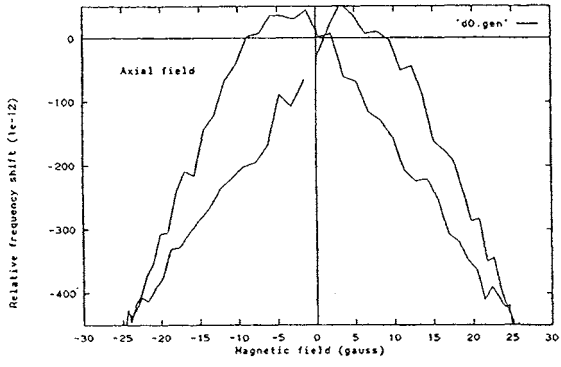
(b)

Fig. 19. *10 MHz SC cut resonators QAS technology: (a) without intermediate plate. (b) with intermediate plate.*

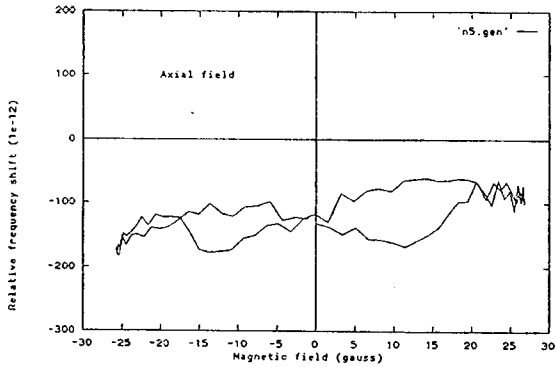
be explained by a mounting dissymmetry revealed by the strong dissymmetry of the magnetic sensitivity curve, indeed it is well known that any dissymmetry in the mechanical assembly of a resonator drastically increases its sensitivity to environmental perturbations (acceleration, pressure, etc.) /4/.

5.2. Comparison between spring materials.

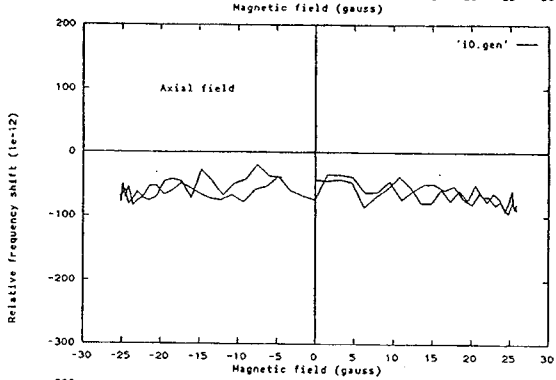
Many efforts have been spent to reduce the magnetic sensitivity of the resonators, among them materials different from the nickel usually used have been tested. A set of five resonators have been investigated, all of them are identical except for the springs made in different materials. The resonators are all 6.144 MHz AT cut QHS units. Because of the particular frequency of these resonators the experimental conditions have been slightly modified, the curves obtained, shown in Fig. 20, are then a bit more noisy, nevertheless they remain worth reading. They are presented from the most to the less magnetic sensitivity. The kovar appears to be much more sensitive than the nickel while stainless steel is noticeably less sensitive. It is not surprising that the resonator made with copper-alloy springs doesn't exhibit any perceptible magnetic sensitivity. This last experiment definitely proves the responsibility of the magnetic properties of the springs in the magnetic sensitivity of resonators. Of course, a reduced magnetic sensitivity is not the only goal to reach, the resonators still have to keep their performance in terms of mechanical behaviour, accelerometric and barometric sensitivity, ageing, etc. This is why this study is in fact only a part of a wider investigation some other aspects of which are presented elsewhere in these proceedings /12/.



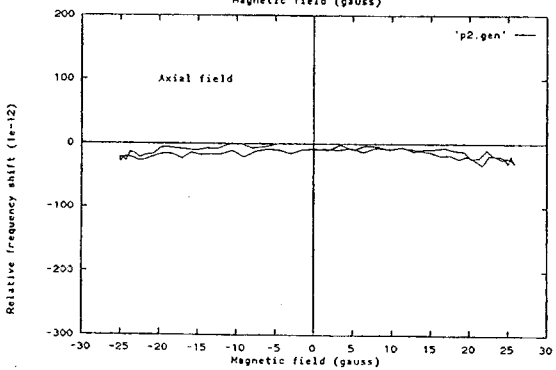
(a) Kovar springs.



(b) Nickel springs.

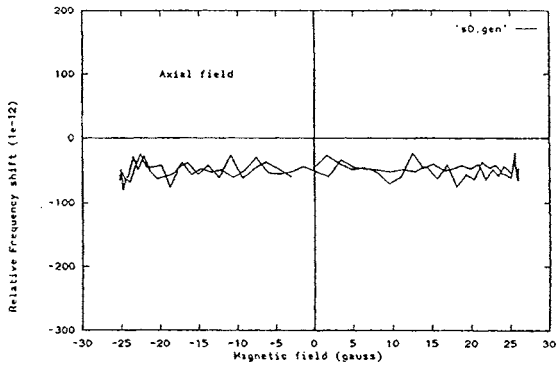


(c) Stainless steel springs.



(d) Other kind of stainless steel springs.

Fig. 20.
6.144 MHz AT cut QHS resonators.



(e) Copper-alloy springs.

6. Conclusion

The magnetic sensitivity of resonators comes undoubtedly from the ferromagnetic properties of the springs used to hold the quartz plate up. This sensitivity can be drastically reduced by using copper alloy springs while kovar and nickel should be avoided. Although the quartz plate itself is not responsible for the magnetic sensitivity, the magnitude of the phenomenon can be strongly increased by the geometrical characteristics of the disc, it is obvious that the thicker the plate is the lower the sensitivity appears. Even in a same lot of resonators, a great variety of magnetic signatures is often observed, some of them being abnormally dissymmetrical in view of the magnetic action and geometric assembly symmetries. It is highly probable that these anomalies betray a geometric dissymmetry in the plate mounting.
In most of the experiments, the higher sensitivity is observed when the magnetic field is directed along the resonator axis that is along the largest dimension of the springs. This fact, in addition with the similarity between magnetostriction curve and magnetic sensitivity curves of resonators, demonstrates that the magnetostriction is the most probable mechanism responsible for the observed phenomenon. It is highly probable that the complexity of the magnetic signature has to be attributed to the spring magnetostriction hysteresis rather than to a complicated stress and strain interaction between plate and springs. This interaction might be as simple as a diametrical tension or compression for which the induced frequency shift is well characterized. Then, the magnetic sensitivity of resonators could constitute an extremely sensitive tool for the experimental investigations on the mechanical actions exerted on the resonator plate. In fact, by knowing the magnetic behaviour of the spring it is possible to determine the kind of the actions the plate is submitted to and the device thus obtained can be easily used for example to experimentally investigate for the lowest sensitivity fastening points on the plate or to study new shapes of springs and supports.

These applications still need theoretical modelling of the magneto-elastic behaviour of the various materials and it is in these directions that the efforts are currently being made.

References

/1/ **R. Brendel, C. El Hassani, M. Brunet, E. Robert,** *"Influence of magnetic field on quartz crystal oscillators"*, Proc. 43rd Ann. Symp. on Freq. Control, pp. 268-274 (1989).

/2/ **R. Brendel, C. El Hassani, P. Krux, E. Robert,** *"Magnetic sensitivity of oscillator components"*, Proc. 5th Europ. Freq. and Time Forum, pp. 390-395 (1991).

/3/ **F. Deyzac,** *"Oscillateur à quartz de bord en environnement sévère. Sensibilité magnétique"*, ONERA. Technical Report # 37/7132 PY (1988).

/4/ **F. Deyzac,** *"Oscillateur à quartz ultrastables à finalités militaire et spatiale. Amélioration du comportement mécanique"*, ONERA. Technical Report # 8/3945 PY (1993).

/5/ **J.-P. Granger,** *"Blindage magnétique de l'OUS de DORIS"*, CNES Technical Report # S2/NT/10 000/3800/MT (1989).

/6/ **R. Brendel, E. Robert,** *"Weak Magnetic Field sensitivity of Quartz Crystal Oscillators"*, Proc. 6th Europ. Freq. and Time Forum, pp. 99-104 (1992).

/7/ **R. L. Sydnor, R. Brendel, W. Wiedemann,** *"IEEE Standards Project P1193, Electric and Magnetic Fields and Effects"*, Proc. 46th Ann. Symp. on Freq. Control, pp. 794-797 (1992).

/8/ **IEEE-SCC-27 Working Group,** *"Guide for Measurement of Environmental Sensitivities of Standard Frequency generators"*, Proposition PAR-P1193 to be published (1994).

/9/ **R. M. Bozorth,** *"Ferromagnetism"*, Van Nostrand, 8th edition, (1964).

/10/ **A. Barlet, E. de Lacheisserie,** *"Analyse de l'aimantation et de la magnétostriction de quelques composants d'un résonateur à quartz"*, Technical Report, unpublished (1993).

/11/ **B. Cretin, W. X. Xie, S. Wang and D. Hauden,** *"Heterodyne Interferometers"*, Optics Communications, **vol 65**, n° 3, (Feb. 1988).

/12/ **J.-P. Michel, M. Zafrany, M. Candelier, J. Lamboley, L. Xavier, F.-A. Kuhnast, M. Nassik,** *"Influence of Part Stability on Ageing of Quartz Crystals"*, These proceedings.

Acknowledgement

This work was supported by Thomson-CEPE.
The authors wish to thank Dr. G. Marianneau for his help.

COMPARISON OF THE PROPERTIES OF THE NEW THERMALLY COMPENSATED MATERIALS WITH TRIGONAL SYMMETRY

J.Schwartzel*, J.Detaint*, B.Capelle**, A.Zarka**, E.Philippot***, A.Ibanez***, J.P.Denis****, D.Cochet-Muchy****.

*FRANCE TELECOM CNET 196, Av.Ravera BP 107, 92225 Bagneux France.
**L.M.C.P. Universités Paris VI & Paris VII; 4, Place Jussieu 75230 Paris France.
***L.P.M.S. Université du Languedoc, Place E.Bataillon 34095 Montpellier France.
**** CRISMATEC 2, rue des Essarts 38000 Gières France.

Abstract:

In this communication, we compare the properties of the four piezo-electric materials presently available or being studied that belong to crystal class 32: quartz, aluminium phosphate, gallium phosphate and langasite. In the first part, the growth methods used to obtain these materials and the perfection of the crystals now obtained are discussed using experimental results. It is observed that all this materials can be obtained with a high crystalline quality.

The second part is devoted to compare the main piezo-electric properties of these crystals. Computed and experimental results concerning the Y rotated cuts and particularly the compensated cuts are given and compared. The most important fact observed is the shift, from quartz, to berlinite, gallium phosphate and langasite, of the angular position of the most interesting compensated cut towards the angles of maximum coupling coefficient. We report also, new determinations of the angular positions of the corresponding compensated cuts of gallium phosphate and langasite.

In the last part, the energy trapping properties of these four materials are compared. We first consider the fundamental and the 3rd overtone of the piezo-electric (shear) modes. For the latter, the conventional energy trapping behaviour disappears for angular regions whose size increase from quartz to gallium phosphate. New features concerning the mode shape of the different overtones of the slow shear mode in the vicinity of the compensated cuts of the more recent of these materials were found and experimentally observed either by electrical measurements or X-ray determination of the mode shape.

We conclude to the important interest of these materials which can possess an unique set of complementary properties: a high cristalline perfection, superior thermal behaviours, and different coupling coefficients adapted to an extremely broad range of applications.

1. INTRODUCTION:

The evolution of electronic systems towards higher frequencies and baud rates has lead to the interest to find new thermally compensated piezo-electric materials permitting to obtain filters with larger bandwidth and oscillators with larger shifts than quartz /1/.

In the recent years it was successively observed that berlinite ($AlPO_4$), lithium tetraborate ($Li_2B_4O_7$), gallium phosphate ($GaPO_4$), and langasite ($La_3Ga_5SiO_{14}$), can fulfil these conditions while equalling or surpassing quartz for other very important properties such as the crystalline perfection, or the acoustic losses. Most of these new materials belong to the crystal class of quartz. Two of them are close structural analogues of quartz.

Berlinite was the first analogue of quartz to be synthesized/2/3/4/5/. Important progresses in the growth methods and in the knowledge of this material were obtained in the past ten years /6/7/8/9/10/. Since few years, crystals with a good crystalline perfection and a sufficient size are available to permit precise evaluation of devices /9/11/ and industrial prototypes.

The second quartz analogue, for which crystals having significative dimensions and perfection were obtained was gallium phosphate. The growth of such crystals/12/13/14/, the measurements of several sets of constants/13/15/, and the first evaluations of devices/15/16/17/19/ were reported in the recent years. Two major advances in the development of high quality crystals were reported more recently: the growth by lateral epitaxy on berlinite seeds/17/18/20/ and the growth using the direct solubility at high temperature/21/.

A very impressive amount of results concerning a lanthanum silico-gallate ($La_4Si_3GaO_{14}$ Langasite or LGS), and several of its modifications (substitution of one or several of the elements of the formula by other elements) or structural analogs has been obtained in Russia since the beginning of the eighties. More recently numerous results were reported: the obtainment of crystals having large dimensions /22/23/24/25/, the measurement of several sets of constants /22/26/ and several studies of devices /22/26/. Langasite and its analogs belong also to the crystal class of quartz.

In this paper we analyse the similarities and differences between the properties of these materials. The results of investigations made to further precise their properties and their possibilities in view of applications to filtering in radiocommunication equipments are reported.

2 CRYSTAL GROWTH and MATERIAL PERFECTION.

2.1 Crystal growth: Quartz and its analogous have phase transitions below their melting points. They display non negligible solubilities in both alcaline and acid aqueous solutions at elevated temperatures and high pressure. So that the natural choice for crystal growth for them is the use of hydrothermal methods. The growth of quartz uses its direct solubility in dilute sodium hydroxide or carbonate solutions at high temperature and pressure [12]. For the aluminium and gallium phosphate most of the recent works have been made using the reverse solubility existing from about 150 to 300 Celsius degrees in different acid solutions.

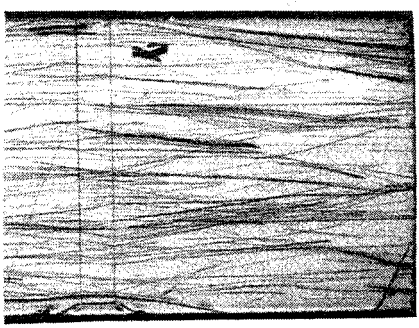
Table 1: Properties and growth methods of the four piezo-electric materials of class 32.

Material	Phase Transition	Solvent	Growth Method	Temperature (°C)	d kg.m-3
Quartz	alpha/beta 573°C	Alkali (Acid salts)	Hydro-thermal	320-450	2655.
AlPO4	alpha/beta 581°C	Acids (Salts)	Hydro-thermal	220-350	2640.
GaPO4	alpha/other. 933°C	Acids + ?	Hydro-thermal	170-400	3570.
LGS	no	Acids Alkali.	Czochral. Pulling	Tf=1470	5764.

The existence of a direct solubility at higher temperature and pressure in the same solutions was demonstrated /21/ and $AlPO_4$ and $GaPO_4$ crystals presenting a very high perfection and interesting properties were obtained recently with such conditions.

Langasite presents no phase transition up to its melting point, so that it is usually grown using the Czrochralsky method/23/24/25/. This pulling technique present the advantage of permitting to obtain large crystals of high perfection without the prerequisited need of disposing of large seeds. The most important properties related to the crystal growth of these materials are summarized in table 1. In this table, we can also observe that quartz and berlinite have a comparable low density whereas gallium phosphate and LGS have higher ones (nearly twice that of quartz for langasite).

2.2 Present state of the crystalline perfection. Figures 1 to 4 are X-ray topographs, obtained using synchrotron radiation (DCI synchrotron LURE Orsay France) that show the typical crystalline quality observed for these materials. The topograph of figure 1 is relative to a Y plate (# 1 mm thick) of synthetic quartz grown by SICN (France). The observations that can be made in this topograph are representative of the most frequently encountered crystalline quality. The dislocation density is about 10/mm2, and is nearly constant from the seed to the end of the Z growth sectors. Most of the dislocations are the prolongation of those existing in the seed, very few are created at the surface of the seed or due to inclusions in the Z growth sector. Quartz crystals having a much reduced dislocation density (a few/cm^2)/27/, and even zero dislocation, can be grown but, to our knowledge, there is presently no steady production of such crystals.



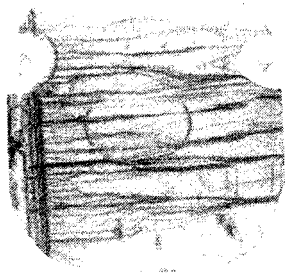


Figure 1: Topograph of Y cut plate of quartz.

Figure 2: Topograph of an AT plate of berlinite.

The topograph displayed in figure 2 is relative to an AT plate of berlinite grown at SICN. It represents also the usual present quality of berlinite crystals/10/. We can observe some fluctuations at the growth restart, but the dislocation density is of the same order as for the quartz sample. Some surface defects due to the polishing process can be observed together with contrasts due to the mounting clips of the plate and to the edges of electrodes deposited on this plate.

In Figure 3, a topograph of an AT $GaPO_4$ plate is presented. The corresponding crystal was obtained at the Montpellier University by two successive growths using the slow heating method, starting from a seed cut in a crystal obtained by lateral epitaxy on a berlinite seed/17/18/. We can observe a density of about 100 dislocations per square millimetre. The crystalline quality is already good because the individual images of the dislocations can be resolved. A fast progress of the quality of gallium phosphate is currently observed. A very high perfection was observed for the crystals obtained directly by lateral epitaxy on berlinite seeds /18/28/ or those grown using direct solubility conditions /21/.

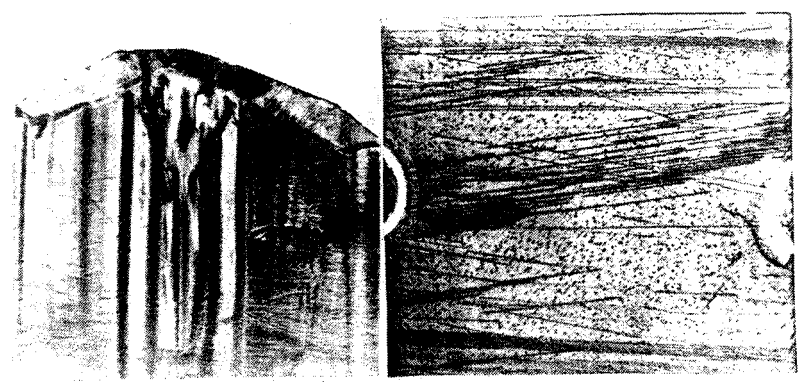


Figure 3: Topograph of an AT plate of gallium phosphate.

Figure 4: Topograph of a Y cut plate of LGS.

The topograph of figure 4 concerns a Y plate of langasite grown in France (CRISMATEC). The dislocation density is extremely small (about 10/cm2) and this sample presents an excellent quality.

3. BULK WAVE PROPAGATION AND EXPERIMENTAL RESONATORS PROPERTIES.

3.1 Computed results: A comparison of the piezo-electric properties of the bulk waves propagating in these four materials was made using computed results concerning the plane piezo-electric (elastic) waves propagating in the Y rotated cuts. These computations were made using the most recent constants published for each material /29/30/15/26/. For the Y-rotated orientations, one two fold axis of class 32, is always contained in the plane of the plate which keeps a monoclinic symmetry. For plates of such symmetry, only one, among the 3 one dimensional modes (a transverse mode) is piezoelectrically excited by an electric field normal to the surface. Depending on the rotation angle, the piezoelectrically active mode is either the slow or the fast shear. The longitudinal mode is never excited with this field configuration. Only the three quantities which determinate the most important properties of the devices will be considered hereafter: the phase velocities (or the related frequency constants NFa=V/2), the coupling coefficients and the temperature coefficients of the resonance frequencies.

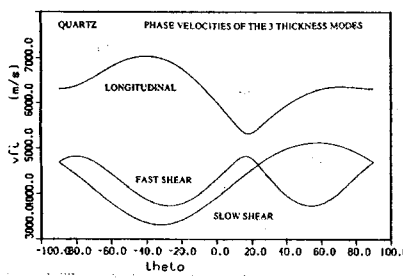


Figure 5a: Velocities of the thickness modes in Y rotated cuts of quartz.

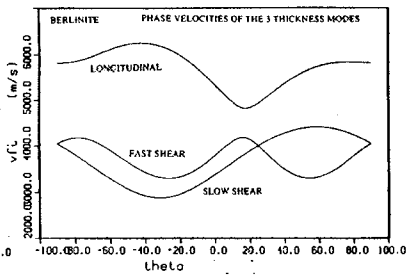


Figure 5b: Velocities of the thickness modes in Y rotated cuts of berlinite.

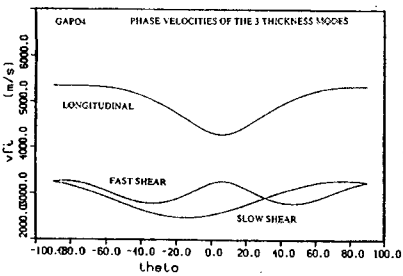


Figure 5c: Velocities of the thickness modes in Y rotated cuts of gallium phosphate.

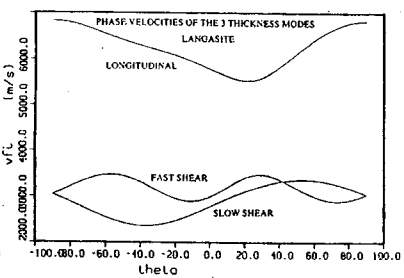


Figure 5d: Velocities of the thickness modes in Y rotated cuts of LGS.

In figures 5 a,b,c,d, the variations of the velocities of the plane waves existing in the Y-rotated cuts of these 4 materials are represented as a function of the rotation angle (theta). We can observe a similar angular behaviour for quartz and berlinite but with some how reduced velocities for the latter material. For $GaPO_4$, the extreme values of the velocities for each mode, appear at quite different angle than for quartz and berlinite. For this material, the velocities are

now quite smaller either for the longidudinal mode or for the shear ones. The shear velocities of LGS are of the same order of magnitude as for gallium phosphate but the longitudinal mode is much faster. The angular variations observed for LGS present some similarities with those computed for quartz and berlinite.

In Figure 6a,c,d we compare the coupling coefficients of quartz, gallium phosphate and langasite. For berlinite, no curve is given since the published sets of constants leads either to too optimistic or too pessimistic results. The exact values for this material would most probably give a curve situated above but quite near to that of quartz. For all the materials considered, we can observe that the angular variations are very similar. This is mostly the consequence of the fact that, in all cases, one of the two independent piezo-electric constants is much larger than the other. We must notice that the coupling coefficient increases in a significant manner from quartz to gallium phosphate. We must also emphasize the fact that for most of the resonators properties, it is the square of the coupling coefficient which must be considered. Then, the differences between the materials become much more important.

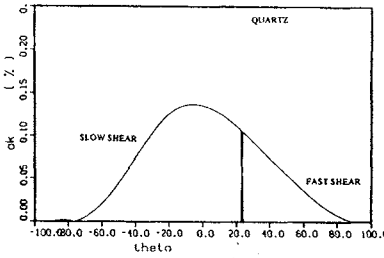


Figure 6a: Coupling coefficient of the piezo-electric shear mode of Y rotated cuts of quartz.

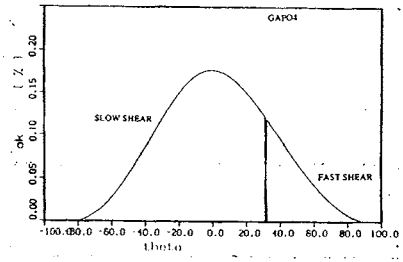


Figure 6c: Coupling coefficient of the piezo-electric shear mode of Y rotated cuts of gallium phosphate.

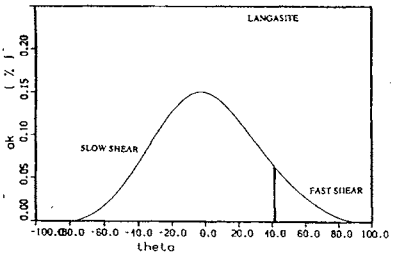


Figure 6d: Coupling coefficient of the piezo-electric shear mode of Y rotated cuts of LGS.

On figures 7a,b,c,d we compare the variations of the first order temperature coefficient of the resonance frequencies (RFTC) of the three modes. Again, the angular variations of the FTC are very similar for quartz and berlinite. Again slightly reduced values are observed for berlinite. For these materials two Y rotated compensated cuts exist. We have demonstrated that the Y-33° cut of berlinite has a 3rd order behaviour very similar to that observed for the AT cut of quartz.

For gallium phosphate, the curves calculated using the last published constants /15/ display a very reduced amplitude of variation around zero. This indicates that gallium phosphate has an outstanding thermal stability. Two Y rotated compensated cuts appear to exist. Only one

of them displays a large coupling coefficient. The angular sensitivity of this cut is much lower than for the AT cuts of quartz and berlinite (reduced value of the slope of the curve at FTC=0).

The angular variations of the FTC of the Y rotated cuts of LGS are very different of the preceding ones. Two compensated cuts exist but again, one only corresponds to a large coupling coefficient. This cut has also a much lower angular sensitivity than AT quartz.

For several of these materials no accurate values of the temperature coefficients of the piezoelectric constants are known. if we suppose that they are of the same order of magnitude as for quartz, slightly different results are found particularly in the angular regions of large coupling coefficient.

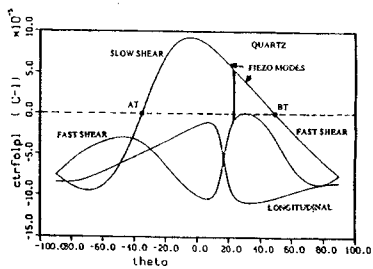


Figure 7a: FTC of the thickness modes in Y rotated cuts of quartz.

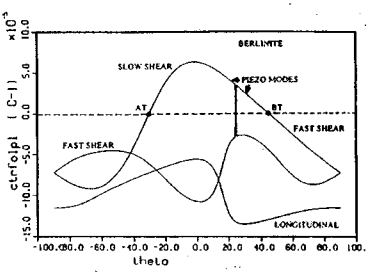


Figure 7b: FTC of the thickness modes in Y rotated cuts of berlinite.

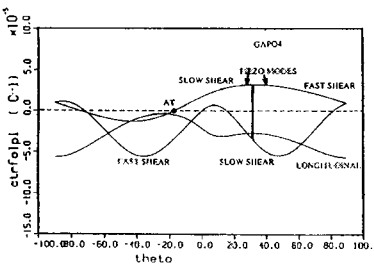


Figure 7c: FTC of the thickness modes in Y rotated cuts of gallium phosphate.

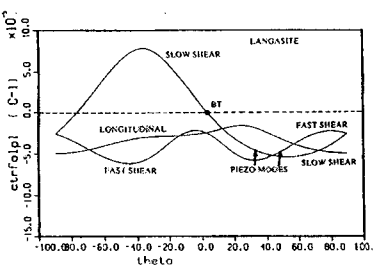


Figure 7d: FTC of the thickness modes in Y rotated cuts of LGS.

As suggested in several recent publications /25/, the numerous possibilities of substitutions of La, Si and Ga atoms of the LGS formula by other elements can permit to vary greatly the angular position and the properties of the compensated cuts.

An important observation is to be made about figures 6 and 7. From quartz to berlinite, gallium phosphate and langasite, the angular position of the most interesting compensated cut shifts towards the small values of theta (in absolute values) and hence towards the region of maximal coupling coefficients.

This fact is probably one of the most important reasons of the large interest of the non-quartz materials of class 32.

3.2 Experimental results:

Many experiments were made to find the compensated cut of berlinite /11/13/. Recently, similar experiments were made concerning gallium phosphate and LGS. Other experiments also were made to evaluate the characteristics of devices made with these latter materials.

In table 2, the experimental results concerning the most interesting Y rotated compensated cuts of berlinite gallium phosphate and langasite are summarized and compared to similar quantities for AT quartz. For all the new materials, the experimental angular position of the compensated cuts is slightly different (some degrees) of the computed value. For berlinite and gallium phosphate this results of the fact that the data used in the computation of these angles are not extremely representative of the better materials now available. For LGS doped crystals were used. These crystals have extremely interesting properties because the compensated cut (inversion at room temperature) is the Y cut, and also, for their energy trapping behaviour (see §4).

The agreement between calculated and experimental results is generally better for the coupling coefficients and the frequency constants than for the temperature coefficients. However, as in previous measurements /6/9/, we have observed that the frequency constants and also the coupling coefficients measured for the best samples of berlinite and gallium phosphate are slightly higher than the computed values. For berlinite, since the material is now stabilized at a high level of quality, a new determination of the constants and of their temperature coefficients will be very useful.

Table 2: Comparison of experimental results (crystals of class 32).

Material	THETA (degree)	k**2	Nfr (kHz mm)	Q (at 5MHz)	FTC2 (10-6/°C2)	FTC3 (10-11/°C3)
Quartz AT	-35.25	.65%	1650	2.7 e6	.0	9-10
					(25°C)	(25°C)
AlPO4 AT	-33.02	1.3%	1471	8.5 e5	.0	8-12
					(80°C)	(80°C)
GaPO4 AT	-13.1	2.5%	1270	7.5 e4	#.0	4-16
					(45°C)	(45°C)
LGS	#.0	2.2%	1381	8.5 e5	-63.	-2.6
					(25°C)	(25°C)

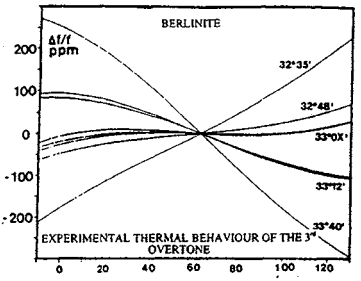


Figure 8: Thermal behaviour near the AT cut of berlinite.

The experimental thermal behaviour observed near the compensated cuts of berlinite, gallium phosphate and langasite are represented in figures 8, 9 and 10. For berlinite, the thermal variations of the resonance frequency near the AT cut are very similar to those of quartz but with a larger inflexion temperature. The value of the third order temperature coefficient of the resonance frequency is given in table 2.

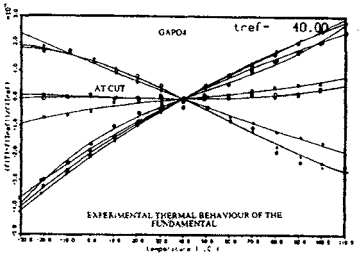


Figure 9: Thermal behaviour for orientations near the AT cut of gallium phosphate.

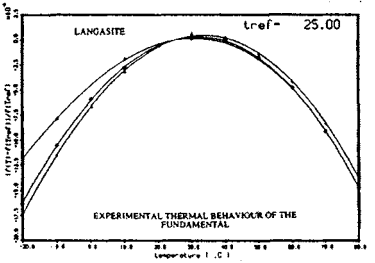


Figure 10: Thermal behaviour for the Y cut of LGS.

For the gallium phosphate, the crystals used for this study display a reduced concentration of OH impurity (of the order of 100 ppm) and the experience acquired with berlinite lead us to believe that the thermal behaviour observed is already quite close of the intrinsic behaviour of high purity $GaPO_4$. The present results indicate a third order behaviour with a cancellation of the second order temperature coefficient in the vicinity of 40°C. In spite of some dispersions, all the examined samples having an orientation near to theta=-13° have displayed an extremely good thermal stability (of the order of 20 ppm in the temperature range [-20,+80°]).

For LGS we have observed parabolic thermal variations similar to those previously indicated by Russian authors /22/. The measured value of the second order temperature coefficient is 63. 10^{-9} °C^{-2}.

During the experiments many cases of strong coupling of the fundamental shear mode with plate modes (mostly flexure) were observed by x-ray topography for gallium phosphate and LGS resonators. Many of the dispersions observed in the measured temperature coefficients may result of this coupling which occur more frequently when rectangular plates are used.

4. ENERGY TRAPPING PROPERTIES.

4.1 Computed results: A comparison of the energy trapping properties of the four materials was made using models of the plane resonators /31/32/33/ based upon the equations given by H.F. Tiersten and co-authors /34/35/36/. On Figure 11 we have represented the angular variations of $\gamma = \sqrt{M_n/P_n}$ the square root of the ratio of the two first coefficients of the equation governing the lateral variations of the main displacement (u_1) in plane resonators. The solutions of this equation, are very dependant on this ratio which condition, when it is real the lateral variations of the mode. When it is imaginary no conventional energy trapping can occur. For the Y rotated cuts of class 32 materials, the M_n coefficients can be real positive or negative numbers or complex number.

For the fundamental mode, we can observe, in figure 11, that γ is always a real (positive) number, greater than unity, which means that the vibration modes excited by circular electrodes are elongated in the direction of the two fold axis x. This lateral anisotropy increases from AT quartz to AT berlinite and AT gallium phosphate while keeping quite moderate values. It is very large for the BT cut of these materials and for langasite. The angular variations of γ are very similar for quartz, berlinite, and, surprisingly, for LGS.

Several resonators properties are very dependant on the mode shape and hence on the value of γ. Among them are the values of the motional elements of the equivalent scheme which depend of a quantity that is approximately the fraction of total acoustic energy confined under the electrodes. For a resonator having round electrodes of a given mass loading, a value of γ very different from the unity leads to a mode shape elongated in one direction and to a value of the inductance which is greater than for γ#1. We have previously demonstrated that the use of elliptical electrodes having an axis ratio equal to γ minimizes the inductance and optimizes the anharmonic spectra (for given values of the electrode area and mass loading). This technique will have a large interest for several of the new materials.

The values of the γ ratio for the third overtone are represented for each material in figure 12. We can observe that there are angular regions where γ is not real so that no conventional energy trapping can occur. For quartz and berlinite resonators, there are two angular regions where energy trapping exists, they are separated by two small regions. For $GaPO_4$ and LGS energy trapping exists only in one angular region which is very small in the case of $GaPO_4$ but include the AT cut. For the third overtones of the Y rotated cuts of the materials of class 32, very large or very small values of γ and hence very anisotropic modes can be encountered.

In figure 13 the values of the γ coefficients computed for the different overtones of LGS in the vicinity of the compensated cut (situated near the Y one) are represented. Very near the Y cut there are several angular regions where no energy trapping exists for a given overtone. Particularly, energy trapping disappears for the third, the 7th, and the 11th overtone at small positive angles, situated close to the Y cut.

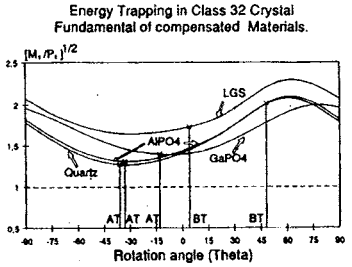


Figure 11: γ coefficient for the fundamental piezoelectric mode (Y rotated cut of class 32 crystals).

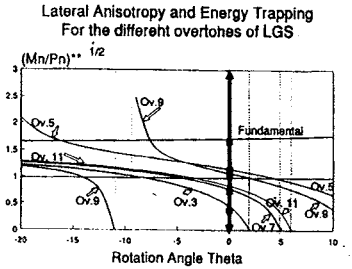


Figure 13: γ coefficient for the different overtones of the piezoelectric mode of Y rotated cuts of langasite.

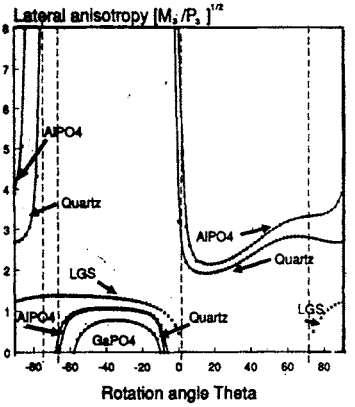


Figure 12: γ coefficient for the 3rd overtone of the piezoelectric mode (Y rotated cuts of class 32 crystals).

For these overtones, this implies rapid variations and hence noticeable angular or "geometric" sensitivities for the anisotropy of the mode. That, in turn, implies noticeable sensitivities for the properties which are directly related to this quantity such as the values of the elements of the equivalent scheme and the anharmonic spectra. This may lead to the need of using more precise orientations than could be required by the low angular sensitivity of the compensated cut.

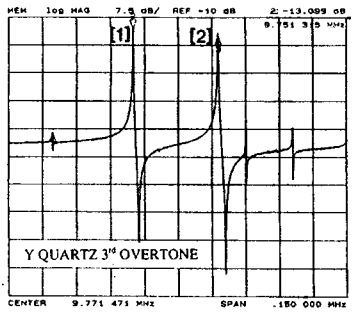
According to the computed values of γ, the fundamental mode extends in the x direction while the 3rd one is elongated in the z direction. The computations predict a more isotropic behaviour for the 5th and the 7th overtones.

On figure 13, we have shown by square dots the experimental observations of the mode shape that were made using X-ray topography.

4.2 Experimentals results:

For the 3rd overtone of Y rotated quartz resonator, it was calculated that the third overtone dissapears very near the Y cut. An example of response curve corresponding to the third overtone for this latter orientation is given in figure 14a (the response curve of a fundamental mode resonator operating at a similar frequency is given in figure 14b). Near the expected resonance frequency of the 3rd overtone, we can observe a non-usual distribution of the modes which present all a large attenuation. The response curve observed for a fundamental mode resonator is very characteristic of energy trapping. X-ray topography observations of the mode shape for all the resonance appearing in figure 14a were made. The topographs indicate very unusual geometries. Four topographs showing the u_1 and u_3 components of the modes of the two most

intense resonances are displayed in figure 15. We can observe that these modes have 2 components of nearly equal magnitude. On the topographs we can also observe that the u_1 component extends up to the edges of the plate. No mode shape similar to those expected for the 1st anharmonic of the 3rd overtone (no nodal lines) was observed.



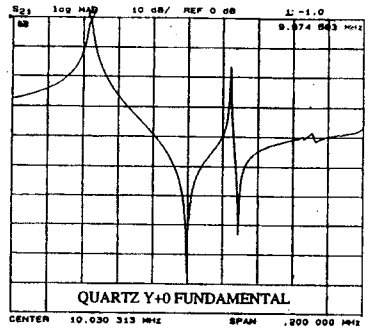


Figure 14a: Response curve near the expected 3rd overtone of a Y cut quartz resonator.

Figure 14b: Response curve near the fundamental mode of another Y cut quartz resonator.

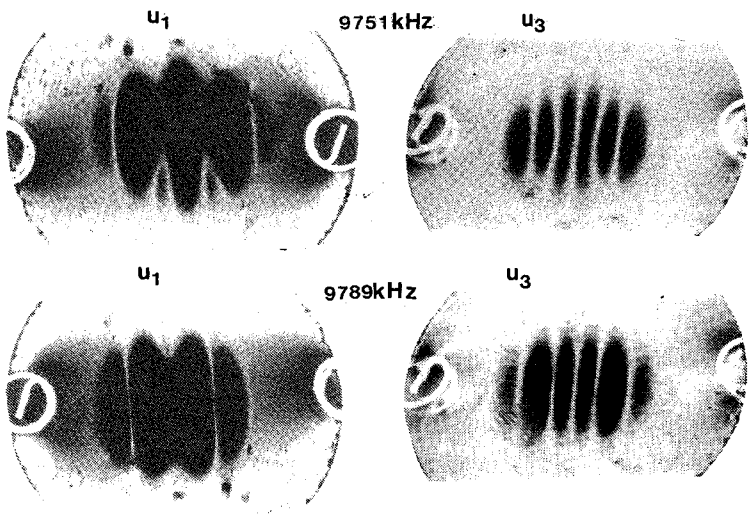
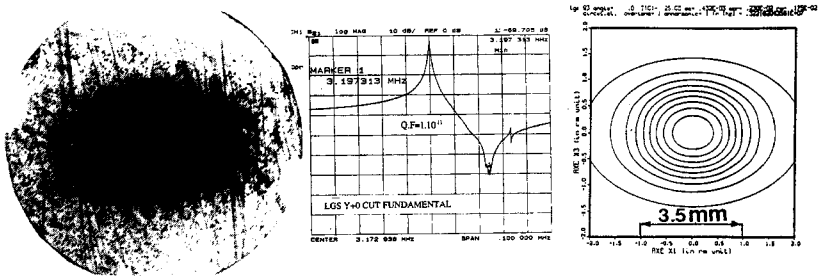


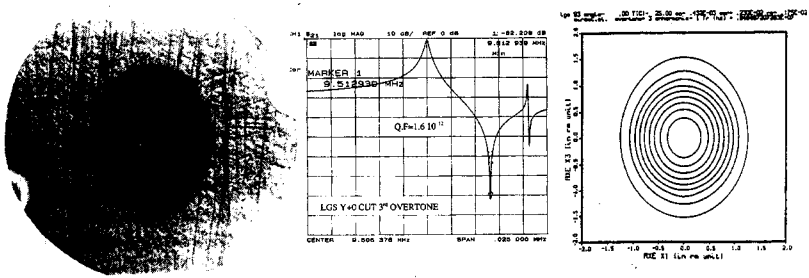
Figure 15: Examples of the u_1 and u_3 components of the modes observed near the 3rd overtone frequency of a Y cut quartz plate (mode at 9751kHZ and mode at 9789 kHz on figure 14a).

Experimental determination of the mode shapes of fundamental and overtone resonators using the compensated cuts of these four materials was also made using synchrotron radiation topography. These observations were compared to mode shapes computed using a model of the plane resonators /31/. It was already observed that these observations for the AT cut of quartz/32/ and the AT cut of berlinite /31/33/ are in a very good agreement with computed ones. It was also observed that the resonance frequencies and the values of the elements of the equivalent scheme are very accurately predicted by the model for quartz. For berlinite, the agreement is generally rather good. For the values of the motional elements of the equivalent scheme, the agreement between computed and experimental values is also good for quartz (some %) and rather good for berlinite, although it seems very sensitive of the parallelism of the plates used for the experiments.

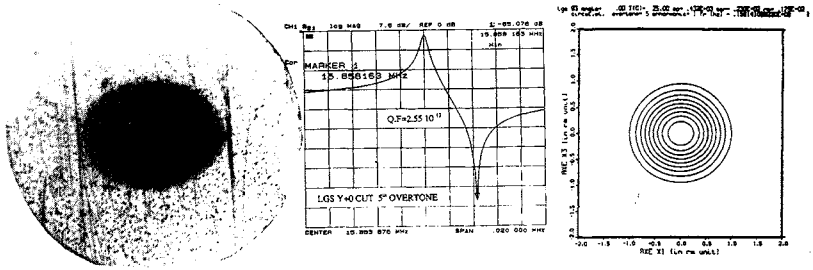
On figure 16 we compare the mode shape experimentally observed for the fundamental, the third and the fifth overtones with the computed ones in the case of a Y cut plane langasite resonator (plate thickness=.432mm, round electrodes 3.5mm diameter, mass loading= .23 %). The experimental response curve is also displayed for each of these modes. For these modes the agreement between the experimental mode shapes, and the computed ones is good. As predicted by the calculations, the direction of maximum extension of the 3rd overtone (z) is different of those existing for the fundamental and the fifth overtone (x). The anisotropy of the fundamental mode is very important. We can notice on the response curve that the Q factor is very high for the 3rd and the 5th overtones.



LGS Y+0 CUT FUNDAMENTAL

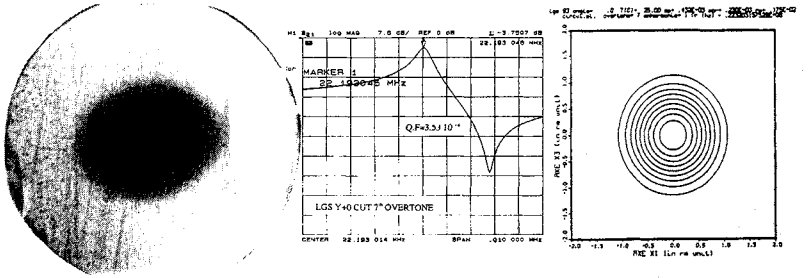


LGS Y+0 CUT 3rd OVERTONE

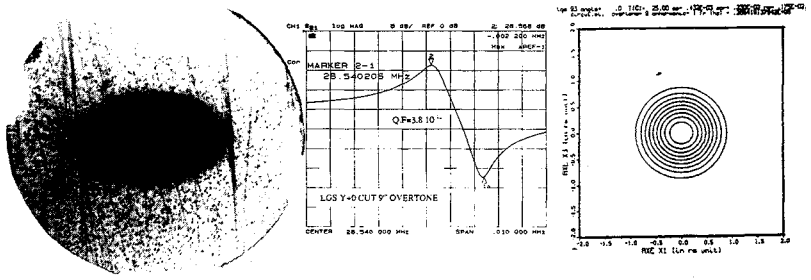


LGS Y+0 CUT 5th OVERTONE

Figure 16: Experimental, computed mode shapes and response curves for the fundamental, the 3rd and the 5th overtones of a langasite plate.



LGS Y+0 CUT 7th OVERTONE



LGS Y+0 CUT 9th OVERTONE

Figure 17: Experimental, computed mode shapes and response curves for the 7th and the 9th overtones of the same langasite plane resonator as in figure 16.

Higher overtones of langasite plane resonators (up to the 9th one) display useful responses and present very large Q.f products. On figure 17, we have represented the electrical responses, the experimental and the computed mode shapes corresponding to the 7th and the 9th overtones of the same plate as in figure 16. The agreement between experimental and computed results is not as good as for the first overtones, may be due to some imprecision in the material

constants (the calculation of the coefficients of the Tiersten equation involves to evaluate weighted difference of several constants, so that, they are sensitive to small imprecisions of the material constants).

CONCLUSION:

The properties of the most important piezo-electric materials of class 32 were compared. It was observed that all these materials can be obtained with a high crystalline perfection. The calculations and the experiments have shown that quartz and its structural analogs possess an outstanding high thermal stability. They have also shown that the energy trapping properties and hence the design rules of devices of the new materials can be very different of those of quartz.

The larger coupling coefficients of $GaPO_4$ and LGS allow to obtain much larger shift for the oscillators and bandwidth for the filters. It permits also to use overtones with much higher ranks than for quartz.

LGS displays extremely reduced acoustic losses as compared to quartz and is the prototype of a series of analogs which may also have very interesting properties (including probably for some of them, very large coupling coefficients). LGS is very promizing for applications requiring very low phase noise.

Gallium phosphate possesses an outstanding set of qualities and must be further developped to obtain the large and highly perfect crystals required for the applications.

On the whole, it appears that the piezo-electric materials of class 32 have very interesting, complementary properties that can cover a very broad range of applications for frequency generation and filtering. The new materials of this crystal class can surpass greatly quartz for all the applications requiring a high coupling coefficient, and also for many other ones demanding a new level of performances on other characteristics (such as: lower losses or reduced phase noise, much reduced dimensions, lower impedance level, high temperature operations, etc....). They are particularly adapted to the new applications to filtering and frequency generation requiring a high level of performances such as in telecommunication equipments/1/.

Acknowledgements: The authors acknowledge the assistance of MM. Caumont, LeThuault and Mrs Daniel for the elaboration of the studied samples and the helpful contributions of MM Joly, Guiglielmoni and Mrs Biblemont .

References:

/1/ J.Detaint, J.Schwartzel; Filtering and Frequency Control for the next Generations of Radiocommunication Systems Proc. 8th European Frequency and Time Forum (these proceedings) Münich March 1994.
/2/ E.D. Kolb, R.A.Laudise, J. Cryst. Growth v.43 p.313 (1978)
/3/ H.Poignant, L.LeMareschal, Y.Toudic; Etude de la solubilité du phosphate d'aluminium dans les solutions d'acide phosphorique; Mat.Res.Bull. vol. 14 p.603-612 (1979).
/4/ B.H.T. Chai, M.L.Shand, E.Buehler, M.A.Gilleo; Experimental data on piezoelectric properties of berlinite; Proc.1979 IEEE Ultrasonic Symposium p.577 (1979).
/5/ J.Detaint, E.Philippot, J.C.Jumas, J.Schwartzel, A.Zarka, B.Capelle, J.C.Doukhan; Crystal growth, physical characterization and BAW devices applications of berlinite; Proc. 39th Annual Frequency Control Symp. p. (1985)
/6/ J.Detaint, A.Zarka, B.Capelle, Y.Toudic, J.Schwartzel, E.Philippot, J.C.Jumas, J.C.Doukhan; Berlinite : Characterization of crystals with a low water concentration and design of bulk wave resonators. Proc. 40th Annual Frequency Control Symp. p.101,114 (1986).
/7/ B.H.T. Chai , J.P.Hou; Evaluation of berlinite growth at high temperature, Proc. 1st European Frequency and Time Forum p.204 (1987).
/8/ E.Philippot, A.Goiffon, J.C. Jumas, C.Avinens, J.Detaint, J.Schwartzel, A.Zarka; New advance in crystal growth of high purity berlinite; Proc. 42nd Annual Frequency Control Symp. p.138-145 (1988).
/9/ E.Philippot, A.Goiffon, M.Maurin, J.Detaint, J.Schwartzel, B.Capelle, A.Zarka. Evaluation of high quality berlinite crystal grown in sulphuric acid medium; J.Crystal Growth v.104 p.713-726 (1990).
/10/ X.Buisson, A.Zarka, R.Arnaud, B.Capelle, Y.Zheng, J.Schwartzel, J.Detaint, Y.Toudic; Crystal growth of berlinite: methods for growing longer crystals and improving the quality. Proc. 6th European Frequency and Time Forum p.287-291 (1992).
/11/ J.Detaint, J.P.Aubry, J.P.Michel, A.Zarka, B.Capelle, R.Lefevre, E.Philippot, J.Schwartzel; Piezoelectric devices using berlinite: Properties and applications. Proc. 1989 URSI symposium on Signals Systems and Electronics p. 433-438 (1989).
/12/ G.F.Engel, P.W.Krempl, J.Stadler Proc. 3rd European Frequency and Time Forum p.50-56 (1989).

/13/ P.Krempl, J.Stadler, W.Wallnöfer, W.Ellmeyer, R.Selic. Proc. 5th European Frequency and Time Forum p.143-147 (1991).
/14/ E.Philippot, A.Ibanez, A.Goiffon, B.Capelle, A.Zarka, J.Schwartzel, J.Detaint. Crystal growth and physical characterisation of $GaPO_4$; Proc. 6th European Frequency and Time Forum p.383-388 (1992).
/15/ W.Wallnöffer, J.Stadler, P.Krempl, Temperature dependance of elastic constants of $GaPO_4$ and its influence on BAW and SAW devices. Proc. 7th European Frequency and Time Forum p.653-657 (1993).
/16/ J.Detaint, J.Schwartzel, C.Joly, A.Zarka, B.Capelle, Y.Zheng, Y.Toudic, E.Philippot; Energy trapping resonators and integrated filters using the AT cut of quartz like materials. Proc. 1992 IEEE Int. Frequency Control Symposium p.639-647 (1992).
/17/ E.Philipot, A.Ibanez, A.Goiffon, M.Cochez, A.Zarka, B.Capelle, J.Schwartzel, J.Detaint A quartz like material: Gallium phosphate $GaPO_4$, crystal growth and characterization. j.Crystal growth v.130 p.195-208 (1993).
/18/ B.Capelle, A.Zarka, J.Schwartzel, J.Detaint, Y.Zheng, A.Ibanez, E.Philippot; X-ray topography study of gallium phosphate crystals and resonators, Proc. 1993 IEEE Int. Frequency Control Symposium p.813-820 (1993).
/19/ W.Wallnöfer, P.W. Krempl; A theoretical study of surface acoustic waves in layered quartz homeotype structures Proc. 8th European Frequency and Time Forum (these proceedings).
/20/ M.Cochez, J.D.Foulon, A.Ibanez, A.Goiffon, E.Philippot, A.Zarka, J.Schwartzel, J.Detaint. Crystal growth and characterisation of a quartz like material: $GaPO_4$; to appear in the Proc. of the 1st European Workshop on Piezoelectric Materials Dec.1993 Montpellier France.
/21/ A.A. Sternberg, O.V.Svereva, Yu.M.Mininzon, L.N. Demianets; Hydrothermal growth of OH-group free crystals $AlPO_4$ and $GaPO_4$: The way of twin reducing. To appear in the Proc. of the 1st European Workshop on Piezoelectric Materials Dec.1993 Montpellier France.
/22/ S.A.Sakharov, I.M.Larionov, A.V.Medvedev; Application of Langasite Crystals in monolithic filters operating in shear modes; Proc. 1992 IEEE Int. Frequency Control Symposium p.713-723 1992.
/23/ A.N. Gotalskaya, D.I. Drezin, V.V.Bezdelkin, V.N.Stassevich Pecularities of Technology, Physical properties and Applications of a new Piezoelectric Material: Langasite ($La_3Ga_5SiO_{14}$); Proc. 1993 IEEE Int. Frequency Control Symposium p.339-346 (1993).
/24/ I.M.Silvestrova, V.V.Bezdelkin, P.A.Senyushenkov, Yu.A.Pisarevsky. Present State of $La_3Ga_5SiO_{14}$ Research. Proc. 1993 IEEE Int. Frequency Control Symposium p348-350. (1993).
/25/ I.M. Silvestrova, Yu.V.Pisarevsky, V.V. Bezdelkin, P.A. Senyushenkov; New Piezoelectric Materials; Proc. 1993 IEEE Int. Frequency Control Symposium p.351-352 (1993).
/26/ V.V. Bezdelkin, E.E. Antonova Fundamental characteristics of thickness vibrations of langasite plates for BAW resonators. to appear in the Proc. of the 1st European Workshop on Piezoelectric Materials Dec.1993 Montpellier France.
/27/ A.Zarka, B.Capelle, Y.Zheng, J.Detaint, J.Schwartzel; Stroboscopic X-ray topography of quartz resonators. Proc. 42nd Annual Frequency Control Symposium p.85-92 (1988).
/28/ B.Capelle, A.Zarka, J.Schwartzel, J.Detaint, E.Philippot, J.P.Denis; Characterization of Piezoelectric Materials: Old and New Crystals; to appear in the Proc. of the 1st European Workshop on Piezoelectric Materials; Dec.1993 Montpellier France.
/29/ B.J. James A new measurement of the elastic and dielectric constants of quartz; Proc. 42nd Annual Frequency Control Symp. p.146-154 (1988).
/30/ D.S. Bailey, J.C.Andle, D.L.Lee, W.Soluch, J.F.Vetelino, B.H.T.Chai Temperature dependance of the material and acoustic wave parameters of berlinite; Proc. 1983 IEEE Ultrasonic Symp. p.335-340.
/31/ J.Detaint, H.Carru, J.Schwartzel, C.Joly, E.Philippot; Optimization of the design of thickness shear devices using the AT cut of berlinite; Proc. 5th.European Frequency and Time forum p.30-40 (1991).
/32/ J.Detaint, J.Schwartzel, C.Joly, E.Philippot, Energy Trapping in Plane and Corrugated Resonators; Proc. 41st Annual Frequency Control Symposium p.314-324 (1987).
/33/ J.Detaint, H.Carru, J.Schwartzel, C.Joly, B.Capelle, A.Zarka. Determination of the cut-off frequency of unelectroded plates and effects of local thickness modifications in plane resonators. Proc. 43rd Annual frequency Control Symposium p.563-573 (1989).
/34/ H.F. Tiersten; Analysis of trapped-energy resonators operating in overtones of coupled thickness shear and thickness twist. J.Acoust. Soc. Amer. v.59 n°4 p.879-883 (1976).
/35/ H.F.Tiersten, R.C.Smythe; An analysis of contoured resonators operating in overtones of coupled thickness shear and thickness twist. J.Acoust Soc. Amer.v.65 n°6 p.1455-1460 (1979).
/36/ D.S Stevens, H.F.Tiersten; An analysis of doubly rotated quartz resonators utilizing essentially thickness modes with transverse variations; J.Acoust.Soc. Amer.vol.79 n°6 p.1811-1825 (1986).

MICROMACHINING OF QUARTZ CRYSTALS : EXPERIMENTS AND THREE-DIMENSIONAL SIMULATION OF ETCHED SHAPES

C.R. TELLIER and D. BENMESSAOUDA
Laboratoire de Chronométrie, Electronique et Piézoélectricité,
La Bouloie - Route de Gray - 25030 BESANCON CEDEX - FRANCE

SUMMARY

Experimental results on the micromachining of mesa and diaphragms on differently-oriented quartz plates are presented. The etched shapes which show marked orientation effects are analysed in terms of the tensorial model for the anisotropic chemical etching. A 3D numerical simulation which works with the dissolution slowness surface of quartz crystal is used to predict final shapes for etched structures.

1 - INTRODUCTION

During the last decade several numerous attempts have been made to apply the photolithographic process to the micromachining of mechanical devices constituting the sensing element of quartz resonators or sensors. Thus the need for simulation tools capable to predict the exact geometrical configuration of structures has considerably increased over the past few years.

Actually these tools are often based on the kinematic model proposed by Frank /1/ which offers the possibility to construct geometrically only the etching shapes composed of limiting planar facets. However new concepts /2, 3/ which use a tensorial analysis of the anisotropic dissolution allow us to derive numerically 2D /4/ and 3D etched shapes. The procedure requires that the dissolution constants involved in the equation for the dissolution slowness surface were determined from experiments. Up to now tentative have been made /3-5/ to predict shapes for two-dimensional dissolution profiles essentially. In this paper emphasis is placed on experimental results and on theoretical 3D predictions related to the wet etching of diaphragms and mesa on various singly-rotated quartz plates.

2 - THEORETICAL BASIS

2.1 Theoretical framework

A crystal surface can be decomposed in a succession of planar surface elements, **ds**, whose orientation are defined by means of two angles of cut φ and θ (IEEE standard). In a tensorial analysis /2/ of the anisotropic chemical attack a dissolution vector, **L**, of magnitude, $L(\varphi, \theta)$ is associated to each surface element. This vector lies parallel to the inward normal, **n**, to **ds** and its magnitude is equal to the reciprocal of the normal etch rate. The magnitude, $L(\varphi, \theta)$, can be expressed in the form /2, 3/ :

$$L(\varphi, \theta) = D_o + D_i n_i + D_{ij} n_i n_j + D_{ijk} n_i n_j n_k + \ldots \qquad (1)$$

involving dissolution tensors D_i, D_{ij}, D_{ijk},... and the cartesian components, n_i, of **n**.

The number of dissolution constants is reduced by the symmetry of the point group related to the crystal into consideration. For the quartz crystal which belongs to class 32 the calculations performed up to tensors of rank $N_{max} = 10$ gives the final compact equation :

$$L(\varphi, \theta) = S_1 + S_2 + S_3 + S_4 \qquad (2)$$

with

$$S_1 = \sum_{j=0} A_j (\cos\theta)^{2j} \quad , \quad S_2 = \sum_{L=0} \sum_{k=1} B_{k,L} (\cos\theta)^{2k+1} (\sin 3\varphi)^{3L}$$

$$S_3 = \sum_{M=0} \sum_{l=0} C_{l,M} (\cos\theta)^{2l+1} \sin\theta (\cos 3\varphi)^{3M} \qquad (3)$$

$$S_4 = \sum_{m=3} D_m \sin\theta (\cos\theta)^{2m} \cos 3\varphi \sin 3\varphi$$

where L and M take non-zero values for $N_{max} \geq 9$.

The coefficient of series A_j, $B_{k,L}$, $C_{l,M}$ and D_m are expressed in terms of the independent dissolution constants which, for tensor of rank 7 to 10, are listed in table 1.

Tensor rank	Independent dissolution constants
7	$D_{5(1)\,1(2)\,1(3)}$, $D_{5(1)\,2(3)}$, $D_{3(1)\,4(3)}$, $D_{7(1)}$
8	$D_{7(2)\,1(3)}$, $D_{5(2)\,3(3)}$, $D_{3(2)\,5(3)}$, $D_{6(1)\,2(3)}$, $D_{4(1)\,4(3)}$, $D_{2(1)\,6(3)}$, $D_{8(1)}$, $D_{8(3)}$
9	$D_{7(1)\,1(2)\,1(3)}$, $D_{7(1)\,2(3)}$, $D_{5(1)\,1(2)\,3(3)}$, $D_{5(1)\,4(3)}$, $D_{3(1)\,6(3)}$, $D_{9(1)}$
10	$D_{9(2)\,1(3)}$, $D_{7(2)\,3(3)}$, $D_{5(2)\,5(3)}$, $D_{3(2)\,7(3)}$, $D_{8(1)\,2(3)}$, $D_{6(1)\,4(3)}$ $D_{4(1)\,6(3)}$, $D_{2(1)\,8(3)}$, $D_{10(1)}$, $D_{10(3)}$

Table 1 : Identification of the independent dissolution constants.

It is obvious that the angles of cut φ and θ vary the extremity of **L** generates in space a surface called the **dissolution slowness surface**.

Moreover in the framework of a kinematic description **/1/** of the anisotropic chemical attack and under the assumption that the dissolution process is only governed by orientation, a moving surface element propagates into the crystal along a linear characteristic trajectory. Thus we can associate to each surface element a propagation vector, **P**, which characterizes the displacement of **ds** within the crystal. The cartesian components, (dx, dy, dz), of **P** can be evaluated **/6/** from the equation, $L(\varphi, \theta)$, expressing the dissolution slowness surface.

2.2 Procedure for the determination of the dissolution slowness surface

In general to derive numerical simulations of etched crystals we start with the equation, $L(\varphi, \theta)$, to calculate vectors **P** related to all the surface elements potentially present at the starting surface. Surface elements which disappear during etching are eliminated in further steps.

Thus in practice we have to extract values for the dissolution constants from experiments. The iterative procedure proposed by C. TELLIER **/7/** uses two types of experimental bi-dimensional etched shapes :

1 **Cross-sectional out-of-roundness profiles** related to initially circular thick plates. Effectively the anisotropic etching induces the formation of limiting planar facets connected with minima in the corresponding polar plot of **L**. As a result the out-of-roundness profile related to a plate (φ_o, θ_o) ressembles crudely to the polar graph in the same plane of cut φ_o, θ_o **/5/**.

2 **Surface profilometry traces** made along different directions on various etched plates. Since the final shape of an etched profile is intimately correlated to the nature and the number of extrema for **L** present in the vicinity of the reference surface **/2, 4/** profilometry traces are used principally in the verification stage to detect the possible presence of minor extrema.

3 - BI-DIMENSIONAL ETCHED SHAPES AND THE DISSOLUTION SLOWNESS SURFACE

3.1 Experimental results

Differently-oriented quartz plates were etched in a concentrated $NH_4F.HF$ solution. In this section we present some data which concern the changes in etched shapes of cross-sections (for starting cylindrical crystals) with the orientation (φ_o, θ_o) of the cross-sectional planes. About 40 different cuts were investigated but here we present some results corresponding to specific orientations which seem critical for the determination of the true shape of the dissolution slowness surface. In the following, plates obtained by a single rotation of $\pm\ \theta_o$ about the X-axis are called respectively AT-θ_o and BT-θ_o cuts whereas the doubly-rotated plates corresponding to a rotation of $\pm\ \theta_o$ about the Y-axis are designated as Y $\pm\ \theta_o$ cuts. Some typical results are shown in Figure 1a, b, c. The out-of-roundness graphs exhibit successive peaks (P) and Valleys (V) where valleys are caused by the development of planar limiting facets. Moreover the etched profiles satisfy the symmetry of point group 32, specially the two-fold symmetry about the X-axis for the out-of-roundness profiles related to AT-35 and BT-50 sections.

3.2 Theoretical shapes and discussion

We make use of all experimental (φ_o, θ_o) out-of-roundness profiles which furnish crude "image" of corresponding (φ_o, θ_o) polar plots of **L** to determine several sets of dissolution constants. We finally choose two sets named laws "18 bis" and "QI" which give different shapes for the disssolution slowness surface.

The discussion is essentially based on a systematic comparison between theoretical and experimental etched shapes. For the prediction of cross-sectional shapes a bi-dimensional simulation described else where **/4/** is sufficient. For the two laws the graphical simulation gives the theoretical shapes shown in Figure 1. For the law "18 bis" we obtain a relative agreement between the theoretical and experimental out-of-roundness profiles related to singly-rotated plates with θ_o in the range [- 90°, 70°].

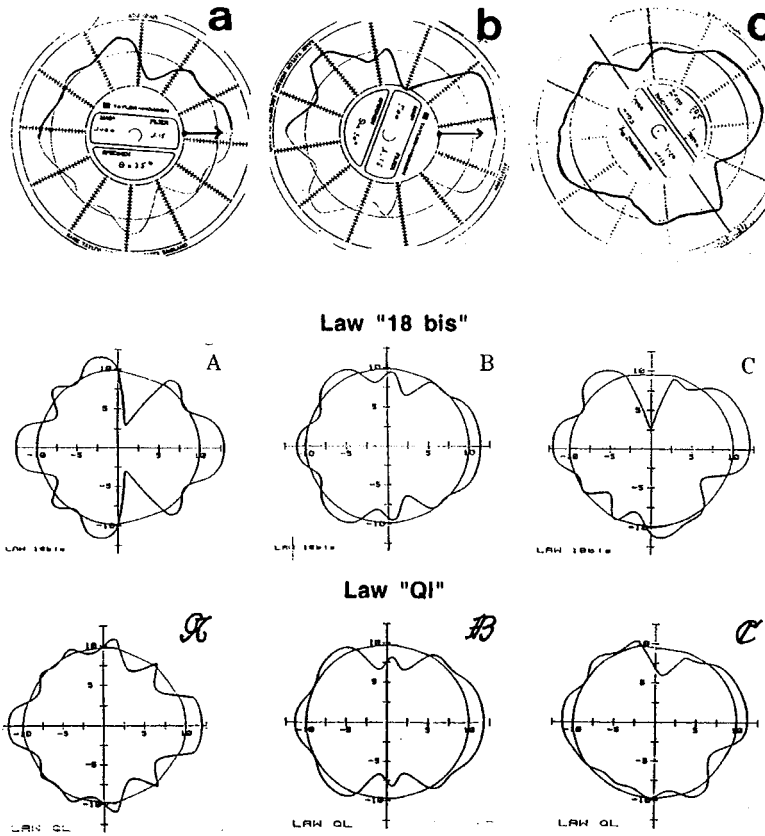


Figure 1 : Experimental out-of-roundness graphs for AT-35 **(a)**, BT-50 **(b)** and Y-40 **(c)** sections. **A**, **B**, **C** : corresponding theoretical profiles as derived from law "18 bis" and 𝔄, 𝔅, ℭ : from law "QI"

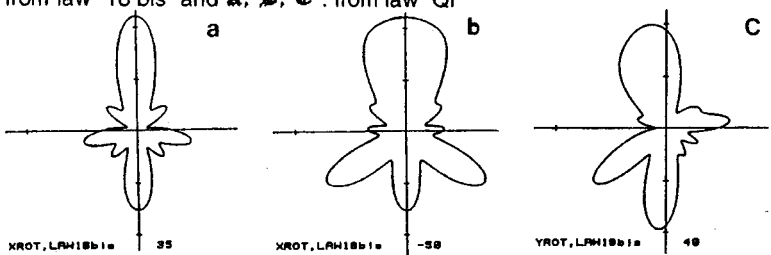


Figure 2 : Polar plots for the law "18 bis" lying in AT-35 **(a)**, BT-50 **(b)** and Y-40 **(c)** cross-sectional planes.

This agreement is conveniently illustrated by Figure 1a, b, A, B where the valleys in theoretical graphs occur for angles whose values coïncides with experimental values. The adequation is not so good for Y-rotated plates the discrepancies affect essentially the accentuation of peaks and valleys rather the number and angular positions of extrema (compare Figure 1 c and C for example).

Some theorectical results for the law "QI" are displayed in Figure 1𝔄, 𝔅, ℭ. For this law we observe several deviations between theory and experiments : some graphs (AT-35 and Y+40 graphs in Figure 1𝔄,ℭ) exhibit more peaks and valleys than in corresponding experimental graphs and relative amplitudes between theoretical successive extrema disagree with experimental observations. Thus from this rapid comparison we infer that the law "18 bis" seems more adequate thant the law "QI". Results on three-dimensional etching shapes will give complementary informations useful for the adjustment of the slowness surface.

4 - THREE-DIMENSIONAL ETCHED SHAPES

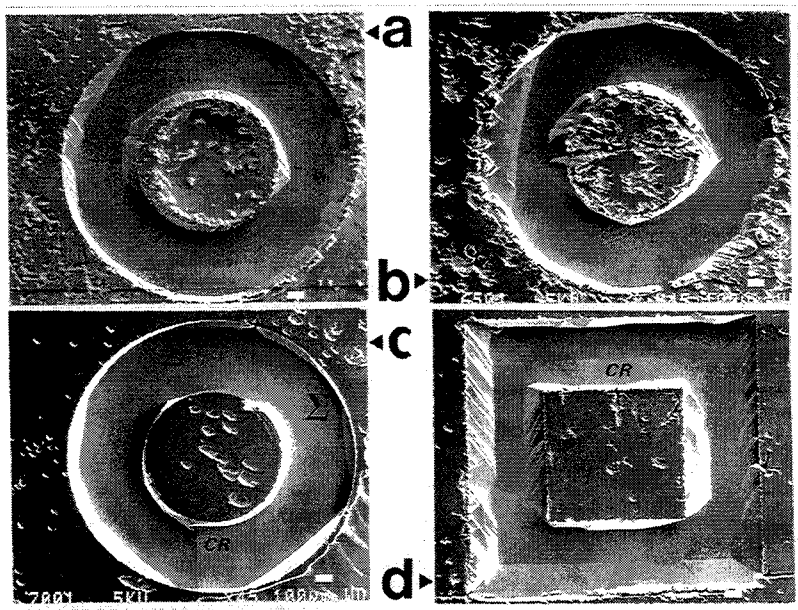
4.1 Micromachined structures

Differently-oriented "AT" and "BT" plates were used in experiments. The masking material was an evaporated Cr-Au film which suffered a specific thermal tratment. Combinations of circular and square-shaped masking patterns were employed for the photolithography process. The samples were etched at a constant temperature 65°C ± 1°C. Figure 3 shows some typical structures for which the depth of etch is about 100 µm. Here we concentrate our attention on the laterally under-etched walls which develop along the edge of an inert mask. The SEM micrographs reveal that the surfaces forming the side of the etched patterns possess slopes and textures which depend strongly on the angle θ_o. In some cases (Figure 4) the facets which limit the side of mesa and hole can be qualified of "planar facet" in many cases the lateral underwall appears as fragmented with a rather vertical "facet" which develops directly under the mask and a somewhat curved shoulder which joints the etched reference surface (see regions CR in Figures 3b and 3d). The extent of these curved regions which cannot be identified with a particular crystal plane can reach several hundred µm because they are generally slightly disoriented with respect to the reference surface. Moreover etch bordering planar and curved facets also form sharp edges which penetrate within the crystal under the mask. We depict such sharp edges on the SEM micrographs of Figure 3a, b, c (regions Σ) but it is difficult to estimate the inclination of such etched sides.

Moreover we also remark that for the various "AT" and "BT" plates the etching does not induce very marked changes in the initial circular shape of pattern. The anisotropy of the chemical micromachining seems moderate because the so-called convex corner underetching plays for quartz crystals a less significant role than for silicon crystals /7/. This behavior can be interpreted in terms of smaller anisotropy ratio between extrema of **L** (of same nature) relative to "crystal planes".

4.2 Three-dimensional simulation of micro-machined structures

The simulation program works with doubly rotated surface elements whose orientations are now defined by four angles : θ_o, φ_o, ψ and α and calculatesthe propagation vector **P**. The time of calculation increases considerably because tests to distinguish between diverging and converging trajectories take now many times.



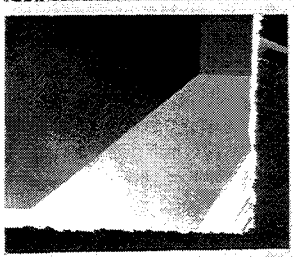


Figure 3 : Mesa and diaphragms etched on AT-55 **(a)**, AT-65 **(b, d)** and BT-70 **(c)** plates

Figure 4 : Formation of planar facets at the corner of a "square" BT-70 hole.

Actually two types of graphical representations are proposed (Figure 5). Firstly we can draw the constant level contours of mesa and holes etched on a reference surface (φ_o, θ_o). A cross-sectional representation of the etched mesa and hole sides allows us to evaluate the lateral underetching and slopes of etched sides. The cross-sectional plane intersects the reference surface along a linear trace L_i which makes an angle ψ_i (Figure 5A) with the rotated axis x'_1.

Theoretical shapes for holes and mesa etched in various singly-rotated plates are given in Figures 5 to 9 as derived from laws "18 bis" and "QI". In a general point of view the 3.D simulation predicts moderate changes in the starting general circular shape of structures in close agreement with experimental observations. Moreover the simulation gives evidence for the development of "planar" facets and curved regions whose orientation and extent is determined by the etching law, i.e. by the slowness surface shape. Thus in the following we undertake separate comparisons of the theoretical predictions with the experimental shapes for the two laws.

4.2.2 The law "18 bis"

Let us turn our attention to Figure 5 which shows theoretical predictions related to the micro-machining of a starting circular hole on an AT-65 plate. Dissolution profiles (Figure 5.1 and 5.2) corresponding to cross-sections which intersect the reference plane along line L_1 and L_2 confirm that edges with obtuse angle α_u are formed in good agreement with the features revealed by the SEM micrograph of Figure 3b. The cross section related to the trace L_4 shows the development of a vertical wall of small extent and of a curved facet which joints this wall to the hole bottom. Moreover the L_5 and L_6 cross-sections are composed of two successive quasi-planar and somewhat curved facets. It is clear that in the SEM micrograph of Figure 3b we can recognize the presence of such facets (facets f_1 and f_2) for directions of alignment which coïncide with the theoretical intersecting lines. The formation of a curved region ("CR" in Figure 5A) for ψ_{Li} in the vicinity of 360° is also in accord with theory.

The numerical simulation predicts a quite similar behavior for an hole etched on an AT-55 plate in particular sharp edges and obtuse angles are formed for quite identical directions of alignment in agreement with the SEM micrograph of Figure 3a. However experiments depart from predictions when we consider the changes in the angle of inclination and in extent of the curved facet related to trace L_3 with the angle of cut. Effectively the theoretical extent of facet "L_3" seems too large as compared with Figure 3a ; but care must be taken to the difficulty to appreciate the "real" extent of a slightly disoriented facet on an SEM image.

Turning our attention to results (Figure 7) related to a BT-70 plate we note that dissolution profiles with obtuse angles are formed (Figure 7.1, 7.2 for example) on a large angular sector i. e. for directions of alignment with ψ_{Li} in the range (20°, 120°). Experimental observations (Figure 3c) corroborate this prediction. Moreover the limiting facets f in Figures 7.1, 7.2, 7.3 develop in theory for intersecting lines L_1, L_3 and L_4 which do not deviate markedly from the experimental traces. A careful examination of the SEM micrograph of Figure 3c shows also the presence of a slightly disoriented curved region (CR) which can be depicted in the theoretical shape for ψ_i around 335°.

Etched shape for mesa can be also derived from the 3D numerical simulation. At this point it should be noticed that in a 3D simulation which gives the etched shape of mesa a main difficulty arisesbecause we are concerned by intersecting surface elements which form intersections of different nature. Effectively only concave intersections are potentially present at a cross-sectional plane whereas in a top view the mesa appears to be composed by a succession of convex intersections. Results presented here are obtained by assuming that the final shape of etched mesa is primarily governed by dissolution criteria /1/ related to concave intersections. A typical result is viewed in Figure 8 for a reference surface BT-70. The agreement between predicted and experimental shapes is equally satisfactory for both an etched hole and an etched mesa specially when we consider the coïncidence of regions where edges with an obtuse angle develop. From this observation we infer that the above assumption is justified.

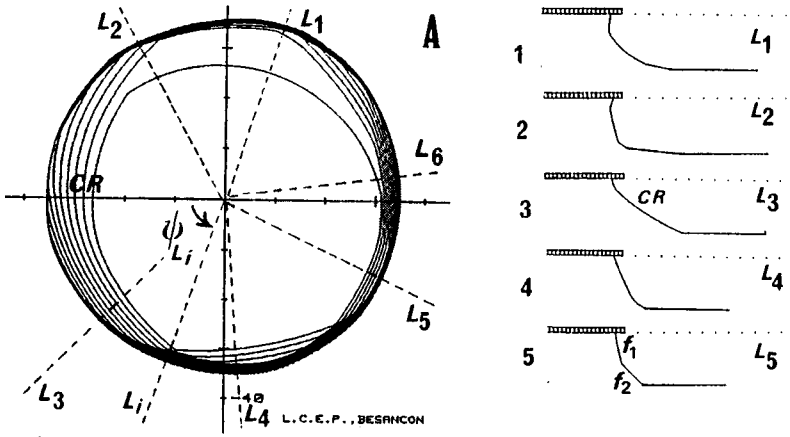


Figure 5 : Theoretical constant level countours and cross-sectional profiles for an AT-65 hole as derived from the law "18 bis".

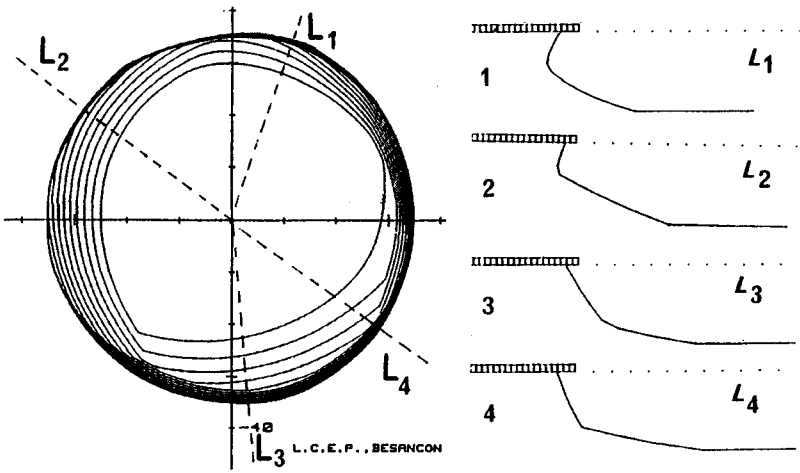


Figure 6 : Predicted shapes for an AT-55 hole as derived from law "18 bis".

4.2.3 The law "QI"

Typical predictions are displayed in Figure 9. We observe that shapes of holes etched on AT-55 and AT-65 plates ressemble to shapes predicted from the law "18 bis" ; in particular we obtain sidewalls characterized by obtuse angles in quite similar regions. Only the extent, E_{CR}, of curved regions seems to be affected by differences in the two etching laws. Turning our attention to the evolution in E_{CR} with the orientation θ_o we observe much departure for law "QI" than for law "18 bis".

Let us now examine the predicted shapes of Figure 9 with the experimental shapes for hole and mesa etched on BT-70 plates. For ψ_{Li} in the range (0°, 180°) laws "18 bis" and "QI" behave rather similarly with for holes the development of edges with obtuse angle and of limiting facets. Only inclinations and directions of alignment of facets, exhibit some differences. Deviations are more marked when we consider, for holes, the presence of curved shoulders in the ψ_{Li}-range (180°, 360°). Effectively the law "QI" leads without ambiguity to the formation of two curved regions S_1 and S_2 (Figure 9) whereas law "18 bis" gives rise essentially to the shoulder S_2 (Figure 7). For this cut a comparison with experiments does not allows us to undertake a judicious choice between the two laws.

5 - CONCLUSION

In conclusion it remains difficult to decide what is the best law. From a systematic comparison of theoretical bi-dimensional and three-dimensional shapes with data the law "18 bis" seems a little more appropriate. The agreement between theory and data concerns in particular the presence of sharp edges and of limiting facets in specific regions whereas deviations are observed when we consider the changes in the extent of the inclined and curved shoulder with the angle of cut. A tentative can be made to explain this departure. For an hole two intersecting surface elements, **ds**, potentially present at the starting surface form concave intersections exclusively. Then according to the stability criteria stated by Irving /1/ the limiting facets which develop when we etch an hole on a plane (φ_o, θ_o) are associated with maxima in L occuring for specific orientations $(\varphi_o, \theta_o, \psi_M, \theta_M)$. It is clear that values of the derivatives $dL(\psi)/d\psi$ and $dL(\alpha)/d\alpha$ for orientation (ψ, α) in the vicinity of (ψ_M, α_M) determine the extent of facets because for accentuated maxima the characteristic trajectories of **ds** diverge very rapidly. Thus relative amplitude and sharpness of successive maxima in L play an important role in the formation and the extent of quasi-planar and curved facets. Most of the departures between theoretical and experimental shapes observed here can be explained on this basis. In these conditions improvement in the 2D and 3D numerical simulation of structures etched in quartz passes through an adjustment of the relative amplitude of successive extrema in the dissolution slowness surface rather than a correction of angular positions of extrema. This adjustment is continously in progress actually and we expect to propose rapidly a dissolution slowness surface in very close agreement with the true surface.

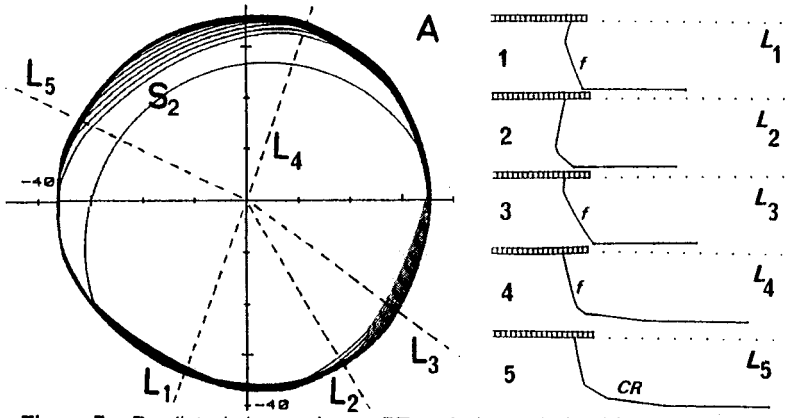


Figure 7 : Predicted shapes for an BT-70 hole as derived from the law " 18 bis".

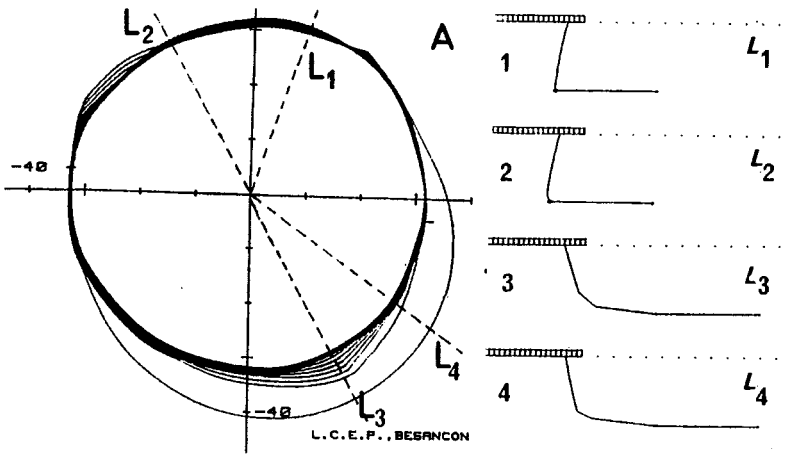


Figure 8 : Predicted shapes for a BT-70 mesa as derived from the law " 18 bis".

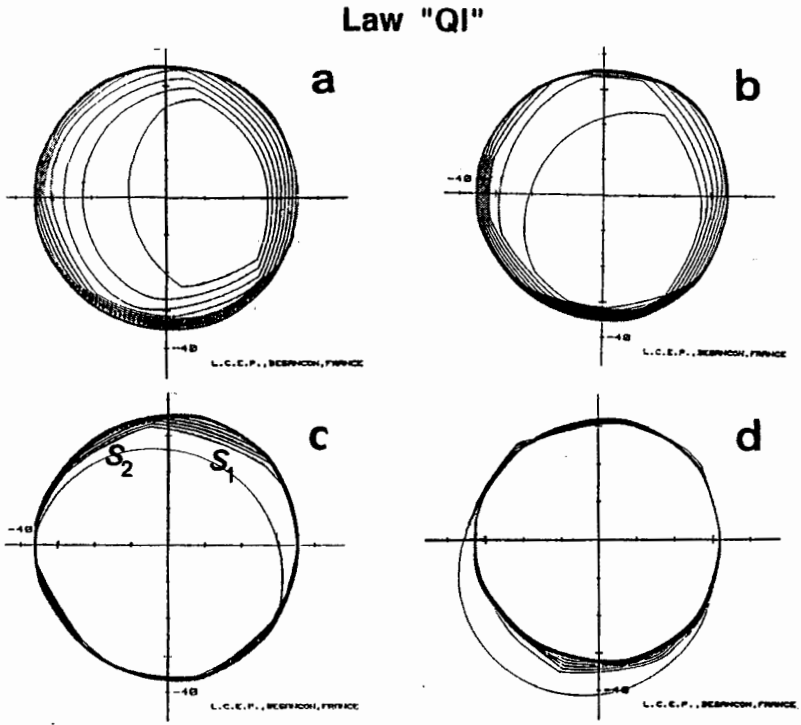


Figure 9 : Theoretical shapes for an AT-55 **(a)**, AT-65 **(b)**, BT-70 **(c)** hole and for a BT-70 **(d)** mesa

REFERENCES

/1/ **K. SANGWAL**, *Etching of Crystals* (North-Holland, Amsterdam, 1987)

/2/ **C.R. TELLIER and J.L. VATERKOWSKI**, J. Mater. Sci., 24 (1989) 1077.

/3/ **C.R. TELLIER, N. VIALLE and J.L. VATERKOWSKI**, Surf. Coat. Technol., 34 (1988) 417.

/4/ **C.R. TELLIER, J.Y. AMAUDRUT and A. BRAHIM-BOUNAB**, J. Mater. Sci., 26 (1991) 5595.

/5/ **T. LEBLOIS and C.R. TELLIER**, J. Phys. III, 2 (1992) 1259.

/6/ **C.R. TELLIER**, J. Cryst. Growth, 100 (1990) 515.

/7/ **C.R. TELLIER, A. BRAHIM-BOUNAB**, J. Mater. Sci. (to be published).

AN INDUSTRIAL QUARTZ MANUFACTURING PROCESS BY CHEMICAL ETCHING

O. CAMBON, JP. MICHEL, J. LAMBOLEY, JP. AUBRY,
Y. BILLON, D. MOUTON, I. LOZACH.

THOMSON/CEPE
44, Avenue de la Glacière - BP 165
95105 ARGENTEUIL - FRANCE

SUMMARY

This paper presents a way to implement an industrial chemical etching process in sodium hydroxide solvent specially to manufacture AT and SC orientations quartz substrates. Previous works [1 to 5] demonstrated that the quartz plates used for BAW devices can be thinned down in this solvent for large thickness removals without damaging the surface texture. The study developped in laboratory led to a process in agreement with the production conditions.

The first part presents the different steps of the process with the used implements and the industrial equipment. Then the performances (frequency adjustment precision and etching homogeneity) of the technic are exposed. Endly, characteristics of resonators for oscillator processed by chemical etching are compared with the performances achieved by conventionnal means.

I. INTRODUCTION

In order to face simultaneously an increase of throughput and a decrease of manufacturing costs, it is imperative to carry out automatic and collective manufacturing means. Reducing the cost of the resonator quartz plates manufacturing has been an important challenge for CEPE. This challenge consists in particular on finding a technic able to thin down very precisely a great number of quartz plates along an important depth without damaging the surface texture. For several years, CEPE has been looking for a good chemical etching method. The acid way (HF, NH_4HF_2...), being unable [6 to 14] to achieve simultaneously deep etch and surface quality, a basic process has been chosen.

II. CHEMICAL ETCHING INDUSTRIALISATION

II.1. The chemical etching solution

Plate thickness and surface finish are the two essential characteristics which determine respectively the resonance frequency and the resonator surtension. The study [2] showed that sodium hydroxide concentrated solutions were the most adapted solvent to reach both characteristics. The chemical etching solution has been described in the patent n°9109170. The kinetics study has shown that quartz dissolution is an Arrhenius thermally activated reaction:

$$\text{Rate } (\mu m/h) = A^{*}\exp(-Ea/RT)$$

where A is a scale factor and Ea the activation energy of the reaction.

Etching temperature must be tightly controlled to ensure a constant etching rate during the process. The etching solution is heated to the etching temperature and the stability is obtained through a thermostated oil bath while a mechanical stirrer ensures the temperature homogeneity. Above the etching solution, the vapours are condensated by a streamer containing a cooling liquid which maintains the solution concentration constant.

II.2. The process (figure 1)

After mechanical grinding or polishing, the plates are directly etched at the final frequency. The complete process includes three main functions :

a. Substrates cleaning before etching: this phase is ensured by aqueous solutions of cleaning agents which allows to eliminate the polishing or grinding abrasive traces.

b. Chemical etching: before etching the wafers and the carriers are heated to avoid a too important change of the etching temperature. Then, the wafers are introduced into the etching bath. The etching process is controlled by a computer. After chemical etching in sodium hydroxide, the substrates are immerged into an acid solution to neutralize the basic solvent. This step is controlled by a pHmeter adjusting automatically the pHvalue.

c. Rinsing and final drying : endly, the wafers are rinsed into desionized water and dried into a hot air flow.

II.2. Industrial equipment and organization

a. The equipment (figure 2)

The industrial equipment consists on several tanks to realize the process described above. A robotic and automatic manipulator conveys each batch into the different baths from the cleaning phase to the drying one. Two microcomputers are interfaced to control the process :

- One computer control all the parameters of the process (the liquid levels into the tanks, the temperatures, the pH value...). All the defects happening during the process are taken into account by this computer through the automate. Only the engineering can access to this computer.

- Another computer calculates the etching time; the operator can start the process cycle from this second computer.

The type of the chemicals and the operating conditions used led us to provide a lot of security systems on the equipment :

- Personnal protections: these protections are necessary only during the renewal of the sodium hydroxide solution (gloves, glasses and clothes).

- Environnemental protections: the equipment is entirely closed. Ventilators extract the vapours which are washed before to be released into the atmosphere. The solutions are neutralized before to be thrown out.

b. The production implements

After mechanical polishing or grinding, the plates with the same frequency are selected into batches. The frequency dispersion of these batches is different in terms of the type of the etched plates. The organization of the classifying operation is complex to optimize the productivity of the equipment. Each batch can contain up to 800 plates.

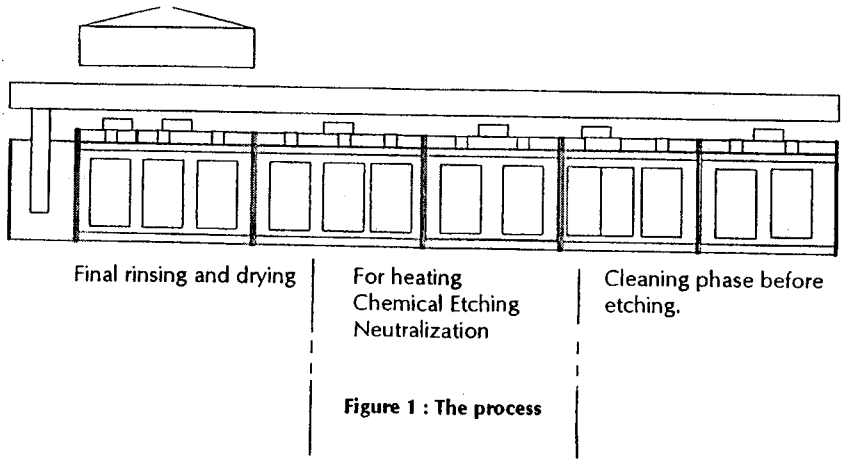


Figure 1 : The process

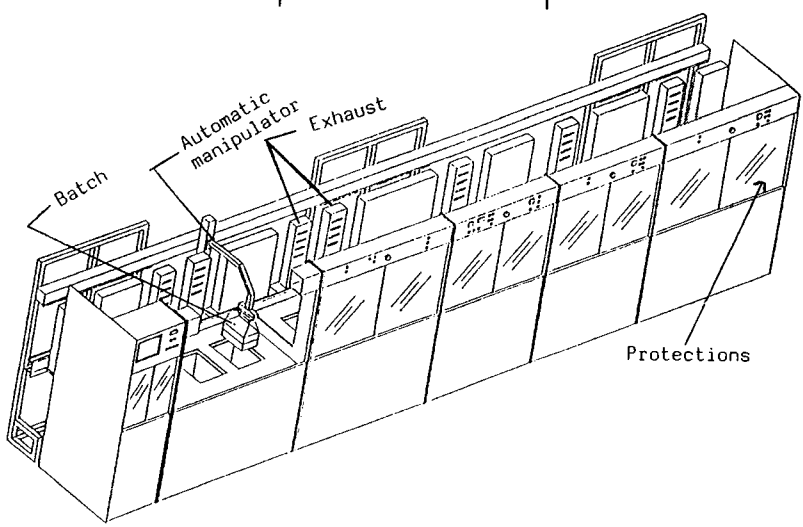


Figure 2 : Industrial equipment

III. TECHNICAL RESULTS

III.1. Frequency precision and homogeneity of chemical etching

Chemical etching must not only ensure a good surface texture but also a precise thickness adjustment of the plates. Moreover, the industrialization of the process requires a collective treatment of the wafers and hence a good etching homogeneity for all plates in a same batch. Table I presents the results after chemical etching of different kinds of resonators.

Cut	Initial surface texture	F (MHz)	Vibration mode	Etching depth (µm)	Frequency adjustment (kHz)	Etching rate (µm/h)	Etching rate homogeneity (µm/h)
AT	ground	4.098	5th overtone	13	+/-0.05	4.620	+/-0.009
AT	polished	4.098	5th overtone	5	+/-0.05	4.390	+/-0.008
SC	polished	10	3rd overtone	5	+/-0.25	7.04	+/-0.01
AT	ground	47	Fundamental	13	+/-5	13.19	+/-0.02
AT	polished	72	Fundamental	10	+/-50	16.05	+/-0.03
AT	polished	146.5	Fundamental	5	+/-150	14.97	+/-0.02
SC	polished	16	3rd overtone	5	+/-3	10.51	+/-0.01
SC	polished	20	3rd overtone	5	+/-5	9.83	+/-0.009
AT	polished	220	Fundamental	5	+/-250	6.07	+/-0.01

Table I

For the ground plates, chemical etching is performed in two phases. During the first one, the superficial layer is removed and the surface texture is improved. The study has shown [2,4] that the etching rate is not a constant at the beginning of the etching of ground plates. During the second phase, the plate frequency is adjusted.

III.2. Characterization of resonators

a. *Resonators for oscillators*

Two kinds of resonators are elaborated from plates adjusted by classical mechanical polishing and chemical etching.

- AT cut high surtension resonators at 4.098 MHz (5th overtone)
- SC cut resonators at 10 MHz (3rd overtone)

Table II gives the electrical parameters of these resonators.

Adjustment means	AT cut (5 MHz 5th overtone)		SC cut (10 MHz 3rd overtone)
	Q*F factor	Aging/ month	Q*F factor
Mechanical polish	$1.24\ 10^{13}$	$3.1\ 10^{-10}$	$1.01\ 10^{13}$
Chemical etching of mechanically polished plates (5µm etching)	$1.27\ 10^{13}$	$3.5\ 10^{-10}$	$1.06\ 10^{13}$
Chemical etching of mechanically ground plates (13 µm etching)	$1.02\ 10^{13}$	$2.35\ 10^{-10}$	

Table II

The results show that the chemical etching does not degrade and even improve the surtension factor value.

In the case of mechanically ground AT plates, the surtension value of $1.02\ 10^{13}$ is obtained after a chemical etching depth of 13 microns. This value, is only a little lower than the one obtained from polished plates.

Several AT resonators have been characterized on oscillators. The relative frequency variations has been measured over several months. The aging rate is maintained on closed value (2.5 to 3.5 10^{-10} per month) for the three kinds of resonators.

b. *HF resonators results*

Resonators (5 µm etching)	70 MHz AT fundamental	139 MHz AT fundamental	220 MHz AT fundamental
Q*F factor	$4.5\ 10^{12}$	$3\ 10^{12}$	$3.7\ 10^{12}$

Table III

Table III shows an average quality factor reached for a significant number of pieces of each kind of frequency. The Q.F. factor is closed to the theoretical one concerning fondamental mode resonators. In order to demonstrate the ability of the process, it must be pointed out that the removed thickness for the 220 MHz resonator is practically equal to the remaining thickness.

III. CONCLUSION

An industrial chemical etching process of quartz plates has been implemented through the use of a sodium hydroxide solvent. The etched depths can reach several ten microns without surface damaging. This process is suitable for AT and SC plates manufacturing. The plate frequency adjustment can be made very precise and the etching homogeneity is sufficient for a collective processing provided the equipment has been defined with a great care regarding the stability of etching control parameters.

An important part of the equipment has also been dedicated to the security of the process and of the operator.

Finally, it has been shown that the performances of the resonators made from the chemical etching equipment (adjusting, quality factor, aging) are at least as good as for classical mechanical polishing.

Bibliography

[1] "Controlled dissolution applied to berlinite and quartz materials", M DELEUZE, O.CAMBON, A. GOIFFON, A.IBANEZ, E. PHILIPPOT, IEEE International Frequency Control Symposium, invited paper, 1993.

[2] "Controlled dissolution of quartz plates in potassium and sodium hydroxide : chemical thinning down of SC cut", M. DELEUZE, O.CAMBON, A. GOIFFON, A. IBANEZ, E. PHILIPPOT, International Workshop, Montpellier, 2-3 Décembre 1993.

[3] "Quartz chemical etching applied to the industrial manufacturing processes", O. CAMBON, M. DELEUZE, JP. MICHEL, JP. AUBRY and E. PHILIPPOT. International Workshop Montepellier, 2-3 Décembre 1993.

[4] "Controlled crystal dissolution applied to quartz", M. DELEUZE, O. CAMBON, A. GOIFFON, A. IBANEZ, E. PHILIPPOT", Journal of Solid State Chemistry, to be published.

[5] "Dissolution contrôlée du quartz", E. PHILIPPOT, O. CAMBON, A. GOIFFON, A. IBANEZ, D. CACHAU-HEREILLAT. International patent n° 9109170.

[6] N. VIALLE, Thesis Universit of Besançon, 18 Mai 1989

[7] J.R. BRANDMAYR, J.R. VIG, 40th Annual Frequency Control Symposium, 1986

[8] J.R. BRANDMAYR, J.R. VIG, 39th Annual Frequency Control Symposium, 1985

[9] C.R. TELLIER, J. of Mat. Sci., 17, 1348-1354, 1982

[10] C.R. TELLIER, Surface Technology, 21, 83-89, 1984

[11] J.K. VONDELING, J. of Mat. Sci., 18, 304-314, 1983

[12] C.R. TELLIER, F. JOUFFROY, J. of Mat. Sci., 18, 3621-3632, 1983

[13] C.R. TELLIER, J.L. VATERKOWSKI, Surface Technology, 26, 275-285, 1985

[14] C.R. TELLIER, N. VIALLE, J.L. VATERKOWSKI, 1st European Frequency and Time Forum, Besançon, 1987

HIGH FREQUENCY QUARTZ RESONATORS MANUFACTURED BY CHEMICAL PROCESS.

G. PENTOVELIS , P. COLLET

Components Quartz and Electronic TEKELEC,
11, rue Keller,
10150 Pont Sainte Marie

ABSTRACT. In the recent past we have observed much interest for high frequency quartz resonators manufactured by chemical process. Ammonium bifluoride has been used for some years, it is however not very convenient, since this process requires high quality material and a good surface finish. Sodium hydroxyde has been another wet medium for removal of quartz material. It has been used since 1978[2] for the fission fragment track detectors, we have tested this process for high frequency quartz resonators manufacture. The quartz characteristics manufactured have been studied: geometry, electrical parameters and frequency-temperature curve. We have used the example of the 139.264 MHz fundamental mode crystal. Our obtained results are shown in the literature.

I. INTRODUCTION.

Chemical etching is very interesting for industrial application because its implementation can be simple and not very expensive. Corrosion reaction of a body is controlled by the adsorption/desorption and diffusion phenomena, so theoricaly chemical etching can improve the surface state of quartz[1], but never can replace mecanical polishing of blanks. On the other hand it is a rival for ion beam etching. Ion beam is used in the high frequency quartz (f>50MHz) manufacture but it is slow and very expensive.

Ammonium bifluoride has been used since a long time as a chemical etchant of quartz. However its industrial application is difficult and the chemical reaction is anisotrope on the cristallographic axes. In addition all the surface defects (produced by the mecanical polish) are reproduced and amplified by the corrosion process. All surface defects are degraded by prolonged acid etching.

In nuclear physic, quartz etching is used for fission fragments track detection[2,3,4,5]. In this discipline, it has been demonstrated that sodium hydroxyde (high concentrated

aqueus solution at high temperature) is a better etchant than ammonium bifluoride. Its chemical rates are more isotropic on the cristallographic axes.

We have applied this basic process in piezoelectricity for chemical manufacture of quartz resonators. From the initial results the advantages are obvious. These first results are mentioned in the following table

Chemical Process	Acid medium $NH_4.HF_2$	Basic medium NaOH
Conditions	Satured solution T≈65°C	High concentration solution T>125°C
Advantages	Low temperature range. High reaction rate Well reported in the literature.	High reaction rate large range of material quality (we can use standard quality material for two sided removement thickness Δd<50μm) Linear Arrhenius relation
Disadvantages	High quality material and surface are required.	T>150°C Solution close concentration control is required. High reaction rate-temperature relation.

II. EXPERIMENTAL CONDITIONS.

All parts of the chemical etching device that are in contact with the solution are made from PTFE. This device stops the vapour loss when the quartz is introduced or removed from the solution. Solution concentration is kept constant. The surface heating is up off 50% of the total solution surface, a device stirer maintains the solution homogeneity.

All blanks are isotropicaly etched (no corrosion rate dispersion is in the solution or during the experience). After etching, the surface planety is maintained (in 16 MHz all blanks have less than one fringe in the UV interferometer). Solution is currently changed to avoid composition modification by the diluted SiO2 or other cause.

After the chemical etching, all blanks are cleaned by the following process:

a) In an acid solution (pH=6) to neutralize all basic impurities on the surface,
b) in cleanser solution (pH=7),
c) in demineralised water
d) in alcohol.

they are then stored at 100°C under vaccum.

For the high frequency resonators, the classical cleaning process is insufficient. All blanks are singly cleaned by a 20% KCN solution (containing abrasive suspensions), demineralised water, and liophylised alcohol. They are kept at 100°C under vaccum for 1 hour before the melting process.

An air gap system is used for blank frequency measurement. Simulations have demonstrated that high frequency measurements are achievable with only small electrode area. We never see zero phase but only minimum impedance frequency value. For frequencies up to 100 MHz we use an electrode area smaller than 0.8 mm^2. Frequency error is smaller than 5 10^{-3} F^2

AT cut quartz is used for the high frequency manufacture. In this paper we will show the results of this study. We will show and discuss only a few results for SC cut quartz .

III. CHEMICAL ETCH PROCESS.

Aqueus solution of sodium hydroxyde at low temperature (T<100°C) have a low chemical reaction rate with the quartz. Use of high temperature solution is limited by the boiling temperature (about 100°C) for all low NaOH concentration aqueus solutions. Reaction rate becomes fast for the concentrated sodium hydroxyde solutions (C>50%) at high temperature range (T>150°C) and it stays constant for the same temperature and concentration values. It can be calculated from the thickness-time ($\Delta d(t)$) graph slope. It is independent of the material used quality (natural, synthetic standard, high quality and swept synthetic)

For AT cut, the reaction rate is independent of the surface quality for lapped (3µm) and polished surface. For a lapped surface up of 3µm, the dissolution is slower but we never study this point. In high concentration and viscosity bath (C>85%), we observe diffusion control near the surface and the reaction rate decreases.

For steady concentration the reaction rate is closely exponential fonction of temperature (Arrhenius equation). Reaction activation energy calculated from the $\ln(v(T^{-1}))$ graph slope is about 0.9 eV.

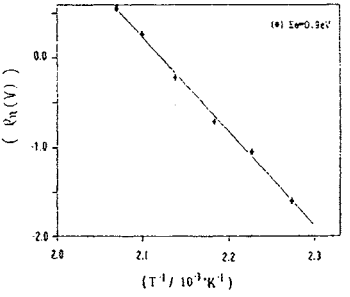


Figure 1.
Quartz dissolution's rate variation with temperature in 80% sodium hydroxyde solution.

For SC cut we observed a slight Arrhenius curve change. AT cut's dissolution rate is higher than the SC cut: 3.8% at 170°C and 3.5% at 190°C. The energy activation of the dissolution calculated for SC cut is 0.2% higher than for AT cut.

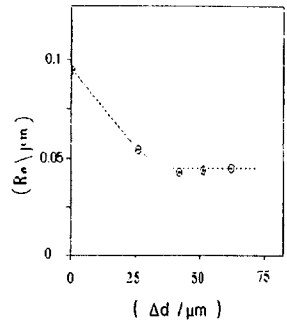


Figure 2.
Rugosity variation for two sides removed quartz thickness in sodium hydroxyde solution.

For a lapped surface the surface quality improved by etching in sodium hydroxyde. The surface quality depends on the amount of quartz removed (see figure 2). For two sides removed thickness greater than 40µm, rugosity becomes constant. On the rugosity spectrum we see that number of peaks decrease with removed thickness. The surface defaults are attenuated, but surface quality will never equal to a mechanically polished surface (Ra≈0.004µm).

For a polished surface: Prior to etching we observed a surface contamination on polished blanks, such as fishbone tracks and material defaults. Following sodium hydroxyde etching all small tracks have disappeared, material defaults and all big tracks have been attenuated on the surface. The dislocations favorize the solution penetration into the quartz and holes or etching pits appears at the surface. After etching the surface quality does not closely depend on the material defaults.

No surface orientation change have been observed in AT cut etched quartz. The blanks surface parallelism remained unaltered.

Chemical etching is more isotropic than in an acid medium. Sodium hydroxyde is a better etchant than ammonium bifluoride medium. For an acid etch it is necessary to have a perfect surface, but it is less important for an NaOH etch. However, deep tracks must be avoided.

IV. HIGH FREQUENCY QUARTZ BY CHEMICAL PROCESS.

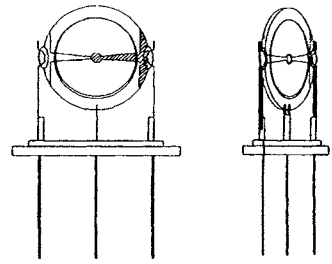
High frequency quartz resonators (f>50MHz) are needed in telecom and military applications. (broad band filter, VCXO with a large band frequency variation,...). It is very important to control the resonators motionals parameters and adapt them to oscillator specifications. We present here our study concerning 139.264 MHz resonators produced by the chemical process. These results can be compared with the results published by others authors on the same resonator produced by ion beam process[6].

We have used unswept and swept high quality material. The first material quality gives low percentage of damaged quartz and the second is perfectly etched.

Motional parameters are greatly influenced by the quartz surface finish. In high frequency resonators series resistance for a chemicaly polished quartz is twice than for a mechanically polished quartz. For high frequency quartz resonators we have chosen an initially polished surface.

Blank's geometry.

Quartz shape is shown on the following figure. The crystal has a reduced thickness in the central area surrounded by a thicker outer ring. We have developped a surface mask specially for the NaOH chemical HF quartz production process. Crystal has the suitable thickness in the centre, after two steps of chemical treatment (see figure 3):



A) bowl etching (using mask)

B) coarse etching over all the surface to increase quartz frequency and frequency adjustment (without mask)

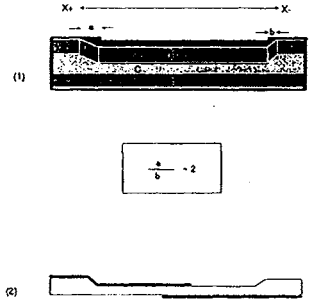


Figure 3.
Chemical etching steps for blanks geometry (1) and Silver electrode deposition (2).

In the frequency range up of 40 MHz, blanks are very fragile and manipulation is difficult. We chose profile etching as first step, and coarse etching is used for surface cleaning of contamination caused by the masking. Bowl depth is not very important for resonator motional parameters. We have chosen 30µm bowl depth on the 100 µm initial quartz thickness (16.7MHz).

The rigid outer ring of quartz exerts stress to the central area. Also series resistance change is observed. Its value is higher than resistance of a plane parallel resonator. We have seen that resistance value depends on the surounding ring diameter.

Figure 4 gives the series resistance variation versus the mass loading. For 3mm rigid ring diameter, resistance become optimal after 0.7 F2 mass loading (in frequency variation units), and δf/f temperature variation is greatly deviated from the theoritical values. For a 4 mm ring diameter the optimum mass loading is 0.18 F2.

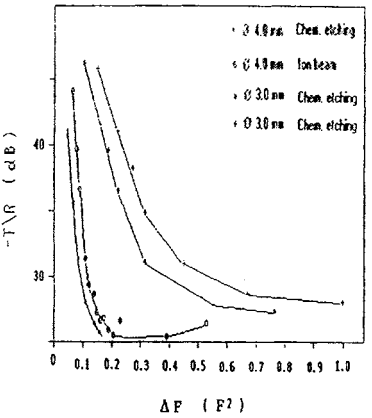


Figure 4. Motional resistance (arbitrairy units)-Mass loading variation for different ring diameters. Electrodes: metal; Ag, surface: 0.2 mm^2 Quartz frequency up of 140 MHz.

There is not any distinction between chemical and ion beam etched quartz.

For 139.264 MHz resonators a 4 mm rigid ring diameter is better than 3 mm for a mass loading upper 0.18 F2. Frequency will be between 142.8 MHz and 143.5 MHz for 0.18 F2 and 0.21 F2 mass loading respectively.

A particularity for AT cut etching should be mentioned: When etching the rigid outer ring formation, after a few microns removed thickness, etch evolution is perpendicular as well as parallel to the quartz surface. Obviously parallel evolution is more advanced near the mask than near the bottom of the bowl (see figure 3). Its advancement depends on the parallel etching rate. So we have observed that on the X axe etching rate depends on the cristallographic orientation. Parallel etching rate on the X^+ axe is twice as higher as on the X^- axe. Although quartz dissolution process is always the same and removed mass rate is controlled by the adsorption/desorption phenomena on the surface. Obviously these effects do not influence the chemical etching of AT cut crystal, because X axe is parallel to the surface. But when surface orientation changes and X axe is not parallel to the surface, etch results will be influenced by the etching rate anisotropy. This explains the different cases of AT and SC cuts behavior observed. Problemes may be observed in the plan-convex or biconvex quartz shape etching.

When quartz is etched in the sodium hydroxyde etching step (b), structural change is observed for few blanks. This behavior is frequently observed for the bath temperatures higher than 195°C. We attribute it to the stress exerted between the rigid ring and the central area. This temperature limits the etching rate to values down of 0.4 μm/min.

Electrode geometry.
Silver electrodes of resonators are vapour deposited under vaccum. Mass loading is greater than 0.18 F^2 and its resistance is less than 1 Ohm.

On the bowl extremity, the edge is not perpendicular to the surface, metal deposition thickness is high and the electrode conduction on this area is good (see figure 3). On ion beam etched quartz bowl edge is perpendicular to the surface and a second metallisation is necessary to improve the electrode conduction.

Electrode area must be adequate for a low motional resistance (high surface area) and for reduced number of unwanted modes (small surface area). For oscillator applications motional resistance of unwanted modes should be twice motional resistance of fundamental mode.

For surface area greater than 0.28mm^2 unwanted modes become high and close to the fundamental mode. For a 139.264 MHz resonator the electrode area must be less than 0.2mm^2.in order to achieve the required performances.

Details of the 139.264 MHz resonators:

Blanks diameter:	5.0 to 5.2 mm
Rigid ring diameter :	4.0 mm
Initial blank thickness:	100 µm
Final thickness:	11.68 µm to 11.62 µm
Bowl depth:	30 µm
Electrodes area :	0.2 mm^2
Electrode material:	silver
Surface mass loading:	0.18F^2 to 0.21 F^2
Blank environment:	vaccum

Motional parameters.
Motional parameters depend on the electrode geometry. If the above blanks geometry is respected the motional parameters should depend only on the electrode area (or surface diameter). For our resonators we have demonstrated the following equations:

$R_S\ (\Omega) = K_{rs} * d^{-2}(mm)$	$K_{rs}=5.6$
$L_S(\ \mu H\ ) = K_{ls} * d^{-2}(mm)$	$K_{ls}=104$
$C_S(\ fF\ ) = K_{cs} * d^2(mm)$	$K_{cs}=12.5$
$C_p(\ pF\ ) = K_{ps} * d^2(mm)$	$K_{ps}=4.65$

d is the electrode diameter.

An example of chemical etched quartz resonator motional parameters are presented in table 1. Motional parameters are comparable with the values earlier published for ion beam etched resonators(6).

TABLE 1. Electrical parameters:
Resonator: 139.264 MHz
Quartz HQ Swept
AT (35°13'30" ±1'30")
Blank diameter: 5.2 mm, surrounding ring diameter: 4mm, electrode diameter: 0.5 mm

Frequency f_0 (kHz) in 25°C	Rs (Ω)	L_s (μH)	C_0 (pF)	Q
139260.605	20.2	396.6	1.27	17169
139260.302	14.2	352.2	1.33	21693
139261.108	15.1	352.2	1.32	20434
139261.291	14.9	361.6	1.29	21226
139262.009	14.9	344.3	1.42	20200
139261.042	15.6	351.3	1.31	19681
139257.379	16.0	343.8	1.37	18826
139261.244	21.3	367.0	1.12	15078
139261.446	18.4	351.4	1.14	16719
139263.810	21.4	341.7	1.73	13993

Unwanted modes.
An example of 139.264 MHz resonator frequency spectrum is given in the figure 5. Two parasite modes appears about 600 kHz and 1150 kHz from the fundamental mode. In the table 2 we show all parasites modes of a number of resonators.

TABLE 2. Unwanted mode vibrations. f_0, R_s, f_n,R_n are frequency and motional resistance of the fundamental and the unwanted mode.
Resonators 139.264MHz
Quartz HQ Swept
AT (35°13'30" ±1'30")
Blank diameter: 5.2 mm, surrounding ring diameter: 4mm, electrode diameter: 0.5 mm

f_1-f_0 (kHz)	R_1/R_s	f_2-f_0 (kHz)	R_2/R_s	f_3-f_0 (kHz)	R_3/R_s	f_4-f_0 (kHz)	R_4/R_s
275	8.3			550	3.5	1255	5.9
270	15.5			610	5.6	1235	3.6
265	10.9	515	12.3	590	4.1	1225	3.3
535	7.1	535	7.1	615	5.1	1250	3.1
265	11.56	505	9.9	590	3.9	1190	2.4
		545	6.8	590	8.3	1255	7.7
255	13.6	575	4.4			1225	4.1
		510	5.3	565	2.9	1195	2.9
255	11.7	564	2.8			1200	3.4
				600	3.6	1175	2.85

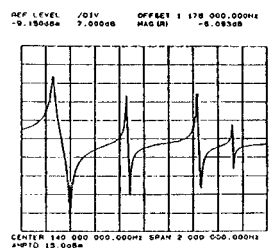


Figure 5. 139.264MHz quartz resonator spectrum with typical unwanted modes.

Frequency-temperature variations.
δf/f-temperature variations must be less than ±10ppm. The results of a number of resonators are resumed on the table 3. Real angle of cut is 35°13'30" AT cut.

TABLE 3. δf/f variation results in the -30°C, +80°C temperature range:
Resonator: 139.264 MHz
Quartz HQ Swept
AT (35°13'30" ±1'30")
Blank diameter: 5.2 mm, surounding ring diameter: 4mm, metallisation diameter: 0.5 mm

Tmin (°C)	δf/fmin (ppm)	Tmax (°C)	δf/fmax (ppm)	theoritical angle
55	-3.0	-1	+4.0	+2'45"
48	-3.0	-2	+4.3	+2'15"
57	-6.0	-10	+8.0	+3'45"
58	-8.0	-15	+10.0	+4'00"
53	-4.0	-8	+6.0	+3'00"
57	-7.0	-14	+9.5	+4'00"
50	-3.5	-7	+5.0	+2'30"
52	-5.0	-8	+6.5	+2'45"
53	-6.0	-11	+8.5	+3'15"
57	-6.5	-10	+7.0	+3'45"

V. CONCLUSION.

Chemical etching in sodium hydroxyde is a good process for high frequency quartz resonator manufacture. This process is better than ammonium bifluoride process. Resonators obtained quality is as good as ion beam etched.

We have mentioned the case of 139.264 MHz quartz resonators, but all these results can extended to others quartz resonators in 50MHz-200MHz frequencies band. For frequencies greater than 200 MHz another study may be necessary, specifically concerning the electrode geometry and deposition process.

Acknowledgments. This work was supported by the French General Direction of Industrial Strategy, Communication and Service Industry division.

We acknowledge Mike Hodgkins for helpfull discussions.

VI REFERENCES.

1) J.R.Vig, J.W.Lebus, R.L.Filter, Proceeding of the 31st Annual Frequency Control Symposium, 1977 p.131,
"CHEMICALLY POLISHED QUARTZ".

2) H.A.Khan, R.A.Akbar, Radit.Phys.Chem. 11(1978)93,
"FISSION FRAGMENT TRACK REGISTRATION AND DEVELOPMENT EFFICIENCIES OF SOME COMMONLY USED CRYSTALLINE TRACK DETECTORS".

3) K.D.Bal, N.Lal,R.S.Waraich, K.K.Nagpaul, Nuclear Instruments and Methods, 173(1980)303,
"ETCHING CHARACTERISTICS OF A NATURAL QUARTZ TRACK DETECTOR".

4) N.P.Singh, M.Singh, S.Singh, H.S.Virk, Nuclear Traks and Radiaton Measurements, 8(1984)41,
"ETCHING STUDIES OF FISSION DAMAGE IN QUARTZ".

5) A.S.Sandhu, R.C.Ramola, L.Singh, S.Singh, H.S.Virk, Indian Journal of Pure & Applied Physics,28(1990)73,
"EFFECT OF NATURE OF THE ETCHANT ON ANISOTROPIC TRACK ETCHING IN QUARTZ".

6) F.M.Stern, J.Dowsett, D.J.Carter and R.J.Williamson, 43th Anual Symposium on Frequency Control - 1989, p.634.
"THE FABRICATION OF HIGH FREQUENCY FUNDAMENTAL CRYSTALS BY PLASMA ETCHING".

Time Transfer (A 3)

TIME-TRANSFER VIA SATELLITES. - A SURVEY.

S. Starker
German Aerospace Research Establishment
Institute for Radio Frequency Technology
Oberpfaffenhofen, D-82234 Wessling

Summary

The transfer of time-scales via satellites requires the solution of special technical problems: Systematic variations of delay times have to be determined or eliminated by special transfer methods. The detection of weak signals in random noise requires efficient methods of averaging. The use of frequency filters limited the signal bandwidth and accuracy in earlier systems. Correlation filters in modern systems allow time signals with unlimited bandwidth and increased accuracies. Improved future systems will be possible by additional geostationary navigation satellites and by H-maser clocks in space.

1. Introduction

Time-transfer is conceived in the following as the transfer of a time-scale from the generating clock to a distant point. Satellites can be used for this purpose as a relay, for regional applications, or as a clock-platform, which allows global time distribution and clock comparisons (Fig. 1). In both cases one-way and two-way methods can be applied. One-way methods make possible the distribution of time to an unlimited number of users; two-way methods allow point-to-point contacts with improved accuracy between a limited number of users.

For a weather independent time transfer with microwave signals, satellite links are superior to terrestrial links, because the propagation delays are more stable, higher carrier frequencies can be used and more bandwidth is available.

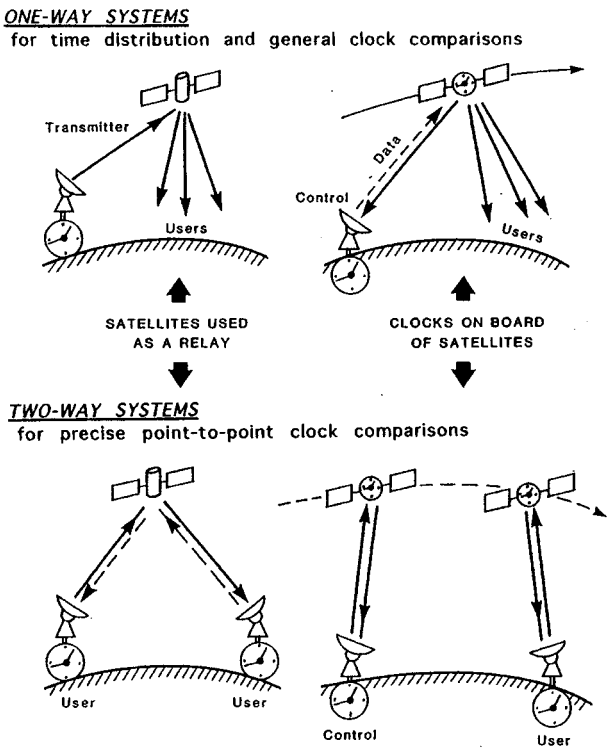


Fig. 1: Different system configurations for time-transfer using satellites.

But several new technical problems arise, as for instance the signal transmission over large distances with limited satellite transmitting power, and delay time variations because of the satellite motion. Existing satellite systems for time transfer are characterized by different solutions of these problems. Thus the following survey shall begin with the problems and their technical solutions and will end with a table of time transfer systems, in which these methods and techniques are applied.

2. A Model of Time Transfer and Error Description

A clock at site A generates a time-scale t_A. The time marks of this scale are transfered to a distant site B with a delay τ_A. The received scale at site B is designated as t'_A (Fig. 2). To mark simultaneous readings at the sites A and B we assume another ideal reference time scale t, which is thought to be available at A and at B. This reference scale is eliminated in the final results. Thus, the connection between the delay time τ_A and the simultaneous readings $t_A(t)$ and $t'_A(t)$ at the sites A and B are described by

$$t'_A(t) = t_A(t) - \tau_A. \tag{1}$$

The length of the time-interval τ_A is measured by units of the scale t_A.

In many applications the received time scale t'_A is compared with another scale t_B, generated at the site B. This additional step is treated in /1/. The following consideration is confined to the different methods of determining τ_A and to the residual errors, which characterize the performance of a transfer-link.

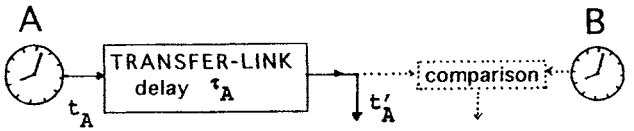


Fig. 2: Model of a time transfer system.

In stationary links, to which most of the terrestrial transfer-links belong, τ_A is often described by a mean value $\bar{\tau}_A$ and by random fluctuations $\pm\delta\tau(t)$, characterized by its standard deviatiation σ_τ.

$$\tau_A = \bar{\tau}_A \pm \delta\tau_A(t) \qquad \text{(for stationary links)} \tag{2}$$

$$\sigma_\tau = \sqrt{\overline{\delta\tau(t)^2}} \qquad \text{(standard deviation of random fluctuations)} \tag{3}$$

The main difference in satellite links is that τ_A is no longer constant but a function of time because of the satellite motion.

$$\tau_A = \tau_A(t) \pm \delta\tau_A(t) \qquad \text{(for satellite links)} \tag{4}$$

Thus, equation (1) has to be modified for satellite links to

$$t'_A(t) = t_A(t) - \tau_A(t) \tag{1a}$$

The function $\tau_A(t)$ must be known a-priori or has to be determined by topical measurements. The difference between the true value $\tau_A(t)$ and the measured or known value $\tau'_A(t)$ is a systematic error.
Thus the uncertainty $\Delta\tau_A(t)$ on the delay time $\tau_A(t)$ consists of two very different parts

$$\Delta\tau_A(t) = \underbrace{\left[\tau'_A(t) - \tau_A(t)\right]}_{\substack{\text{systematic} \\ \text{error}}} \pm \underbrace{\delta\tau_A(t)}_{\substack{\text{random} \\ \text{error}}} \tag{5}$$

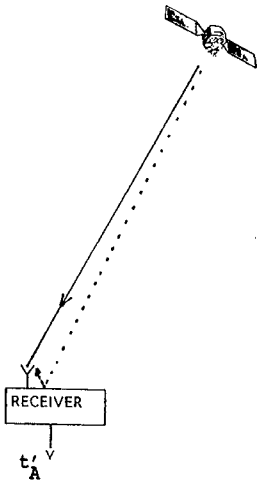
The value of the totat systematic error depends on several influences in a complex, systematic way, e.g. on the relative position of the satellite to the ground stations, the environment and antenna diagrams of the ground stations, the calibration procedure of measurement and time-transfer equipment etc. Therefore it appears reasonable, to describe the sum of all these different contributions to the systematic error by a Root Mean Square value, similar to the standard deviation

$$\hat{s} = \sqrt{\overline{\left[\tau'_A(t) - \tau_A(t)\right]^2}} \quad . \tag{6}$$

This value is an estimate of the sum of residual errors, which is left after all calibration and correction procedures and which cannot be reduced by smoothing, contrary to the random errors.
With this, an estimated value of the overall uncertainty of the delay time τ_A can be written as

$$\sigma^2_{\tau A} = \hat{s}^2 + \sigma^2_\tau . \tag{7}$$

A rough overview on the main sources of systematic and random errors is given in Fig. 3.



ERROR SOURCE	SYST. ERRORS	RAND. ERRORS	METHODS FOR ERROR REDUCTION
SATELLITE POSITION	X		2-WAY-RANGING
SATELLITE VELOCITY	X		2-WAY-DOPPLER
TRANSMITTER DELAY	X	X	(EQUIPMENT CALIBRATION
CABEL	X	X	(EQUIPMENT CONTROL
ANTENNA DIAGRAM	X		PHASE-DIAGRAM + ATTITUDE MEASUREMENT
IONOSPHERE	X	X	2 SIGNAL FREQUENCIES IONOSPHER.-MODELLING
TROPOSPHERE	X	X	2-WAY RANGING TROPOSPH.-MODELLING N-MEASUREMENT
MULTIPATH	X	X	ADAPTION ANTENNA-ENVIRONMENT SIGNAL WAVEFORM
ANTENNA DIAGRAM	X		CALIBRATION AND CONTROL
CABEL	X	X	" " "
RECEIVER DELAY	X	X	" " "
NOISE		X	MATCHED FILTER, CORRELATION FILTER
OVERALL SYST.ERROR	X		SYSTEM CALIBRATION DIFFERENTIAL MEASUREMENTS

Fig. 3: Main error sources in satellite time transfer systems and methods for error reduction.

3. Reduction of Systematic Errors by Diverse Transfer Methods

In one-way systems the actual distance between the satellite and the ground stations has to be determined from their positions. The usual way to provide the users on the ground with the necessary information on the actual satellite position is the transfer of orbit data from the satellite to the ground together and on the same carrier frequency as the time signal.
Two-way systems do not require this position information because the distance, or rather the delay time, is determined directly and more accurately from the measured transmission time.

This is true for satellites used as a relay as well as for satellites used as clock platforms. The latter case is illustrated in Fig. 4.
The systematic influence of the ionosphere on a one-way signal may be corrected by using available a-prior knowledge on ionospheric parame-

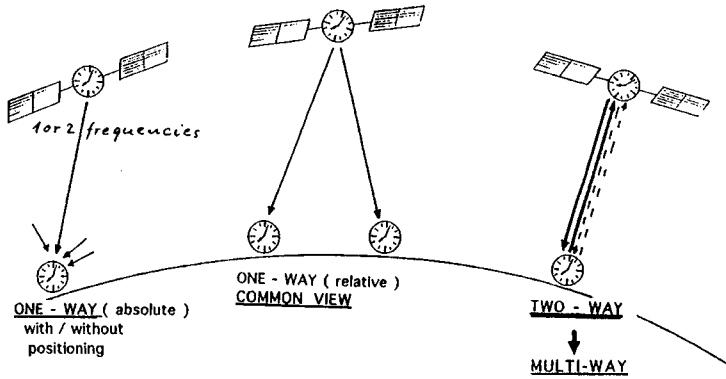


Fig. 4:

TIME TRANSFER METHODS USING CLOCKS IN SPACE

$$\tau_{f1} - \tau_{f2} = \frac{40,3 \cdot TEC}{C} \left[\frac{1}{f_1^2} - \frac{1}{f_2^2} \right]$$

$$TEC = \frac{c[\tau_{f1} - \tau_{f2}]}{40,3 \cdot \left[\frac{1}{f_1^2} - \frac{1}{f_2^2} \right]}$$

$$\Delta\tau_{f1} = \frac{40,3}{c \cdot f_1^2} \cdot TEC$$

$$= \frac{[\tau_{f1} - \tau_{fs}]}{[1 - (\frac{f_1}{f_2})^2]}$$

Fig. 5:

Determination of the 'TEC' and the ionospheric influence on the delay-time by the two-frequency method.

ters or by measurement of the Total Electron Content (TEC) along the propagation path by using two different carrier frequencies for the signal transmission. The equations for this method are composed in Fig. 5.

Further systematic errors can be eliminated by the well known 'common-view' method, if only ground clocks have to be compared by simultaneous reception of the same satellite signal. In this case all correlated position-, propagation- and equipment errors are reduced.

As to the delay-time determination by two-way measurements a careful analysis is necessary of the asymmetric delays in the downward and upward link (Fig. 6). Such asymmetries are caused by the relative velocity between satellite and ground stations, by using different carrier-frequencies in the down- and upward-link and by different transmitters and receivers in these two links, which may differ in their delay times as well as in their frequency-characteristics. Equations of Fig. 7 describe these asymmetric terms. For further information see /2/.

4. The Reduction of Random Errors by Averaging

The random error of a single measurement consists of different contributions from different error sources

$$\sigma_{\tau 0} = \left[\sum_{i=1}^{N} \sigma_{\tau i}^2 \right]^{1/2} ; \qquad i = 1 \ldots N \text{ for} \quad \begin{array}{l} \text{- thermal noise} \\ \text{- tropospheric noise} \\ \text{- ionospheric noise} \\ \text{- etc.....} \end{array} \tag{8}$$

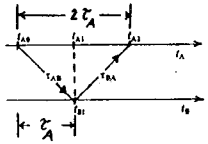
Averaging over n measurements with independent random errors reduces this error to

$$\sigma_\tau(n) = \frac{\sigma_{\tau 0}}{\sqrt{n}} \tag{9}$$

An averaging interval T contains different numbers n of independent values for different error contributions $\sigma_{\tau i}$, if these have different correlation times τ_{ci}

TIME DIFFERENCE BETWEEN CLOCKS A AND B IS THE DIFFERENCE OF SIMULTANEOUS READINGS AT THESE CLOCKS.

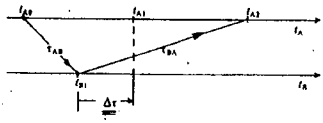
a) symmetric case:
readings t_{A1} and t_{B1} are simultaneous events:



$$\Delta t_{AB}(t_{A1}) = t_{A1} - t_{B1}$$

$$t_{A1} = \frac{1}{2}\,(t_{A0} + t_{A2})$$

b) asymmetric case:
readings t_{A1} and $(t_{B1} + \Delta\tau)$ are simultaneous events:



$$\Delta t_{AB}(t_{A1}) = t_{A1} - (t_{B1} + \Delta\tau)$$

$$\Delta\tau = \Delta\tau_d + \Delta\tau_a + \Delta\tau_c$$

relative velocity $\vec{v}$ between clock A and B — atmospheric propagation — equipment delays

Fig. 6

SIMULTANEOUS EVENTS AND ASYMMETRIES IN A TWO-WAY TIME TRANSFER

$$\Delta\tau_d = \frac{1}{2}\,(\tau^d_{AB} - \tau^d_{BA}) = -\frac{T_r^{*)}}{2} \cdot \frac{f}{c} = \frac{T_r}{2} \cdot \frac{f_{DA}}{f_A}$$

T_r = travel time between antennas

===> TWO-WAY DOPPLER MEASUREMENT

$$\Delta\tau_a = \frac{1}{2}\,(\tau^a_{AB} - \tau^a_{BA}) = \frac{1}{2} \cdot 1{,}34 \cdot 10^{-7} \cdot \mathrm{TEC}\,(f_{AB}^{-2} - f_{BA}^{-2})$$

===> TWO-FREQUENCY 'TEC'-DETERMINATION

$$\Delta\tau_e = \frac{1}{2}\,((\tau^e_{AB} - \tau^e_{BA}) + (\tau^r_{AB} - \tau^r_{BA})) = \frac{1}{2}\,(\tau^e_{AB} - \tau^e_{BA}) + \frac{1}{2}\,(\rho_{AB} - \rho_{BA}) \cdot \frac{f_n}{f}$$

with $\tau^e + \rho \cdot \frac{f_n}{f} = \tau^e + \tau^r$; $(\rho = \text{const.})$

frequency dependence of equipment delay

===> EQUIPMENT CALIBRATION AND CONTROL

Fig. 7: Asymmetric terms in a two-way time transfer (from /2/).

$$n_i = \frac{T}{\tau_{ci}} \; ; \qquad T > \tau_{ci} \; ; \tag{10}$$

τ_{ci} is the minimum distance of statistically independent random errors. For white noise of bandwidth B the correlation time is about

$$\tau_{ci} \approx \frac{1}{2\,B} \tag{11}$$

With this, the standard deviation of the averaged total random error can be approximated by

$$\sigma_\tau(T) = \left[\frac{1}{T} \sum_{i=1}^{N} \sigma_{\tau i}^2 \cdot \tau_{ci} \right]^{1/2} \; ; \qquad \text{(valid for } T > \tau_{ci}\text{)} \tag{12}$$

The correlation times τ_{ci} may be very different, e.g. $\tau_c \ll 1s$ for thermal noise, and $\tau_c \gg 1s$ for ionospheric fluctuations.

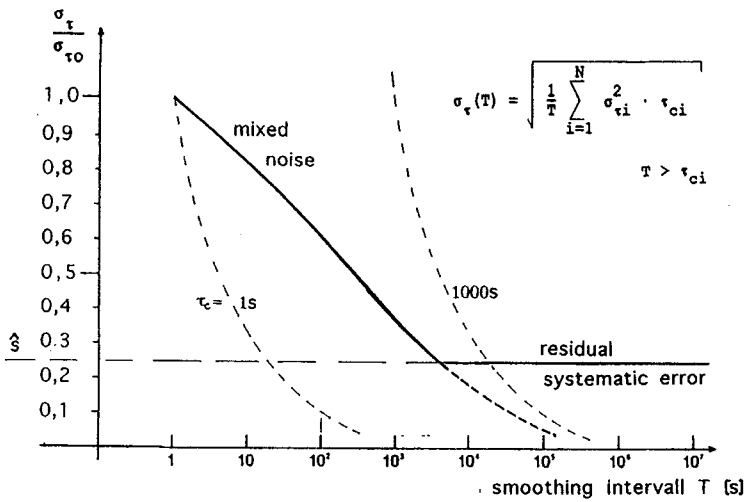


Fig. 8: The reduction of random errors with different correlation times τ_c by averaging.

Fig. 8 shows the effect of smoothing over a time interval for noise contributions with $\tau_c = 1$ and $\tau_c = 1000$ as dotted lines. The traced

line describes this effect for a mixture of both. The possible improvement of the total error is limited by the residual systematic error, which cannot be reduced by further averaging.
Practical results of time transfer deviations for different averaging times are given in /1/ for LORAN-C, GPS-Common View and 2-way satellite links.

If the random errors of two measurement samples in a distance T_s are not independent, the sum or difference of these errors has to be described using the correlation function $\rho(T_s)$.

$$\overline{\left[\Delta\tau(t) \pm \Delta\tau(t+T_s)\right]^2} = 2\,\sigma_\tau^2\left[1 \pm \rho(T_s)\right] \tag{13}$$

$$\text{with } \rho(T_s) = \frac{\overline{\Delta\tau(t) \cdot \Delta\tau(t+T_s)}}{\sigma_\tau^2}\,; \qquad (-1 \le \rho \le +1)$$

With these equations, the standard deviation of the difference of the total random error in a sample distance T_s is

$$\sigma_{diff}(T_s) = \left[\sum_{i=1}^{N} 2\,\sigma_{\tau i}^2\,[1 - \rho_i(T_s)\right]^{1/2} \tag{14}$$

and the fractional frequency uncertainty can be described by

$$s_y(T_s) = \frac{\sigma_{diff}(T_s)}{T_s}\,. \tag{15}$$

This equation is equivalent to the Allan variance, and describes the accuracy of a frequency transfer on the basis of noise parameters.

Fig. 9 illustrates the possible improvement of frequency transfer accuracy with increasing transfer time T_s for a mixed noise influence, compared with the frequency stabilities of different clocks.
Practical results of frequency transfer accuracies as a function of sample time T_s are given in /1/. The different slopes, which are

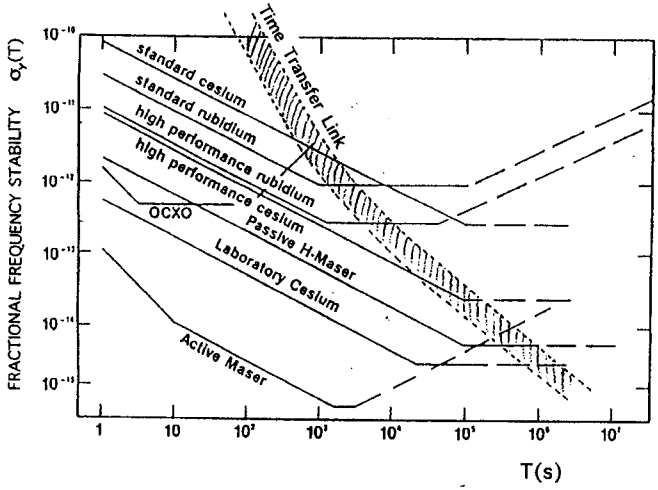


Fig. 9: The accuracy of frequency transfer with increasing sample time T_s is illustrated by the hatched curve. The change of the slope is caused by random errors with different correlation times.

shown there for different systems, are caused by different error sources with different correlation times.

5. Transfer Systems with Band-limited Signals

A basic condition for the reliable detection of signals in noise is a sufficiently high signal to noise ratio of about

$$\frac{S}{N} > 100 \quad (\hat{=} \ 20 \text{ dB}); \tag{16}$$

$$N = k \, T_e B \qquad \text{(for white noise)} \tag{17}$$

S = received signal power

N = noise power

$k = 1{,}38 \cdot 10^{-23}$ Ws/K

T_e = equivalent noise temperature

B = noise bandwidth.

The conventional method for achieving a high S/N is the use of frequency-filters, which confine the noise bandwidth B to the signal bandwidth B_s

$$B \rightarrow B_s .$$

Thus it is necessary to concentrate the limited signal power to a relatively narrow bandwidth B_s, to achieve the desired S/N ratio of equ. (16).

The allowable signal bandwidth B_s depends on the received signal power S and on the equivalent noise temperature T_e as follows:

$$B_s < \frac{S}{100\ k\ T_e} . \tag{18}$$

Because of this condition the bandwidth of satellite time signals was limited to a few KHz until recently; and because signals of such small bandwidths possess relatively poor slopes, time marks can be transferred by this detection method only with low accuracy, according to the equation

$$\sigma_\tau = \frac{1}{2\ B_s} \sqrt{\frac{N}{S}} . \tag{19}$$

A quantitative estimation with usual numbers, as given in Fig. 10, shows that the uncertainty of time transfer with this kind of systems will be generally larger than 1 μsec.

Typical systems of this kind are TRANSIT, GOES and the application of early geostationary communication satellites for clock comparisons. Some main characteristics of these systems are collated in Fig. 11.

First experiments with clocks in space required the utmost improvement of this transfer technique: the transfer of a Rb-clock time scale from the TIMATION-satellites to the ground was performed by a set of narrow-band time signals, called side-tones, which were spread over a large frequency bandwidth. They were detected successively in the ground receiver and evaluated in a combined process, to achieve a residual time transfer uncertainty of $\sigma_t < 1$ μs.

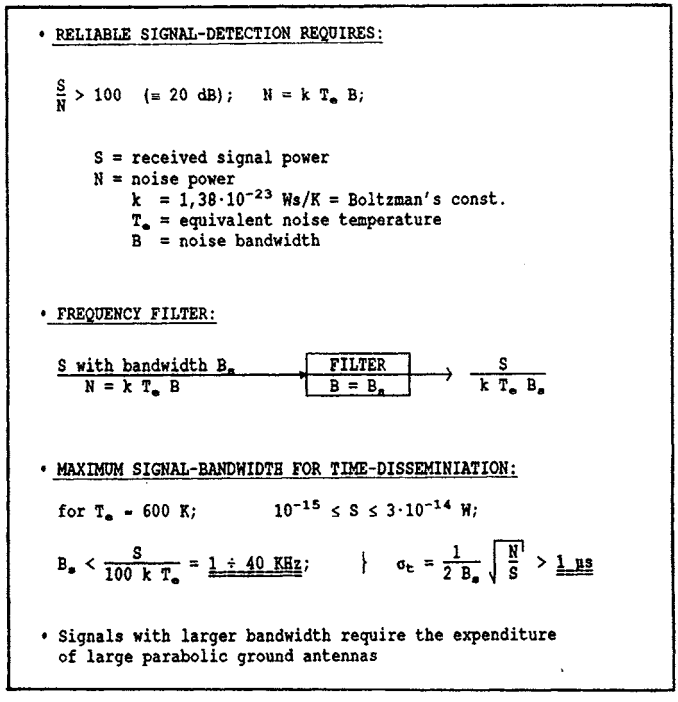


Fig. 10: An estimation of allowable signal bandwidth and achievable time transfer accuracy with frequency filters.

SYSTEM-CATEGORY	EXAMPLE	ACCURACY	CHARACTERISTICS
Time distribution via geostationary satellites	GOES (ATS-1)	σ_t = 100-10 µs	• one-way • satellite as relay • regional, continuously available
Time distribution by navigation satellites	TRANSIT	σ_t = 20 µs	• one-way • two frequencies • on-board crystal clock; • satellite as transmitter • global, intermittend availability
Terrestrial clock-comparisons via geostationary satellites	experiments with • ATS - 1,3 • SYMPHONIE • etc.	σ_t = 5µs - 20µs	• two-way • satellite as relay • large ground antennas • point-to-point contacts
Clocks in space	1. TIMATION 2. Red-shift experiments	σ_t = < 1 µs $\Delta f/f = 1{,}5\cdot10^{-14}$	1. • one-way • side-tone modulation • Rb-clock on board 2. • H-maser on board of a rocket • 3-way doppler-cancelling • 3-frequencies • 2 hours measurement time

Fig. 11.: Some transfer systems, which use band-limited signals.

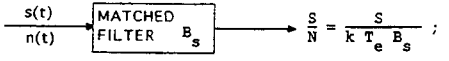
A H-maser clock on board of a rocket was compared to a ground-maser by this method in the red-shift experiment 1976 /3/. An extremely high frequency-transfer accuracy was achieved by using extremely narrow time signals (= sinus-waveforms) and by eliminating ionospheric and Doppler-influences by a special 3-way Doppler-cancelling method. This is a combination of a 2-way Doppler-measurement and a 2-frequency measurement of the ionospheric electron content. For further information on this sophisticated transfer-method see /3/.

6. Correlation-Detection of broad-band Signals

A very sensitive method for the detection of time-signals in noise is the use of a correlation filter. This is a logic advancement of the matched frequency filter and works as shown in Fig. 12. In this case it is necessary to know exactly the waveform of the expected time-signal and to generate a replica of the expected signal in the receiver. By multiplying this replica with the received signal and smoothing the product by a low pass filter, an output voltage is achieved, which is proportional to the autocorrelation function R_{ss} of the time signal and which depends on the time shift τ between the received signal and the reference signal. Thereafter a control device has to shift the reference signal in such a way, that the output voltage R_{ss} becomes a maximum; this implies that $\tau = 0$ and that the reference signal has exactly the same position in time as the received time signal. Thus, the reference signal can be led to the receiver output instead of the received signal. The random noise on this output signal is determined by the bandwidth B_L of the low pass filter, which is in every case much smaller than the signal bandwidth. By this way, weak time signals of unlimited bandwidth can be detected in noise, because a high S/N ratio can be achieved, independently of the signal bandwidth.

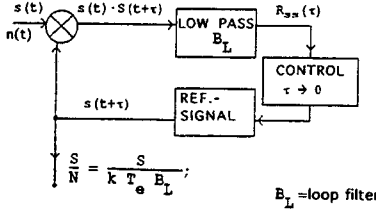
But there arises a new systematic error source, because of an unavoidable small time-lag between the received time signal and the controlled reference signal. This lag depends on different parameters of the control-loop and on the slope of the function $R_{ss}(\tau)$. To minimize this error, it is necessary to choose a signal with an optimum auto-

a) CONVENTIONAL FREQUENCY FILTER:



B_s =signal bandwidth

b) CORRELATION - FILTER:



B_L =loop filter bandwidth;

S/N is independent of signal bandwidth, but

c) A SIGNAL WAVEFORM IS NECESSARY WITH A SUITED AUTO-KORRELATION FUNKTION (ACF)

$$R_{ss}(\tau) = \frac{1}{T_P}\int_0^{T_P} s(t) \cdot s(t+\tau)\, dt$$

$$= \max. \text{ with } \tau = 0 \pm n \cdot T_P;$$

Fig. 12: T_P = signal period

SIGNAL DETECTION WITH A CORRELATION-FILTER

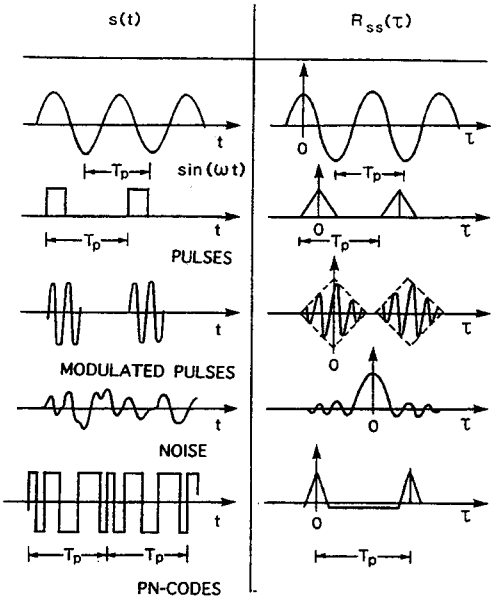


Fig. 13: TIME-SIGNALS AND THEIR AUTOCORRELATION-FUNCTIONS

correlation function (ACF) for this purpose. Fig. 13 shows such functions $R_{ss}(\tau)$ for different time signals. It is obvious that pulses and Pseudo-Noise (PN) codes meet this requirement very well. The latter ones are preferred in practical systems, because the problems of high pulse peak power can be avoided.

7. Transfer Systems using broad-band Signals

Without going into the details on PN-code signals and their application in satellite navigation and time transfer systems only three basic aspects shall be mentioned here: The desired ACF can be achieved only with codes of a special length, called "Maximum-Length Codes". Fig. 14 shows an example with the length $L = (2^n - 1) = 15$ chips ($n = 4$). The Global Positioning System (GPS) of the USA has "Clear-Acquisition" codes of the length $L = (2^{10} - 1) = 1023$ chips.

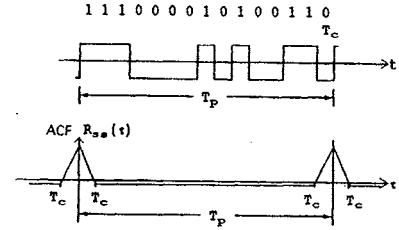
Another advantage of PN-codes is that time signals with different PN-codes on the same carrier frequency can be distinguished from each other, if the used codes are orthogonal, that is to say, that the Cross-Correlation-Function (CCF) of these two codes is about zero for each time shift τ.

The navigation system GPS uses a family of such orthogonal codes (Gold-codes), whereby each satellite transmits at the same carrier frequency his own PN-code, which is orthogonal to the code of any other satellite.

This possibility of code-multiplex is not used in the Russian navigationsystem GLONASS, which uses the same PN-code for each satellite, but modulated on different carrier frequencies.

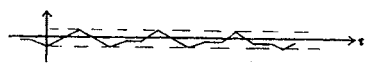
PN-coded time signals were developed and applied at first in satellite navigation systems, but the same technique can easily be transferred to other time-transfer or ranging systems: remarkable improvements in the two-way time transfer via geostationary satellites can be achieved by using broad-band signals, and a direct comparison of precise clocks in space and on the ground will require this techni-

MAXIMUM LENGTH CODE: $L = (2^n-1)$ bit; n = even
PERIOD $T_P = (2^n-1) \cdot T_c$;
e.g. L = 15



GOLD CODES = ORTHOGONAL ML-CODES

$$CCF = |R_{ij}(\tau)| < \frac{2^{(n+2)/2} + 1}{(2^n - 1)} \quad ; \quad n = \text{even}$$



GPS-SIGNALS = FAMILY OF GOLD CODES

	T_P	T_c	L	f_o [MHz]	P_E[dbW]
C/A-CODE	1 ms	10^{-6}s	1023	l_1= 1575,42	≥ - 160
P-CODE	267 days	10^{-7}s	$2,3 \cdot 10^{14}$	l_2= 1227,6 l_1= w.o.	≥ - 166 ≥ - 163

Fig. 14: PSEUDO-RANDOM-NOISE-CODES

APPLICATION	EXAMPLE	ACCURACY	CHARACTERISTICS
TIME DISSEMINATION VIA GEOSTATIONARY SATELLITES	ENVISAGED: - TDRS	σ_t = 10 ns?	• ONE-WAY • ONE FREQUENCY • PN-CODES + DATA
NAVIGATION SATELLITES - FOR TIME DESSIMINAT. - FOR CLOCK-COMP. - PROJECT	GPS GLONASS COMMON VIEW GNSS	σ_t = 30 100 ns σ_t = 30 ns σ_t = 2 - 10 ns σ_t 1 - 10 ns (?)	• ONE-WAY • ONE OR TWO FREQUENCIES • CS-, RB-CLOCKS ONBOARD • GEOSTAT. SATELLITE AS A RELAY • PHASE-COHERENT TRANSPONDER
TERRESTRIAL CLOCK-COMPARISONS VIA GEOSTAT.SATELLITES - WITH MICROWAVES - WITH LASER	EXPERIMENTS: • ATS - 1, 3 • SYMPHONIE • INTELSAT • LASSO	σ_t = 10 100 ns σ_t = 1 ns	• TWO-WAY / PN-CODE-SIGNALS • SATELLITE AS RELAY • LARGE GROUND ANTENNAS • POINT-TO-POINT CONTACTS TWO-WAY
CLOCKS IN SPACE - SHUTTLE EXPERIMENT - PROJECT	NAVEX SPACE-MASER	σ_t = 5 - 10 ns σ_t < 0,1 ns < 0,01 ns $\Delta f/f < 5 \cdot 10^{-15}$ (T = 12 h)	• TWO-WAY; L-BAND • PN-CODES + DATA • CODE-PHASE MEASUREMENT • MICROWAVE X-, S-BAND • CODE PHASE + CARRIER PHASE MEASUREMENT • 2 SIGNAL FREQUENCIES

Fig. 15: Some time transfer systems using broad-band signals.

que. Both applications, mentioned above, were investigated experimentally during the last years. Some of these systems and experiments are compiled in Fig. 15.

Two projects are mentioned in this table, which may be of some importance in the future and which should be explained shortly as an outlook to the future.

8. Future Developments

There is a strong interest in aeronautics as in other civil organisations, to have a satellite navigation system like GPS or GLONASS under international control, which should be more adapted to civil requirements than the existing systems, because these are provided and operated mainly for national, military purposes.

One of the most promising developments in the direction of an international Global Navigation Satellite System (GNSS) is the concept of an 'Overlay' to GPS and GLONASS, which will consist of some additional geostationary navigation satellites, which work as a relay for GPS/GLONASS-like time signals and for the distribution of additional data, necessary for the improvement of reliability, integrity and accuracy of the system /4,5/. These time-signals will be free of intentional accuracy limitations like S.A. ("selective availability") and will have several other advantages for future time transfer, as compiled in Fig. 16 /6,7/.

Another progress in global time transfer can be expected from more precise clocks in space. The one-way ranging method, which is used for the control of clocks on navigation satellites, is a combined measurement of time deviation and satellite position. The accuracy of this method is limited by the fact that a wide spread net of ground stations with synchronized clocks is required.

The first investigation of a two-way time transfer with broad-band PN-code signals between clocks in space and on the ground was performed in the navigation experiment NAVEX during the first German Space-

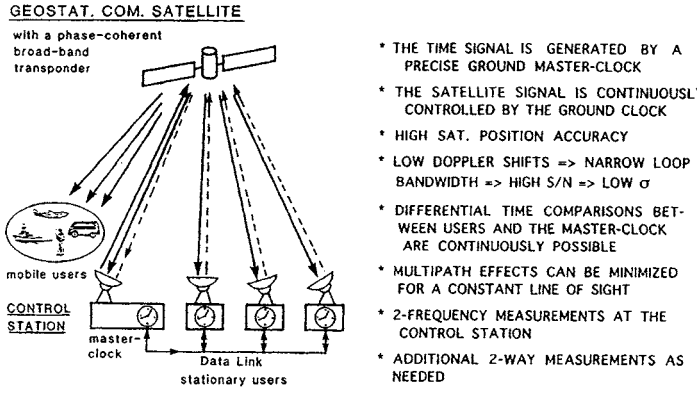


Fig. 16 ADVANTAGES OF FUTURE GEOSTATIONARY NAVIGATION SATELLITES

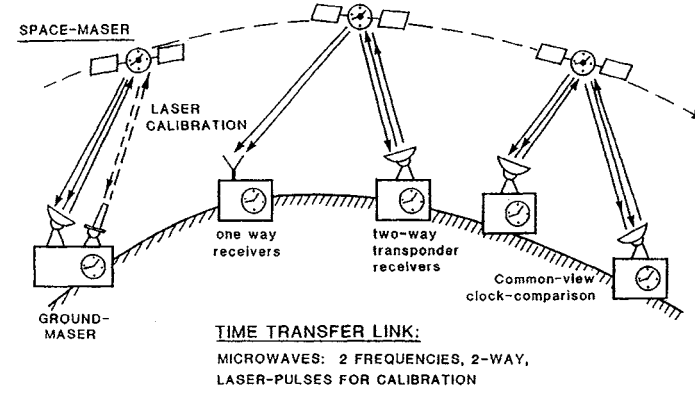


Fig. 17 H-MASERS IN A POLAR ORBIT CAN, TOGETHER WITH A PRECISE TT-LINK, CONNECT THE BEST CLOCKS ON THE EARTH TO A UNIQUE SET.

Shuttle mission 1985 /8/. Since that time the development of H-maser clocks for space applications made progress in the USA and in Russia. ESA started such a development recently in view of future common experiments and applications in this field /9,10/. The time transfer for the control of these space-clocks will be performed by a two-way microwave link, combined with a two-frequency measurement of ionospheric effects and a LASER calibration-link (Fig. 17). The accuracy will be improved relative to NAVEX by using higher carrier frequencies and an additional carrier phase measurement. The realization of this link shall be based on an existing equipment for precise range and range rate measurements (PRARE), which was space-proofed just last month /11/.

The combination of H-masers in space with such an optimized transfer link will improve the global time transfer to about 0,1 ns and the frequency transfer to about 10^{-15} per day or better.

9. References

/1/ D.W. Allan, A. Lepek: Trends in international timing. Proc of the 7th European Frequency and Time Forum, 16-18 March 1993; pp. 221-230.

/2/ S. Starker, E. Sappl, W. Schäfer: Microwave links for precise time- and frequency transfer between ground- and space-based clocks. Proc. of the 7th European Frequency and Time Forum, 16-18 March 1993; pp

/3/ R.F.C. Vessot e.a.: Tests of relativistic gravitation with a space-borne hydrogen maser. Phys. Rev. Lett., Vol 45, pp. 2081-2084, Dec. 29, 1980.

/4/ S. Starker: A GPS-NAVSTAR-Supplement as a possibility for autonomous civil satellite navigation. 35th Congress of the IAF, October 7-13, 1984, Lausanne.

/5/ J. Barbier, T. Tremas: European Complement to GPS, ... Proc. of the 7th EFTF, 16-18 March 1993.

/6/ S. Starker, H. Nau, G.V. Kinal, J. Willner: Geostationary Navigation Satellites as a Means of Time Transfer. Proc. of the 4th Eurepean Frequency and Time Forum 13-15 March 1990.

/7/ A. Brown, D.W. Allan, R. Walton: Precise Time Dissemination using the INMARSAT Geostationary Overlay. IEEE Frequency Control Symposium, Salt Lake City, June 1993.

/8/ S. Starker, D. Arko, J. Hammesfahr, H. Nau, E. Sappl, H. Schild, R. Schimmel, F. Stelzel: Navigation Experiment NAVEX on Spacelab Mission D1. Final Report, DLR, Institute of Radio Frequency Technology, D-82230 Wessling, 1990.

/9/ G. Busca, S. Starker, S. Feltham: In Orbit Demonstration of a H-Maser Clock System. Proc.: 23rd Anual Precise Time and Time Interval (PTTI) Applications and Planning Meeting, NASA-CP 3159, 1992, 11 p., pp. 385-399, 4 Figs., 11 refs.

/10/ R.F.C. Vessot, E.M. Mattision, G.U. Nystrom, R. Decher, S.J. Feltham, G. Busca, S. Starker and S. Leschiutta: Hydrogen Masers for Space. Proc. of the 6th European Frequency and Time Forum, ESA-SP 340 (1992), 7 p., pp. 19-25, 8 figs., 1 tab., 17 refs.

/11/ W. Schäfer
The Application of the PRARE Principle for Precise Two-way Time Transfer to Space-based Atomic Clocks.
6th European Frequency and Time Forum, Noordwijk/The Netherlands, 17-19 March, 1992, ESA-SP 340 (1992), pp 41-47, 8 figs., 3 refs.

Two-Way Satellite Time Transfer Experiments Between Six European Laboratories Using The INTELSAT (VA-F13) Satellite

J A Davis and P R Pearce
National Physical Laboratory
Queens Road, Teddington, Middlesex, United Kingdom TW11 0LW

D Kirchner
Technical University Graz, Graz, Austria

H Ressler
Space Research Institute, Graz, Austria

P Hetzel
Physikalisch-Technische Bundesanstalt, Braunschweig, Germany

A Soering
Forschungs- und Technologiezentrum, DBP Telekom, Darmstadt, Germany

G de Jong
NMi Van Swinden Laboratories, Delft, The Netherlands

P Grudler and F Baumont
Observatoire de la Cóte d'Azur, Grasse, France

L Veenstra
International Telecommunications Satellite Organization
Washington, D.C. United States of America

Abstract

The results of Two-Way Satellite Time and Frequency Transfer (TWSTFT) experiments made between six European laboratories, using the INTELSAT (VA-F13) satellite located at 307 °E, are reported. Time transfers were performed between all of the laboratories over an extended period. The use of the spread-spectrum technique enabled up to three time transfers to be performed simultaneously. The characteristics of the time transfers made between each pair of ground stations were examined. Possible interference effects due to multiple transmission of spread-spectrum signals and additional clean carrier signals were examined and shown to be minimal.

Keywords
Time transfer, frequency transfer, satellite communications, spread spectrum techniques.

1. Introduction

Two-Way Satellite Time and Frequency Transfer (TWSTFT) offers the potential of both time and frequency transfer at high accuracy. Ultimately, sub-nanosecond accuracy time transfers may be realised. Also, frequency comparisons with uncertainties of parts in 10^{14} or less may be possible with averaging times as short as a few hours. At present, routine comparisons between atomic clocks contributing to Universal Coordinated Time (UTC) are realised using the Global Positioning System (GPS) in the satellite 'common view' mode /1/. TWSTFT is, however a method of time transfer which is independent of GPS, and offers higher precision, and potentially higher accuracy. TWSTFT may be used as a check on the performance of the GPS common view method.

The TWSTFT method requires the two laboratories undertaking a time transfer to transmit and receive signals simultaneously via a satellite. The earth stations should preferably be located close to the clocks under comparison. Small Ku-band earth stations are used to communicate via a geostationary satellite /2/. The use of a Code Division Multiple Access (CDMA) spread spectrum technique enables multiple transmissions to be made to the satellite at the same frequency, and it limits the power spectral density transmitted by each of the earth stations. TWSTFT modems have been developed /3/, their delay stability has been measured /4/, along with the delay stability of the associated earth station /5/, /6/, /7/. The results of several time transfer experiments have been reported including accounts given in /8/, /9/ and /10/.

In this paper, the preliminary results of a time transfer experiment between six laboratories are reported. The simultaneous accessing of a common space segment by six earth stations enabled detailed measurements to be made of the relative performance of each station, and of the common characteristics of the TWSTFT system.

2. Laboratory Earth Stations

The six laboratories participating in this experiment are the Technical University Graz (TUG), National Physical Laboratory, Teddington (NPL), Van Swinden Laboratories, Delft (VSL), Physikalisch-Technische Bundesanstalt, Braunschweig (PTB), Forschungs- und Technologiezentrum, DBP Telekom, Darmstadt (FTZ), and the Observatoire de la Cote d'Azur, Grasse (OCA). Of these laboratories TUG, FTZ and OCA require precise time for research, telecommunications and astronomical applications respectively, while the remaining laboratories generate and maintain their own national time scales.



Figure 1 Locations of the Earth Stations

Table 1 Earth Station Characteristics

Laboratory	TUG	NPL	VSL	FTZ	PTB	OCA
Latitude Longitude Altitude	47°04'04"N 15°29'39"E 493 m	51°25'15"N 359°39'37"E 23 m	52°00'02"N 04°22'56"E 20 m	49°52'08"N 08°37'33"E 126 m	52°17'50"N 10°27'38"E 143 m	43°45'00"N 06°55'00"E 1260 m
Antenna diameter	1.8 m	2.4 m	3.0 m	1.8 m	1.8 m	1.8 m
Maximum EIRP	52.5 dBW	52.0 dBW	47.5 dBW	53.0 dBW	53.0 dBW	52.6 dBW
Receive G/T at 11.0 GHz	21.4 dB/K	21.9 dB/K	22.6 dB/K	22.9 dB/K	22.9 dB/K	20.3 dB/K
Elevation Angle	5.8°	13.6°	10.9°	9.4°	7.3°	12.8°
Clocks available	HP 5071 Cs	Active H maser HP 5071 Cs	HP 5071 Cs	HP Cs	Active H maser, Primary Cs standard	HP Cs
Frequency agility	1 kHz	1 kHz	1 kHz	1 kHz	1 kHz	1 kHz
MODEM	MITREX (Original)	MITREX (Original)	MITREX (Original)	MITREX (Modified)	MITREX (Modified)	MITREX (Modified)

The locations and characteristics of the earth stations are shown in Figure 1 and Table 1 respectively. The earth stations at PTB and FTZ were purchased as commercially available Dornier systems, the stations at TUG and OCA are based on commercially available Very Small Aperture Terminals (VSAT), with additional hardware added to provide 1 kHz frequency agility /11/. The stations at NPL and OCA have been constructed using commercially available satellite communications equipment. The clocks under comparison were located close to the earth stations. The greatest separation between the timing centre and the earth station occurs at NPL and PTB, where 250 m cable links are required.

3. Space Segment

The INTELSAT (VA-F13) satellite located at 307°E was chosen for these experiments. This satellite may be used to perform time transfers within Europe and between Europe and North America within a single measurement session. This eliminates the requirement to re-position the earth station antenna, during a program of measurement sessions. The motorisation of the antenna is therefore not essential for this work, this reduces the cost of hardware. The provision of cost free satellite time by INTELSAT enabled the time transfer experiment to take place.

A disadvantage of using the INTELSAT (VA-F13) satellite at 307°E is the relatively low elevation angles which occur both for stations within Europe (Table 1) and for stations within parts of North America. The characteristics of time transfers undertaken using low elevation angle stations is examined in this paper.

4. Establishment and Operation of the Earth Stations

4.1 Establishment of a TWSTFT station

The establishment of an earth station for TWSTFT work requires both the station's construction and the obtaining of permission to access the relevant satellite operator's space segment.

Satellite communications hardware is available commercially. The obtaining of a TWSTFT modem is however a problem, as the MITREX modems /3/ used in these experiments are no longer commercially available. Alternative modems, which are compatible with the MITREX modem are under development /12/, /13/, /14/. The assembly of the earth station equipment is relatively straightforward.

The earth stations must be approved for operation by the relevant satellite operator, which in this case is INTELSAT. This requires verifying the performance characteristics of the station. There are two approaches to obtaining earth station verification. Firstly, verification tests may be performed at the station, where test transmissions are monitored by an assisting earth station. The use of 'type approved' hardware may reduce the number and complexity of the required tests. Alternatively, the transmit and receive characteristics of the earth station may be determined without making satellite transmissions, these measurements may then be submitted to the satellite organisation. In both cases the verification test procedure is administered by the earth station operator's National Signatory on behalf of the satellite organisation. Clearance must also be obtained from the earth station operator's Radio Regulatory Authority.

4.2 Transmission of clean carrier signals

Line-up tests were performed using the INTELSAT (VA-F13) satellite at 307°E. Clean carrier signals were up-linked from all of the earth stations, these transmissions were adjusted to operate at frequencies and EIRP levels agreed in the INTELSAT transmission plan. Initially the up-link EIRPs were set to 47 dBW, with each clean carrier signal spaced 20 kHz from its neighbour. The relative EIRP levels and frequencies of the transmissions were then equalised within 1 dB and 1 kHz respectively. An example of the simultaneous transmission of four clean carrier signals is shown in Figure 2.

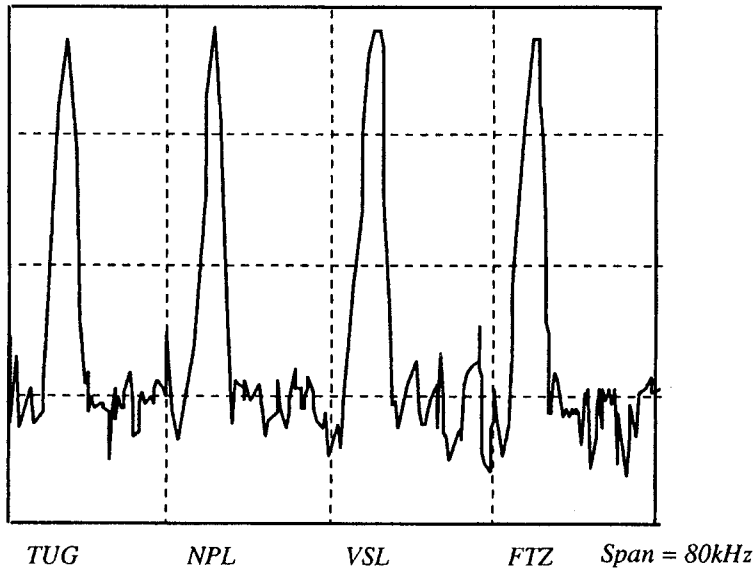


Figure 2 Simultaneous transmission of four clean carrier signals

4.3 Power stability of the received clean carrier signals

The stability of the clean carrier signals were simultaneously monitored at NPL, an example is shown in Figure 3. Correlation between variations in the signal power levels may be due to the common down-link. The relatively large power level variations of the signal transmitted by TUG is due to rain fade, the reduction in signal level between 13:15 UTC and 14:10 UTC correlates with a local storm. The Earth stations with low elevation angles are more sensitive to rain fade. Occasionally, rain fade may also severely attenuate the transmissions from the earth stations with relatively high elevation angles.

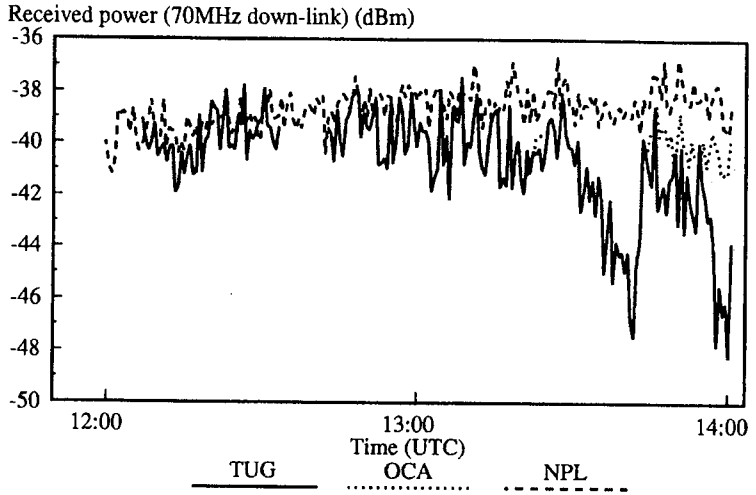


Figure 3 Stability of the clean carrier signals received at NPL.

4.4 Operation of the earth station

The operation of the earth station for TWSTFT experiments is relatively simple. Pre-arranged measurement schedules were easily followed without any telephone contact being made between the transmitting earth stations. INTELSAT Operations Centre (IOC) was informed before the start of transmissions. The up-link frequency and EIRP level were checked before starting transmissions. Satellite station keeping was good, re-positioning of the earth station antenna between measurement sessions was not required. During these European experiments the earth stations were able to receive their own transmitted signals. This enabled the carrier to noise power density ratio C/N_0 to be checked by performing ranging experiments, again without the need to make telephone contact with other stations. Up-linking of coded ranging signals usually resulted in the modem locking within approximately 20s of the start of transmissions. Drifts in the satellite translation frequency and up-link frequencies between measurement sessions were usually small and were corrected after the modem has locked.

5. Two-Way Time Transfers

5.1 The dependence of measurement noise on the EIRP level

An example of a typical short time transfer is shown in Figure 4. The variation of the short term link noise with transmitted EIRP was examined by making a series of successive ten minute ranging experiments within the same measurement session. Examples are shown in Figure 5.

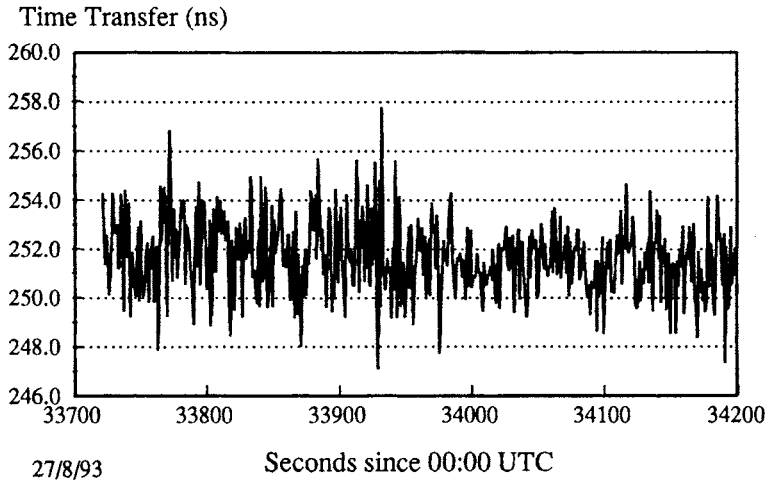


Figure 4 FTZ-TUG Time transfer

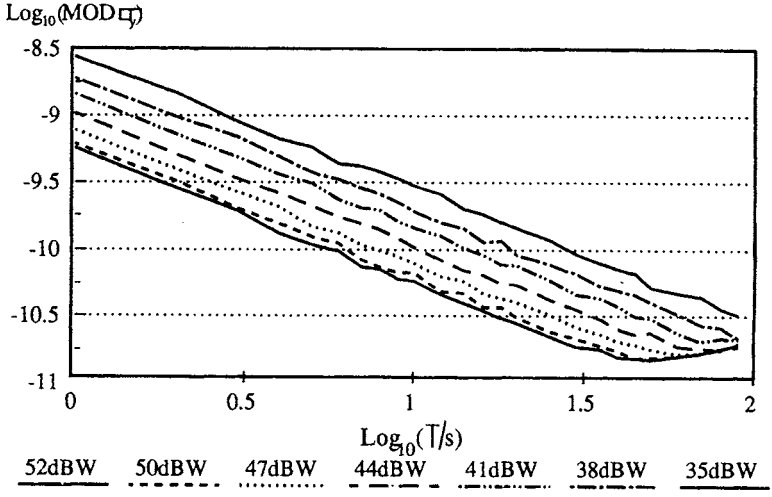


Figure 5 Plots of $Log_{10}MOD\sigma_y$, against $Log_{10}\tau$ for ranging measurements made in the same measurement session at different EIRP levels

5.2 Correlation effects within a time transfer

Two-way time transfers are calculated from two data sets of link measurements, one set recorded at each earth station, L_1 and L_2. Short-term noise present on these links may originate in both the TWSTFT system and in the generation of the one Pulse Per Second (1PPS) clock signals. The value of $MOD\sigma_y$ for the two component links for short averaging times τ may often differ significantly. The time transfer T is determined by calculating the half difference of the link measurements and adding a correction for systematic delay asymmetries /15/.

The half sum, H of the link measurements may also be calculated. If there is no correlation between the noise on each of the links then

$$MOD\sigma_y(T) = MOD\sigma_y(H) = 0.5\left[(MOD\sigma_y(L_1))^2 + (MOD\sigma_y(L_2))^2)\right]^{0.5} \quad (1)$$

which is normally the case.

For the PTB-NPL time transfers shown in Figure 6 $Log_{10}MOD\sigma_y(H)$ was less than $Log_{10}MOD\sigma_y(T)$ over short averaging times. Noise within the two link measurements correlated which was probably due to variations in the generation of chain of 1pps signals supplied to the time transfer modems.

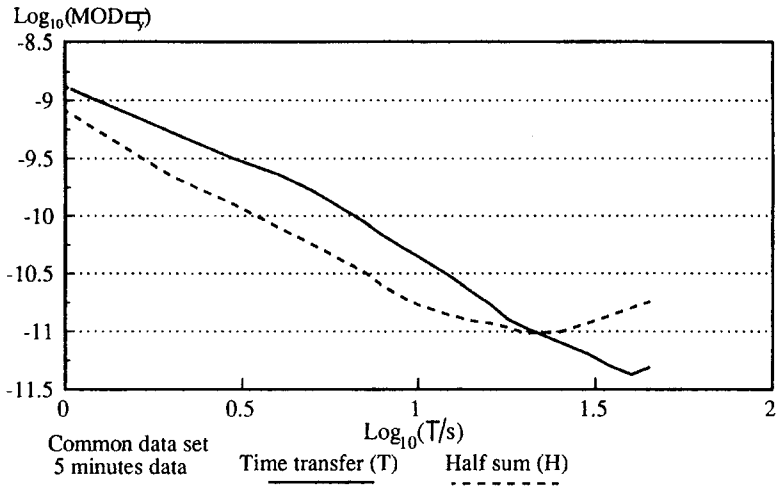
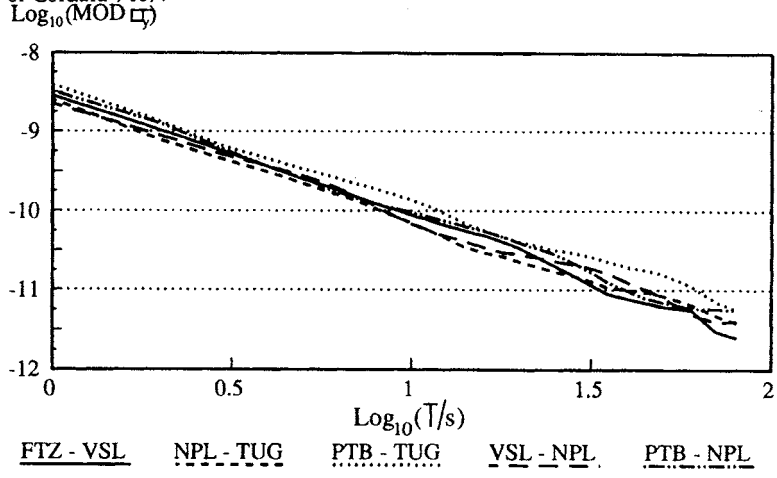


Figure 6 Plots of $Log_{10}MOD\sigma_y$ against $Log_{10}\tau$ for time transfer (T) and half sum (H) values, calculated from the same data set

5.3 Comparisons of time transfers made between different pairs of earth stations

The CDMA spread spectrum technique enables simultaneous time transfers to be made between several pairs of Earth stations. Up to three time transfers were simultaneously performed enabling the measurement noise of each time transfer to be examined. Curves of $Log_{10}MOD\sigma_y$ against $Log_{10}\tau$ are shown in Figures 7a and 7b, two figures are used for ease of presentation only. All ten curves are similar, there is no time transfer which is substantially better or worse than any other, all agree within measurement error. The data spread for time transfers involving stations with

low elevation angles eg (PTB-TUG) were comparable with time transfers made between laboratories with high elevation angles eg (VSL-NPL). The measurement noise may be slightly lower for time transfers where both stations are using the original MITREX modems for example (VSL-TUG). This is consistent with the work of Cordara /16/.



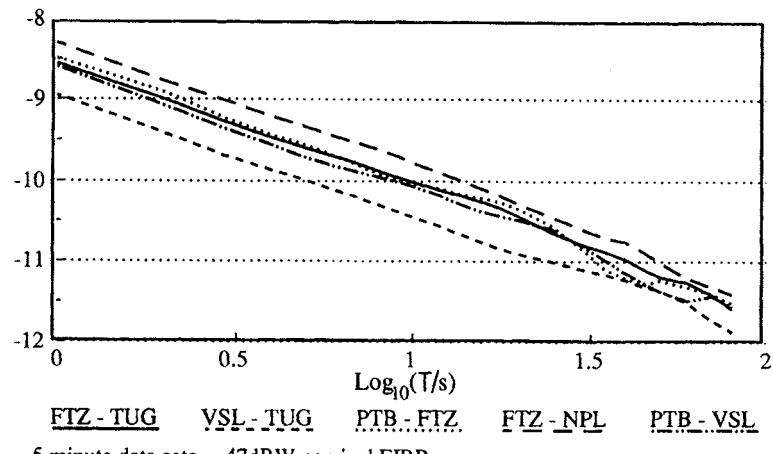


Figure 7a and 7b Plots of $Log_{10}MOD\sigma_y$ against $Log_{10}\tau$ for all time transfer made in a single measurement session

5.4 Interference effects

The degree of interference between simultaneously transmitted coded signals may depend on both the signal levels and on the orthogonality of the modem codes. If the coded signals all have approximately the same level, then possible interference effects will depend principally on any limitations to the orthogonality of the codes.

A time transfer was performed continuously between NPL and PTB. Three additional coded signals were successively superimposed on the time transfer by the remaining earth stations, the short term noise of the (PTB-NPL) time transfer was then compared. The experiment was repeated using additional clean carrier signals as interference.

Examples of curves of $Log_{10}MOD\sigma_y$ against $Log_{10}\tau$ are shown in Figure 8. With signals transmitted at the 47 dBW level there was no correlation between the transmission of additional clean carrier or coded signals and an increase in the noise level of the original signal.

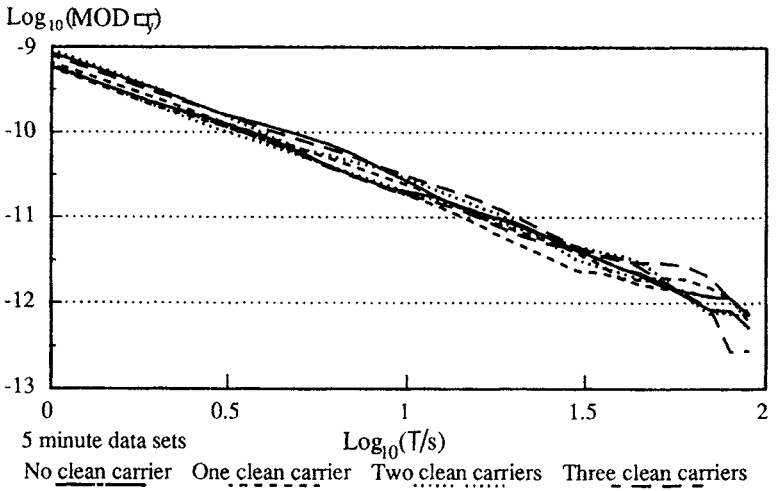


Figure 8a The addition of clean carrier signals to a PTB-NPL time transfer. Curves of $Log_{10}(MOD)\ \sigma_y$ against Log_{10} are plotted.

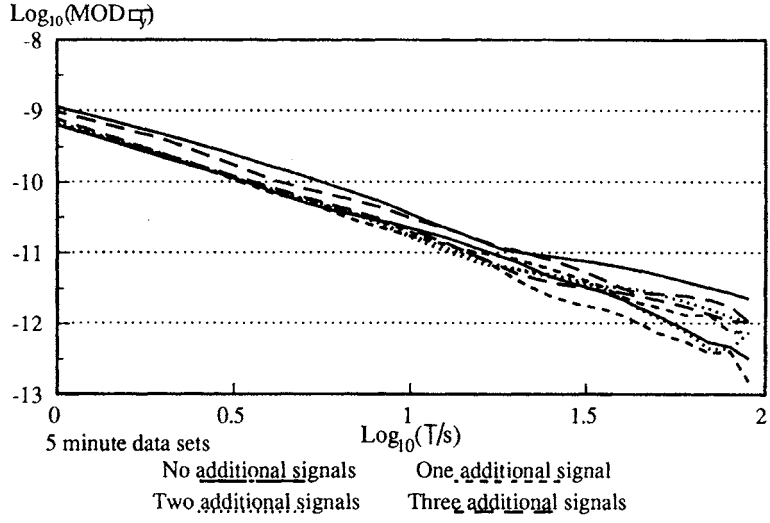
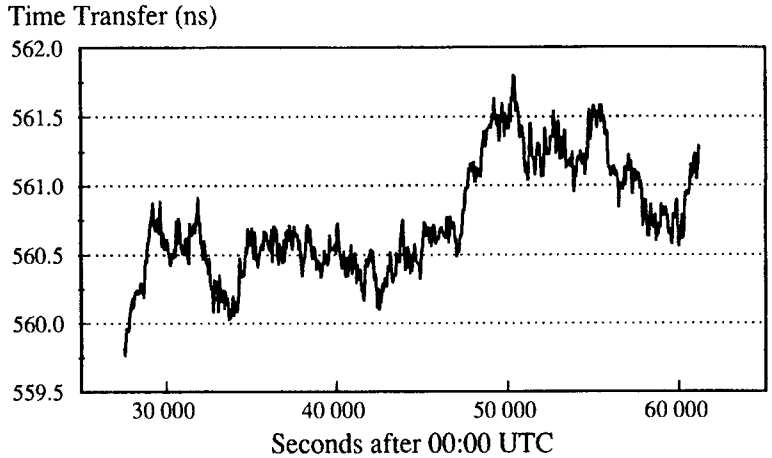


Figure 8b The addition of coded signals to a PTB-NPL time transfer. Curves of $Log_{10}MOD\sigma_y$ against $Log_{10}\tau$ are plotted

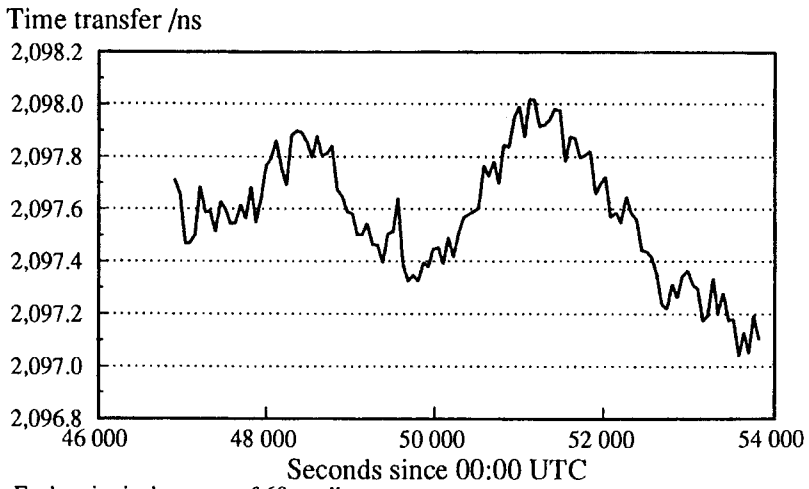
5.5 Time transfers over longer periods

Time transfers have been performed over extended periods. Figures 9 and 10 show plots of NPL-TUG and PTB-NPL time transfers recorded over periods of eight and two hours respectively. Curves of $Log_{10}\sigma_x$ against $Log_{10}\tau$ for both time transfers are shown in Figure 11. The curves of $Log_{10}\sigma_x$ against $Log_{10}\tau$ for the NPL-TUG time transfer show a sharp gradient change from -½ to ½ at an averaging time of 100s. This is attributed to the principal source of noise changing from the TWSTFT system to the clocks under comparison. A HP 5071 Cs standard and an active hydrogen maser were used at TUG and NPL respectively. The later portion of the curve is consistent with the transfer σ_x being limited by clock noise. This result is significant because the performance of atomic clocks may be examined in detail, after only a few minutes of time transfer operation. The averaging time at which the gradient change occurred was found to depend on both the clocks under examination and the uplink EIRPs.



Each point is the mean of 60 readings

Figure 9 NPL-TUG Time transfer



Each point is the mean of 60 readings

Figure 10 PTB-NPL Time transfer

Plots of $Log_{10}\sigma_x$ against $Log_{10}\tau$ for the PTB-NPL time transfer also show a sharp gradient change. Active hydrogen masers were used at both locations. The sharp gradient change occurring when τ = 200s and the following increase in σ_x is due to delay instabilities in the TWSTFT system rather than to the clocks under examination.

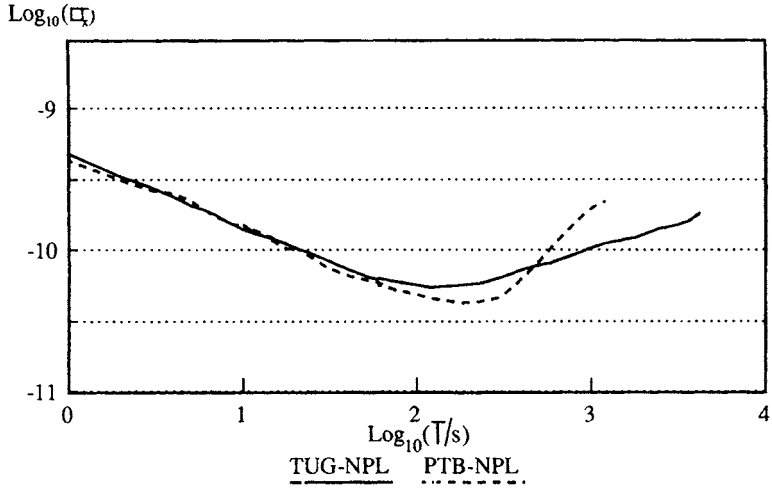


Figure 11 Plots of $Log_{10}\sigma_x$ against $Log_{10}\tau$ for both the (NPL-TUG) and (PTB-NPL) time transfers

A possible explanation for the difference in time transfer characteristics over the two links may be made in terms of the differences in the delay stabilities of the modems which are used. A modified MITREX modem is used at PTB while TUG and NPL used the original MITREX modems. The delay stability of the modified modems are much lower than the delay stability of the original modems. This is due in part to the signal delays within the modified modems having a much higher temperature dependence. There is also an indication of periodic delay changes occurring within the modified modems /16/.

6. Future Work

6.1 Transatlantic TWSTFT

An immediate objective is to extend the TWSTFT work to perform transatlantic time transfers. Experiments have recently started. Transatlantic TWSTFT differs from the European experiment in that separate satellite transponders carry the two transmissions. It is not usually possible to observe ones own transmission. The principal concern is the possible delay asymmetries occurring between the two transponder paths, and any changes in this asymmetry.

6.2 Absolute delay calibrations

The differential delay of the TWSTFT earth stations must be determined for both European and transatlantic time transfers in order to determine the absolute time transfer. A portable 'calibrating' earth station is being prepared for this work. The earth station calibration using a satellite simulator will also be performed in parallel /17/ as an alternative method. Accurately calibrated earth stations may form the basis of a detailed comparison of the TWSTFT and GPS common view methods over the same multiple time transfer links.

7. Conclusions

The feasibility of using TWSTFT for routine comparisons between a group of six laboratories within Europe has been demonstrated. Multiple time transfers may be undertaken without any significant interference effects. Averaging times of only a few hundred seconds are required for the accurate comparison of commercial Cs clocks. Initial indications suggest that the time transfer accuracy is limited at the nanosecond level by delay instabilities within the TWSTFT instrumentation. Further evaluation is required.

Acknowledgements

This work has been supported under the UK National Measurement System Program for Time and Frequency, by the Austrian Academy of Sciences and the Jubilee Fund of the Austrian National Bank.

References

/1/ Lewandowski, W and Thomas, C: 'GPS time transfer', IEEE Proc. Vol.79, No.7, pp 991-1000, 1991.

/2/ Kirchner, D and Riedler, W: 'Timing by satellite: Methods, recent developments and future experiments' Proc. Int. Symp. Satellite Transmissions, ESA SP-245, 1985 pp 175-186.

/3/ Hartl, Ph, Gieschen, N, Mussener, K M, Schafer, W, Wende, C M: 'High accuracy global time transfer via geosynchronous telecommunication satellites with MITREX', J. Flight Sciences and Space Research, Vol.7, No.5, pp 335-342, 1983.

/4/ Kirchner, D, Ressler, H. and Soering, A: 'Messungen mit dem MITREX-modem' Acte due Congres de Chronometrie CP 6, 1986, pp 71-75.

/5/ Davis, J A and Pearce, P R: 'Characterisation of the signal delays in a ground station designed for satellite two way time transfer'. Proc. Seventh European Frequency and Time Forum, pp 113-118, March 1993.

/6/ Kirchner, D, Ressler, H, Riedler, W and Soering, A: 'A two-way time transfer experiment via ECS-1 using the MITREX Modem', IEEE Trans. Instrum. Meas., Vol.37, No.3, pp 414-417, 1988.

/7/ Delay stability of satellite earth stations as used in two-way transfer, CCIR Radiocommunication study groups, document 7-2/7-E.

/8/ Howe, D A: 'High accuracy time transfer via geostationary satellites: Preliminary results', IEEE Trans. Instrum. Meas. Vol.37, No.3, pp 418-423, 1988.

/9/ Hetzel, P and Soering, A: 'Two-way time comparisons via DFS-Kopernikus 2 between FTZ and PTB', Proc. Seventh European Frequency and Time Forum, pp 125-128, March 1993.

/10/ Takahashi, F et al: 'Two way time transfer experiments using an Intelsat satellite in an inclined orbit', Proc. IEEE Trans. Instrum. and Meas. Vol.42, No.2, pp 498-504, 1993.

/11/ Kirchner, D: 'Design considerations for KU-band two-way satellite time transfer terminals', Proc. Fourth European Frequency and Time Forum, pp 631-638, 1990.

/12/ Landis, G P, Galysh, I: 'NRL/USNO two way time transfer modem design and results', IEEE Freq. Control Symp. pp 317-326, 1992.

/13/ Baumer, C, Dienert, M: 'High accuracy two-way time transfer and ranging experiments using SATRE modems', Proc. Seventh European Frequency and Time Forum, pp 345-348, 1993.

/14/ Davis, D D and Howe, D A: 'A direct sequence spread-spectrum modem design for time transfers via geostationary satellites', Proc. Fifth European Frequency and Time Forum, pp 89-95, 1991.

/15/ Kirchner, D: 'Two way time transfer via communication satellites' IEEE Proc. Vol.79, No.7, pp 983-990, 1991.

/16/ Pettiti, V, Cordara, F, Galliano, P G: 'Two-way time comparisons on Olympus Satellite. Proc. 8th European Frequency and Time Forum.

/17/ De Jong, G: 'Accurate delay calibration of satellite ground stations for two-way time transfer', Proc. 3rd European Frequency and Time Forum, pp 1898-203 and Proc. 21st PTTI meeting, 1989, pp 107-116, 1989.

Two-way time comparisons via Olympus satellite

V. Pettiti, F. Cordara, P.G. Galliano (*), G. Tignanelli (**)

(*) Istituto Elettrotecnico Nazionale Galileo Ferraris
Torino, Italy.

(**) Istituto Superiore delle Poste e Telecomunicazioni
Roma, Italy.

Summary

A two-way satellite time transfer between IEN and ISPT, using Olympus satellite in the Ku band and 2500A MITREX modems, started in July 1992. For this purpose, two mobile Alenia earth stations equipped with 3,3 m dish antennas were used.

The campaign was interrupted in September 1992 for modifications to the modems, it was resumed in March 1993 and continued up to the end of life of the satellite.

Instabilities at the 10 to 15 ns level in the synchronisation results together with long term nearly sinusoidal effects of the same order were observed and investigated.

To separate the contribution of the different pieces of equipment, delay measurements of the two station co-located at the IEN were performed. Moreover, a characterization in laboratory conditions of the two 2500A MITREX modems used, as regards to their sensitivity to the temperature, to the input signal level and to the long term stability of their delays was also carried out.

1 - Introduction

A two way synchronization experiment in the Ku band between the Istituto Elettrotecnico Nazionale (IEN) in Torino and the Istituto Superiore delle Poste e Telecomunicazioni (ISPT), in Roma, using Olympus geostationary satellite (19° W) and two mobile Alenia stations, namely TDS B1 and TDS B3, equipped with 2500A digital MITREX modems, started in July 1992.

The stations were available at the two institutes only for periods of about three days, when the satellite was scheduled for the synchronization experiment.

Seven measurement sessions performed up to the end of September 1992 showed both periodical variations of about 10 ns peak-to-peak and time jumps of the order of 50 ns in the synchronization results, mostly due to the digital modems used.

The experiment was resumed in March 1993, after the modification of the MITREX modems performed at the Institute of Navigation in Stuttgart, and the measurement sessions were interrupted after July 1993 when the satellite became unusable.

The two measurement sessions of May and July 1993, were performed in Torino with the two stations co-located.

2 - The two-way time transfer

This method is well known to the time and frequency community and has been recommended by the Comitè Consultatif pour la Definition de la Seconde (CCDS - Rec. S6) in 1989 and by the International Radio Consultative Committee (CCIR - Rec. 582-1) in 1990. A complete description of its principle is given in the literature /1,2/, and its strenght lies in the near simultaneity of the transmission to a satellite of the local time

references, PRN coded, of the stations involved and in measuring at each station, in its own time scale, the time of arrival of the second pulses received. The time difference between the UTC references maintained in the two laboratories, in our case IEN and ISPT, is given by the following equation:

$$\begin{aligned} UTC(IEN) - UTC(ISPT) = {} & 0{,}5\cdot(\Delta_{TO} - \Delta_{RM}) + \\ & + 0{,}5\cdot[(\tau^{U}_{TO} + \tau^{D}_{RM}) - (\tau^{U}_{RM} + \tau^{D}_{TO})] \\ & + 0{,}5\cdot(\tau'_{S} - \tau''_{S}) \\ & + 0{,}5\cdot[(\tau^{TX}_{TO} - \tau^{RX}_{TO}) - (\tau^{TX}_{RM} - \tau^{RX}_{RM})] \\ & + \Delta\tau_{R} \end{aligned}$$

where:

Δ_{TO} and Δ_{RM} are the counter reading at IEN - Torino and at ISPT - Roma that are exchanged to compute the difference between the two time scales; the second term at the right-end side represents the difference of the sums of the signal delays in the up and down links which at the frequencies used in our experiment (14 and 12 GHz), because of the insignificant influence of the ionosphere on the differential delays, can be neglected; the third term refers to the signal delays of the satellite transponder for both directions: as a unique transponder was used in our case for the signals transmission and reception, it can be neglected; in the fourth term, the differential delays of the earth stations transmitting and receving parts and in the equipment used in the two sites are represented. These last include the differential delays of the MITREX modems and of the IF cables used. In this experiment it was not possible to measure separately the delays of the transmitting and receiving parts, but only the differential delay of the two stations, that was found equal to 84 ns; concerning the delays of the cable connecting the earth stations to the modems, the sum of the Tx and Rx contribution has been measured by means of the MITREX modems signals in both location amounting to 419 ns for IEN and to 1860 ns for ISPT. The fifth term is a correction for the non-reciprocity of the path due to the rotation of the earth (Sagnac effect) that can be computed from the satellite position and the earth stations coordinates, and has been found equal to 16 ns for the case of IEN and ISPT.

The synchronisation results have been computed taking into account the corrections reported above and amounting to a total of 100 ns.

3 - Earth stations and equipment used

The equipment set up used for the experiment at IEN and ISPT is shown in Fig. 1. The time scales were generated by a HP 5061A/004 cesium clock plus a phase microstepper at IEN, and a single cesium HP 5061B/004 in the case of ISPT. No routine GPS link exists between the two scales, but daily measurements using the passive television method are performed.

The digital MITREX modems mod. 2500A, modified to work as the analog ones, were supplied with the local UTC frequency and sychronized at the beginning of each

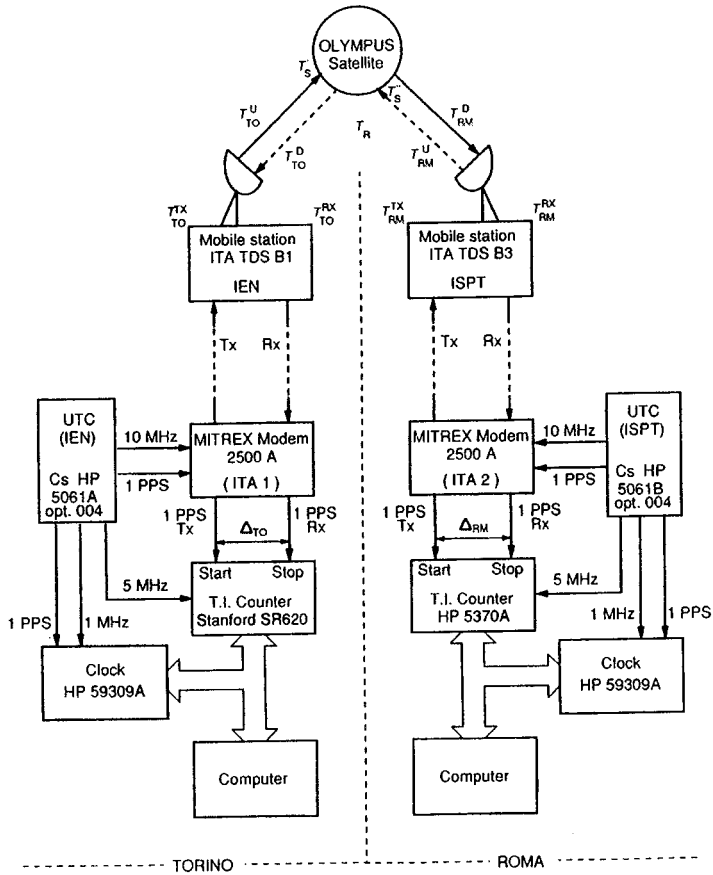


Fig. 1 - Equipment set-up used at IEN and ISPT.

measurement session to UTC(IEN) and UTC(ISPT). The external time interval counters used were namely a Stanford SR 620 with 25 ps resolution at IEN and an HP 5370A with 20 ps resolution at ISPT. To start and stop the synchronization measurements, the 1PPS/T_X and 1PPS/R_X signals supplied by the modems were used.

The earth stations were mobile Alenia stations designed for Olympus satellite and operated at 14,145 GHz (T_X) and at 12,750 GHz (R_X). Their main specifications are reported in table 1.

Table 1 - Earth stations specifications

Stations coordinates	Torino:	45°01' N 7°38'	Roma:	41°50' N 12°28' E
Elevation		33°		44°
Antenna	parabolic of Gregorian type			
Diameter	3,3 m			
Mount	steerable (el. 0° to 60°, az. -90° to +90°)			
EIRP	74 dBW			
G/T	26,5 dB/K			
Frequency agility	T_X = 125 KHz, R_X = 125 KHz			
Frequency range	T_X:13,00-13,25 GHz and 14,00-14,50 GHz R_X :10,95 to 12,75 GHz			

The spot beam 2-B of Olympus satellite was used and the channel bandwidth was 27 MHz. The tipical carrier-to-noise ratio (C/No) in clear sky conditions, was of the order of 70 dB, and the frequency offset of the 70 MHz IF received from the satellite and supplied to the MITREX modems was less than 1 KHz.

Due to the technical problems occurred to Olympus satellite in its launching phase and a consequent lack of fuel to perform orbit manoeuvres, the satellite movement during this experiment was very irregular.

4 - Synchronization results

The two-way comparisons via Olympus were resumed on March 24 and interrupted on July 2, 1993.

The satellite was available for the experiment, on the average, 5 - 6 hours per day, three consecutive days per month, and mostly synchronization measurements were performed. Measurements block duration ranged from 5 minutes to 1 hour with repetition rates of 1 second, 10 seconds and 1 minute. The PRN codes used were 0 for IEN and 1 for ISPT. At the beginning of each session, the difference UTC(i) - $1PPST_X(i)$ was also measured. In practice, only the measurement sessions beginning on March 24 and April 27 were devoted to the time scales comparisons. In Fig. 2, are reported the residuals of the one-second differences between UTC(IEN) and UTC(ISPT), as measured on March 26; the typical standard deviation is of the order of 0,1 ns. Fig. 3 shows the modified Allan variance computed for the same set of data where it can be seen that its dependance on τ fits well with the $-3/2$ slope, typical of white noise PM. Assuming that this behaviour is on the average at the $1{,}2 \cdot 10^{-10} \cdot \tau^{-3/2}$ level, the time necessary to reach the noise level of an HP Clock option 004, expressed in Mod $\sigma_y(\tau)$, is at least 20 s. In figures from 4 to 7 are reported the results of four synchronization sessions, one

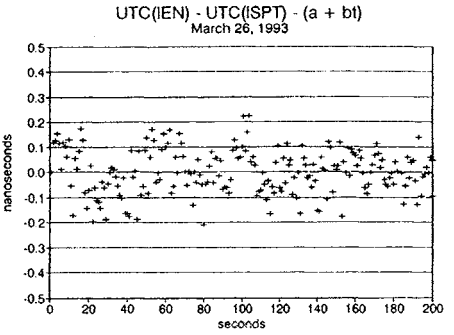


Fig. 2 - Residuals of the one-second differences.

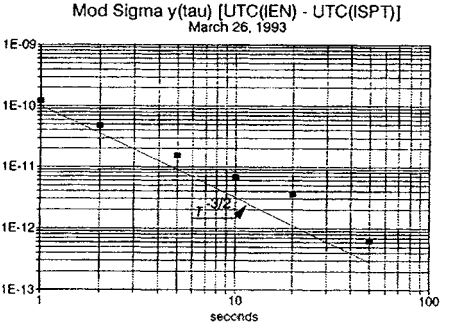


Fig. 3 - Mod. Allan deviation of the residuals.

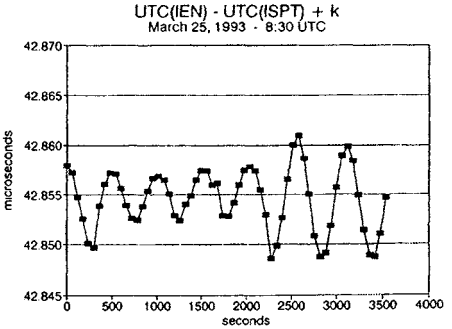


Fig. 4 - Synchronization results.

measurement per minutes, performed on March 25 and 26 and on April 29 and 30; from these graphs it can be noticed that nearly-sinusoidal instabilities of the order of 15 ns (peak-to-peak) are always present. Their period varies from about 500 s to 1000 s.

To try to explain these anomalies, some tests were made to verity if it was due to the strict coherence between the frequencies of the clocks compared, supplied to the MITREX modems /3/. The tests consisted in supplying the 10 MHz reference to the modem with different frequency offsets at both stations, but no improvement was obtained in the synchronization results. It was not also found a correlation with the periodicity of the antenna automatic tracking system (step track).

During the synchronization sessions of 29 April 1993, a cesium clock transportation to the ISPT laboratory in Roma was performed by IEN. In the same occasion, a Stanford 502B GPS receiver was available at ISPT and common view measurements with IEN were performed, once per day, on 28 and 29 April. The results obtained with these methods, have been compared with the two-way data for the same date and reported in table 2. Some

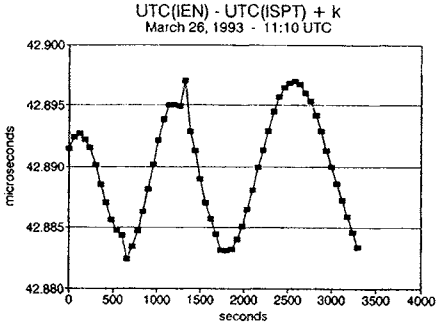


Fig. 5 - Synchronization results.

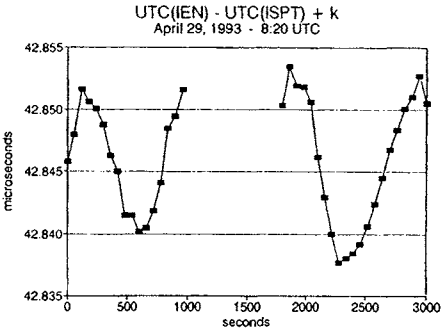


Fig. 6 - Synchronization results.

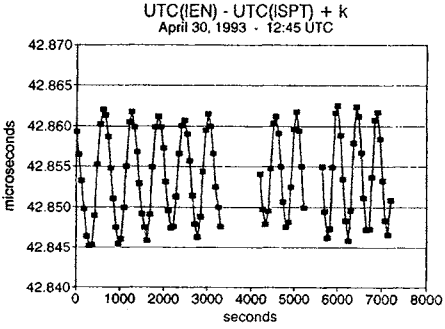


Fig. 7 - Synchronization results.

Table 2 - Synchronization results comparison

Date	UTC(IEN) - UTC(ISPT)		
	Clock transport	GPS	two-way
1993-04-28	44,000 μs (10:19)	44,065 μs (10:19)	--
1993-04-29	44,000 μs (09:30)	44,063 μs (12:23)	44,034 μs (08:20)

comments are necessary to the data reported. The estimated uncertainty of the clock trasportation is 20 ns (1σ) meanwhile it is not possible to evaluate precisely the uncertainty of the GPS results because the ISPT coordinates are not know with a precision comparable with that of the IEN, determined by BIPM, and the delay of the ISPT-GPS receiver has never been calibrated. Concerning the two-way value, it has been obtained by averaging 40 measurements, 1 minute apart, and applying the corrections for the equipment and stations differential delays (84 ns) and for the Sagnac effect (16 ns). It must be pointed out that the differential delay correction applied was determinated during the July 1993 measurements with the stations co-located in Torino and that the stations and cable configuration could have not been the same used in the April sessions.

5 - Station delay measurements.

To investigate on the anomalies encountered in the synchronization sessions and to determine the differential delay, the two stations were co-located at IEN from June 30 to July 2. The same reference of frequency and time was supplied to the modems and a unique time interval counter (Stanford SR620) was used to measure directly the difference 1PPS/R_X(ITA1) - 1PPS/R_X(ITA2). The reference frequency for the synthesizer of the stations up and down converters, was chosen either externally (cesium frequency) or internally. A sample of the results obtained is shown in figures from 8

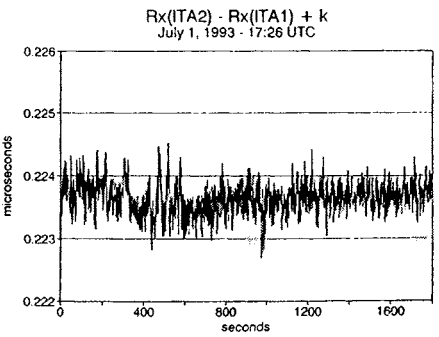


Fig. 8 - Differential delay (ext. freq. to synth.).

to 10 where small periodical variations and sudden jumps in the differential delay can be seen. These anomalies are present either with the external or the internal reference supplied to the stations synthesizers. The standard deviation of a measurement block of about 1800 one-second differences, was 200 ps (best case), but reached also 3 ns in the case of Fig. 9 and 10. The repeatability of the differential delays measurements was not good and differences up to ±10 ns have been found; to correct the synchronization data, the value of 84 ns, obtained by averaging the data of Fig. 8, was chosen because very near to the mean value of the differential delays measured in all sessions. A correction for the differential delays of the 1PPS/T_X signals has also been applied. Both during the synchronization sessions in Torino and Roma and with the stations co-located, some delay measurements with the earth stations TLT (Test Loop Translator) were performed. In Fig. 11 are reported the results of about 5 hours measurements that show both small jumps (~1 ns) and larger ones (~10 ns); this fact occurred in several sessions. Some measurement sessions were performed analyzing the behaviour of a single station: to do this, the PRN signals of the two modems were sent to one station T_X input via a mixer, transmitted to the satellite, and separated at the station R_X output with a similar device. The measurement set-up was the same used with the stations co-located. For both earth stations, no jump has been detected nor slow variations of the measured delay was found for observation times of 30 minutes.

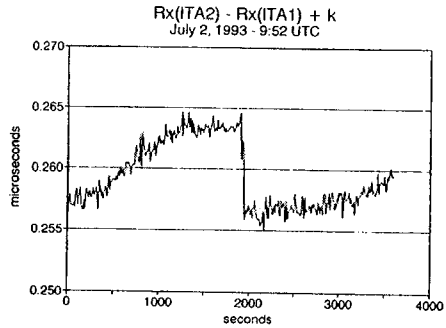


Fig. 9 - Differential delay (int. freq. to synth.).

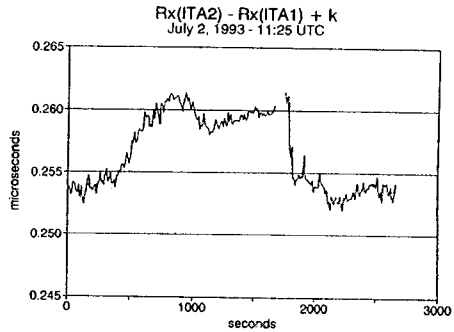


Fig 10 - Differential delay (ext. freq to synth.).

6 - Mitrex modems characterization

The MITREX modems sensitivity to the temperature and to other operative parameters has been reported in the literature /4,5,6/ therefore, after the conclusion of the two-way experiment, it was decided to investigate the instability of the modems delays versus the input signal level, C/No ratio, the offset of the received frequency and the temperature. The time and frequency reference to the modems and to the time interval counter (Stanford Research SR 620) was supplied by an HP 5071/001 cesium. All the measurements were performed with the modems 1 PPS/T_X synchronized once for all by the external time reference. The measurements of the signal levels and of the C/No ratios, were taken in the "clean carrier" condition, using a spectrum analyzer.

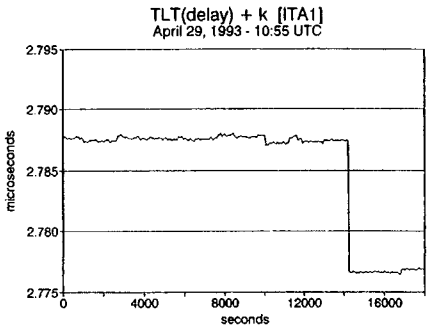


Fig. 11 - Test Loop Translator delay.

The equipment used to verify the influence of the locking range in the modems delays and to characterize the delay dependance from the input signal level and from the C/No, is reported in Fig. 12. To generate the 70 MHz IF frequency offsets, a double conversion technique with two synthesizers was used. The signal from a white noise

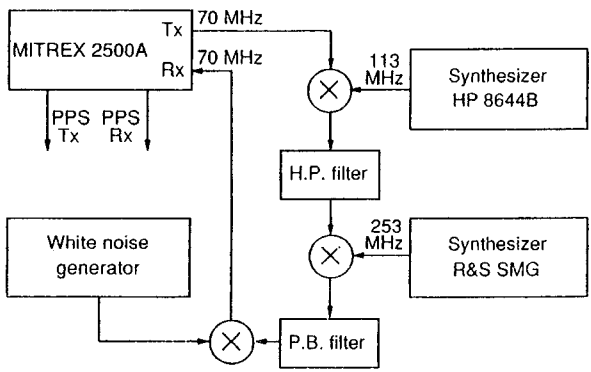


Fig. 12 - Equipment setup for the Mitrex modem characterization.

generator has been mixed with the IF frequency for the C/No measurements.

First, the locking range of each modem, without noise, has been found with the associated delay variation. The results are reported in table 3.

Table 3 - Locking range of the modems.

Mitrex 2500A	T_X signal	R_X attn. [dB]	70 MHz ± Δf [KHz]
ITA1	-20 dBm	0	+13, -13
(Torino)	-20 dBm	20	+13, -13
ITA2	-20 dBm	0	+14, -13
(Roma)	-20 dBm	20	+9, -9

In these conditions, the maximum variations of 1PPS/T_X - 1PPS/R_X, with an external loop, have been measured and found of 1,2 ns and 2,4 ns for ITA1 modem (R_X attn. equal to 0 and 20 respectively), and of 1,4 ns and 3,7 ns for ITA2 modem (same R_X configuration). More details on this dependence can be seen in Fig. 13 for a Δf range of ± 2 KHz and for different signal levels and C/No ratios. It has been verified, for an RF signal level of −32 dBm, that a minimum C/No of 54 dB was requested for the modems to lock.

The effects of the input signal level on the modems delay are shown in Fig. 14; in the range from −20 to −50 dBm, the slope of the delay variation is + 80 ps/dB (ITA1) and +8 ps/dB (ITA2).

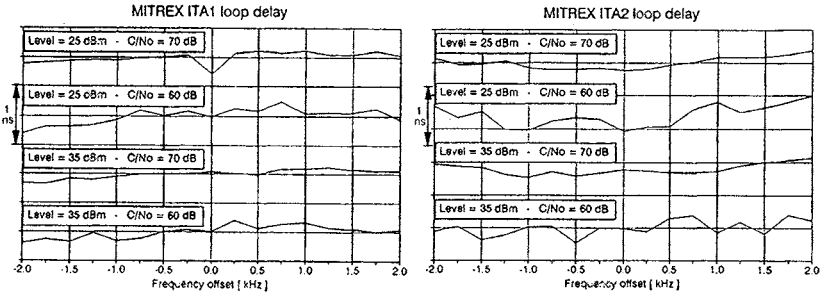


Fig. 13 - Loop delay variations for IF frequency offsets.

To characterize the modems versus the temperature changes, a thermostatic cell, operating in the range from 15 to 35 °C, was used and the differences $1PPS/T_X$ - $1PPS/R_X$ were measured at about 2 degrees steps, waiting for the modems to reach a thermal stabilization. The temperature coefficients (K_t) obtained, in the range from 20 to 30 °C and for an input signal level of –20 dBm, are:

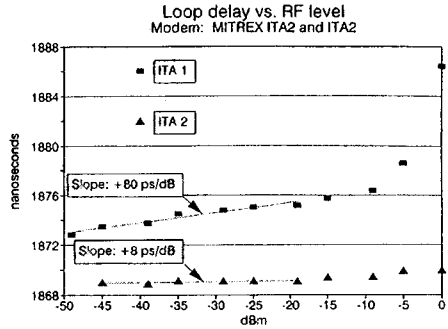


Fig. 14 - Loop delay variations vs. input level.

K_t(ITA1) = –66 ps/°C (R_X attn. = 0) and –78 ps/°C (R_X attn. = 20)
K_t(ITA2) = –229 ps/°C (R_X attn. = 0) and –235 ps/°C (R_X attn.= 20)

In the range from 15 to 30 °C, the variation of the difference 1PPS(5071A)–$1PPS/T_X$ have also been measured and the following temperature coefficients have been found:

K_t (ITA1) = +563 ps/°C and K_t (ITA2) = +446 ps/°C

The stability of the electronic counter used in all these measurements has been determinated in the same temperature range (15 to 30°C), measuring the delay of a coxial cable external to the cell, and its temperature coefficient found equal to –5 ps/°C.

Finally, the long term stability of the modems and of the electronic counter delays, in laboratory conditions, was investigated for a seven days period measuring $1PPS/T_X$ - $1PPS/R_X$ signal. The results are reported in Fig. 15, together with the laboratory temperature; in the case of ITA1 modem, a correlation between delay and temperature changes can be observed.

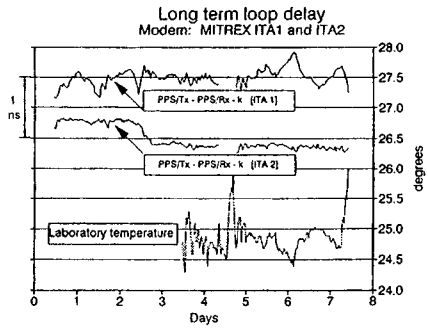


Fig. 15 - Long term loop delay variations.

All the delay values reported in the graphs are means over 100 readings, 1 second apart, with a tipical standard deviation (1σ) of 40 to 80 ps, except for the case of the temperature coefficients of the modems where the mean value was computed over 30 time interval measurements.

7 - Conclusions

A two way synchronization experiment in Italy between IEN and ISPT via Olympus satellite, using spread-spectrum signals, was performed in 1993.

For measurement sessions lasting 3 minutes, with a repetition interval of one second, the data show a white noise PM behaviour with an average level of about $1{,}2 \cdot 10^{-10} \cdot \tau^{-3/2}$.

The accuracy of the time scales comparisons, heavily limited by the medium and long term instabilities of the mobile earth stations used and possibly by the Olympus satellite anomalies, is of the order of 20 ns (1σ). This value is very far from the capabilities, well demonstrated in other experiments, of this method, and on the basis of the investigations performed afterwards on the equipement used, it can be concluded that fixed earth stations are madantory for attaining the nanosecond accuracy in these comparisons.

To reach this goal nevertheless, much care must be devoted to the enviroment where the modems and the measuring system are installed because of the influence of the temperature on these devices.

The same care must be taken towards the input signal IF frequency and level and its C/No that also affects the modems delays.

Acknowledgements

The authors wish to thank the Telespazio personnel in charge of the mobile earth stations for the cooperation given.

This work was supported by the Italian PTT Administration in the frame of the Olympus experiments.

References

/1/ D. Kirchner: Two-way time transfer via communication satellites. Proc. of the IEEE, Vol. 79. n. 7, July 1991, pp. 983-990.

/2/ D. W. Hanson: Fundamentals of two-way time transfers by satellite. Proc. of 43rd Annual Symposium on Frequency Control, June 1989, pp. 174-178.

/3/ D. A. Howe: Time tracking error in direct-sequence spread-spectrum networks due to coherence among signals. IEEE Transactions on Communications, vol. 38, n. 12, December 1990.

/4/ Politecnico di Torino - Thermal characterization of Mitrex 2500 modem - Technical Report (in Italian), May 1985.

/5/ H. Ressler, D. Kirchner: Experience with the Mitrex modem. INW Report 8501, Thechnishe Universitat-Graz, 1985.

/6/ P. Banerjee, H. K. Thiel: A study on the performance of the Mitrex modem for two way time comparison. Research Report NPL-93-A.1-0004, NPL India, 1993.

SYNCHRONIZATION OF RIRT'S SECONDARY TIME/FREQUENCY REFERENCE STANDARD VIA GLONASS SIGNALS

Yuri G.Gouzhva, Arvid G.Gevorkyan, Pyotr P.Bogdanov, Anatoly S.Perlin

Russian Institute of Radionavigation and Time,
2 Rastrelli square, St.Petersburg, 193124, Russia

ABSTRACT

The Secondary Time/Frequency Reference Standard (STFRS) of the Russian Institute of Radionavigation and Time (RIRT) is a part of the State Time/Frequency Service and belongs to the country's secondary time and frequency standards. The STFRS consists of several hydrogen and cesium beam frequency standards providing its relative frequency accuracy with respect to the State Time and Frequency Standard (STFS) no worse than $3 \cdot 10^{-13}$.

Until recently, the transportable clock and the television time signals comparison equipment were the main means of the high-precision referencing of the STFRS to the STFS.

Using transportable clock, however was not regular, and the use of TV signals provides accuracy at the level of 80 ns.

To increase the accuracy of STFRS comparison, from February, 1992 GLONASS signals are regularly used for comparing STFRS against the STFS in direct-synchronization mode. In this paper we present the results of the STFRS comparisons during 1992-1993 and determine the ways to improving an accuracy of STFRS's synchronization against STFS.

1 INTRODUCTION

At present using GPS and GLONASS satellite radionavigation systems signals is one of the most precise methods of clock synchronization . Diverse aspects of clock synchronization via GPS signals are widely observed in various references /1,2/. During last 2-3 years several experiments of clock synchronization via GLONASS signals took place /3,4/. But for a long time use of GLONASS signals was limited because of delays in deployment of the system and lack of special time recivers.

That's why RIRT held a number of experiments to estimate possibilities of serial airborne GLONASS ASN-16 navigation receiver as a time receiver /5/. The achieved results have shown that the use of ASN-16 receiver provides clock synchronization accuracy within 20 ns for 1 day in direct synchronization mode and within 10 ns for 1 day in common-view mode.

That enabled to start in February , 1992 with regular comparisons of RIRT's secondary time/frequency reference standard against the State Time and Frequency Standard via GLONASS signals with ASN-16 receiver. The results of this comparisons are presented in this paper. Results achieved proved an efficiency of using GLONASS signals for clock synchronization and provided an increase in accuracy and reliability of STFRS vs.STFS comparison.

2 MAIN CHARACTERISTICS OF RIRT's SECONDARY TIME/FREQUENCY REFERENCE STANDARD

The RIRT Secondary (Operational) Time/Frequency Reference Standard (STFRS) is meant to solve the tasks as follows:

- participation in the State Time/Frequency Service as applied to the check of time transfer accuracy when using LF and VLF radio stations and TV-channels;
- regular checking and certification of various

time/frequency standards designed in RIRT;

- research and working out design principles for various STFRS used in diverse platforms.

The STFRS comprises:

- equipment for keeping a dimention of frequency and time unit. It is built up of two hydrogen frequency standards provided with the automatic cavity tuning system;

- time-keeping equipment consisting of four cesium beam frequency standards, as well as time generating/keeping equipment;

- internal comparison equipment made on the basis of automated measuring/computing complex;

- external comparison equipment using television time signals, as well as time signals of LF and VLF radio stations;

- support equipment (high-reliability power supply; air conditioning system; environmental parameter measuring system; monitoring, control and signalling panel and other devices).

The STRFS belongs to the country's secondary time and frequency standard.The total STFRS relative frequency accuracy with respect to the STFS does not exceed $3 \cdot 10^{-13}$.The relative STFRS frequency shift does not exceed $5 \cdot 10^{-14}$ per month. The frequency instability is $5 \cdot 10^{-14}$ when measured during 1 hour and $3 \cdot 10^{-14}$ when measured over 1 day,the maximum measurement error being $2 \cdot 10^{-14}$ and $5 \cdot 10^{-15}$ respectively.

Until recently the most precise methods of external comparison of STFRS relative to STFS were transportable clock and television time signals. But regular use of transportation clock was limited because this method is very expensive. At the same time, the error of STFRS comparison relative STFS via TV signals is about 80 ns for 1 day due to radio relay transmission line between Moscow and St.Petersburg.

The provisional continious Atomic Time - TA (STFRS),

is taken as STFRS's time scale. It is computed on the basis of physical time of the most precise hydrogen frequency standard taking into account the results of internal comparisons of all hydrogen and cesium frequency standards. The coordinated STFRS's time - UTC (STFRS), corresponds to TA (STFRS) taking into consideration the corrections which provide its synchronization accuracy of 2 μs relative UTC (SU), generated by STFS. External comparisons are held relative to operational STFRS's time - TOP (STFRS), generated from a signal of one hydrogen or cesium frequency standards. The transformations of comparisons' results against operational STFRS's time to atomic time is realized with the relationship:

$$\Delta T_{TA(STFRS)} = \Delta T_{TOP(STFRS)} - \Delta T_{ST'} \quad (1)$$

where $\Delta T_{TA(STFRS)}$ - offset between atomic STFRS time and UTC (SU);

$\Delta T_{TOP(STFRS)}$ - offset between operational time and UTC (SU) according to external comparisons' results;

ΔT_{ST} - correction for mutual offset of operational and atomic times.

3 CLOCK SYNCHRONIZATION VIA GLONASS SIGNALS

As it well known there are four main methods of clock synchronization via satellite navigation system signals, i.e.: direct synchronization, common-view, clock transportation and very-long base interferometry (VLBI) Techniques Over Short Baselines /1/. But the methods of the utility are the first two ones, which principles are shown in Fig.1.

The direct synchronization method is the most simple. It provides the global coverage and requires no other data, but those received from satellite navigation message. The time offset between the user clock and UTC(SU) being GLONASS reference time is given by the

relationship:

$$\Delta T = S - \left(\frac{D}{C} - \tau_{ion} - \tau_{TROP} - \tau_{rec} \right) + \Delta T_{sat} + \Delta T_{SCT}, \quad (2)$$

where S - measured pseudorange from the satellite to the user, i.e. the difference between two identical codes, one received and the other generated by the receiver; each of these codes is synchronized by its own clock;

D - range from the satellite to the user, computed on the basis of broadcasted ephemerides X_i, Y_i, Z_i and the known coordinates of the receiver antenna X_A, Y_A, Z_A;

C - speed of light;

τ_{ion} - ionospheric propagation delay;

τ_{TROP} - tropospheric propagation delay;

τ_{rec} - receiver delay;

ΔT_{SAT} - difference between satellite clock and System Common Time (SCT) computed on the basis of broadcasted time and frequency corrections (τ_i, γ_i);

ΔT_{SCT} - difference between SCT and UTC (SU) contained in the navigation message as a set corrections τ_s.

The common-view method is a technique of a mutual clock synchronization. It presupposed simultaneous measurements at the points of clock location via one of the satellite's signals and further exchange of the results of the measurements. In this case, a mutual difference of clock times is determined from the relationship

$$\Delta T_{A-B} = \Delta T_A - \Delta T_B, \quad (3)$$

where $\Delta T_A, \Delta T_B$ - the results of determination of each clock's offset relative UTC (SU) in conformity with (2).

It is obvious that, due to elimination of several components, stipulated by GLONASS characteristics which are common for A and B clock, its mutual synchronization accuracy increases.

The ASN-16 unit, designed in RIRT, is one-channel one-frequency navigation user equipment. Therefore the ionospheric delay is computed on the basis of parameters which are autonomously stored in the receiver. The tropospheric delay is determined in a similar way. The receiver delay is calculated by means of its regular calibration.

While solving the time task, ASN-16 provides the operation via one chosen satellite with output of 1 Hz and the result of determination of the time difference between this signal and UTC (SU). That's why for clock synchronization via GLONASS signals using ASN-16, an additional counter of 1 Hz signal difference in ASN-16 output and user's clock is required. The measurement configuration using ASN-16 is shown in Fig.2.

4 THE RESULTS OF STFRS COMPARISONS

Regular use of GLONASS signals for STFRS comparisons against STFS in direct synchronization mode with ASN-16 receiver began from February, 1992.

The procedures of comparisons were following. Three 15-minute sessions of measurements were held over the day. For this purpose the satellite closest to zenith is chosen. The sessions measurements are averaged for a day (the drift of the difference between TOP (STFRS) and UTC (SU) was eliminated), and the value of the offset between TOP (STFRS) and UTC (SU) is determined - $\Delta T^{GL}_{TOP(STFRS)i}$. Later it is re-calculated according to (1) into offset between TA (STFRS) and UTC (SU) - $\Delta T^{GL}_{TA(STFRS)i}$.

From October, 1992, the GPS signals with GEM SRS-01 navigation receiver are also used for STFRS' comparison. SRS-01 provides, simultaneously with the navigation

determination, time task solving . It has an output 1 Hz signal synchronized against UTC (USNO). The measurement configuration using SRS-01 is also shown in Fig.2. The procedure of comparisons are similar to those for GLONASS measurements.

The results of STFRS comparisons via GLONASS signals, GPS signals and TV signals for October-December , 1992 are presented in the Tabl.1 as ten-day estimations. There are also the offsets between UTC(SU) and UTC(USNO) calculated on the basis of STFRS comparisons via GLONASS and GPS signals. The similar data were published by Bureau International des Poids et Measures (BIPM). The results for 1993 are shown in Fig.3.

Table 1

The estimations of STFRS comparisons via GLONASS signals, TV signals and GPS signals

Measurements interval	$\Delta T_{TA(STFRS)}$, μs			UTC(SU)-UTC(USNO)	
	GLONASS	TV	GPS	Calcul.	BIPM
01-10.10.92	-4.68	-5.05	-6.45	-1.77	-1.71
11-20.10.92	-4.68	-5.11	-6.38	-1.70	-1.57
21-31.10.92	-4.72	-5.14	-6.34	-1.62	-1.46
01-10.11.92	-4.84	-5.33	-6.34	-1.50	-1.34
11-20.11.92	-4.89	-5.36	-6.28	-1.39	-1.24
21-30.11.92	-5.02	-5.51	-6.29	-1.27	-1.12
01-10.12.92	-5.13	-5.59	-6.26	-1.13	-1.05
11-20.12.92	-5.16	-5.66	-6.20	-1.04	-0.95
21-31.12.92	-5.25	-5.71	-6.10	-0.85	-0.85

This data shows that the results of STFRS comparisons via GLONASS and TV signals have a systematic difference varying from 0.68 μs in the begining of 1993 to 0.85 μs after April 20, 1993. It is obvious that this difference

is caused by irregularity of calibration of TV signals transmission line. The UTC(SU) - UTC(USNO) values calculated on the basis of STFRS comparisons via GLONASS and GPS signals and the values published by BIPM have a good coincidence (discrepancy from zero to 0,16 μs).

Processing STFRS comparisons results via GLONASS signals gives following estimations for its accuracy characteristics in 1993: relative STFRS frequency accuracy against STFS does not exceed $7.2 \cdot 10^{-14}$; relative STFRS frequency change does not exeed $1.1 \cdot 10^{-14}$ for 1 month.

5 PROSPECTS FOR INCREASING STFRS SYNCHRONIZATION ACCURACY

Further increasing STFRS synchronization accuracy is connected with realization of common-view mode via GLONASS signals. For this purpose RIRT's specialists reworked ASN-16 unit in order to use it as a GLONASS time receiver interfacing with PC. This provides fully automatic process of measuring and data processing. When this equipment is installed in VNIIFTRI and other State Time/Frequency Service platforms, it will enable to realize comparisons of country's standards according to a program similar to the one developed by BIPM for comparisons via GPS signals.

Preliminary experiments held in RIRT with two ASN-16 units has shown that the instrumental error of clock synchronization in a common-view mode is about 10 ns over a day /5/.

Increase of clock synchronization accuracy in a common-view mode with ASN-16 can be expected from using direct "raw" measurements and its PC processing with improved algorithms for ionospheric and tropospheric delay, along with more precise data for satellite ephemerids and receiver antenna coordinates.

CONCLUSION

The RIRT's Secondary (operational) time/frequency reference standard is a part of the State Time/Frequency Service and belongs to the country's secondary time and frequency standards. The relative STFRS frequency accuracy with respect to the STFS during 1993 was no worse than $7.2 \cdot 10^{-14}$.

For STFRS comparisons relative to STFS, transportable clock, TV signals and, since February, 1992, GLONASS signals in a direct-synchronization mode are used also. The accuracy of STFRS comparisons via GLONASS signals with ASN-16 receiver is 20 ns.

Further increasing STFRS synchronization accuracy via GLONASS signals is connected with realization of a common-view mode expected accuracy of standards' comparisons is less than 10 ns.

ACKNOWLEDGEMENTS

The authors are pleased to express their gratitute to their collegues in RIRT for participation in experiments and data analysis.

REFERENCES

/1/ D.V.Allan and M.Weiss. Accurate time and frequency trasfer during common-view of a GPS satellite. Proc. 34th Ann.Symp. on Frequency Control. May 1980, pp. 334-346.

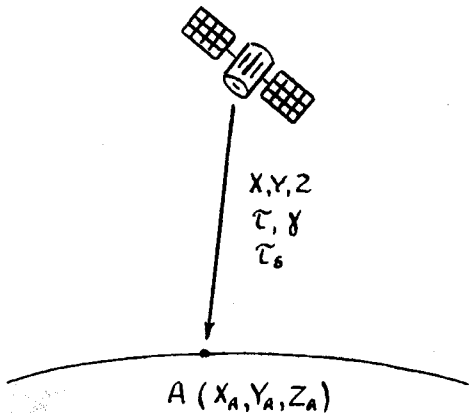
/2/ W.Lewandowski, C.Thomas. GPS time transfer. Proc. of the IEEE, Special Issue on Time and Frequency. July, 1991, pp. 991-1000.

/3/ P.Daly, G.T.Cherenkov, N.B.Koshelyaevsky, S.B.Pushkin. Satellite time transfer between UTC (USNO) and UTC (SU) using NAVSTAR GPS and GLONASS. Proc. 4th

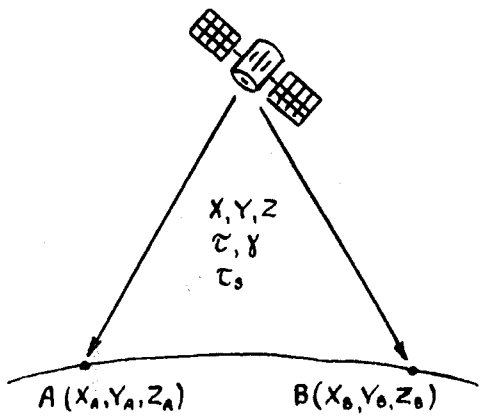
Institute of Navigation Technical Meeting.September, 1991, pp. 199-206.

/4/ P.Daly, N.B.Koshelyaevsky, W.Lewandowski, G.Petit, C.Thomas. Comparison of GLONASS and GPS time transfer between two West European time laboratories and VNIIFTRI. Proc. 23rd PTTI meeting,1991, pp.341-350, 1991.

/5/ Y.G.Gouzhva, A.G.Gevorkyan, P.P.Bogdanov, V.V.Ovchinnikov. Using GLONASS signals for clock synchronization. Proc. 25th PTTI meeting (accepted).



a) The direct-synchronization technique



b) Common-view method

Figure 1. The direct-synchronization technique and common-view method via GLONASS signals.

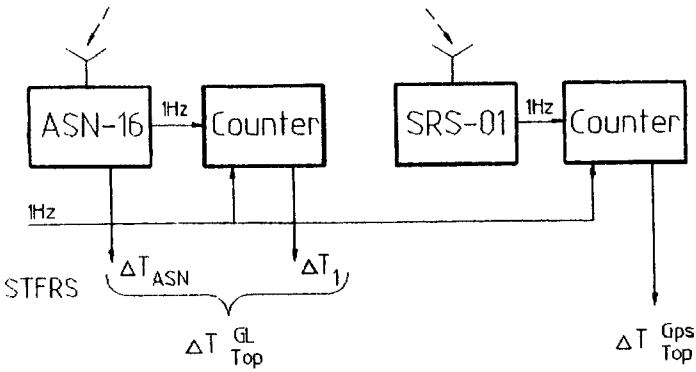


Figure 2. Measurement configuration.

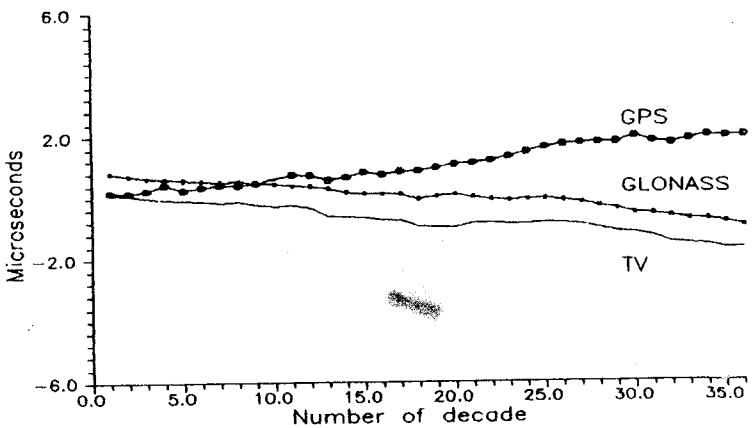


Figure 3. TA(STFRS)-UTC(SU) by GLONASS signals and TV signals and TA(STFRS)-UTC(USNO) by GPS signals.

LASER RANGING SUPPORT FOR TV TIME TRANSFER USING GEOSTATIONARY SATELLITES

J.McK. Luck and J.R. Woodger

Orroral Geodetic Observatory
Australian Surveying and Land Information Group
Canberra, ACT Australia

Abstract

TV signals broadcast from the Australian communications satellite AUSSAT K1 in geostationary orbit have been used for clock comparisons within the country since 1986. Orbit corrections derived from radar tracking by the master control station in Sydney yield time transfer precision of 100ns. An orbit improvement scheme using GPS time transfer between several major timing centres converts the TV pseudo-range differences to range differences, improves principally the orbit's inclination and eccentricity values and station biasses, and has demonstrated precision potential better than 10ns.

The first satellite OPTUS B1 of the current generation was launched in August 1992 and is equipped with an array of 14 retro-reflectors, each 38mm diameter, for laser ranging. In an experiment in July 1993 at the Orroral Laser Ranging System which normally supports geodynamics research, an avalanche diode detector kindly provided by Wettzell Observatory, Germany achieved an astonishing return rate of over 100 echos per minute with single shot precision 25mm on several nights. This technique will yield range accuracies sub-nanosecond, and is being built in to the TV method supporting the National Time System embodied in UTC(AUS).

1. INTRODUCTION

The terrestrial 'line-10' method of time transfer in Australia [1,2] was replaced in 1986 by TV transmissions from AUSSAT K1, the Australian domestic communications satellite [3]. The principle is illustrated in Figure 1 and the measurement setup in Figure 2 [4]. The system has been in regular use since 1987. Figure 3 shows the performance of a hydrogen maser in Hobart, Tasmania compared with UTC(AUS) during 1993 using satellite TV alone.

The AUSSAT series was replaced by the OPTUS series, the first of which (OPTUS B1) was launched on 24 August 1992 and came into operation on 16 December 1992, at 160^{o} East longitude. The time transfer results improved markedly due to use of two radar ranging stations and better orbits.

OPTUS B1 takes advantage of the laser ranging facility at Orroral Geodetic Observatory just outside Canberra, by being equipped with laser retroreflectors for assistance in determining the orbit by adding high quality range measurements. A pilot project was conducted in July 1993, and it is expected that routine laser ranging will commence in April 1994.

2. ORBIT DETERMINATION

2.1 TV TIME TRANSFER OBSERVATION EQUATION

If t_o is the time of transmission of a TV sync pulse from the satellite at any time, x_i the measurment of the time of reception against the clock at station i, T_i the clock error, d_i the receiver delay or station bias, e_i the unmodelled random and systematic errors, and p_i the propagation delay from satellite to station i corrected for refraction and ionospheric effects, then:

$$x_i = t_0 + T_i + d_i + p_i + e_i$$

whence, subtracting the results from two stations i and j:

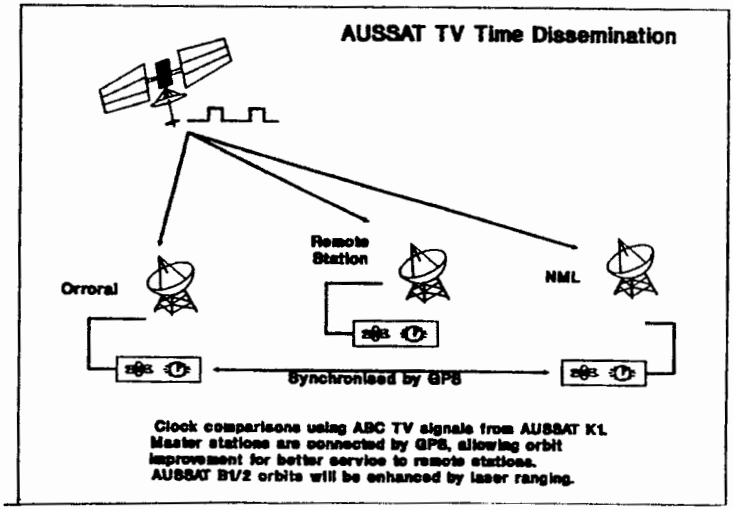


Figure 1: From Ref [4]. The current satellite is OPTUS B1.

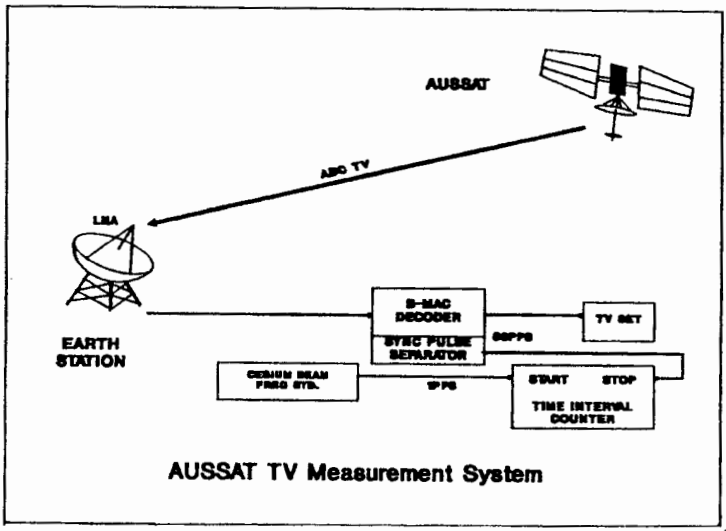


Figure 2: From Ref [4]. The current satellite is OPTUS B1.

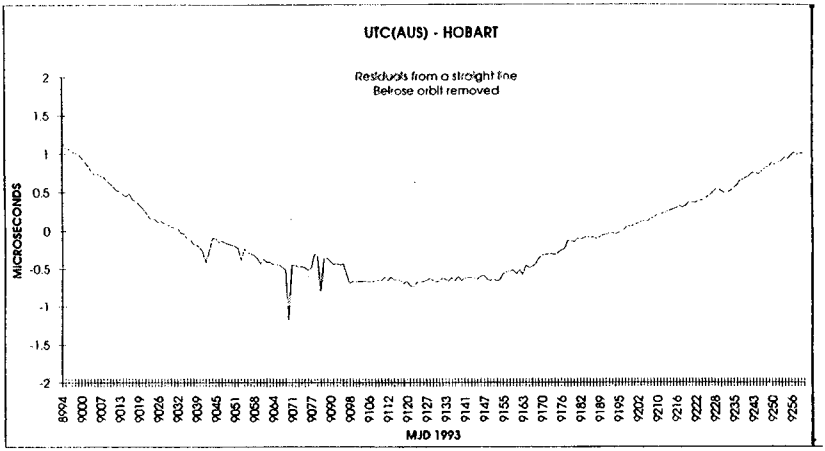


Figure 3: Residuals after removal of a trend of the clock differences (T_i-T_j) between the Hobart hydrogen maser and Sydney, using an orbit computed from raw elements.

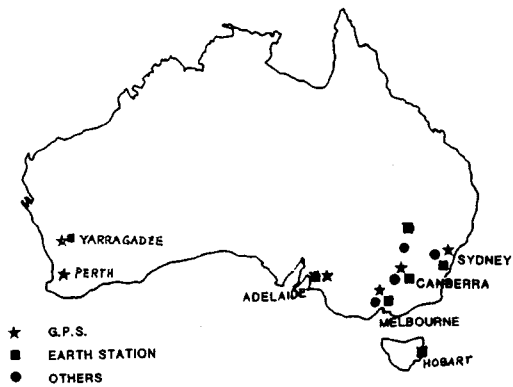


Figure 4: Distribution of stations using GPS, OPTUS TV ('Earth Station') and terrestrial TV ('Others'), March 1994.

$$x_i - x_j = (T_i - T_j) + (d_i - d_j) + (p_i - p_j) + (e_i - e_j). \quad (1)$$

When the unmodelled errors are assumed zero, the clock difference is:

$$T_i - T_j = (x_i - x_j) - (d_i - d_j) - (p_i - p_j). \quad (2)$$

The relative station biasses d_i, d_j can be calibrated by carrying a standard receiver and antenna cable to each station. The propagation delays p_i, p_j need to be computed from precise orbits.

2.2 RADAR RANGING

The OPTUS Major Earth Stations at Belrose, Sydney NSW and Lockridge, Perth WA perform regular orbit determination for on-orbit station keeping purposes, and kindly provide the resulting Kepler elements. The nominal accuracies are 30m in range and 0.05° in azimuth and elevation. The elements are fed into a numerical orbit integration program written in-house, which produces the predicted ranges at each TV observation epoch. Details of the integrator are given in Table 1. Time transfer precision from these 'raw' orbits is currently better than 100ns as estimated by straight-line fits through 30 daily readings, which is sufficient for the majority of users.

Table 1: Orbit Integrator

Feature	Description
Type	: Cowell-Crommelin 7th order
Interval	: Selectable. Normally 60 seconds
Reference Frame	: True of Date
Gravity Field	: GEM10N to (8,8)
Lunar Gravity	: Brown analytical theory, truncated
Solar Gravity	: Newcombe theory
Solar Radiation	: Nominal cross-section and solar flux. Includes effects of eclipse
Coordinate Transformations	: Obliquity, precession, nutation, GAST from IERS Standards
Earth Rotation	: IERS Bulletin A supplied through USNO

2.3 ADDING GPS TIME TRANSFER

Suppose that the clock differences $T_i - T_j$ at several stations are already known by GPS time transfer. Then equation (2) can be rearranged to permit solution for the range differences:

$$p_i - p_j = (x_i - x_j) - (T_i - T_j) - (d_i - d_j). \quad (3)$$

These range differences, provided that there is suitable geographical geometry among the GPS-equipped 'Master' stations, can be used to improve the orbit. The distribution of stations in Australia is shown in Figure 4. Using hourly TV readings and linearly interpolating between daily measurements on GPS SV 12, we have found that good solutions are obtainable for eccentricity and inclination, and that station biasses relative to a 'Grand Master' station are also quite well-determined. Thus, $(d_i - d_j)$ can be treated as solution parameters

rather than as calibration constants. Some results of this orbit improvement scheme in 1986 were reported in [3].

Having improved the orbit by a combination of TV and GPS measurements at the Master stations, the time transfer to 'Remote' stations is considerably improved. The Remote station only needs to purchase a standard inexpensive domestic Earth Station.

2.4 LASER RANGE MEASUREMENTS

The use of range differences from a limited geographical area gives an intrinsically weak solution for a geostationary orbit in terms of Kepler elements - the angular elements mean anomaly, argument of perigee and ascending node cannot be determined (and could not be even if eccentricity and inclination were appreciably non-zero). It is therefore desirable to add extra types of measurement in order to separate all the parameters. Laser ranging can provide these, giving range measurements with 1-2cm precision and accuracy, and (although we have not incorporated them yet) azimuth and elevation angles precise to 0.001^{o}.

If p_0 is the laser ranging observation from some station '0', and the GPS-corrected TV differences are (pi-pj) as in equation (3), the combined observation equation set becomes:

$$\begin{vmatrix} p_i - p_j \\ \\ p_0 \end{vmatrix} = \mathbf{F}[a,e,i,M,\omega,\Omega;\ SRP;\ d_2,..,d_n;\ t] \qquad (4)$$

where SRP is a solar radiation pressure parameter (not implemented), the biasses d_i are relative to the Grand Master station (i=1), and there are in general different observation epochs. Equations (4) are solved by the Method of Least Squares.

3. LASER RANGING TO OPTUS B1

3.1 SPACE SEGMENT

The Laser Retroreflector Array (LRA) was commissioned by OPTUS P/L. The design and construction project was managed by Hughes Aircraft, with British Aerospace Australia (BAeA) as contractors. The individual retro-reflectors were designed by Beattie Steel of Macquarie University, Sydney, ground and polished by James Optics, Melbourne, coated with indium tin oxide by AG Thompson, Adelaide and tested and mounted by BAeA.

The LRA was designed to yield a laser return rate of 1 photoelectron per shot at Orroral Observatory when the ground segment operates at its maximum efficiency. This is midway between the rates expected from lunar ranging (0.01PE/shot, range 380,000km) and LAGEOS (100PE/shot, range 6000km).

There are 14 individual corner cubes, each 38mm diameter aperture, mounted in a rectangular tray as shown in Figures 5 and 6 (from [5]). The rear faces are mutually offset from perpendicular by 0.8 arcseconds to account for the velocity aberration of 3.6 arcsecond induced by the 2.7km/sec relative speed between satellite and station. Special interferometer tests were devised to measure the dihedral angle offsets between the rear faces, and the reflected intensity distribution was measured by a novel radiometric test [6]. Each corner cube is mounted in the tray in such a way that thermal/mechanical stresses are relieved yet electrostatic buildup (which would interfere with the spacecraft's radio systems) is eliminated.

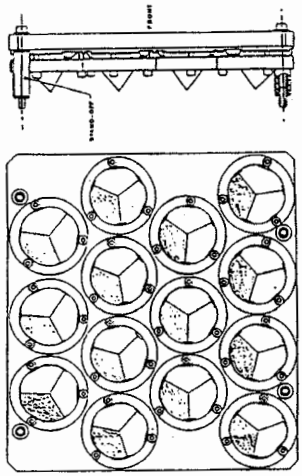


Figure 5: From Ref [5]. Arrangement and mounting of corner cubes in the Laser Retroreflector Array.

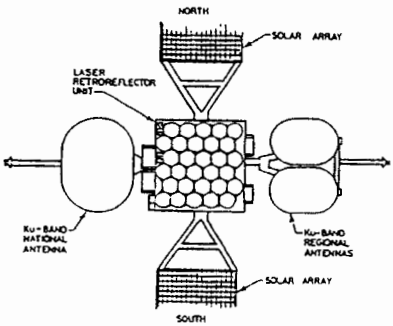


Figure 6: From Ref [5]. Position of the Laser Retroreflector Array on the spacecraft.

3.2 GROUND SEGMENT

The Orroral Laser Ranging System (OLRS) is about 60km south of Canberra in the Australian Capital Territory. Its principal equipment is a 1.5m Ritchey-Chretien telescope for transmitting and receiving laser pulses, a Quantel Nd:YAG laser producing frequency-doubled pulses at 532nm wavelength at 7-14 shots/sec, each transmitted pulse having width 110ps and energy 100-200mJ. The detector is normally an ITT F4129f Micro Channel Plate (MCP) Photomultiplier Tube, but for the experiment described was a SP114 silicon Avalanche PhotoDiode (APD) brought out especially by Dr Ulrich Schreiber from Satellitenbeobachtungsstation Wettzell, Germany.

The range timing system is derived from an Oscilloquartz 3200 crystal, with the high resolution for laser ranging provided by LeCroy 2229A TDC verniers. Time tagging is assisted by a HP5071 Cesium Standard, which since 1 January 1994 acts unambiguously as UTC(AUS), and three HP5061A Cesium Standards intercompared hourly with UTC(AUS). UTC is accessed via two single-frequency GPS Time Transfer Units (Trimble 4000 and FTS 8400), and the building also contains the TV receivers.

The precision of laser ranging measurements to geodynamics satellites such as LAGEOS I & II is 1cm, the accuracy being no worse than 5cm. Details of the equipment and its performance can be found in [7].

3.3 THE JULY 1993 EXPERIMENT

Seventeen successful ranging sessions were accomplished in the period 12-20 July 1993, taking advantage of the presence of Schreiber and his APD. Predictions were derived from orbital elements provided by OPTUS P/L. Each session lasted 30-60 minutes, before which the satellite was found and centred visually in the telescope eyepiece to get the pointing correct to better than 10 arcseconds (and this was no mean feat as such a geostationary satellite is very faint - we estimate its stellar magnitude as 11-12 - there was much tripping around in the dark).

The first echos were received at 16:42 UTC on 12 July. The pattern of residuals (C-O) from the best session are shown in Figure 7, where the problem of separating true returns from background noise and thermal noise in the detector is clearly revealed - a movable gate of variable width is used to minimise noise as far as possible.

Under good weather and seeing conditions, a return rate of over 100/min was achieved for long periods. The one-sigma single-shot standard error of each session after filtering and subtracting a suitable polynomial was 3cm, which was comparable with ranging results to our regular targets at the time. However, the observed results disagreed with the predictions by up to ±3km.

The raw results have been compressed into Normal Points (5-min means), corrected for internal system delays and atmospheric refraction, and applied to orbit determination.

This experiment splendidly verified the design and manufacture of the LRA, and the efficiency of the OLRS under good conditions.

4. ORBIT SOLUTIONS

4.1 DATA AND THEIR LIMITATIONS

Not all the laser ranging data were used, since the earlier sessions were not properly calibrated. 42 5-minute Normal Points were used in the following analysis.

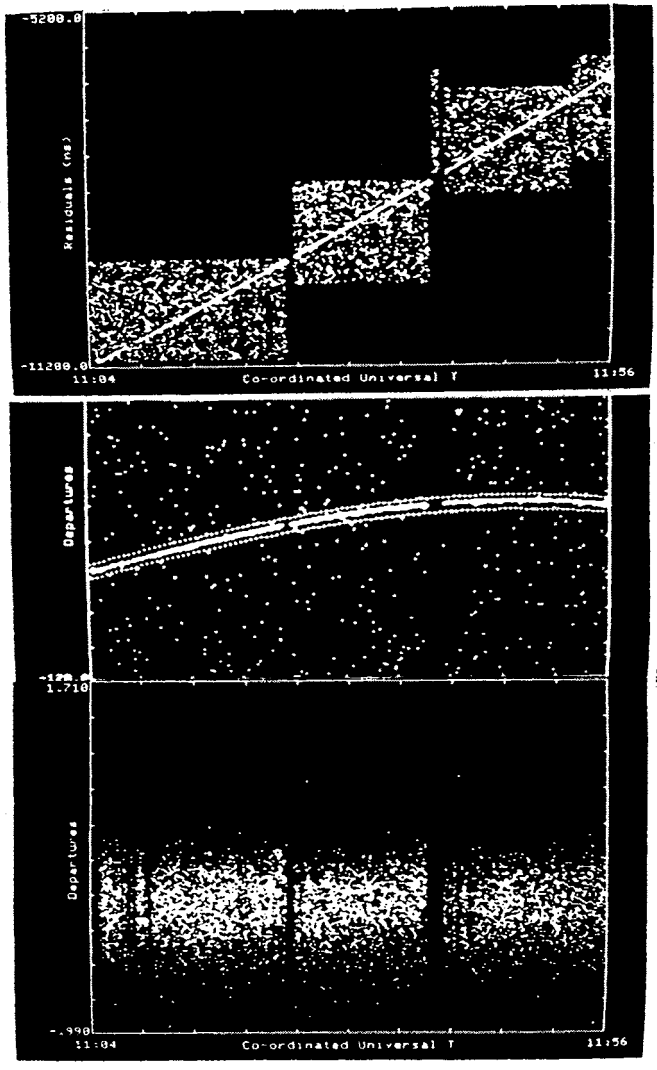


Figure 7: Range residuals during processing of a 53-minute session observed on 18 July 1993. The Y-scale is nanoseconds.

(Top) - Raw residuals. A 2μs range gate was moved twice to maintain track. Note the amount of noise present within the gate.

(Middle) - Expanded scale after removing a polynomial and filtering. Note the operator-entered filter limits.

(Bottom) - The Y-scale now spans 2.61ns after further polynomial removal and filtering, and Normal Points can be formed.

Only 3 timing centres contributed TV/GPS results : Orroral, CSIRO National Measurements Laboratory (NML) in Sydney, and MOBLAS 5 SLR station at Yarragadee,WA. The hourly TV system at Telecom Research Laboratory in Melbourne,Vic was unfortunately out of action at the time so the geometry was strongly biassed in the East-West direction.

All GPS receivers are single frequency using Broadcast Ephemeris. Only the NML TTU conformed to the BIPM Common View schedule at the time, and only SV 12 was received at the other stations.

OPTUS B1 underwent an orbit manouevre at 05:32UTC on 16 July. This report accordingly excludes TV, GPS and laser data before then. The TV/GPS data used spans five days, but there was laser data on only the first three days, and always within ±4 hours of local midnight so the full diurnal variation was not sampled.

4.2 SOLUTION PARAMETERS AND STRATEGY

The solution program solves for osculating Kepler elements at epoch, which is the date of the most recent manouevre, station biasses relative to the Grand Master, and laser ranging bias as shown in Table 2. Routines to replace the Kepler elements with the "non-singular" Lyddane elements suitable for circular equatorial orbits are being tested (but don't work at the time of writing). An algorithm to solve for a solar radiation pressure parameter is available but not yet implemented.

Table 2: Orbit Solution Parameters

Parameter	Description	Lyddane Element
a	semi-major axis	a
e	eccentricity	e.sin(M)
i	inclination	e.cos(M)
M	mean anomaly at epoch	sin(i/2).sin(Ω)
w	argument of perigee	sin(i/2).cos(Ω)
Ω	right ascension of ascending node	M + Ω + w
d_i,i=2,n	station biasses (Orroral and Yarragadee w.r.t. NML)	
S	solar radiation pressure (k.C_r.A/m) (not implemented)	
R	laser range bias (rejected after testing)	

The solution starts by integrating the orbit from the OPTUS P/L raw elements, and computing residuals and partial derivatives which are used with the observations to get a first solution. The orbit is re-integrated to get a second solution, and so on until convergence of the standard errors in station biasses is reached. At each iteration, any combination of parameters can be solved for - it soon becomes evident which one to choose before proceeding.

4.3 RESULTS AND CONCLUSIONS

Table 3 shows the standard error of each solved parameter, the standard error of unit weight and condition number for 4 combinations of parameters under 3 scenarios:

(1) Laser, TV and GPS data with TV weighted for 100 nanoseconds;

(2) Laser, TV and GPS data with TV weighted for 10 nanoseconds;

(3) TV and GPS data only.

The laser data is weighted at 0.2 ns when used.

These results show that it is pointless to solve for any of the angle variables M, ω or Ω when there is no laser data, and only one of them can be solved for with laser data present. The best solutions are the 6-parameter solutions for a, e, i, M and two station biasses, with the bias standard errors being proportional to the TV weights. It is also clear that, with a suitable choice of parameters, iteration makes little difference.

Table 3: Statistics of Orbit Solutions

Ordinary characters show 1st solution, boldface show final (4th) solution.
L = Laser + TV + GPS, T = TV + GPS only.
.1 = TV/GPS weighted for 100ns, .01 = TV/GPS weighted for 10ns σ.

Scenario	Standard Errors									Condition
	a (m)	e (10^{-9})	i (μ^o)	M (μ^o)	ω (μ^o)	Ω (μ^o)	Biasses (ns) Orr	Biasses (ns) Yar	σ_o (ns)	Number (ideal≈1)
L,.1, 4-param		46	24				435	435	3300	19
		33	**17**				**318**	**318**	**2400**	**19**
L,.01,4_param		47	24				44	44	340	19
		35	**18**				**33**	**33**	**250**	**19**
T, 4-param		456	223				8	8	62	1.3
		524	**256**				**9**	**9**	**71**	**1.3**
L,.1, 5-param	1.0	33	18				313	313	2400	20
	0.3	**9**	**5**				**83**	**83**	**640**	**20**
L,.01,5-param	1.0	34	18				32	34	243	20
	0.3	**10**	**6**				**10**	**11**	**76**	**20**
T, 5-param	1.2	269	130				5	14	36	12
	1.2	**269**	**131**				**5**	**14**	**36**	**12**
L,.1, 6-param	0.2	18	4	18			56	56	430	176
	0.2	**18**	**4**	**18**			**55**	**55**	**420**	**179**
L,.01,6-param	0.3	25	6	25			8	9	61	175
	0.3	**26**	**6**	**25**			**8**	**9**	**61**	**178**
T, 6-param	1.2	340	212	87217			1009	17086	36	10^7
	1.2	**340**	**213**	**87356**			**1011**	**17113**	**36**	10^7
L,.1, 8-param	0.2	18	8	953	Big	Big	38	38	289	10^8
	0.2	**11**	**5**	**573**	**Big**	**Big**	**28**	**28**	**214**	10^8
L,.01,8-param	0.4	33	14	1715	Big	Big	7	10	53	10^8
	0.5	**26**	**12**	**1327**	**Big**	**Big**	**7**	**11**	**51**	10^8
T, 6-param	1.2	535	214	Huge	Huge	Huge	1512	Big	36	10^8
	1.1	**517**	**204**	**Huge**	**Huge**	**Huge**	**1432**	**Big**	**35**	10^{10}

Figure 8 shows the measurements residuals directly after integration from the raw orbit elements, for both TV stations relative to NML and for laser ranging, while Figure 9 shows them after re-integration from the third iteration of the 6-parameter solution selected above. The biasses have reduced, but slopes remain which need to be investigated. Statistics of straight-line fits through the two sets of data are given in Table 4, and show that the effect of orbit improvement on the TV difference data is equivalent to a simple linear trend, at least during this experiment. The fit standard error for laser ranging improved considerably, however. This analysis, together with the evidence in Figure 3, reveals that the raw orbits can satisfy users certainly at the 1 microsecond level - which was not the case when the laser ranging project was initiated in 1987.

It remains to be seen what conclusions can be drawn when a good distribution of TV/GPS stations and round-the-clock laser ranging become available.

Table 4: Results of Fitting Straight Lines through Orbit Residuals

Observed - Integrated = A + B.(t - t') where t' is mean epoch.

Quantity	Orr'l-NML TV		Y'dee-NML TV		Laser Ranging	
	Initial	3rd	Initial	3rd	Initial	3rd
A (μs)	-0.024	-0.292	4.146	-0.370	1.341	-0.024
B (μs/day)	-0.0431	-0.0493	0.0062	-0.0959	0.0547	-0.0117
σ (μs)	0.027	0.028	0.017	0.016	0.020	0.003

5. FURTHER DEVELOPMENTS

5.1 GROUND SEGMENT

- An avalanche diode detector is currently being installed, to ensure the required sensitivity.

- The concentration of light on the detector will be enhanced by a re-designed secondary mirror due for installation in March 1994.

- An optical filter with 70% transmission and 0.1nm bandwidth is to be ordered, to minimise background noise especially during daylight.

- New bearings have been installed, to improve pointing accuracy.

5.2 GPS TIME TRANSFER

- A project is in hand to transform the coordinates of all antennae from WGS84 to ITRF91 (epoch 1988.0), by reference to fundamental SLR and VLBI sites forming ITRF91, and by re-survey where necessary. This project will include the OPTUS antennae feed horns, too.

- The Australian National Time Committee is encouraging all stations to conform to BIPM Common View schedules and CCDS format.

- Procedures for obtaining Precise Ephemerides from IGS are being established, in order to make post-processed corrections from Broadcast Ephemeris results.

- A dual-frequency GPS TTU is included in Orroral's budget requests, while a network of geodetic Turbo-Rogue receivers forming the

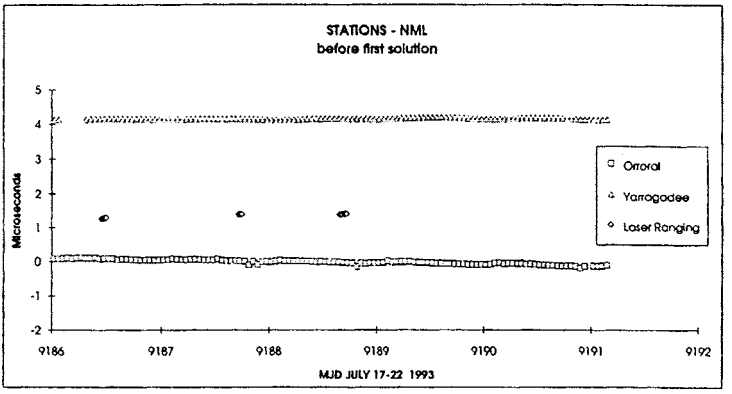


Figure 8: Observation residuals (without any trend removal) with respect to quantities computed from raw orbital elements, for the two GPS-corrected TV differences and for laser ranging.

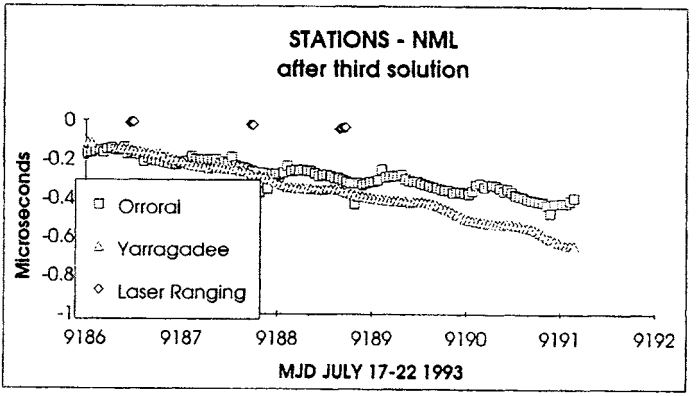


Figure 9: As for Figure 8, with respect to quantities computed from the third iteration in the orbit determination process.

Australian Fiducial Network is expected to provide local ionospheric data for application to the single-frequency TTUs.

- An improved method of interpolating between GPS measurements is being studied.

- The problem of receiver calibration is still to be addressed.

5.3 SOFTWARE

- At present, TV measurements are single readings taken every hour at any site. The effects of averaging over, say, 10 seconds should be studied.

- Evaluation of the effect of using Lyddane rather than classical Kepler orbital elements is under way.

- Relativistic corrections and solution for a solar radiation parameter are under discussion.

All the above developments are designed to enhance the precision and accuracy of the method, and to promote reliable, routine operation. Most of them have valuable application in other areas of Orroral Observatory's work.

Acknowledgements

Duncan Henry of OPTUS Pty.Ltd. managed development, construction and testing of the LRA which was funded by OPTUS through Hughes Aircraft Co. Anthony Belo and Andrew Edwards of OPTUS kindly provide regular orbit predictions and much advice. Nelson Evans and Jeff Boiko of BAeA were instrumental in mounting and testing the LRA. Bob Schaeffer at AG Thompson delivered excellent coatings. Many other staff of these companies provided valuable technical and managerial expertise. It was a privilege to be associated with Beattie Steel in the corner cube optical design and testing.

Bill James spent endless hours making the corner cubes, polishing them and getting the dihedral offset angles correct. He also designed and is making the new secondary mirror, and has been a continual pillar of support. Ulrich Schreiber of Wettzell provided the photon detector which was entirely responsible for the great success of the experiment.

The staff at Orroral Observatory has given outstanding technical and observational support. In particular, one of the authors wishes to acknowledge the other (JRW) for his superb management of the data both in this experiment and for many years in the Australian National Time System.

References

1. Miller,M.J. "Synchronization of Distant Clocks by Television Pulse Comparisons",Proc.Astron.Soc.Aust.,Vol.1,No.7,p.352 (1970).
2. Luck,J.McK. "Comparison and Coordination of Time Scales",Proc. Astron.Soc.Aust.,Vol.3,No.5,p.357 (1979).
3. Luck,J.McK.,Harvey,I.K.,Harris,R.W.,Harvey,G. "Status of Activities and Plans for Precise Time Comparisons in Australia",Proc.18th PTTI,
   Washington,D.C.,p.177 (1986).
4. Luck,J.McK.,Dunkley,J.,Armstrong,T.,Gifford,A.,Landis,P.,Rasmussen, S.,Wheeler,P.,Bartholomew,T.,Stein,S. "A Proposed Time Transfer Experiment Between the USA and the South Pacific",Proc.23rd PTTI, Pasadena,CA,NASA Conf.Publ.3159,p.351 (1991).
5. Evans,N.O. "Laser Retroreflector Unit Development",unpublished note

(1990).
6. James,W.E.,Steel,W.H.,Evans,N.O.J.L. "Design and Testing of a Cube-Corner Array for Laser Ranging",SPIE Vol.1400 Optical Fabrication and Testing,p.129 (1990).
7. Luck,J.McK. "Performance of the Upgraded Orroral Laser Ranging System",Proc.8th International Workshop on Laser Ranging Instrumentation,NASA Conf.Publ.3214,p.11-6 (1992)

Atomic Frequency Standards I
(B 3)

$\Delta m_f = \pm 1$ TRANSITIONS IN CESIUM BEAM FREQUENCY STANDARDS

V. Giordano[#] , L. Pichon * , B. Boussert, P. Cérez, P.Blot[#]

Laboratoire de l' Horloge Atomique
Unité Propre de Recherche du CNRS associée à l'Université de Paris-Sud
Bat. 221, Université de Paris-Sud, 91405 Orsay cedex , FRANCE

INTRODUCTION

It is well known that spurious $\Delta m_f = \pm 1$ transitions could be excited in cesium beam resonators. Such transitions are due to the presence in the Ramsey cavity of microwave magnetic inductions perpendicular to the static magnetic field direction. In the case of E-plane Ramsey cavities, there undesirable orthogonal components result from the distorsion of the field lines in the vicinity of the cavity holes. In H-plane cavities, the natural curvature of the field lines leads to transverse components of the microwave field.

$\Delta m_f = \pm 1$ transitions could induce a frequency pulling (Ramsey pulling) of the clock transition [1] and have to be minimized. It is then important to clarify how $\Delta m_f = \pm 1$ transitions are produced in order to choose an appropriate design for the Ramsey cavity.

In this paper, we present a study of the $\Delta m_f = \pm 1$ transitions. It is shown that their resonance features are derived from a pure 2 level atomic system interacting with four spatially separated oscillating fields. The transition probability of the $\Delta m_f = \pm 1$ resonance is calculated with the actual variation of the microwave field along the atomic beam obtained with a tridimensional computing method. Experimental results are compared with the theoretical prediction.

(#) LPMO , 32 avenue de l'Observatoire, 25044 Besançon Cedex, FRANCE
(*) LGEP, Plateau du Moulon, 91192 Gif s/Yvette Cedex, FRANCE

EXPERIMENTAL OBSERVATIONS

Figure 1 presents the microwave spectrum recorded with the LHA optically pumped cesium beam resonator CsIII described in a previous publication [2]. The static magnetic field is set to 80 milliGauss.

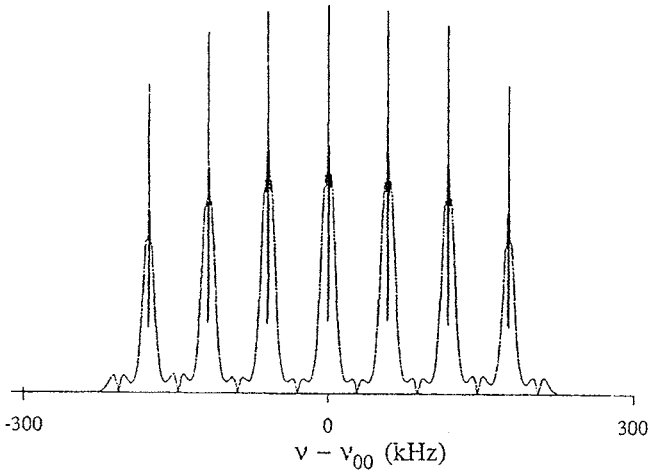


Fig.1 Microwave spectrum delivered by CsIII. B_C=80mG

The microwave spectrum shows the seven familiar Ramsey type resonance lines corresponding to the $\Delta F=\pm 1$; $\Delta m_f=0$ or σ microwave transitions. In addition we can see eight small resonance profiles surrounding the seven main lines. They present a typical shape characterized by a dip at resonance and two wide bumps symmetrically located on each side of the resonance frequency. Such resonances are also present in almost all commercial and primary cesium beam tubes. They are interpreted as $\Delta F=\pm 1$; $\Delta m_f=\pm 1$ or π microwave transitions, which are induced when the oscillating magnetic field B_{RF} has a component perpendicular to the static magnetic field B_C. The CsIII cavity is an H-plane cavity, ie the atom trajectories are parallel to the large side of the waveguide (see fig.2). In this case, the natural curvature of the field lines in the cavity near the short circuit leads to transverse components of the RF field. Figure 3 shows typical records performed with a static magnetic field set to 500mG in order to avoid any overlap with adjacent transitions. Around the power which optimizes the clock signal the bump amplitude depends linearly on the microwave power fed into the cavity. The position of the bumps maxima does not depend on the microwave power

at least within the considered power variation. At the optimum power, the bump amplitude is of the order of 4% of the clock signal amplitude.

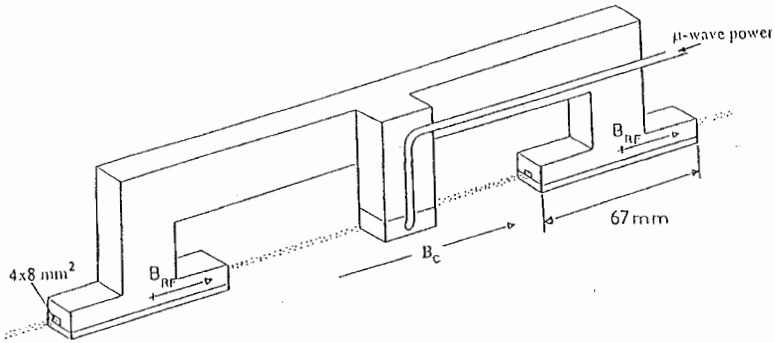


Fig. 2 The CsIII H-plane Ramsey cavity.

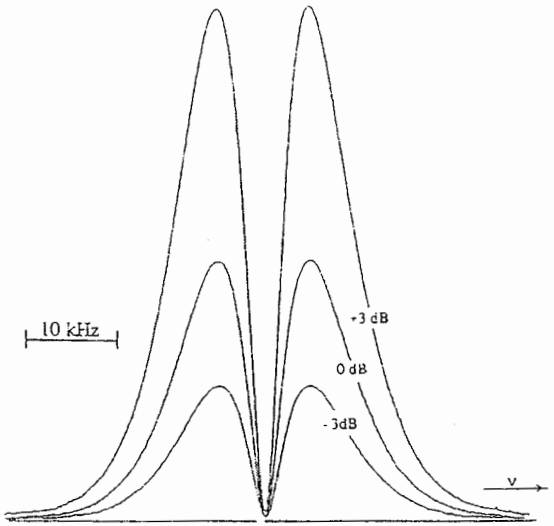


Fig.3 π microwave transition at several microwave powers

MICROWAVE π TRANSITION PROBABILITY

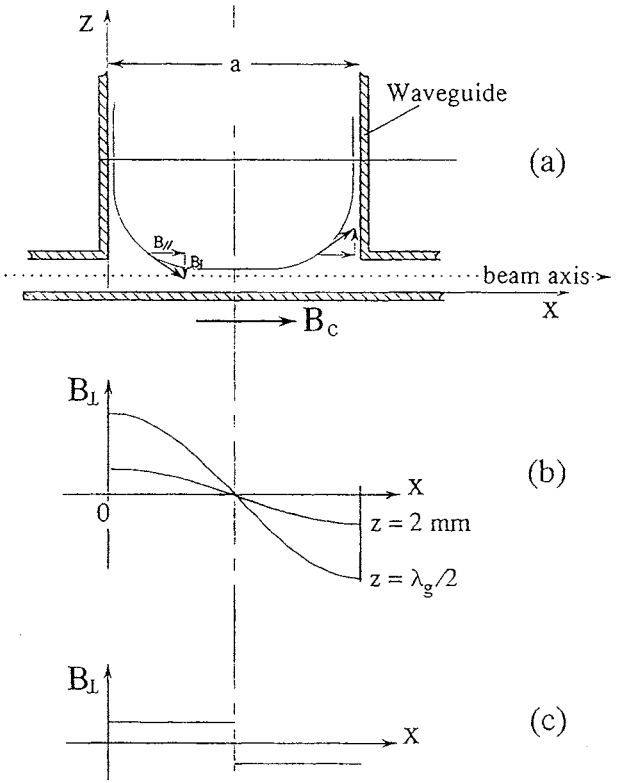


Fig.4 Microwave magnetic field lines in an arm of the Ramsey cavity.

Figure 4a clearly shows that the curvature of the magnetic field lines in the arm of the H-plane Ramsey cavity induces two out of phase perpendicular components of B_{RF}. If we neglect the distorsion of the field lines in the vicinity of the beam apertures, the variation of the transverse component $B_\perp$ is proportional to $\cos(\pi x/a)$; a is the largest internal dimension of the rectangular waveguide. A rough approximation of this cosine profile is reported on fig. 4c. This square wave profile is used to derive an analytical expression of the π transition probability .

Let us consider first monokinetic two levels atoms and the square wave profile shown on figure 5.

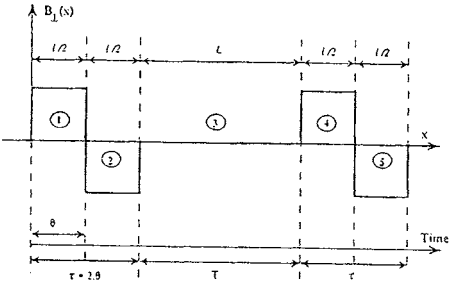


Fig. 5 The simplified 5 zones profile used in the transition probabilty calculation

According to the formalism developed in reference [3] for the Rabi method of magnetic resonance, the state of a two-level quantum system is represented by a three component density state vector $\mathbf{a}(a_1;a_2;a_3)$. a_1,a_2 and a_3 being respectively the real, the imaginary part of the coherence and the population difference between the two considered quantum states. The time evolution of **a** in space region with constant RF-field is described by a 3x3 matrix M. The expression of M is given in [3]. In the present case, the atom crosses successively five magnetic regions as shown in fig. 5. We have:

$$\mathbf{a}(\tau,T,\tau) = M_5M_4M_3M_2M_1\,\mathbf{a}(0) \qquad (1)$$

with $M_5=M_2$ and $M_4=M_1$. τ is the cavity arm interaction time and T the drift time between the two cavity arms. **a**(0) is the state vector after the atomic preparation. We assume that $a_1(0)=a_2(0)=0$ and $a_3(0) \neq 0$ being the population difference induced by the atomic preparation.The transition probability at the output of the Ramsey cavity is given by:

$$P(\tau,T,\tau) = P = \frac{1}{2}\left[1 - \frac{a_3(\tau,T,\tau)}{a_3(0)}\right] \qquad (2)$$

After straight forward calculation, the following expression for the transition probability is obtained:

$$P = \frac{16b_\perp^2\Omega_0^2}{\Omega^4}\sin^4\left(\frac{\Omega\tau}{4}\right)\left[\left(\frac{b_\perp^2}{\Omega^2} + \frac{\Omega_0^2}{\Omega^2}\cos\frac{\Omega\tau}{2}\right)\cos\frac{\Omega_0 T}{2} - \frac{\Omega_0}{\Omega}\sin\frac{\Omega\tau}{2}\sin\frac{\Omega_0 T}{2}\right]^2 \qquad (3)$$

$b_\perp$ is the Rabi angular frequency proportional to the amplitude of the constant microwave field $B_\perp$.$\Omega_0 = \omega - \omega_0$; ω is the angular frequency of the oscillating field and

ω_0 the transition angular frequency between the two state. $\Omega = \sqrt{b_\perp^2 + \Omega_0^2}$

The transition probability has to be averaged over the Maxwellian velocity distribution in the beam [4].Taking into account of experimental parameters the averaged transition probability has been computed. The result is shown on figure 6.

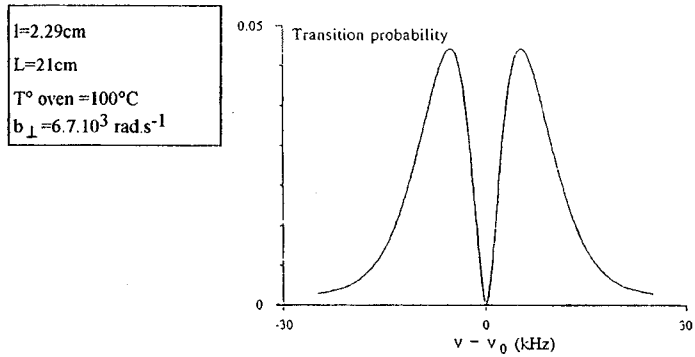


Fig.6 Calculated π transition probability

The general shape of the computed profile matches the recorded pattern shown in figure 2. However a discrepancy exits between the calculated and measured amplitudes. As shown below, this last vanishes if we consider more accurate models for the atomic system and for microwave magnetic field variations along the atomic beam trajectories.

ACTUAL FIELD PROFILES

One of us has developped electromagnetic field computations based on the finite element method. It allows the study of complex microwave structures such resonant cavities with holes [5].A determination of the actual $B_\perp$ profile taking into account the field perturbations caused by the holes in the cavity structure is then possible [6]. The studied structure and the corresponding finite element model are represented on figure 7a and 7b. The figure 7c shows, in a cut plane the distribution of magnetic field vectors in the middle of the cavity. The resulting magnetic field profile along an atomic trajectory passing in the middle of the beam aperture is shown on fig. 8a. It is rather different from pure cosine which should exist in the absence of holes.

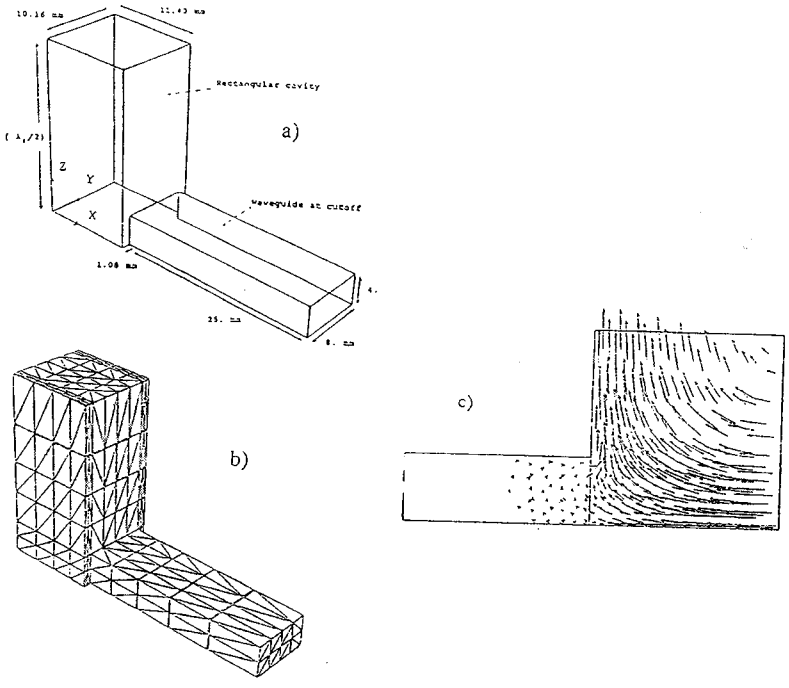


Fig.7 Studied struture and result.

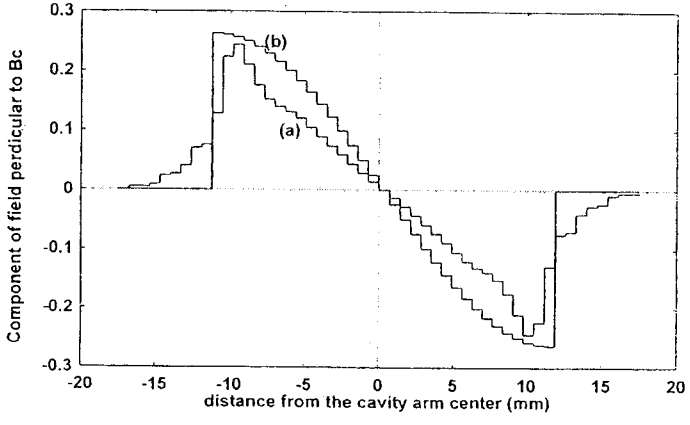


Fig.8 a- actual field profile b- cosine field profile.

π MICROWAVE TRANSITION DOUBLETS

According to the quantum selection rules, 14 π transitions are allowed between the two ground-state hyperfine levels of Cs^{133}. 12 of them are associated in 6 closely spaced doublets. The doublet separation is 1.1 kHz /Gauss leading to a very small frequency separation at the static magnetic field intensities used in atomic beam resonators. The other six doublets are unresolved and only 8 microwave π transitions are visible in practice. We consider the 4 level system shown on figure 9 .The two transition probabilities are treated independently because we implicitly assume that the static field is high enough to decouple the Zeeman sublevels in each multiplet. Each monokinetic probability is first computed using the standard matrix technique described previously. The actual field profile is approximated by a stair case function; $b_\perp$ is constant on each step. Then the probability is averaged over the maxwellian interaction time distribution. Finally the two transition probabilities are summed to obtain the signal delivered by the resonator. Figure 10 shows the evolution of the experimental π transition probability compared with our computation

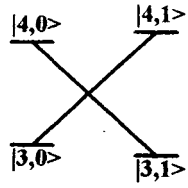


Fig. 9

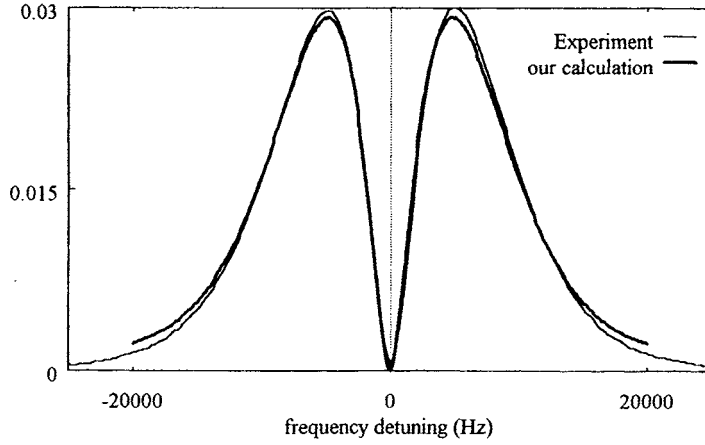


Fig. 10 π transition probability:
Experimental data and Calculation with actual field profile

The good agreement between our calculation and the experimental data strengthens our confidence in the model which has been used in the treatment of π transitions in H-plane cavities.

CONCLUSION

We have explained the shape of the π microwave resonances which are present in cesium clocks equipped with H-plane cavities. The model which has been developed considers a two level syatem interacting in 4 zones with the actual microwave transverse field profile existing in the cavity. This last has been obtained with tridimensional electromagnetic field computations. A good agreement is found between theory and experimental records. The model is valid when the static magnetic field is high enough to decouple all the atomic levels. In that case, π and σ transition probabilities can be treated independently.

REFERENCES

[1]L.S. Cutler, C.A. Flory, R.P. Giffard and A. De Marchi.
J. Appl. Phys. , **69**, (5), 1 March 1991.

[2]V. Giordano, A. Hamel. P. Petit, G. Theobald, N. Dimarcq, P. Cerez, C. Audoin.
IEEE Trans. on Ultrasonics, Ferroelectrics and Frequency Control, UFFC-38, n°4, 1991, p. 350.

[3]J. Vanier and C. Audoin
The quantum physics of atomic frequency standards, Adam Hilger, Bristol 1989, Vol. 2, p. 624.

[4]A. Hamel, V. Giordano, P. Petit, C. Audoin.
Proceedings of the 4th European Frequency and Time Forum, Neuchâtel Switzerland, 1990, p. 169.

[5]L. Pichon, A. Razek.
ACES Journal, Vol. 6, n°2, 1991, p. 133.

[6]V. Giordano, L. Pichon, P. Cerez, G. Theobald (to be published)

RECENT RESULTS OBTAINED WITH THE PTB's OPTICALLY PUMPED CESIUM ATOMIC CLOCK CSX

S. Weyers, R. Schröder, and A. Bauch
Physikalisch-Technische Bundesanstalt (PTB), Braunschweig

SUMMARY

The cesium atomic clock CSX of the PTB has been provided with two systems for excitation of the atoms with laser light. After a complete evaluation of all systematic frequency shifting effects the frequency values measured with the optical detector and with the conventional Langmuir-Taylor detector agree within $1 \cdot 10^{-14}$ for different optical excitation schemes and laser powers. Our measurements show that a light shift due to the thermal radiation of the Langmuir-Taylor detector is smaller than $1 \cdot 10^{-14}$. An improvement of the frequency instability of the clock operated with optical state preparation and the Langmuir-Taylor detector is discussed.

1. INTRODUCTION

In order to compare the possibilities of an optically pumped cesium atomic clock with a conventional one the clock CSX of the PTB was provided with two systems for the excitation of atoms with laser light (Fig. 1).

The first one serves for state preparation of the atoms behind the polarizing magnet P. The second one serves for the analysis of the atomic quantum state behind the second irradiation zone of the Ramsey cavity.

In chapter 2 we will show our results concerning the frequency measurements with the optical and with the Langmuir-Taylor detector. In connection with these measurements we were able to prove that in the CSX there is no light shift due to the Langmuir-Taylor detector at the level of $1 \cdot 10^{-14}$ (chapter 3).

When in the first interaction zone the atoms are optically pumped into the hyperfine level $F_g = 3$, the CSX is operated in the so-called "flop-in" mode in which the frequency instability should be considerably lower than in the normal "flop-out" operation mode (chapter 4).

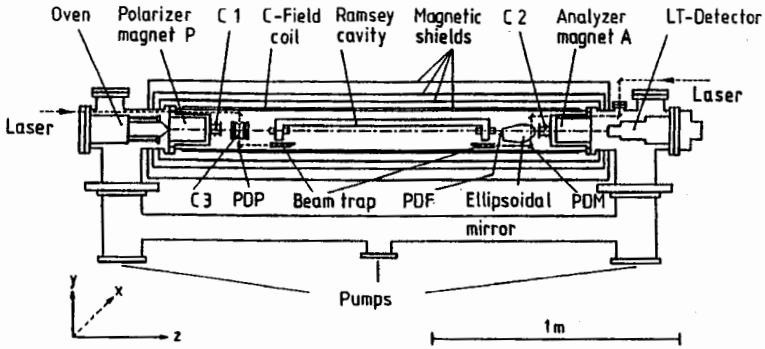


Fig. 1: Vertical section of the PTB's experimental frequency standard CSX with the optical systems for atomic state preparation and detection. PDP and PDM are photodiodes monitoring the incident laser intensity. The photodiode PDF collects the fluorescence light. C1 — C3 mark supplementary coils.

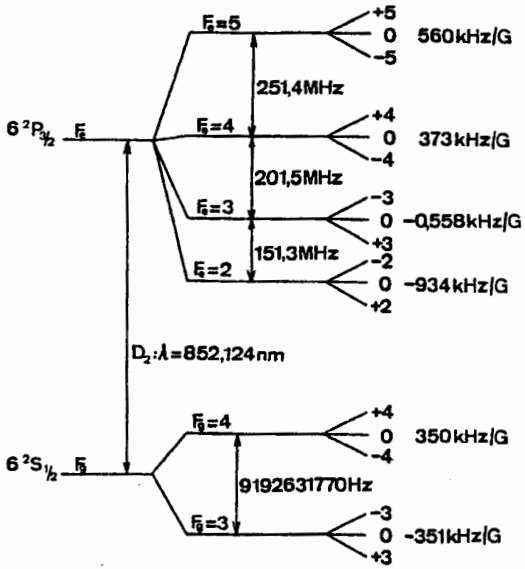


Fig. 2: Part of the energy level diagram of ^{133}Cs (not to scale).

2. FREQUENCY MEASUREMENTS

With the optical detection system (Fig. 1; details are described in /1/) it is possible to detect the fluorescence light of the atoms which have made the clock transition ($F_g = 4$, $m_F = 0 \longrightarrow F_g = 3$, $m_F = 0$), when these atoms are excited with laser light at the cesium D_2 line (Fig. 2), linearly polarized, perpendicularly to the direction of the atomic beam. The fluorescence light detected with the photodiode PDF is used for frequency locking the VCXO. We measured the quantity

$$F_P(CSX) = \frac{f_0(CS2) - f_P(CSX)}{f_0(CS2)}, \qquad (1)$$

where $f_0(CS2)$ is the clock transition frequency of the primary clock CS2 of the PTB, and $f_P(CSX)$ is the clock transition frequency of the CSX not corrected for the Ramsey cavity phase difference.

For a frequency measurement with the CSX it was advantageous to set the microwave power -1,5 dB below the value which yields maximum resonance signal. Thus the contribution of a velocity group of slow atoms to the signal was minimized, otherwise it would lead to a distortion of the Ramsey pattern /1/.

The values of $F_P(CSX)$ measured with the optical detector referenced to the Langmuir-Taylor detector data are shown in Figs. 3a and 3b, where $F_P(CSX) = 0$ corresponds to the mean frequency value obtained with the Langmuir-Taylor detector. The statistical uncertainty of this mean frequency value is $4 \cdot 10^{-15}$. The atoms were excited with three different excitation schemes ($F_g \longrightarrow F_e$) and with different powers of the laser beam in order to investigate a possible light shift originating from the fluorescence light or the stray light.

	$F_P(CSX) \cdot 10^{15}$
LT	2015 ± 4
3 -> 2	2009 ± 5
3 -> 3	2020 ± 9
4 -> 4	2015 ± 16

Tab. 1: Frequency values $F_P(CSX)$ (1) measured with the Langmuir-Taylor detector (LT) and with the optical detector using the excitation schemes ($F_g \longrightarrow F_e$) as indicated.

The mean values of the data and their uncertainties are listed in Tab. 1. Obviously there is no light shift at the level of $1 \cdot 10^{-14}$. This result is confirmed by the following: when the $F_g = 3 \longrightarrow F_e = 2$ excitation scheme was employed, we obtained $F_P = (2008 \pm 4) \cdot 10^{-15}$ with a laser power of 0,44 mW and with a laser power of 0,06 mW we got $F_P = (2009 \pm 13) \cdot 10^{-15}$.

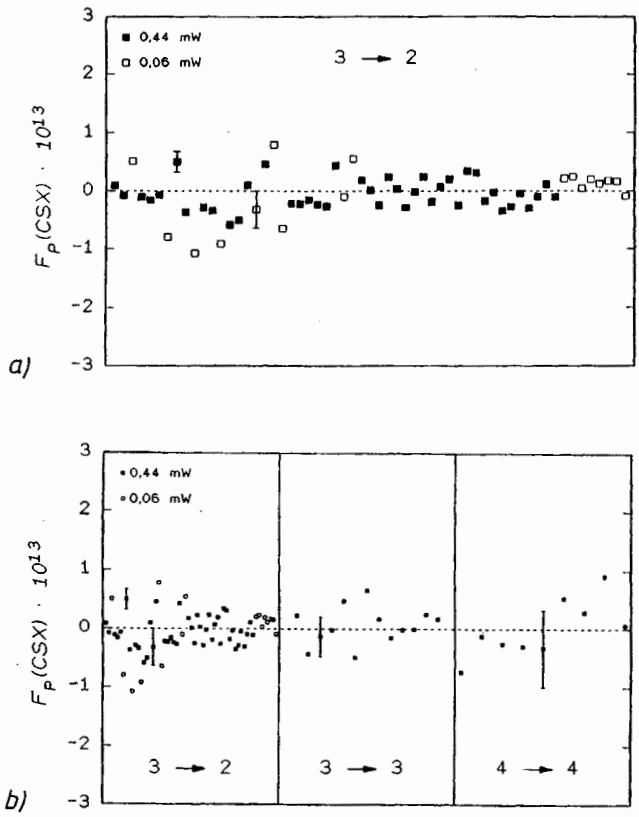


Fig. 3a, b: Relative frequency F_p(CSX) obtained with the optical detection system at two different laser powers using different optical excitation schemes $F_g \longrightarrow F_e$. F_P(CSX) = 0 corresponds to the value obtained with the conventional detector. The statistical uncertainty of this value is $4 \cdot 10^{-15}$. The data points collected in a random sequence are mean values averaged over 8 to 18 hours. Typical error bars are indicated.

3. LIGHT SHIFT DUE TO THE LANGMUIR-TAYLOR DETECTOR

The frequency measurements with the CSX can be carried out in three modes: in the conventional mode with the Langmuir-Taylor detector and in two optical modes with the filament of the Langmuir-Taylor detector heated or not heated.

We measured F_P(CSX) with the optical detection system, when the heating of the Langmuir-Taylor detector was switched on or off. The results are shown in Fig. 4. The mean values of F_P(CSX) are $(2011 \pm 7) \cdot 10^{-15}$ and $(2011 \pm 6) \cdot 10^{-15}$, respectively. At a level of 10^{-14} there is no indication of a light shift caused by the light emitted from the heated filament.

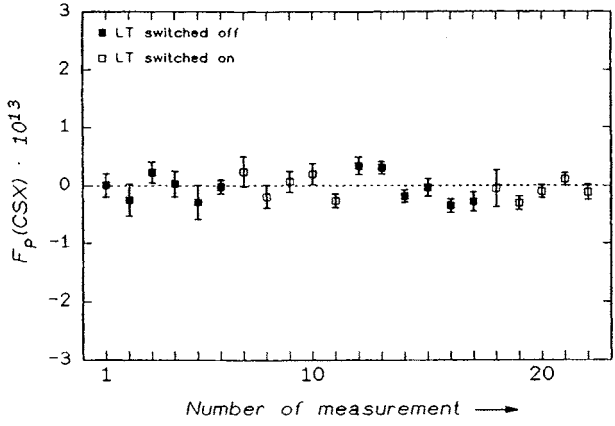


Fig. 4: Relative frequency F_p(CSX) obtained with the optical detection system measured with the Langmuir-Taylor detector (LT) switched on or off. F_P(CSX) = 0 corresponds to the value obtained with the conventional detector. The statistical uncertainty of this value is $4 \cdot 10^{-15}$. Excitation scheme $F_g = 3 \rightarrow F_e = 2$. Laser power: 0,44 mW.

In the following we compare this result with a theoretical evaluation. For the alkali atoms it is a good approximation to assume that only the first excited p state makes a significant contribution to the light shift. Thus, according to /2/ the frequency shift df/f_0 due to the electric field E(t) of the thermal radiation of a blackbody can be estimated from

$$\frac{df}{f_0} = \frac{-k \cdot 10^{-6} \langle E^2(t) \rangle \, [1 + q]}{f_0} \; \mathrm{Hz\,(V/cm)^{-2}}, \qquad (2)$$

where $<E^2(t)>$ given by the Planck radiation law is $(8{,}319\ V/cm)^2\ [T(K)/300]^4$. T(K) is the temperature of the blackbody given in Kelvin.

For ^{133}Cs the experimental value of $k = 2{,}25$ /3/ and q is given by $q = 1{,}4 \cdot 10^{-2}\ [T(K)/300]^2$ /2/. Here we consider the thermal radiation of the heated filament (1500 K) of the Langmuir-Taylor detector. Taking into account the fractional solid angle seen from the second microwave interaction zone, which is $3{,}4 \cdot 10^{-6}$, we obtain $df/f_0 = 4{,}9 \cdot 10^{-17}$, which is in fact too small to be measured in the CSX.

To sum up, measurements as well as a theoretical evaluation show that the light shift in a conventional atomic clock due to the thermal radiation of the Langmuir-Taylor detector should be at least smaller than $1 \cdot 10^{-14}$.

4. PROSPECTS OF OPERATING THE "FLOP-IN" MODE

Actually the atomic clocks of the PTB are equipped with state selecting magnets of lens-like properties. Thus the clocks are operated in the so-called "flop-out" mode, i.e. atoms which have made the clock transition ($F_g = 4$, $m_F = 0$ —> $F_g = 3$, $m_F = 0$) are subsequently "flopped out" by the analyzing magnet A. The disadvantage of this operation mode is the high background consisting of all atoms in the states $F = 4$, m_F ($m_F \neq 0$, $m_F \neq -4$), which contribute only to the noise and not to the signal. Hence the frequency instability is considerably larger than it would be if the clock was operated with the same parameters but in a "flop-in" mode.

In order to realize such a flop-in mode the CSX was provided with an optical interaction zone between the polarizer magnet P and the first microwave interaction zone of the Ramsey cavity (Fig. 1). At the pumping zone a pair of Helmholtz coils C3 creates a magnetic induction of 0,6 G to prevent coherent population trapping which may lead to incomplete pumping /4/. To optically pump the atoms into the state $F_g = 3$ we use the pumping transition $F_g = 4$ —> $F_e = 4$ (σ-polarization) of the cesium D_2 line (Fig. 2). It has been shown that for to obtain a high signal-to-noise ratio it is crucial to pump almost all atoms /5/: in the expression for the noise a small amount of unpumped atoms may lead to a term proportional to the square of the atomic flux measured at the detector and therefore to a considerable increase of the noise.

With a digital spectrum analyzer we measured the noise of the detector signal. For Fourier frequencies < 1 kHz the noise spectrum is essentially flat. We took the

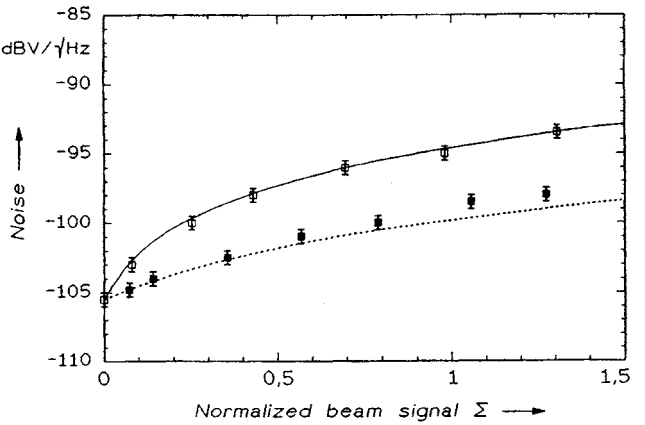


Fig. 5: Beam signal noise as a function of the normalized beam signal Σ (see text). The full line and the broken line represent theoretical calculations without and with optical pumping, respectively. The measured values are indicated by open squares (without optical pumping) and full squares (with optical pumping).

level of this most important part of the noise spectrum to determine the instability of the CSX.

As the actual atomic flux at the first interaction zone with the laser light is unknown, we define a normalized beam signal Σ. $\Sigma = 1$ holds if a beam current of 1000 pA is detected with the Langmuir-Taylor detector without optical pumping applied. A plot of the measured noise values versus Σ enables a comparison of the achievements at the same atomic flux emerging from the cesium oven.

The full line in Fig. 5 shows the noise calculated for the flop-out mode (without laser pumping) for different normalized beam signals Σ at the interaction zone, where the detector noise (essentially the (4kT/R)-noise) and the shot-noise of the atomic beam have been taken into account. For the pumping of the atoms in the interaction zone with laser light the calculated noise is shown by the broken line. The squares indicate the results of the noise measurements using the digital spectrum analyzer. Without optical pumping the measured values (open squares) agree well with the theory. If optical pumping is applied, theory and experiment (full squares) differ at normalized beam signals $\Sigma > 0{,}5$. This is explained by the quadratic dependence of the additional noise contribution upon the atomic flux, if the pumping of the atoms is incomplete.

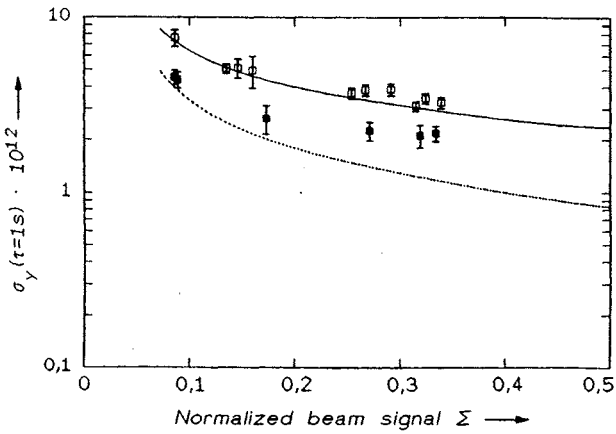


Fig. 6: Two-sample standard deviation $\sigma_y(\tau = 1s)$ as a function of the normalized beam signal Σ (see text). The full line and the broken line represent theoretical calculations without and with optical pumping, respectively. The measured values are indicated by open squares (without optical pumping) and full squares (with optical pumping).

For this reason we restricted the frequency instability measurements to $\Sigma < 0{,}5$. In Fig. 6 the full line shows the two-sample standard deviation $\sigma_y(\tau = 1s)$ calculated for the flop-out mode (without laser pumping) as a function of Σ. The open squares show the results of the measurements in this mode. With optical pumping of the atoms in the optical interaction zone, the calculated $\sigma_y(\tau = 1s)$ is shown by the broken line. The full squares are the measured values.

Obviously there is a considerable reduction of $\sigma_y(\tau = 1s)$ in the case of optical state preparation. The deviations of the measurements from the theoretical predictions result probably from an actual malfunction of the CSX, which is under investigation.

5. CONCLUSIONS

As demonstrated above, the conversion of the PTB's experimental cesium clock CSX from a conventional to a hybrid machine in which the optical pumping can be studied has been successful. The stray light does not cause a light shift at the level of $1 \cdot 10^{-14}$.

As a first application of the newly gained experimental capabilities the light shift caused by the thermal radiation of the heated filament of the Langmuir-Taylor detector was analyzed. It is at least smaller than $1 \cdot 10^{-14}$, in agreement with a theoretical estimation.
Finally, first results concerning optical preparation of the atomic beam show the possibility of operating a previously conventional cesium clock (flop-out mode) in the flop-in mode and thus to obtain a lower instability.

This work has been supported by the Deutsche Forschungsgemeinschaft and the European Community.

REFERENCES

/1/ S. Weyers, R. Schröder, T. Heindorff, and A. Bauch, Proceedings of the 7th European Frequency and Time Forum, Neuchâtel, 473 - 477 (March 1993)
/2/ W. M. Itano, L. L. Lewis, and D. J. Wineland, Physical Review A **25**, 1233 - 1235 (1982)
/3/ J. R. Mowat, Physical Review A **5**, 1059 - 1062 (March 1972)
/4/ E. de Clercq and P. Mangin, Proceedings of the 3rd European Frequency and Time Forum, Besançon, 281 -292 (March 1989)
/5/ N. Dimarcq, Thesis, Université de Paris-Sud (1992)

Light-shift saturation in a laser pumped rubidium frequency standard

Gaetano Mileti and Pierre Thomann

Observatoire Cantonal - Rue de l'Observatoire 58 - CH - 2000 Neuchâtel - SWITZERLAND

Abstract

It is shown that under certain conditions the light-shift is not a linear function of the number of photons. This non linear behaviour has been observed previously but is incompletely understood. The geometrical origin of the effect is demonstrated. The results show that the main cause of the light-shift saturation are the radial gradients of intensity in the laser beam. The axial gradients due to attenuation of light have only a reduced influence, and the optical saturation is negligible. A simple model explaining the measured effects is developed. This model underscores the physical parameters, such as light broadening, which are critical.

Keywords : Double resonance, Laser pumping, Light-shift saturation, Rubidium standards.

1. Introduction

The present work reports on our investigation on the possibility of replacing the discharge lamp with a laser in a rubidium frequency standard. This research is connected with the recent developments performed in ON of rubidium clocks for both industrial and space applications /1/. We focus here on the well-known problem of light-shift saturation, observed already in 1975, but not fully understood yet /2/.

The paper is organised in the following way. Section 2 describes the laser and resonator modules. Section 3 gives a general overview on the problem of light shift saturation, describes well-known and other related effects, and recalls the theory of double resonance experiments. Our physical interpretation of the observed phenomena is given in Section 4, and in Section 5 two experiments verifying this interpretation are presented. The conclusion is in Section 6.

2. Practical implementation

The experimental set-up consists of two separate modules placed on an optical table : the laser part, and the absorption cell part. They are both controlled with independent breadboard electronics contained in standard racks.

The laser

The pumping light is emitted by a single-mode laser diode tuned on one of the two Rb87 D_2 resonance lines at 780 nm. We had the possibility of locking the laser either on the absorption profile of the vapour where optical pumping was produced, or in a separate cell containing a different buffer gas. With the laser locked to the resonance cell, the light-shift was minimal because the pumping light was tuned to the resonance of the interrogated atoms, whereas with the laser locked to the separate cell, the pumping light was tuned 190 MHz higher, thus inducing a strong light-shift.

The results shown in this paper have been all obtained with Sharp single mode laser diodes presenting roughly 80 MHz line-width and a maximal power of 7 mW. However, the average light power has never been greater than 3 mW/cm^2, in beams of less than 1 cm^2 section.

The laser module also contains a collimator and a Peltier element to stabilise the diode temperature.

The cell and the resonator

We used cylindrical glass cells. Their diameter and length were 14 mm and 25 respectively. These absorption cells were heated between 55 and 75°C. They contained Rb87 with 30 mB of nitrogen. The microwave interrogation was provided with a magnetron type resonator producing a radially homogeneous field with a sine-square longitudinal dependence. In such a resonator, the useful RF field area is determined by the vapour volume enclosed by the electrodes. The magnetron is used also as the cell oven, and is contained in a 4 cm length cylinder. The C-field coil, producing approximately 200 mG is wound around this cylinder. Two cylindrical magnetic shields enclose the resonator, in order to reduce the perturbations of the external magnetic fields.

3. Historical and theoretical background

In this section, we present how the light-shift saturation appears in our practical implementation. As the phenomenon was thought to be the consequence of inhomonegeities in the interrogated vapour, and so dependent of the particular set-up, this step was necessary.

Light-shift saturation

Figure 1 shows a typical result obtained with a laser beam which is much narrower than the cell diameter. In this measurement it is important to determine precisely the unperturbed 0-0 transition frequency. In fact, the saturation of light-shift occurs at very low light intensity (< 0,1 mW/cm^2), and in this region, the resonance signal is small. The zero-light-shift microwave frequency was determined in the following way. The light-shifted frequency was measured versus light intensity in three different conditions : (A) Laser tuned to separate cell on F=1 (negative light-shift), (B) Laser tuned to separate cell on F=2 (positive light-shift), (C) Laser tuned to resonance cell on F=2 (near zero light shift). The corresponding curves were extrapolated to zero light intensity.

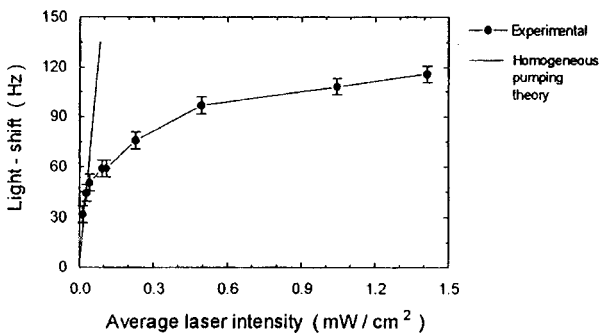


Fig. 1. The displacement of the centre of the Rb87 0-0 hyperfine transition versus the laser intensity. In this case, the beam had a diameter of 2,5 mm, and the cell 14 mm. The laser had a wavelength of 780 nm (D2 line of rubidium) and depopulated the F = 2 hyperfine level. The detuning was estimated to +190 MHz.

The result shown in Figure 1 illustrates a typical behaviour of light-shift saturation. The displacement increases linearly as a function of the light intensity only for average light intensities lower than 0,1 mW/cm^2. For higher light intensities, the slope is strongly reduced. In practice, we observed that if a narrow laser beam is used, a significant signal is obtained only for light levels in the saturated light-shift regime.

Other characteristics of the signal

For light intensities producing light-shift saturation, we observed that the other resonance characteristics (amplitude and width) also showed an unusual saturation behaviour, as illustrated by Figure 2 and 3. The theoretical curves show the expectations assuming a model where homogeneous pumping occurs in the vapour. They are the solutions of standard rate equations describing a double resonance experiment involving optical pumping with low light intensity and that we discuss below.

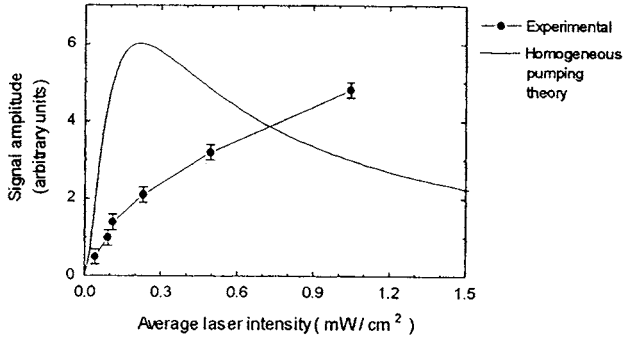


Fig. 2 Signal amplitude versus laser intensity. (c.f. Fig. 1)

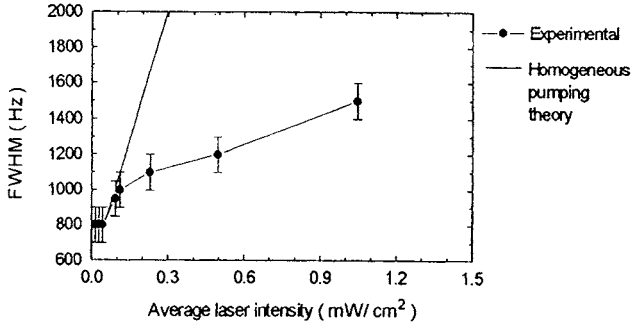


Fig. 3. Line-width versus laser intensity. (c.f. Fig. 1)

Atomic response and theoretical expectations

It is well-known that the response of the atomic medium to the optical pumping and to the microwave interrogation has to be developed in a local way, since most of the involved physical parameters are not homogeneous in the vapour. Furthermore, the description using the density matrix formalism is complicated by the number of atomic sublevels involved, so that numerical calculations are often necessary. Such computations have been made many times in the past, but unfortunately they did not reveal the basic physical phenomena which were predominant for a given effect.

With this work we wanted to extract as much information as possible with the simpler atomic model. In this model we tried to outline at least qualitatively the observed effects and understand clearly the physics. Through a more elaborate description, we expected to get the quantitative information, but no new fundamental effects. We neglect optical saturation, which is entirely justified by the optical intensities we used, and by the Doppler and collisional broadening.

A standard three-level atom model (two ground hyperfine levels and an optically excited level) leads to the following rate equations for the population difference ($\Delta = \rho_{11} - \rho_{22}$), and the coherence ($\rho_{12}$) between the two ground levels /3/ :

$$\dot{\Delta} = -\gamma_1 (\Delta - \Delta_0) + 4\ \omega_1\, \mathrm{Im}(\rho_{12}\ e^{-i\omega t}) \tag{1}$$

$$\dot{\rho}_{12} = -\ (\gamma_2 - i\,\omega_{Rb})\ \rho_{12} + i\ \omega_1\, \Delta\ e^{i\omega t} \tag{2}$$

The different parameters have the following standard meanings :

ω_{Rb} is the 0-0 hyperfine transition where we included all the shifts (light-shift and collisional shift);

ω_1 is the microwave Rabi frequency;

Δ_0 is the population difference without microwave;

$$\Delta_0 = \frac{\Gamma}{2\gamma_1} \tag{3}$$

$$\gamma_i = \gamma'_i + \frac{\Gamma}{2};\ (i = 1,2) \tag{4}$$

γ'_1 and γ'_2 are the longitudinal and transverse relaxation rates;

Γ is the absorption rate, corresponding to the number of photons absorbed per second and per atom (we assumed that the laser spectrum had a centre ν_L and a width such that only one of the two hyperfine sublevels could absorb light. The absorption rate depends on the pumping light spectral density $I(\nu)$ and the frequency dependent atomic absorption cross section $\sigma_{Rb}(\nu)$:

$$\Gamma = (h\,\nu_L)^{-1} \int I(\nu)\ \sigma_{Rb}(\nu)\, d\nu \tag{5}$$

Γ is the physical parameter which plays the essential role in understanding the saturation of the light-shift, the amplitude and the broadening of the resonance.

The stationary solutions of (1-2) can be easily found. Calculating the transmitted light, the resonance signal is given by (6) where $\Omega^2 = 4\ \omega_1^2\ \gamma_2/\gamma_1$.

$$Signal \propto \frac{\Gamma^2}{(4\gamma'_1 + 2\Gamma)} \frac{\Omega^2}{(\omega - \omega_{Rb})^2 + (\gamma'_2 + \Gamma/2)^2 + \Omega^2} \tag{6}$$

If the light intensity is assumed to be homogeneous, a single value of Γ results from equation (5) and equation (6) gives the theoretical behaviour shown in Figure 1 - 3. We refer to these predictions as "homogeneous pumping theory".

Summary of the discrepancies between experiments and "homogeneous pumping theory"

Fig 1-3 illustrate very well the main features connected with light-shift saturation. For an atomic frequency standard, the most important result is of course given by the frequency behaviour (Fig. 1). For a physical interpretation however, Fig. 2 and 3, representing the amplitude and the width of the resonance signal are of greater interest. We observe on Fig 1-3 that :

- For the three parameters, theory and experiments are in good agreement for light intensities lower than 0,2 mW/cm^2. When the light becomes stronger, however, experimental results departs drastically from the "homogeneous pumping theory".
- We can not reach a negative slope of the resonance signal amplitude (Fig 2).

Previously published works revealed already some of these effects, but their correlation does not seem to have been clearly recognised /4/.

4. Physical interpretation of the observed effects

Atomic response and theoretical expectations

As proposed by other authors, all the inhomogeneities of physical parameters have to be incorporated in order to understand the above described effects. We suggest that these effects have as the main cause the signal dependence on the absorption rate.

The local signal amplitude is given by the expression (6) putting $\omega=\omega_{Rb}$, and is represented by the theoretical curve of Fig. 2.

$$Signal\,amplitude \propto \frac{\Gamma^2}{(4\gamma'_1+2\Gamma)} \frac{\Omega^2}{(\gamma'_2+\Gamma/2)^2+\Omega^2} \tag{7}$$

The total detected signal is given by the integral of (6) over the resonance cell volume, taking into account that ω_{Rb} , Γ, and ω_l depend on position. The resonance signal is then given by a weighted sum of lorentzians centred at ω_{Rb} (which depends linearly on the light intensity). The relative weight of each volume element for the total signal is given by (7). It can be shown however, that the relative weight of each volume element for the determination of the light-shifted resonance frequency is the ratio of the amplitude (7) to the square of the line-width. Both weighting functions, represented on Fig. 4, are very selective on the pumping rate, specially the second. As a result, for a fixed laser detuning, the resonance frequency is mainly determined by the atoms subjected to a light intensity close to Iopt.

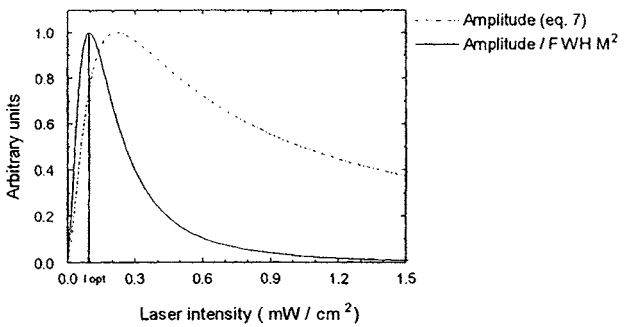


Fig. 4. The weighting functions discussed in the text.

Case of a gaussian beam : light-shift saturation

In most of the experiments where light-shift saturation was observed, a laser beam much narrower than the absorption cell was used. In such a situation, the interrogated Rb atoms are subjected to a wide distribution of light intensities between 0 and the maximal power. If we assume that the beam has a gaussian profile in the radial direction, the results of Fig. 1-3 can be understood as follows.

When the maximal light intensity is lower then Iopt, the signal comes mainly from the atoms in the centre of the beam. In this region, the light-shift and the signal amplitude are linear, and the line-width is constant because no broadening occurs. When the maximal light intensity is larger than Iopt, the situation is different. The atoms in the central part of the beam give a minor contribution, and the detected signal comes from those atoms which are subjected to Iopt. These atoms are located between two closely spaced concentric cylinders whose diameters increase with increasing light intensity.

As the resonance signal is produced by atoms feeling approximately the same light intensity, the light-shift, the amplitude and the line-width tend to saturate to a constant value. When the light intensity is further increased, it is easily seen that all the signal comes from the wings of the beam.

5. Experimental results

Selection of the interrogated atoms

The following experiment was performed in order to check the explanation of light-shift saturation given in the previous section. The gaussian laser beam irradiating the absorption cell was enlarged in order to fill the whole cell while being strongly inhomogeneous. A diaphragm placed between the cell and the detector allowed to vary the diameter of the beam which reached the detector. The results are presented in Fig. 5.

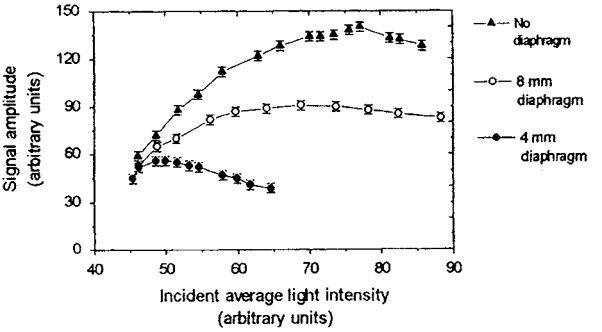


Fig. 5. The signal amplitude as a function of the light intensity. In all curves the vapour is illuminated with the same light, diaphragms of different diameters were placed between cell and detector.

We notice that these results are in good agreement with the expectations. With a reduced light intensity, all the signal comes from the centre of the beam (4 mm diameter diaphragm), and when the light is stronger, this part of the beam gives less signal. In this case, the signal comes from the atoms on a ring.

Homogeneous beam

If the laser beam is radially homogeneous, we expected that the experimental result would be closer to the theoretical results of Fig. 1-3. With the same practical implementation of the experiments described in Chapter 3, but with an homogeneous beam, we obtained the results given by Fig. 6-8.

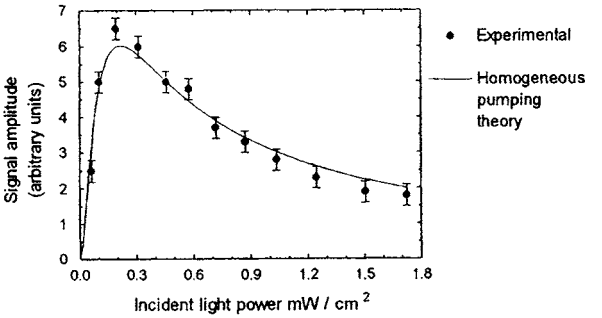


Fig. 6. Signal amplitude as a function of laser intensity with homogeneous light.

Figure 6 exhibits a dependence of the signal amplitude on light intensity very different from the case of a narrow beam (Fig. 2). For light levels higher than 0,2 mW/cm², the signal decreases if the intensity increases. This result is in full agreement with the "homogeneous pumping theory".

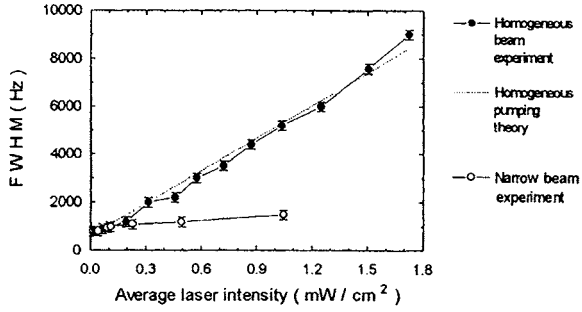


Fig. 7. Comparison of the line-width versus light intensity with homogeneous and narrow beams.

As expected, we notice on Figure 7 that the line-width increases much more strongly with the homogeneous beam than with the narrow beam, as function of the light intensity. If all the interrogated atoms are illuminated with the same light, they are all broadened in the same way.

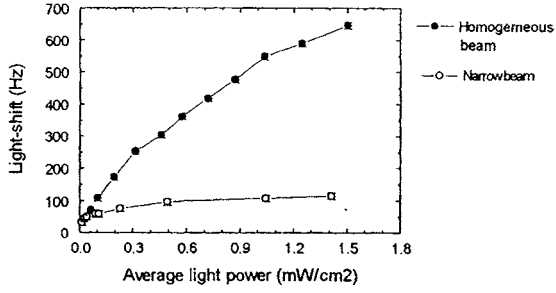


Fig. 8. Comparison of the light-shift versus light intensity with homogeneous and narrow beams.

The result shown in Figure 8 demonstrates that for a fixed laser detuning, the detected clock transition is very dependent on the type of beam. The curves show the two extreme situations of a narrow beam, and an homogeneous beam. With an average light intensity larger than 0,2 mW/cm², in the first case all the signal comes from the edges of the beam, while in the second case, all the atoms still contribute to the signal.

The non complete linearity of the light-shift with the homogeneous beam is probably due to the residual inhomogeneities in the beam, and in particular to the longitudinal inhomogeneity of light due to its attenuation by the rubidium vapour.

6. Conclusion

The problem of light-shift saturation in a laser-pumped rubidium frequency standard has been studied. A physical interpretation based on a simple three-level model allowed to understand most of the features observed. The main characteristics of the resonance signal : centre, amplitude and width could be described in different geometrical situations.

In the past, several publications have mentioned the relevance of light inhomogeneities in the rubidium vapour. This study shows quantitatively how the light inhomogeneities - in a simple model - to account in a consistent way for the light-shift, the line-width and amplitude of the microwave resonance as a function of the light intensity.

The atoms which are subjected to a high spectral density of light (>0,2 mW/cm^2 in typical conditions) have only a minor contribution to the resonance signal. In fact, in a double resonance experiment, the ratio *signal amplitude /(signal width)*^{2} is very selective with respect to the absorption rate Γ. Even if through the vapour atoms are subjected to wide range of light intensities, only those which feel a light power within a narrow range give a significant signal. It is important to determine where these atoms are, because they finally determine the characteristics of the resonance, in particular its centre. In most of the practical situations, for example with a narrow beam, if the light is too strong the signal will come mainly from the edges. That was probably the case in most of the previous works published in the past.

7. Acknowledgement

We want to thank Professor G. Busca for stimulating discussions and the collaborators in ÖN who offered their technical assistance, specially C. Couplet, J.-F. Lechenne and P. Scherler.

8. References

/1/ G. Mileti, I. Rüedi, and H. Schweda, (1992), Proc 6th EFTF, Noordwijk , pp 515/19.

/2/ M. Arditi and J.-L. Picqué, J. Phys. B : Atom. Molec. Phys., Vol. 8 (1975) No 14 , pp L331/35.
J. C. Camparo, R. P. Frueholz, and C. H Volk, Phys. Rev. A, Vol. 27 (1983) No 4, pp 1914/24.
M. Hashimoto and M. Ohtsu, IEEE J. Quantum Electr., Vol. QE-23 (1987) No 4, pp 446/51.
M. Hashimoto and M. Ohtsu, J. Opt. Soc. Am. B., Vol. 6 (1989) No 10, pp 1777/1789.
J. T. Liu, P. Thomann, L. Zhang, G. Busca, (1990), Proc. 4th EFTF, Neuchâtel, pp 157/59.

/3/ P. Thomann and G. Busca, J. de Physique, C8, Tome 42 (1981), suppl. No 12, pp. C8-189/197

/4/ L. L. Lewis and M. Feldman, Proc. 35th AFCS (1981), pp. 612/24, see also /2/

Miniaturized Rubidium Clocks for Space and Industrial Applications

Pascal Rochat, Hartmut Schweda, Gaetano Mileti, G. Busca

Observatoire Cantonal, Rue de l'Observatoire 58
CH-2000 Neuchâtel, Switzerland

Abstract

This paper describes the development of a novel miniaturized Rubidium clock for space applications based upon a magnetron cavity.
Two spaceborne Rubidium clocks, issue from the present development will constitute the main time reference for the Russian RADIOASTRON I space VLBI satellite and will be used for calibration purpose and as back-up local oscillator in case of failure of the main local oscillator system (which is based upon a quartz oscillator locked to ground H-masers via appropriate microwave links). In such a case the Rb clock will be used for observation of strong radio sources.
For this particular application the short time stability and temperature sensitivity are very important.
The measured short time stability is $\sigma_y(\tau) = 4 \bullet 10^{-12} \bullet \tau^{-1/2}$ *and the temperature sensitivity is* $1{,}6 \bullet 10^{-10}$ *for -20°C to 65°C operating temperature range. Other specifications like g-sensitivity and magnetic sensitivity which could be important for future missions have been optimized and tested. These performances have been obtained by a careful trade-off between the light-shift and the buffer gas shift in order to minimize the temperature coefficient of the physics package.*
Based upon small size physics package, this unit requires only 8W of power at room temperature including the spacecraft interface and a fully insulated power supply. The overall volume including the spacecraft interface is 1 liter.
A similar design approach was applied in the realisation of the industrial version which has equivalent characteristics but with an overall volume of only 0.25 litres.

INTRODUCTION

Since 1988 the Observatory of Neuchâtel has performed studies and developed breadboards of miniaturised Rubidium physics packages for ground applications.
The principle of the physics package based upon a magnetron type resonator was already described at EFTF 1992. /1/
In February 1992, we started to redesign this industrial physics package in order to reduce significantly the material and production costs. At the same time the development of our associated miniaturized industrial type electronics was started at the Observatory.
The development of a space compatible Rubidium standard started in October 1992 in the framework of ESA's In Orbit Technology Demonstration Programme (TDP).
This paper presents the development activities in the field of lamp pumped Rubidium Ultra Stable Oscillator (RUSO) during last two years at the Observatory of Neuchâtel starting with a description of the industrial version.

1. Industrial miniaturized Rubidium oscillator

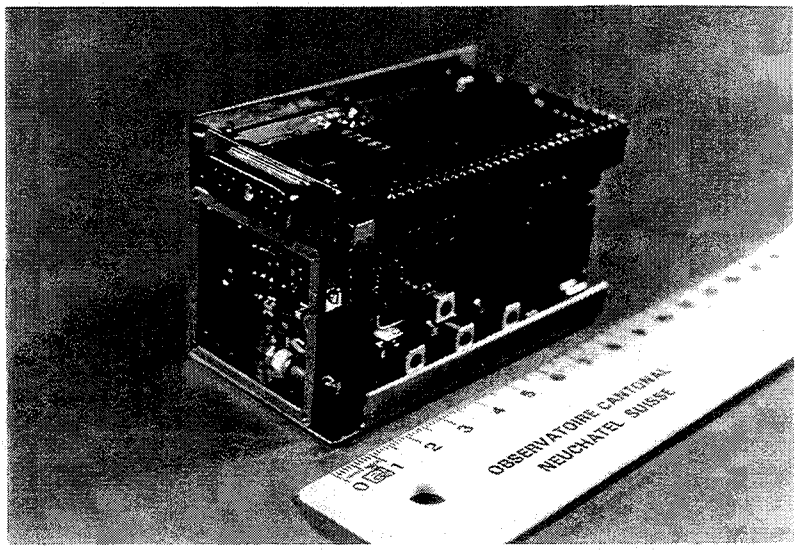


Fig. 1

The principal characteristics of this unit are its very small volume (0.25 l) and the built-in data interface usable for monitoring and control.

The typical performances of this Rubidium standard are

- 3 to 4 x 10^{-10} over - 20 to 65°C op. temp. range
- 10 MHz output, adjustable by step of 2x 10^{-10} with internal synthesizer
- fine adjustment by analog or RS232 command in step of 1x 10^{-11}
- <1x $10^{-11}\tau^{-1/2}$ short term stability
- 15W (25W optional) power consumption during warm-up
- <8W@25°C steady state power consumption
- < 5 min. warm-time to reach a frequency < 5 x 10^{-10}
- working parameters stored in EEPROM memory

Intensive tests are presently performed on 10 prototype units, starting with long term stability measurements in order to define the final specifications of the final product.
Fig. 2 presents the general block diagram of the industrial unit showing the physics package + interface assembly. The two principal interface functions to the physics package are the lamp exciter RF oscillator and the Step-Recovery driver circuit. These two critical circuits are directly mechanically attached to the physics package for stable impedance matching.

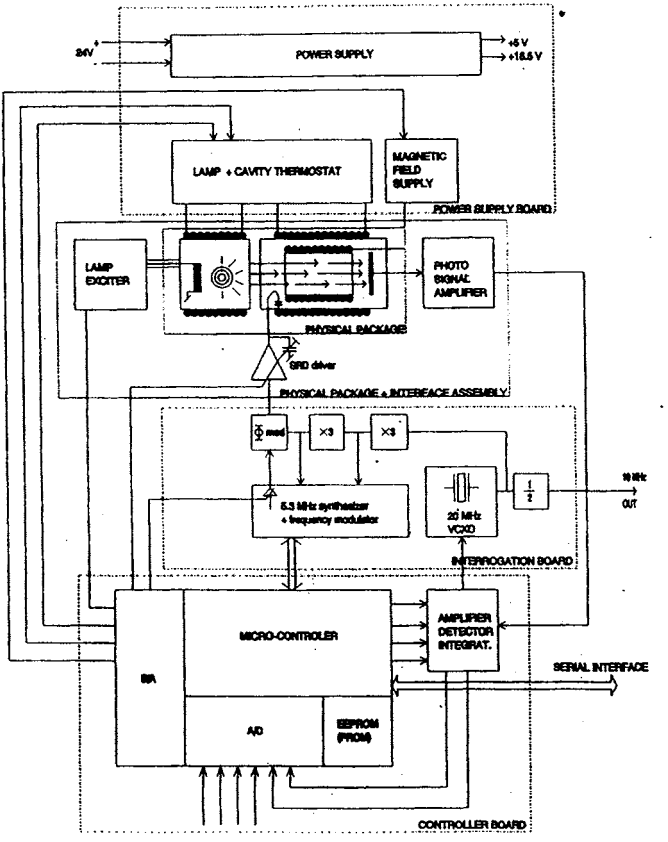


Fig. 2: Industrial Rubidium block diagram

The electronics package is implemented on 3 different boards,

- The power supply board includes the lamp and the cell oven control circuits and the necessary sub-voltages generation.
- The controller board contains a microcomputer unit which is associated with multichannel D/A converters and the servo loop circuits.
- The interrogation board which generates the microwave signal corresponding to the hyperfine rubidium atom transition.

1.1 Electronics package description

The interrogation signal generation starts with a fundamental mode 20 MHz crystal multiplied by 9 in order to produce 180 MHz which is phase modulated by a 5.3 MHz synthesizer. The output frequency is directly derived from the 20 MHz VCXO.
The 10 MHz center frequency adjustment is made by changing the 5.3 MHz synthesizer frequency. The step size of frequency change is normalized to
2×10^{-10}/step at 10 MHz level.
The controller board contains a microcomputer unit which is associated with multichannels D/A converters. The microcomputer also controls the interrogation and synchronous detection sequence as well as Rb absorption "Dip" search. The multichannel D/A converter is used for parameters setting of the instrument .
The built-in serial interface allows an automatic parameter adjustment procedure during the manufacturing process as well as a coarse and fine adjustment of the center frequency. All the working parameters can be stored in a built EPROM memory. In addition an 8 channels A/D converter is used for monitoring of principal internal signals like: light level , signal level , thermostats voltages, RF power etc....
This feature was intensively used for the characterisation and tests of the space physics package also.
The power supply board includes the lamp and the cell oven control circuits and the necessary sub voltages generation.

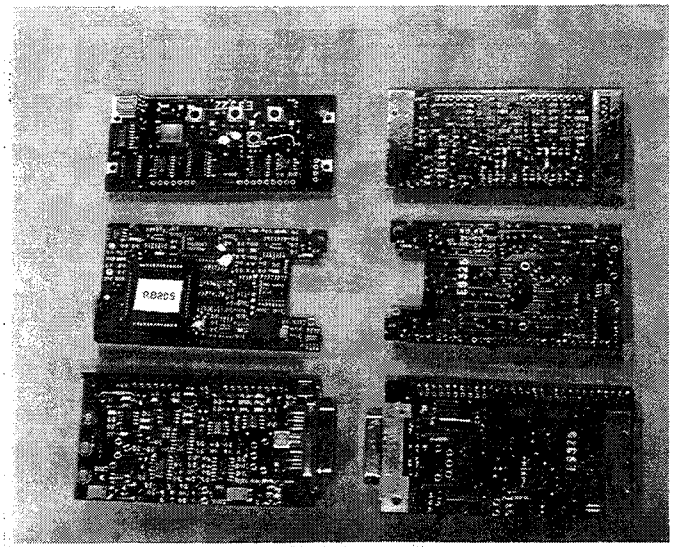


Fig. 3: Industrial RUSO electronics package

1.2 Physics parameters optimisation

To find the optimal working point for the Rb lamp and absorption cell three main parameters have to be adjusted in an iterative manner: the lamp temperature, the cell temperature and the oscillator power.

It is well known, that the light-shift depends in a non-linear way on these three parameters because they all affect the light characteristics (intensity and spectrum). Furthermore, the pressure shift and inhomogeneity effects taking place in the absorption cell also have to be taken into account in order to understand the observed dependence of the clock frequency on these critical parameters. /1/, /2/.

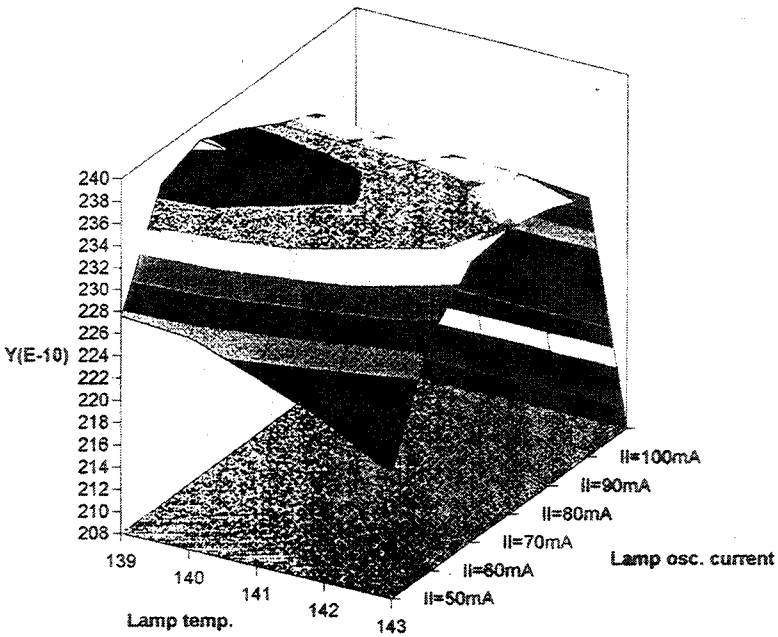


Fig. 4: Relative frequency offset versus lamp RF oscillator current

Figure 5 is a sample plot of the frequency shift due to lamp oscillator current and cell temperature. For a given cell temperature, the behaviour corresponds to the lamp characteristic. The parabolic behaviour is a combined effect of light-shift connected to the light absorption and N_2 buffer gas pressure shift.

The optimal operating point in this particular case corresponds to about 70 mA lamp oscillator current and 83.5°C oven cell temperature.
This physical parameters optimisation can be made in an automatic manner by using the previously described electronics package. This provides a very good flexibility and a cost effective solution for the adjustment of the critical physics parameters.

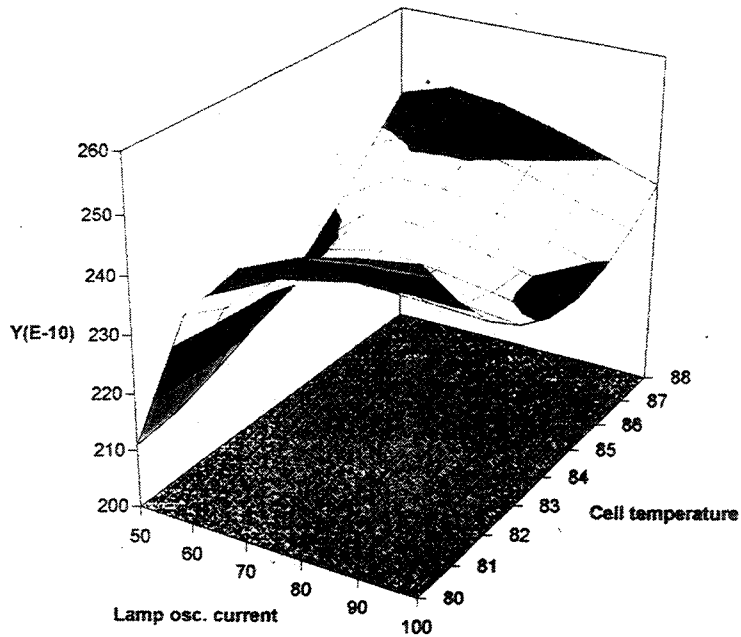


Fig. 5: Frequency versus Cell temperature for various light intensities

1.3 Typical test results

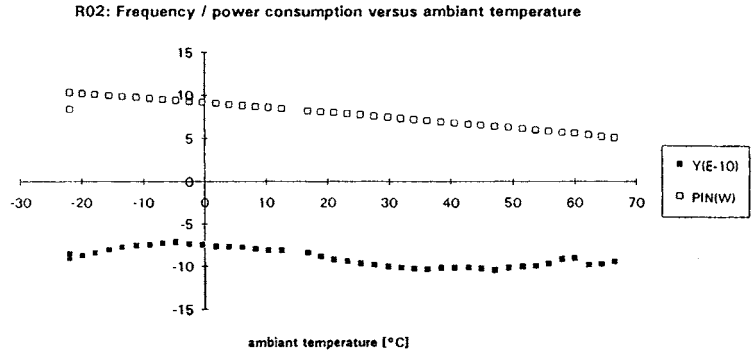


Fig. 6: Frequency versus ambient temperature characteristic

Fig. 6 gives a typical frequency behaviour (Y=3E-10) and power consumption (PIN) characteristics (7.5 W at 25°C) versus ambient temperature.
Fig. 7 shows a typical warm-up characteristic with 25W input power. We reach $3x10^{-10}$ frequency stability after 4 min. of operation.

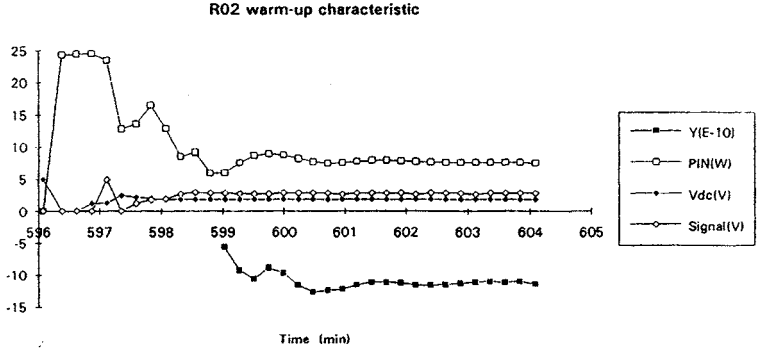


Fig. 7: Warm-up characteristic

Note : for the final industrial version the input power will be limited to 15W. In this case the warm-up time will be 5 to 6 min. instead of 4 min.

Y [E-10]	: relative frequency in 10^{-10} units
PIN	: input power in W
VDC	: light level intensity in V
Signal	: 2nd harmonic signal level

Long term frequency behaviour

Several industrial units were used for long term tests as well as previously developed 19 inch format prototype units.
The picture below shows the ON long term measurement test-set.

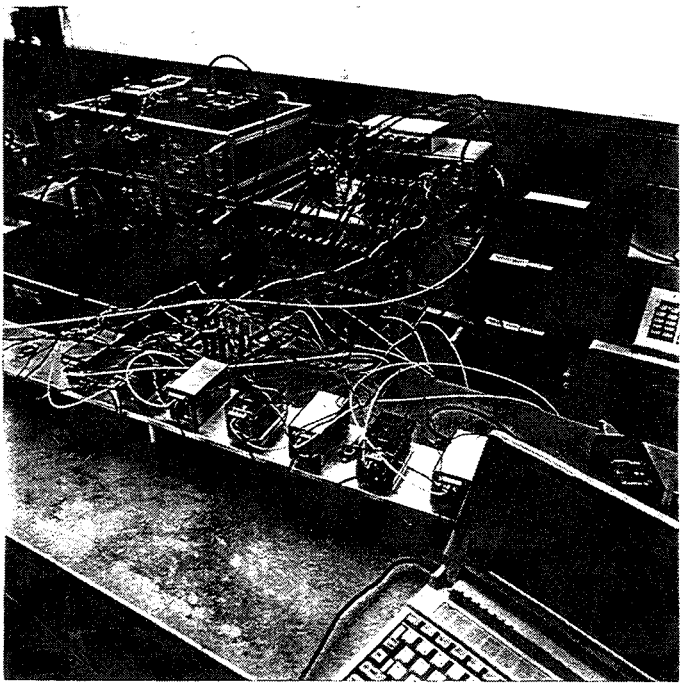


Fig. 8: Long term measurement test set

Up to now we built and measured three breadboards in 19 inch format (These breadboards are representative as far as the cell and the lamp are concerned.), 7 industrial 0.25 l units, and one engineering model for space application.

Long term test summary:

This summary list the different long term frequency behaviour for the various units built at ON up to the end 1993. The drift is expressed in 1e-11/month unit, associated with the number of days of operation , from the first ignition.

		Unit:	drift (1e-11/month)	measured after xxx days of op.
Breadboards:		Rb4	+1.07	700
		Rb6	-0.4	590
		Rb8	3.35	570
Industrial 0.25L	R01		+14	54
		R02	+9.36	46
		R03	not measured (used as test bench)	
		R05	+0.96	150
		R06	-2.6	150
		R07	+3.2	150
		R08	+0.6	150
		R09-11	under test	
Space engineering model	PEM		+12.8	60
			+5.8	150
			-0.19	260

Note : The long term measurement set-up has only 10 channels, for this reason. It was possible to measure the 15 working units continuously.

Conclusions:

The general trend is a positive drift, decreasing versus time ,associated with an increase of the light level transmitted through the absorption cell.

For a continuous operation, all the tested units reached the drift specification goal of 4×10^{-11}/month after at least 150 days of operation.

The future plan is to accelerate the preaging procedure consisting in a baking the critical elements (lamp and cell) in order to reduce this burn-in period.

The Figure 9 represents the long term frequency behaviour under normal lababoratory conditions of the industrial unit R08.

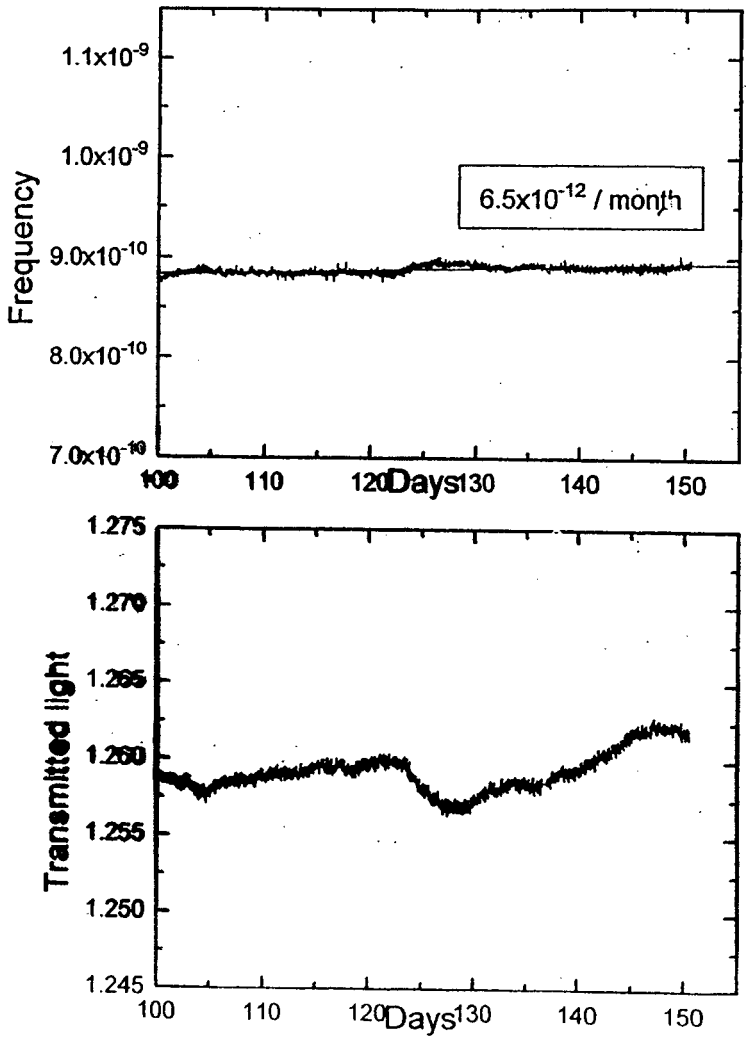
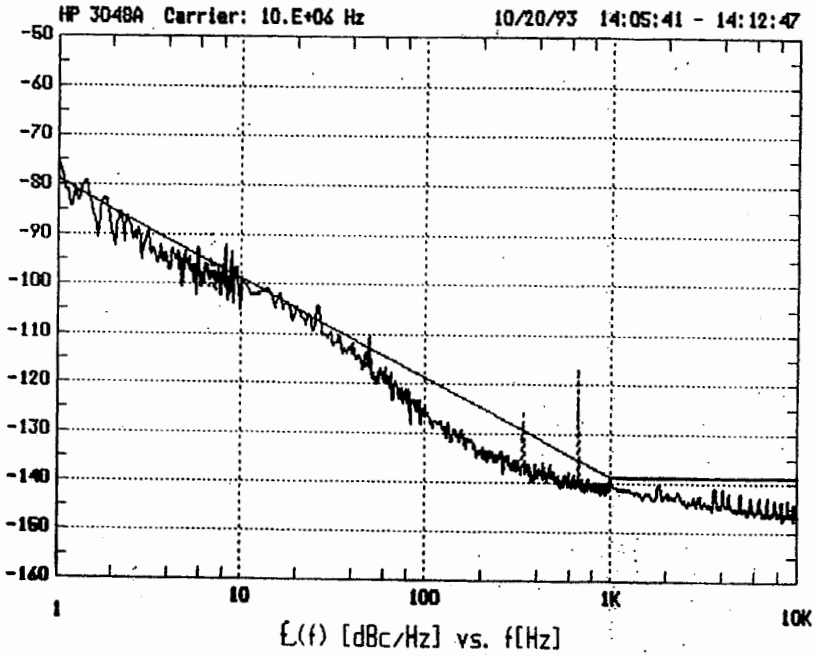


Fig. 9: Long term frequency record of industrial unit R08



RO2 Phase noise characteristic

2. Space RUSO

2.1 Physics package

The principle design characteristics of the physics package are its low power consumption, small size and mass, along with minimal environmental sensitivities and mechanical ruggedness.

A unique feature of the design is that all parts of the physics package are incorporated in a sealed titanium tube. Inside this tube, lamp and cell sections form two separate blocks which operate at well-defined but different temperatures. The cylindrical tube is filled with xenon for low thermal conductivity. Each block is connected to one of the end plates by a thin-walled tubular titanium spacer. This configuration greatly reduces the thermal flow between the blocks and to the tube envelope. It allows a very compact design with low power consumption, short warm-up time and minimal environmental sensitivities. Apart from the improved stability upon outside temperature variations, it is obvious that the sealed container makes the physics package virtually insensitive to barometric influences.

Other design features which contribute to the compact design are:

- use of the integrated filter technique (IFT)
- use of a magnetron-type microwave resonator

The integrated filter technique which combines the optical filtering and pumping in one cell contributes also to the reliability since the configuration is simplified and the number of components reduced. The thermal capacitance of the cell assembly is relatively low. As a consequence the necessary power during warm-up is greatly reduced.

The magnetron resonator is a cylindrical cavity loaded with a concentric capacitive-inductive structure (annular metal electrodes). It allows smaller cavity dimensions and concentrates the microwave field at the right region of the cell.

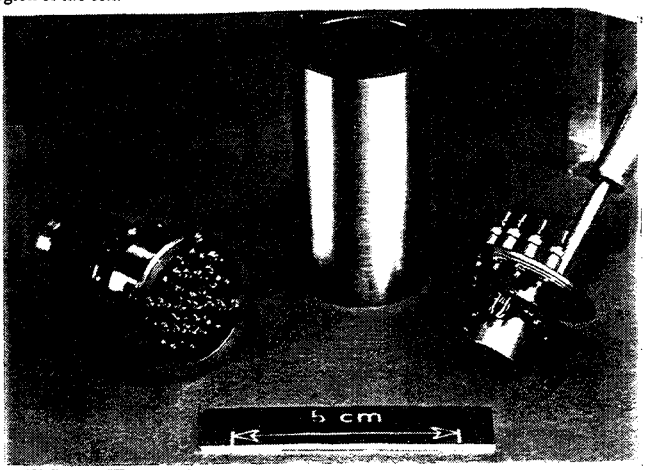


Fig. 10 and 11 show the physics package tube before and after assembly. The attached copper tube for pumping and filling the tube is pinched off in the final sealing process

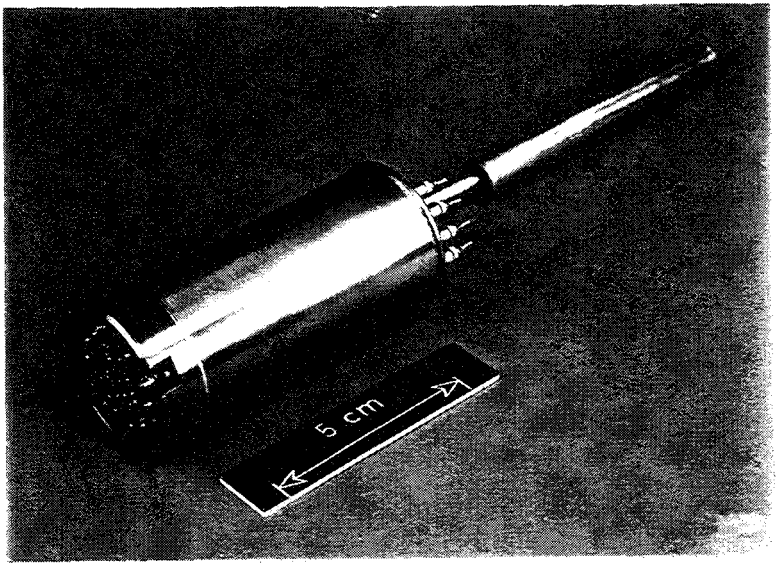


Fig. 11: Sealed titanium tube

Space RUSO physics package parameters:

SENSITIVITIES:

Light Shift:	$\approx \pm 5 \cdot 10^{-11} / \%$
Lamp TC:	$\approx 0 \pm 1 \cdot 10^{-10} / ^{\circ}C$
Cell TC:	$\approx 0 \pm 1 \cdot 10^{-10} / ^{\circ}C$
RF Power Shift:	$+1.5 \cdot 10^{-10} / dB$

The light shift coefficient of $\pm$ 5×10^{-11}/% of light can be optimized within this range as well as the lamp and cell temperature coefficients.

2.2 Space Electronics:

The second phase of the program consisted in the development of a space compatible electronics. The main differences to ground electronics are :

- the use of high REL components
- the use of a field programmable gate array for the control electrics synthesizer section
- a full galvanic insulation of the power supply
- a spacecraft interface for Radioastron.

Picture below shows the complete space compatible electronics package integrated onto a flex-rigid-flex PCB including DC/DC converters.

A block diagram of the spaceborne RUSO is given in Fig. 13

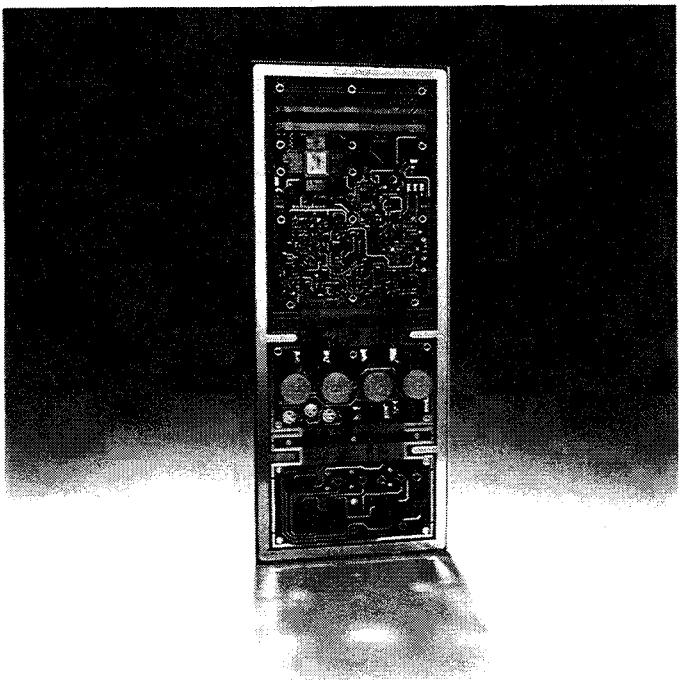


Fig. 12 Space Main Board

The following figures show the overall configuration of the unit in its housing.

SPACE RUSO SYSTEM BLOCK DIAGRAM

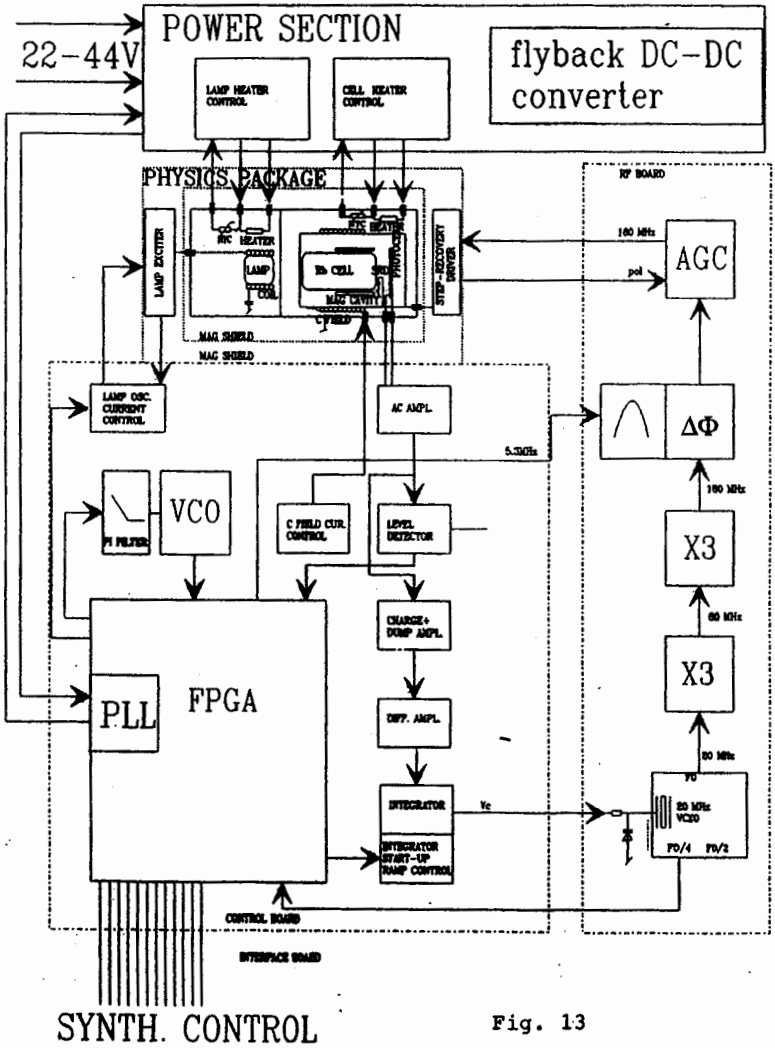
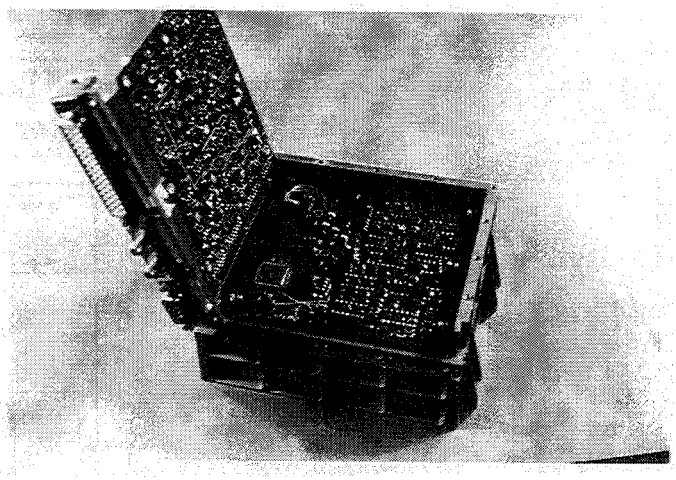
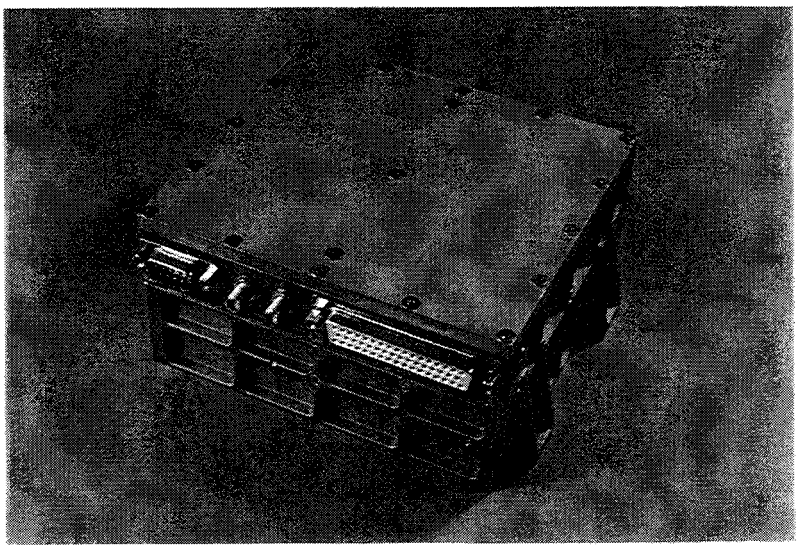


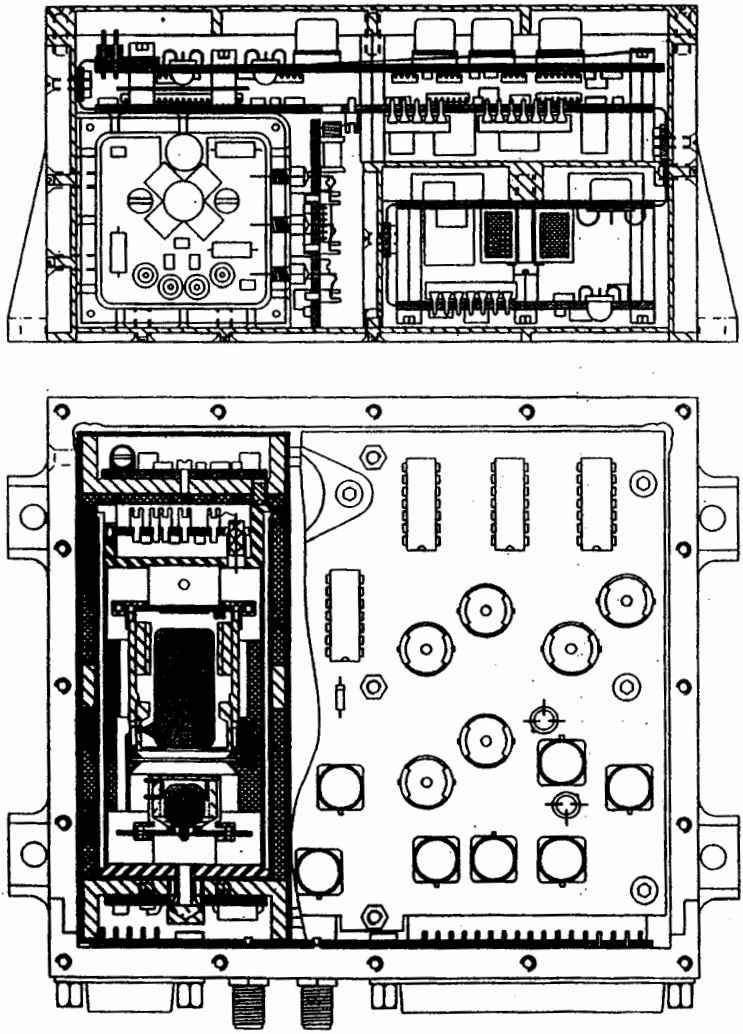
Fig. 13



Space RUSO showing the control board



Rubidium Ultra-Stable Oscillator for Space Applications



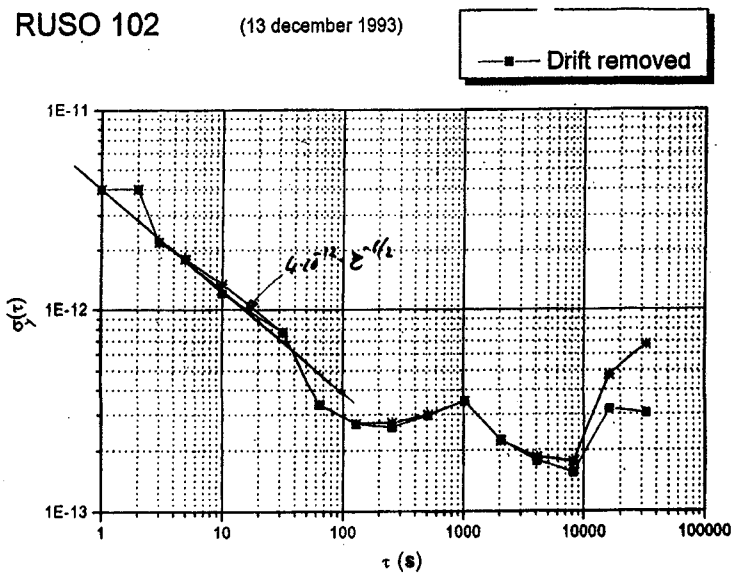
Space RUSO Top assembly drawing

2.3 Test results

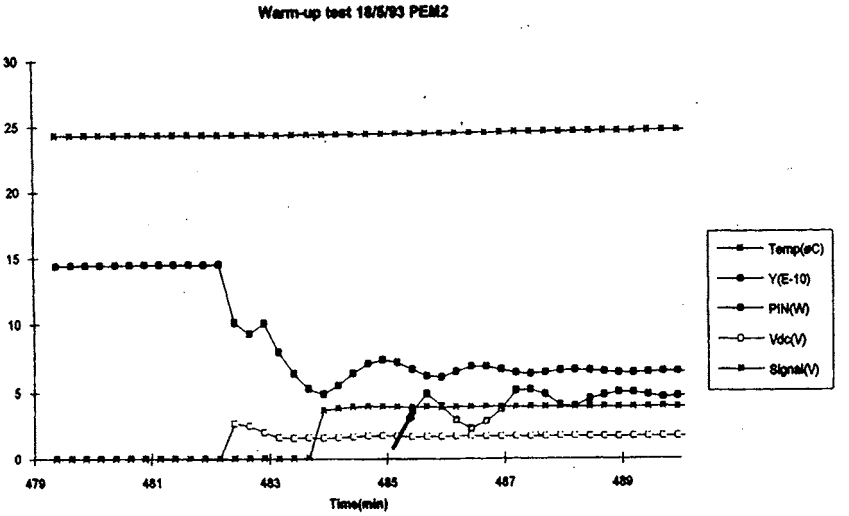
The signal parameters measured on the engineering model are the following :

Absorption "dip" amplitude:	0.8mA
Photo cell current:	144μA
Shot noise :	$8pA / \sqrt{Hz}$
Total noise:	$16pA / \sqrt{Hz}$
Line Q:	$3.6 \cdot 10^6$
Discriminator slope:	0.84nA / Hz
Predicted stability (with total noise and 50% dead time):	$2.8 \cdot 10^{-12} \cdot t^{-1/2}$
Measured stability: (Pre-Engineering Model)	$4 \cdot 10^{-12} \cdot t^{-1/2}$

Short term stability test of space RUSO engineering model:



Warm -up test:



Temp (°C) : ambient temperature
Y [E-10] : relative frequency offset in 10-10
PIN : output power in W
VDC : light level
Signal : 2nd harmonic signal level

The following table is a condensation of the specifications and test results measured on the PEM physics package associated with an industrial electronics package.

PARAMETER	ESA spec.	TEST **Results**
Long term frequency stability ($\Delta f/f_0$)	**< 4e-11/m**	**<1e-11/m**
Short term stability $\sigma(\tau^{-1/2})$ for $1 \leq \tau \leq 100s$	**<1e-11 $\tau^{-1/2}$**	**4e-12 $\tau^{-1/2}$**
Temperature stability ($\Delta f/f_0$) in - 25 to + 60°C range	**< 3e-10**	**1.6e-10**
Static g-sensitivity	**< 4e-12/g**	**3.5e-12/g**
Magnetic sensitivity	**< 2e-11/G**	**<1e-11/G**
SSB phase noise (L (f_m) (f_0 = 10 MHz) in dB		
f_m = 1 Hz	**-80**	**-84**
= 10 Hz	**-100**	**-104**
= 100 Hz	**-120**	**-124**
= 1.000 Hz	**-140**	**-142**
= 10.000 Hz	**-140**	**-145**
Warm-up time to reach stability < 5×10^{-10}	**<6 min**	**<3e-10 / 6min**
Retrace ($\Delta f/f$)	**<2e-11**	**<1.2e-11**
Power consumption during warm-up	**< 15W**	**14.5W**
Power consumption during normal operation	**< 8W**	**7.5W (25°C)**
Volume	**<1liter**	**1.2 liter**
Mass	**<1kg**	**1.3kg**
DC power supply voltage range	**22 to 43V**	**16 to 45V**

Acknowledgements:

This work was in part supported by ESA under The Agency's In-Orbit Technology Demonstration Program (TDP). We would like to thank ESTEC expects for their efficient continuing support during this on-going program as well as our colleges P. Thomann, M. Frosio, C. Couplet, M. Dürrenberger and P. Scherler and J.-C. Sapin.

References:

/1/ G. Mileti, I. Ruedi and H. Schweda, Proc. 6th EFTF, 515, (1992)
/2/ P. Willemin, G. Busca, J.T. Liu and P. Thomann, Proc. 5th EFTF, 18, (1991)
/3/ J. Rabian, P. Rochat, Proc. 2th EFTF, 461, (1988)

Time Scales (A 4)

Relativistic theory for picosecond time transfer in the vicinity of the Earth

G.Petit[1] and P.Wolf[1,2]

[1]Bureau International des Poids et Mesures,
Pavillon de Breteuil, 92312 Sèvres CEDEX, France
[2]Queen Mary and Westfield College, School of Mathematical Sciences,
Mile End Road, London E1 4NS, Great Britain

Summary. *The problem of light propagation is treated in a geocentric reference system with the goal of ensuring picosecond accuracy for time transfer techniques using electromagnetic signals in the vicinity of the Earth. We give an explicit formula for a one way time transfer, to be applied when the spatial coordinates of the time transfer stations are known in a geocentric reference system rotating with the Earth. This expression is extended, at the same accuracy level of one picosecond, to the special cases of two way and LASSO time transfers via geostationary satellites.*

1. Introduction

It is well known that in relativity the notion of simultaneity is not defined a priori so that a conventional choice of a definition has to be made. This choice will then lead to a corresponding definition of clock synchronization as synchronized clocks must simultaneously produce the same time markers. A widely used definition is that of coordinate simultaneity and corresponding coordinate synchronization, as given, for example, by Klioner /1/:

"Two events fixed in some reference system by the values of their coordinates (t_1,x_1,y_1,z_1) and (t_2,x_2,y_2,z_2) are considered to be simultaneous with respect to this reference system, if the values of time coordinate corresponding to them are equal: $t_1 = t_2$. In the following this definition of simultaneity (and corresponding definition of synchronization) we shall call coordinate simultaneity (and coordinate synchronization)."

Clearly, the synchronization of two clocks by this definition is entirely dependent on the chosen reference system and is thus relative in nature, rather than absolute.

In practice, coordinate synchronization between two distant clocks can be achieved by the exchange of an electromagnetic signal. From the knowledge of the positions of the clocks at emission and reception of the signal in the reference system of synchronization and the laws of light propagation in the same reference system, the coordinate time elapsed during transmission T_t can be calculated.

For the construction and dissemination of international reference time scales, coordinate synchronization in a geocentric reference system is required. We choose the geocentric non-rotating reference system as defined by the resolution A4 of the IAU /2/, with asymptotically flat spatial coordinates, but with Terrestial Time TT being the coordinate time. TT is an ideal form of the International Atomic Time TAI, which is the basis of the measurement of time on the Earth. By its definition, TT differs from the coordinate time of the IAU by a constant rate.

The clocks that are to be synchronized are usually fixed on the Earth, and have their spatial positions given in a rotating reference frame. Using the metric equation of the non-rotating system and taking into account the displacement of the clocks (in the non-rotating system) resulting from the relative movement of the two systems during signal propagation, the transmission coordinate time T_t can be calculated.

Recently, the precision of clock synchronization between remote clocks on the surface of the earth has reached the sub-nanosecond level /3,4/ with further improvements expected in the near future. For these applications it seems sensible to develop the theory to the picosecond accuracy level. Recent theoretical studies in this field claim an accuracy of 0.1 nanosecond /1/, and in some cases the provided formulae are expressed in terms of path-integrals making them more difficult to use than explicit expressions /5,6/. In this article we provide explicit equations for synchronization in a geocentric system of two clocks that have their positions given in the rotating system. All terms that in the vicinity of the Earth (within a geocentric sphere of 200000 km radius) are greater than one picosecond are included. Outside this sphere terms due to the potential of the Moon may amount to more than 1 ps and need to be accounted for separately. We also present formulae (to the same accuracy) for the special cases of two way time transfer (section 3) and LASSO (LAser Synchronization from Stationary Orbit, section 4) time transfers via a geostationary satellite. Here a possible small

residual velocity of the satellite (< 1 m/s) results in further terms contributing some tens of picoseconds for two way- and LASSO time transfers.

We will assume that all clocks are rate corrected for the gravitational potential at their positions, and their velocities in the reference system of synchronization, and hence run at the rate of TT.

2. Formula for a one way transfer

We consider a rotating frame $(t,\bar{x}_r)$ which rotates at a constant angular velocity ω with respect to a fixed star oriented one $(t,\bar{x})$ with two clocks a and b, at $\bar{x}_{ra}$ and $\bar{x}_{rb}$ at time t_0 when the two frames coincide. The two clocks are to be coordinate synchronized by the transmission of an electromagnetic signal from a (emission at t_0) to b (reception at t_1). To this end the coordinate time interval $T_t = t_1 - t_0$ elapsed between emission and reception of the signal needs to be calculated.

The metric of a geocentric non-rotating system in the first post-Newtonian approximation with TT as coordinate time and asymptotically flat spatial coordinates (for $r \rightarrow \infty$ the components of the spatial metric $g_{ij} = \delta_{ij}$) is:

$$ds^2 = -(1 - 2U/c^2)(1 + L_g)^2c^2dt^2 + (1 + 2U/c^2)(dr^2 + r^2d\theta^2 + r^2\sin^2\theta d\phi^2) \qquad (1)$$

where: ds is the relativistic line element.
t is the coordinate time TT.
r, θ (colatitude), ϕ (longitude) are spherical coordinates in a non-rotating geocentric coordinate system.
U is the gravitational potential of the Earth (positive sign).
$L_g = 6.969291 \times 10^{-10}$.

The scaling factor $(1 + L_g)$ results from the choice of TT as coordinate time. L_g is equal to U_g/c^2, where U_g is the value of the gravitational potential on the geoid including the centrifugal potential due to the rotation of the Earth.

Introduction of post-post-Newtonian terms and terms due to the tidal potentials of the moon, sun and planets into the metric leads to a correction to the propagation time of a light signal in the vicinity of the Earth of less than one picosecond. Hence (1) is sufficient for our purposes.

Transforming to the rotating frame by $d\phi = \omega dt + d\phi_r$, then setting $ds^2 = 0$ for a light signal and solving the resulting quadratic for dt provides an expression for the transmission coordinate time T_t:

$$T_t = \int \{du/c - U_g du/c^3 + \omega r^2 \sin^2\theta d\phi_r/c^2 + [1+r^2\sin^2\theta(d\phi_r/du)^2]\omega^2 r^2 \sin^2\theta du/2c^3 + 2Udu/c^3\} + O(c^{-4}),$$

where du is the increment of coordinate length along the transmission path and the integral is to be taken from a to b along the transmission path in the rotating frame.

Evaluating the above expression (for a detailed derivation see /7/) gives an explicit formula for the transmission time:

$$T_t = R_0/c + \delta$$

$$= R_0/c - U_g R_0/c^3 + \bar{R}_0.\bar{v}_b/c^2 + (v_b^2 + \bar{R}_0.\bar{a}_b + (\bar{R}_0.\bar{v}_b)^2/R_0^2)R_0/2c^3 + 2GM_E \ln\{[x_{rb} + \bar{n}.\bar{x}_{rb}]/[x_{ra} + \bar{n}.\bar{x}_{ra}]\}/c^3 \quad (2)$$

where:

δ is the total relativistic correction.

$\bar{R}_0 = \bar{x}_{rb} - \bar{x}_{ra}$

$\bar{v}_b = \bar{\omega} \times \bar{x}_{rb} + \bar{v}_{rb}$

$\bar{a}_b = \bar{\omega} \times (\bar{\omega} \times \bar{x}_{rb})) + \bar{\omega} \times \bar{v}_{rb} + \bar{a}_{rb}$

$\bar{n} = \bar{R}_0/R_0$ is the unit vector along the transmission path

$\bar{v}_{rb}$ is the satellite velocity in the rotating frame

$\bar{a}_{rb}$ is the satellite acceleration in the rotating frame

and the two frames coincide at $t = t_0$.

The above expression provides the coordinate transmission time for a light signal travelling from station a to station b in the vicinity of the Earth (within a geocentric sphere of 200000 km radius) with the coordinates of the two stations given in an Earth fixed rotating frame. All terms that are greater than one picosecond are included. Note however, that atmospheric delays which can amount to several tens of nanoseconds are not considered and need to be taken into account separately.

3. Two way time transfer

We consider a two way time transfer between two stations c and d, fixed on the surface of the Earth, via a geostationary satellite s (as shown in Fig. 1). Two signals are transmitted in opposite directions leaving c and d at t_0 and $t_0+\Delta t$ respectively. They reach the satellite at t_1 and t_3, where they are immediately retransmitted, and arrive at the opposite stations at t_2 and t_4. We shall assume that the clocks have been synchronized previously to within 0.1 s, a typical station to satellite transmission time (which can be achieved without difficulty in practice), and that the satellite has a residual velocity v_r smaller than 1 m/s and a residual acceleration in the rotating frame of less than 10^{-5} m/s^2. These values have been chosen as typical but the derived formulae could accomodate values larger by two orders of magnitude while maintaining picosecond accuracy.

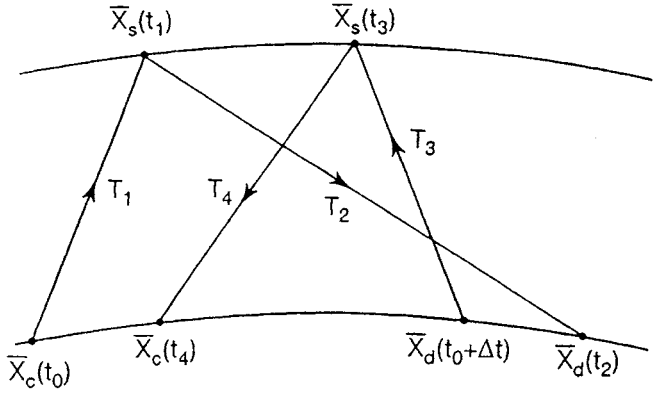


Fig. 1. Two way time transfer in the non-rotating frame.

From the clocks two coordinate time intervals are obtained (assuming that the clocks are rate corrected as mentioned in section 1):

$$t_c = t_4 - t_0$$
$$t_d = t_2 - t_0 - \Delta t$$

For synchronization the interval Δt is required. The derivation, details of which can be found in /7/, yields:

$$\Delta t = (t_c - t_d)/2 + \delta$$

$$\delta = \{\bar{R}_{cd}.(\bar{\omega} \times \bar{x}_{rs}) + [(R_{cs} - R_{ds} - c\Delta t)(R_{ds}\bar{R}_{cs} + R_{cs}\bar{R}_{ds}).\bar{v}_r] /(2R_{cs}R_{ds})\}/c^2 + O((v/c)(v_r/c)\Delta t) \quad (3)$$

where:
$$R_{cs} = \bar{x}_{rs} - \bar{x}_{rc}$$
$$R_{ds} = \bar{x}_{rs} - \bar{x}_{rd}$$
$$R_{cd} = \bar{x}_{rd} - \bar{x}_{rc}$$

The satellite position is given at $t = t_0$. The relativistic correction δ arises from the motion of the stations and the satellite in the frame of synchronization and the gravitational delays for the individual transmissions. The first term in (3) is equivalent to $2\omega A_E/c^2$ with A_E being the area of the equatorial projection of the quadrangle whose vertices are the centre of the Earth and the positions of the satellite and the stations in the rotating frame.

Table 1: Value of the second term of formula (3) for the relativistic correction of two way time transfers. The last column gives the values of Δt for which this term vanishes.

Transfer between	Second term of (3)(ps)		Δt(ideal) (ms)
	Δt = 0 s	Δt = 50 ms	
Toulouse(France) - Paris(France) via INMARSAT2 at 15.5°W	-5	-172	-1,57
FTZ(Germany) - PTB(Germany) via Kopernikus2 at 28.5°E	-2	-169	-0,57
Kourou(F.Guyana) -HBK(S-Africa) via INMARSAT2 at 15.5°W	-11	-177	-3,28
Graz(Austria) - USNO(USA) via INTELSAT at 53°W	35	-131	10,47

The second term in (3) varies with v_r and Δt, and can amount to several hundred picoseconds. If $\Delta t \sim 0$, it can amount to several tens of picoseconds, depending on the residual velocity which is in general not well known. However, one can compensate for it by intentionally introducing a desynchronization in order to drive this term towards zero, which is the case when the two signals arrive at S at about the same time (ie. $t_1 \approx t_3$).

This is illustrated in table 1 which shows the value of this term for several time transfers, for a residual velocity of 1 m/s. Repeating the time transfer at different times with different residual velocities will introduce apparent discrepancies unless this term is taken into account, or the transmission of signals is intentionally desynchronized, by setting Δt to the value Δt(ideal) of table 1 to within 50 microseconds.

4. LASSO

In this method laser pulses emitted from the stations c and d at t_0 and $t_0+\Delta t$ respectively are reflected by the geostationary satellite at t_1 and t_3, and return to the stations at t_2 and t_4 respectively (as shown in fig. 2).

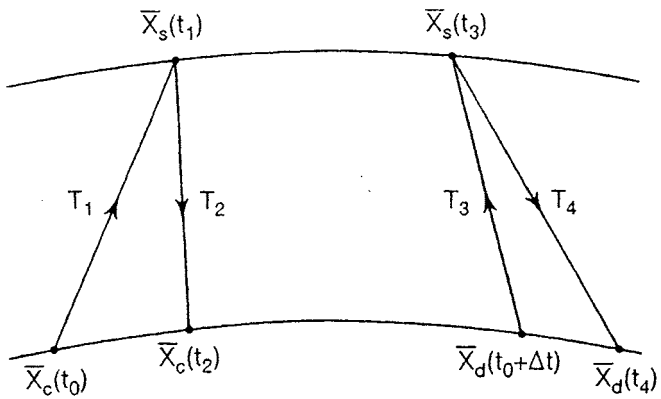


Fig. 2. LASSO time transfer in the non-rotating frame.

The satellite is equipped with a clock which measures the time interval between arrival of the signals. Hence three coordinate time intervals (after rate correction of the clocks) are obtained:

$$t_c = t_2 - t_0$$
$$t_d = t_4 - t_0 - \Delta t$$
$$t_s = t_3 - t_1$$

Similarly to the two way case the derivation of Δt, detailed in /7/, yields:

$$\Delta t = (t_c - t_d)/2 + t_s + \delta$$

$$\delta = [\vec{R}_{cd}.(\vec{\omega} \times \vec{x}_{rs}) + \Delta t(\vec{\omega} \times \vec{v}_r).\vec{x}_{rd}]/c^2 + O((v/c)(vr/c)(R_0/c)) \quad (4)$$

As in (3), the satellite position is given at $t=t_0$ and the first term is equivalent to $2\omega A_E/c^2$. The second term varies with $\vec{v}_r$ and Δt. This term is smaller than 10^{-2} ps for $v_r \sim 1$ m/s and $\Delta t \sim 0.1$ s, which is the case for a two way transfer and hence it does not appear in (3). However, for LASSO Δt can amount to several minutes in practice /4/, and therefore the second term in (4) can contribute up to 10 ps. Note also that while the second term of (3) can be minimised by an appropriate choice of Δt, this is not the case in (4).

5. Constraints for practical applications

For picosecond accuracy, the relativistic correction δ contains terms in c^{-2} and in c^{-3} in the case of one-way time transfers (2), and terms in c^{-2} only in the case of two-way (3) and LASSO (4) transfers.

The term in c^{-2} can amount to a few hundred nanoseconds, depending on the relative positions of the transmission and reception points. For example, between the Earth and a geostationary orbit, the maximum value is about 200 ns for the one way- and 400 ns for the two way case. In order to compute this term with picosecond accuracy, it is sufficient for all quantities in the term in c^{-2} to be known with a relative uncertainty of one or two parts in 10^6. This requires coordinates known to within 6-12 m for the Earth stations, including uncertainties in the realization of the reference frame which are below ~1 m for e.g. WGS84 and ITRF. This is generally the case for time laboratories. The satellite position should be known to within some tens of metres, depending on its orbit, and this is generally

not the case a priori for a satellite without geodesic objectives. In addition the velocity of the satellite should be known to the same relative uncertainty of one or two parts in 10^6, which is also not the case in general. Typically the position of a geostationary satellite is known to an accuracy of ~1 km which results in an error in the computation of the c^{-2} term of ~10 ps. Similar arguments can be made to set constraints in the case of higher orbits or satellite to satellite time transfers.

In the real case of a non-perfect geostationary orbit, the constraint on the knowledge of the velocity of the satellite is transferred to the residual velocity v_r. For the one way and two way techniques, this constraint is about 1 cm/s for picosecond accuracy but in the two way technique it can be completely relaxed by an intentional desynchronization of the emission of the signals at the two stations, as mentioned in section 3. For LASSO, the constraint on v_r is about 10 cm/s if one wishes to use laser pulses from the two stations separated by Δt of several minutes. The constraint on v_r can be relaxed by decreasing Δt.

When one of the stations is on the Earth, propagation through the atmosphere is the major problem for one way time transfer. It leads to delays that can reach several tens of nanoseconds and can certainly not be calibrated to picosecond accuracy. This problem is not considered in this study. However the effects cancel to the picosecond level in the two way (provided the up and down frequencies are close enough) and LASSO techniques.

6. Conclusion

We have derived the relativistic formula for a one way time transfer between two stations that have their position given in a geocentric reference frame rotating with the Earth including all terms in c^{-3} and larger. For time transfer with a geostationary satellite the terms in c^{-3} can amount to around 10 ps for the Sagnac correction and 80 ps for the gravitational delay. At present, one way time transfers are not accurate enough to necessitate the consideration of these terms. However, with accuracy expected to increase in the near future, and in view of possible satellite to satellite transfers (which would eliminate uncertainties due to atmospheric delays) these terms might well become significant.

We also provided expressions for the relativistic treatment of two way and LASSO time transfer techniques. We have shown that the main errors in computing these expressions are due to the uncertainties in the position and the residual velocity of the satellite. The uncertainty in the position leads to an error in the

computation of $2\omega A_E/c^2$ of the order of 10 ps for both techniques. The uncertainty in the residual velocity affects the two techniques differently. For LASSO the second term in (4) is typically of the order of 10 ps, hence reducing the overall uncertainty for LASSO requires better knowledge of the satellite position as well as consideration of the additional term. For two way time transfers, on the other hand, the second term in (3) can reach 80 ps (for $\Delta t = 0$). Hence reducing this term by an appropriate choice of Δt will improve the overall accuracy of the two way time transfer even in the case where $\bar{v}_r$ is unknown.

In both techniques, the precision of experiments repeated over periods of several weeks could be affected by the variation of the residual velocity of the satellite, if the corresponding terms are not accounted for.

This shows that the time community is rapidly approaching levels of precision and accuracy that will necessitate a more exact development of the theory. We consider the present paper a step in that direction.

Acknowledgements. The authors are grateful to Dr. Claudine Thomas and Prof. Bernard Guinot for helpful discussions and support.

References

/1/ Klioner S.A., 1992, Celestial Mechanics and Dynamical Astronomy 53, p.81

/2/ IAU, 1992, Information Bulletin 67, p. 7

/3/ Hetzel P., Soring A., 1993, Proceedings 7th European Frequency and Time Forum Neuchâtel 1993, Swiss Foundation for Research in Microtechnology, p.125

/4/ Veillet C., Fridelance P., 1993, Proceedings 7th European Frequency and Time Forum Neuchâtel 1993, Swiss Foundation for Research in Microtechnology, p.637

/5/ CCDS, 1980, Comité Consultatif pour la Définition de la Seconde, 9th Session

/6/ CCIR, 1990, International Radio Consultative Committee, Reports, Annex to Volume VII, p.150

/7/ Petit G., Wolf P., 1994, accepted for publication, Astronomy and Astrophysics

PRELIMINARY EVALUATION OF TIME SCALES BASED ON HYDROGEN MASERS

Marc A. Weiss and Fred Walls
NIST Time and Frequency Division, 847
325 Broadway, Boulder, CO 80303, USA

Abstract
Two experimental time scales based on ensembles of hydrogen masers were generated and compared with a variety of other references both internal and external to NIST. The masers all had some type of active cavity control to reduce frequency drift due to cavity changes. The first experimental time scale, TA2M, was generated over the interval from MJD 49050 to 49190 (4 March 1993-22 July 1993) using measurements between the masers every 6 hours. TA2M was compared to the NIST AT1 time scale, primarily based on commercial cesium frequency standards, and via once-per-day GPS common-view measurements to the USNO unsteered master clock and to UTC. The drift of TA2M was small compared to the uncertainty of $2 \cdot 10^{-16}$/d in estimating linear frequency drift.

The masers at NIST were then linked to two masers at USNO using GPS common-view time transfer to study the performance of a maser ensemble, TA2M1, at measurement times from a few days to a few months. Again we found that the drift of TA2M1 was smaller than the estimation uncertainty of $0.6 \cdot 10^{-16}$/d. The stability of the ensemble at 1 month appeared to be about 1-2 parts in 10^{15}.

From this work we see that the use of auto-tuned hydrogen masers in time scales is very promising. A single such maser rivals the 1993 stability of UTC. The typical fractional frequency stability of the masers showed flicker frequency modulation at $4 \cdot 10^{-15}$ or less for measurement times of 10 d-20 d. Linear frequency drift was measured for the masers against the TA2M1 time scale at levels under $1 \cdot 10^{-16}$/d with uncertainties of $0.4 \cdot 10^{-16}$/d. Measuring frequency drift against the definition of the SI second at this level pushes the limits of current technology.

Introduction
Active hydrogen masers (H-masers) with cavity tuning now maintain fractional frequency stability with little drift. In the past, active H-masers were not useful for time scale generation because of linear drifts of fractional frequency of the order of 10^{-14} to 10^{-15}/d. Frequency drift in itself does not perturb a time scale, but rather the error in estimation of this deterministic effect [1]. The problem

Contribution of the US Government, not subject to copyright.

with estimation is basically that, due to the presence of flicker and random walk frequency modulation in the reference cesium standards, in the time it takes to attain the needed precision, the true value for the drift will have changed. The fractional frequency stability of several devices is shown in Figure 1. The stability of the masers shown is without any estimation of frequency drift. Clearly, the drift must be estimated very precisely for these devices to compete with the best cesium standards in long term.

A technique for interrogating the resonant cavity to measure its electrical size in real time was first developed at NIST in 1976 [2]. Our studies confirmed that the change in the cavity size was the major factor causing frequency drift in masers [3]. Somewhat different techniques are now used in a number of active H-masers. Frequency drift is now controlled in such masers to a level that is difficult to measure. We will see here that measuring the drift of these devices requires integrating over about 1 year against some of the world's best current cesium frequency standards.

We have combined a number of these devices at NIST and at USNO into experimental time scales to illustrate their potential. We find that the stability of a single maser is greater for measurement times out to several months, than most national time scales based on cesium beam frequency standards. Further, the drift at one year is competitive with some of the best ensembles in the world. This experiment was designed to document this performance and to better characterize the masers. H-masers with cavity autotuning can never be a primary reference for producing accurate frequency, due to limits in evaluating the wall shift of $5 \cdot 10^{-13}$ at best [4]. However, they can maintain a frequency once they have been evaluated.

Short-Term Stability

The short-term frequency stability of the masers at NIST were measured using a dual mixer system [5]. The noise floor of the standard system is far above that of active masers for measurement times below about 1000 s. To circumvent this problem, data between three masers were taken on two channels each. Cross correlation was then used to reduce the contribution of both the reference masers and the measurement system to the estimation of the Allan variance for the maser under test [6]. In this way the measurement noise was reduced to approximately 1 part in 10^{14} for measurement times from 0.5 to 20 s. For measurement times from 10 s to 10^5 s, we also used the modified Allan variance to estimate frequency stability. The fractional frequency stability of three masers at NIST, CH-A, CH-B, and H1, operating in a cavity autotuned mode are shown in Figure 2. The stabilities estimated using these improved techniques are plotted for the three masers for integrations times from 1 s to 10,240 s. The data points for longer times are measurements against the TA2M scale as described below using the NIST dual mixer time scale system which records time differences between clocks every 2 hours.

Masers CH-A and CH-B are cavity tuned by using a magnetic field gradient to modulate the atomic linewidth, and hence to modulate frequency pulling due to cavity mistuning. In this mode the atomic line width of one maser is modulated and the cavity is frequency-adjusted under servo control until its oscillation frequency as measured against the other maser is independent of atomic line width. The reference and tuned masers are then switched so that the cavity frequency of both masers are autotuned. In principle this should remove the effect of cavity mistuning which is the major contribution to frequency drift in conventional masers [3]. Maser H1 uses a different scheme for cavity autotuning in which the cavity frequency is modulated. A servo system then tunes the cavity frequency until the oscillation frequency is independent of small mistunings. This approach does not require an external frequency reference.

Description of the Time Scales

We report here the generation of two experimental time scales. We used the TA2 time scale algorithm to generate both these scales. TA2 is a smoothing algorithm, in which a time scale is run both forwards and backwards in time. The results from the two directions are then combined to obtain a post-processed scale optimized for frequency stability [7,8]. The one-directional time scale algorithm used, AT2, is an extension of AT1, the algorithm used to generate the real-time unsteered time scale at NIST, with the addition of an estimate of confidence of the frequency estimate. This confidence estimate allows the forward and backward AT2 estimates of frequency to be combined into a smoothed estimate of frequency. A third pass in the forward direction uses this smoothed estimate of the frequency offset of each clock to estimate the time offsets.

The first scale we generated is called TA2M. It used measurements among masers at NIST over the interval from MJD 49050 to 49190, 4 March 1993 to 22 July 1993. Measurements taken every 6 hours among the three masers, CH-A, CH-B, and H1, discussed in the previous section at NIST were combined to generate TA2M. This experimental scale was compared to the NIST AT1 time scale primarily based on commercial cesium frequency standards, and via GPS to the USNO unsteered master clock and to the international UTC, as generated by the BIPM and published in Circular T.

Measurements taken once per day at NIST were linked to once-per-day measurements at USNO using the GPS common-view time transfer technique to generate the TA2M1 scale, with scale estimates once per day. The CH-A, CH-B, and H1 masers at NIST were then linked to the N4 and N5 masers at USNO to study the performance of a maser ensemble at longer measurement times. This ensemble was run from MJD 49050-49349, 4 March 1993 to 28 December 1993, though maser CH-A at NIST left after 49190, and maser CH-B was not cavity-tuned for most of this run. Thus, the long-term performance of the scale is due primarily to maser H1 at NIST and masers N4 and N5 at USNO. These

masers all use the same type of cavity autotuning mentioned above which does not require an external frequency reference.

Results from the TA2M Scale

Since there are only three clocks in the scale and they are all similar in long term, they are each weighted with ⅓ of the scale. There is a brief period where maser H1 is unavailable, because adjustments were being made. For this period, the scale becomes a simple average of the other two masers. There is no measurable shift in the scale during this period. Figure 3 shows the fractional frequency of the maser H1 versus TA2M, and the scale TA2M versus AT1 and USNO A1r on the same plot. Frequency smoothing and offsets were applied to separate them on the plot. We can see the period MJD 49136-49150 when the maser H1 is absent, and there is no apparent shift in TA2M against either AT1 or USNO A1r at these times.

Figure 4 shows the frequency of TA2M versus the international UTC. The drift of the maser scale was negligible compared to the GPS common-view noise and frequency stability of the scale over the period of the measurements. The confidence on an overall drift estimate due to these factors was $1.5 \cdot 10^{-16}$/d. The drift of the ensemble was small compared to this uncertainty.

The stability of the masers against the TA2M scale is shown in Figure 2, extending the short term measurements. The values against TA2M begin at measurement times of 6 h = 21,600 s. These stabilities have been adjusted for correlation with the scale by the following formula relating the true variance to the measured variance [9]:

$$\sigma^2_{true} = \frac{\sigma^2_{measured}}{1-w}, \qquad (1)$$

where w is the normalized weight of the clock in the ensemble. Since each clock had weight ⅓, the true variance is 1.5 times the measured variance. We see that the white FM level of the masers is about $7 \cdot 10^{-13} \tau^{-1/2}$, τ in s, reaching a flicker floor of about $3 \cdot 10^{-15}$ at about 1 d.

The drift of the three masers was measured against TA2M. The CH-B masers had a linear frequency drift value of $-0.9 \cdot 10^{-16}$/d with a confidence of $\pm 0.4 \cdot 10^{-16}$/d. The H1 and CH-A masers had negligible drift versus TA2M.

Results From the TA2M1 Scale

Figure 5 shows the fractional frequency of the masers H1 at Boulder and Figures 6 and 7 show the masers N4 and N5 at USNO versus TA2M1. Since the masers CH-A and CH-B at Boulder were removed from the scale at MJD 49190, the GPS common-view noise became absorbed into the scale after that. Hence the apparent noise of H1 went up and the GPS noise of masers N4 and N5 went down at this time. To minimize this effect we added another maser at NIST, H11, to the ensemble. Since this maser had no cavity autotuning we

modelled its drift and weighted it very low in long term. We obtained drift estimates by first including it in the scale with weight 0, allowing the scale to estimate its drift, then re-running the scale with H11 included and drift modelled. In this way, when the drift values were used for predicting H11 in the TA2 algorithm, TA2M1 was not steered to any other scale, as could have occurred if the drift was originally measured against another scale.

We again found here that the drift of the ensemble was too small to be measured against any available external frequency standard. The value of frequency drift for TA2M1-UTC over the 280 d interval MJD 49059-49339 for which data were available for the scales, was $-0.1 \cdot 10^{-16}$/d from a three-point second-difference estimator. If we assume the model of linear frequency drift in the presence of a flicker FM stochastic process at a level of $3 \cdot 10^{-15}$, the confidence on this drift estimate, comparing 150 d frequency averages is [10]

$$\frac{\sqrt{2}}{\tau}\sigma_y(\tau)|_{\tau=150d} = 0.3 \cdot 10^{-16}/d \qquad (2)$$

Thus the drift of TA2M1 against UTC was negligible.

Figure 8 shows the stability of the masers against the ensemble. Comparing to Figure 2 we see that the GPS time transfer noise dominates out to about 10 d. N4 was the only the maser with any apparent drift. It was estimated to be $1.0 \pm 0.3 \cdot 10^{-16}$/d using a three point overall second difference, with the confidence estimated as above. We want to point out that, although there is statistical significance to the magnitude of the N4 drift against UTC, we may not have a frequency standard accurate enough to determine whether the maser or UTC was drifting. The overall second difference estimator compares average frequencies over two successive 150 d intervals. To determine average frequency offset from an SI accurate frequency over 150 d at a level of $(3/\sqrt{2}) \cdot 10^{-16}$ we need a frequency standard accurate to about $3 \cdot 10^{-14}$. This approaches the limits of the best available standards today [11,12,13]. Thus we can only measure the change in frequency between the masers and UTC.

We were able to observe another significant effect using the TA2M1 scale. The maser CH-B lost maser CH-A as its reference for cavity tuning on 49190. We decided to use the H1 maser as a reference on MJD 49250, and the autotuning came back on for a few days, but unfortunately came unlocked again due to a power outage. Some time later, after moving masers to another room, we again relocked the CH-B autotuning to maser H1. Figure 9 shows the performance of maser CH-B against the TA2M1 scale over the entire 340 d interval. When cavity tuning was reactivated, the maser came back to a frequency very close to what we would have predicted by frequency linearly using the constant drift of $-0.9 \cdot 10^{-16}$/d from the period 49050-49190.

Conclusions

One implication of these results is that a time scale based on three active autotuned hydrogen masers would be significantly more stable for integration

times less than 1-2 months than our present NIST cesium ensemble. Such a maser ensemble would compete with the stability of UTC as it was in 1993. The stability of UTC will also change as it includes more new standards. In the very long term AT1, TA2M1, and UTC seem to flicker at a few parts in 10^{15}. In any case, UTC must be used for frequency accuracy and consistency with the rest of the world.

Another conclusion we can draw is that the cavity autotuning systems eliminate the major cause of frequency drift in active masers. There may be some residual drift due to other effects, but it is smaller than about 10^{-16}/d. It is particularly interesting that maser CH-B came back to the predicted frequency after 100 d without autotuning, after being moved to a new room outside of its environmental chamber, and being locked to a different frequency reference.

Further, to make good use of the stability of these devices, advances in time and frequency comparisons will be necessary. Here the GPS common-view time transfer noise dominated over the masers out to about 10 d. This measurement noise can be lessened by using precise ephemerides and ionospheric measurements [14,15], but over the approximately 1000 km baseline between NIST and USNO there will not be more than a factor of 2 or 3 improvement. Since UTC is generated based on 10 d measurements, we can see these new frequency standards, but barely. Being able to use such frequency standards reliably in UTC at 10 d will require better frequency comparison techniques.

Acknowledgements
The authors would like to particularly thank Dr. Gernot Winkler for supplying the hydrogen-maser and commercial primary cesium-beam clock data from the USNO. This data was invaluable to the results obtained in the paper. Several significant suggestions have been made by Dr. Judah Levine, Dr. Sam Stein, Dr. Donald Sullivan, and Dr. Matt Young.

References

[1] David W. Allan, "Time and Frequency (Time-Domain) Characterization, Estimation, and Prediction of Precision Clocks and Oscillators," *IEEE Trans. UFFC,* vol. 34, no. 6, pp. 647-654, Nov. 1987.

[2] F. L. Walls, H. Hellwig, "A New Kind of Passively Operating Hydrogen Frequency Standard," *Proc. 30th Annual Symposium on Frequency Control,* pp. 473-480, Ft. Monouth NJ, June 1976.

[3] F. L. Walls, D.A. Howe, "Timekeeping Potential Using Passive Hydrogen Masers," *J. Physique,* vol. 42, no.12, pp. C8-151 - C8-158, Dec 1981.

[4] F. L. Walls, "Frequency Standards Based on Atomic Hydrogen," *Proc. IEEE,* vol.74, no.1, pp.142-146, Jan 1986.

[5] S. Stein, D. Glaze, J. Levine, J. Gray, D. Hilliard, D. Howe, L .A. Erb, "Automated High-Accuracy Phase Measurement System," *Proc. IEEE Trans. Instrm and Meas.,* vol.32, no.1, pp.227-231, March 1983.

[6] A. Lepek, F. L. Walls, "Cross Correlation Analysis Improves Time Domain Measurements," *Proc. 47th Annual Symposium on Frequency Control,* pp. 312-320, Salt Lake City, UT, June, 1993.

[7] M. A. Weiss and T. Weissert, "TA2, A time scale algorithm for Post-Processing: AT1 plus frequency variance," *IEEE Trans. Instrum. Meas.,* 1991, vol. IM-40, pp. 496-501.

[8] M. A. Weiss and T. Weissert, "Sifting Through Nine Years of NIST Clock Data WITH TA2," *Metrologia,* to be published, Spring 1994.

[9] P. Tavella, J. Azoubib, C. Thomas, "Study of the Clock-Ensemble Correlation in ALGOS Using Real Data," *Proc. 5th European Frequency and Time Forum,* Besançon, pp. 435-441, March 1991.

[10] M. A. Weiss, C. Hackman, "Confidence on the Three Point Estimator of Frequency Drift," *Proceedings of the 21st Annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting*, Nov. 1993, McLean, VA.

[11] R.E. Drullinger, J.H. Shirley, J.P. Lowe, D.Glaze, "Error Analysis of the NIST Optically Pumped Primary Frequency Standard," *IEEE Trans. Instrum. Meas.,* 1993, vol. IM-42, pp. 453-4 56

[12] A. Bauch et al., "CS2: The PTB's New Primary Clock," *IEEE Trans. Instrum. Meas.,* vol. IM-36, pp. 613-616, June, 1987.

[13] A. Bauch et al., "The New PTB New Primary Cesium Clocks," *IEEE Trans. Instrum. Meas.,* vol. IM-42, pp. 444-447, April, 1993.

[14] W. Lewandowski, G. Petit, C. Thomas, "Precision and accuracy of GPS time transfer," *IEEE Trans. Instrum. Meas.,* Vol 42, No 2, p 474, 1993.

[15] W. Lewandowski, M. A. Weiss, "Precise Ephemerides for GPS Time Transfer," *Proc. 21st Precise Time and Time Interval Applications and Planning Meeting*, 1989.

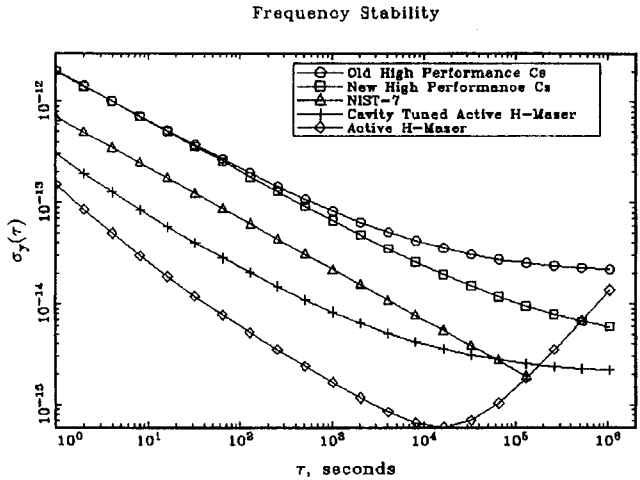


Figure 1: The frequency stability of current and recent state-of-the-art clocks [4]. The stability of the autotuned masers now exceeds the long term performance of commercial cesium beam standards.

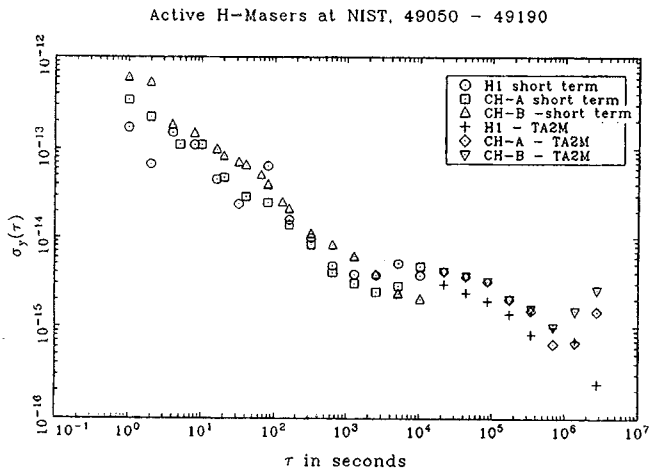


Figure 2: The short term stability of three auto-tuned H-masers at NIST. The stabilities from 1 to 10,240 s are from the enhanced measurement system and analysis technique, the stabilities from 21,600 s = ¼ d to 16 d are against TA2M, corrected for correlation.

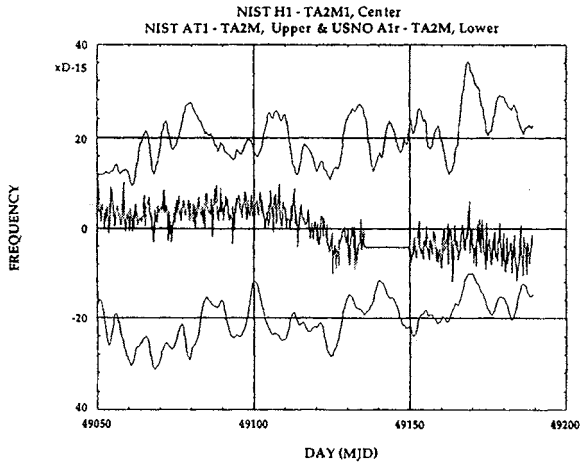


Figure 3: The residual fractional frequencies of the NIST H1 maser, the NIST AT1 scale and the unsteered USNO A1r scale all against the TA2M scale. Mean frequencies have been removed, AT1-TA2M has been offset by $2.0 \cdot 10^{-14}$, and A1R-TA2M AT1-TA2M and A1r-TA2M has been offset by $-2.0 \cdot 10^{-14}$. In addition AT1-TA2M and A1r-TA2M have been smoothed to show slower frequency changes.

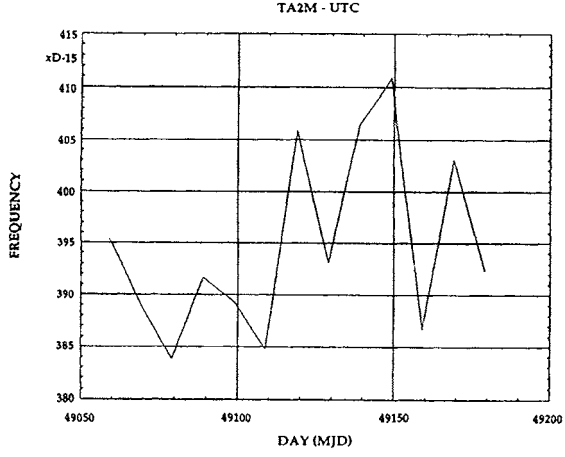


Figure 4: The frequency of TA2M versus the international UTC in 10 d frequency averages. The drift of the maser scale was negligible compared to the GPS common-view noise and frequency stability of the scale over the period of the measurements. The confidence on an overall drift estimate due to these factors was $1.5 \cdot 10^{-16}$/d. The drift of the ensemble was small compared to this.

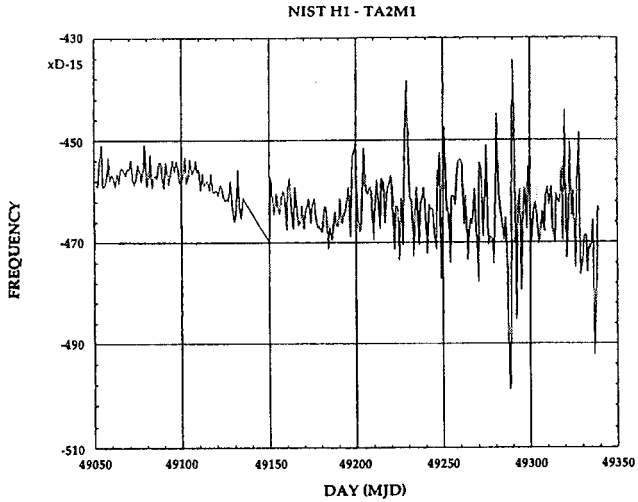


Figure 5: The fractional frequency of the NIST H1 maser versus TA2M1.

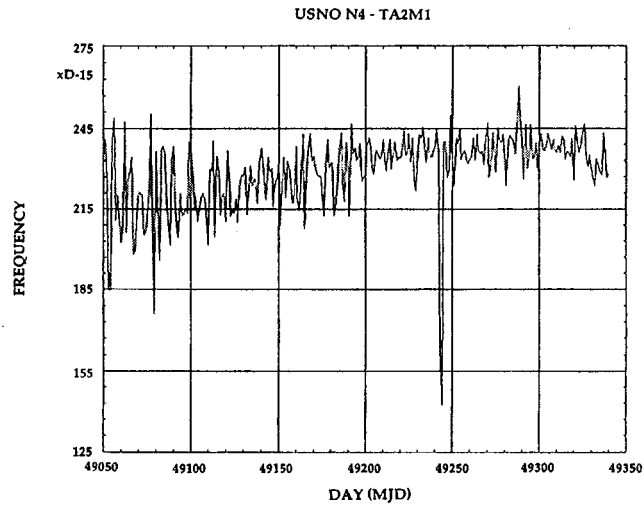


Figure 6: The fractional frequency of the USNO N4 maser versus TA2M1.

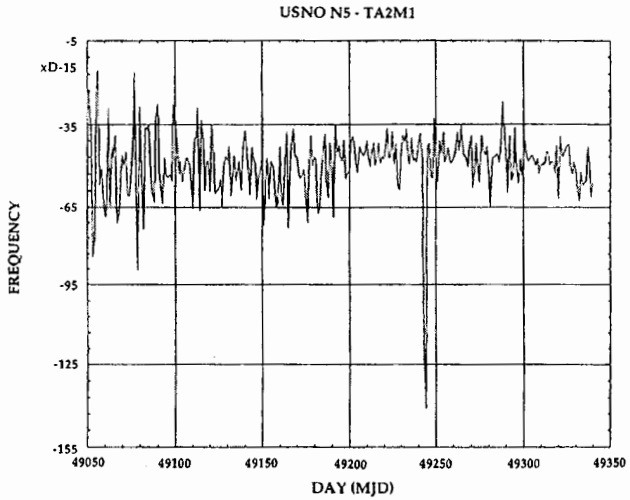


Figure 7: The fractional frequency of the USNO N5 maser versus TA2M1.

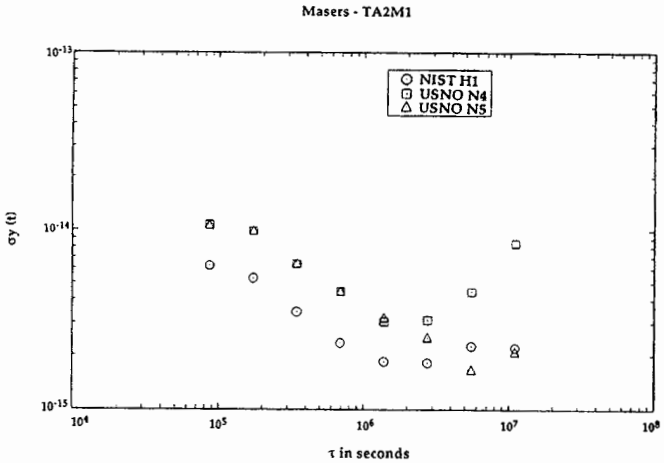


Figure 8: The fractional frequency stability of the H1 maser at Boulder and the masers N4 and N5 at USNO versus TA2M1. The noise of GPS covers the maser performance out to about 10 days. The N4 maser had a linear frequency drift of $1.0 \pm 0.3 \cdot 10^{-16}$/d.

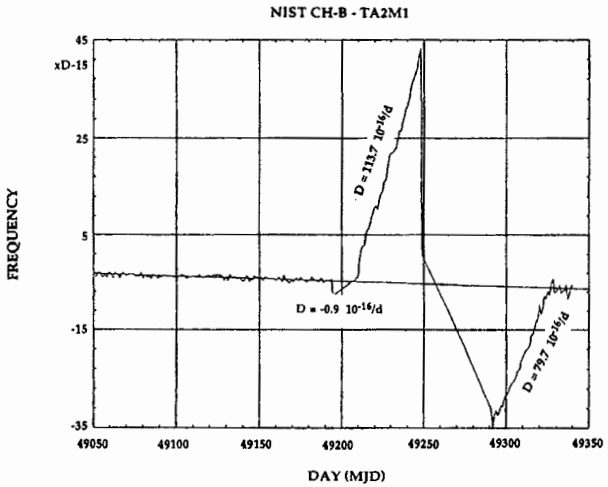


Figure 9: The fractional frequency of the NIST CH-B maser versus TA2M1. Cavity tuning was turned off at 49190 and back on at 49336 with a different reference and outside its environmental cavity. In spite of this the standard returned to its predicted frequency within the uncertainty, as illustrated by the drawn line.

The Hydrogen Maser Clocks of DLR: Relative Stability and Time Comparisons to Clocks operated at PTB Braunschweig and IMVP, NPO, "VNIIFTRI" near Moscow

H. Nau, C. Lehner, R. Manetsberger, J.Hahn
DLR - Deutsche Forschungsanstalt für Luft- und Raumfahrt e.V., Institut für Hochfrequenztechnik
Weßling, Germany;
A. Bauch, PTB - Physikalisch-Technische Bundesanstalt,
Braunschweig, Germany;
N.B. Koshelyaevsky, Institute of Metrology for Time and Space, IMVP "VNIIFTRI", Moscow, Russia

Summary

In the tracking station near Weilheim the DLR is running two hydrogen maser clocks. The monitoring system of these clocks will be described and the relative frequency stability between these clocks will be reported. Furthermore first results of time comparisons via GPS "Common View" between these H-Maser Clocks and those which are operated at PTB Braunschweig and the "Institute of Metrology for Time and Space" in the Moscow region will be presented.

1. The Hydrogen Maser Clocks of DLR

In summer 1992 two Hydrogen Masers were procured by the Department for Navigation and Time Transfer of the Institute for Radio Frequency Technology of DLR/Oberpfaffenhofen and installed in the Satellite Ground Tracking and Telecommand station near Weilheim as new ultra-stable time and frequency source. The active H-Masers are CH 1-75 types from the Russian company KVARZ.

The Masers can be operated independently of each other or in a Cavity Auto Tuning Mode (CAT). The stable frequency from one of the Masers is then used as a reference signal for that one operating in the

auto tuning mode. The method of line quality modulation is used. Dependent on the mode of operation you will have different frequency stability values as can be seen from the diagram figure 1.

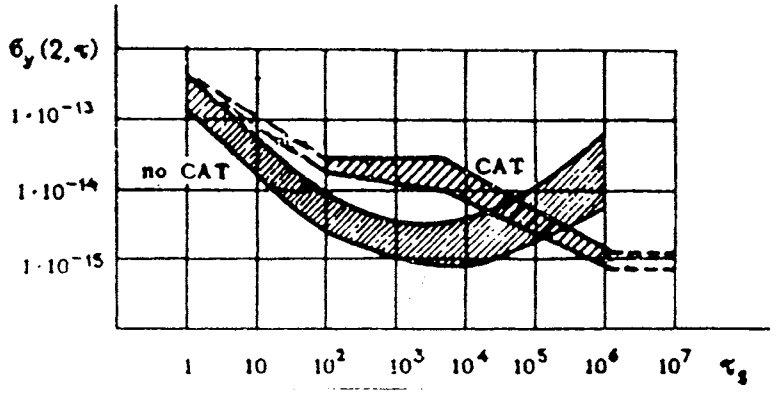


Figure 1: Frequency stability with Cavity Autotuning on and off.

The cavity auto tuning mode is good for long term stability and can be used to produce a very stable timescale. A long term frequency stability of better than 10^{-14} can be achieved.

Each Maser is located in a separate airconditioned room of one building. The registration and measurement equipment is installed in a room beside the Masers' room.

2. Real time Monitoring of the H-Masers

In our station the Masers are compared to each other and against a cesium frequency standard from Oscilloquartz continuously via a phase comparator, which multiplies each phase difference by a factor of 10 000 and operates on 5 or 100 MHz. These multiplied phase differences are then measured by a time interval counter and used by a computer for a real time monitoring display of the time intervals and the frequency differences beside the meteorological informations: temperature, humidity and pressure.

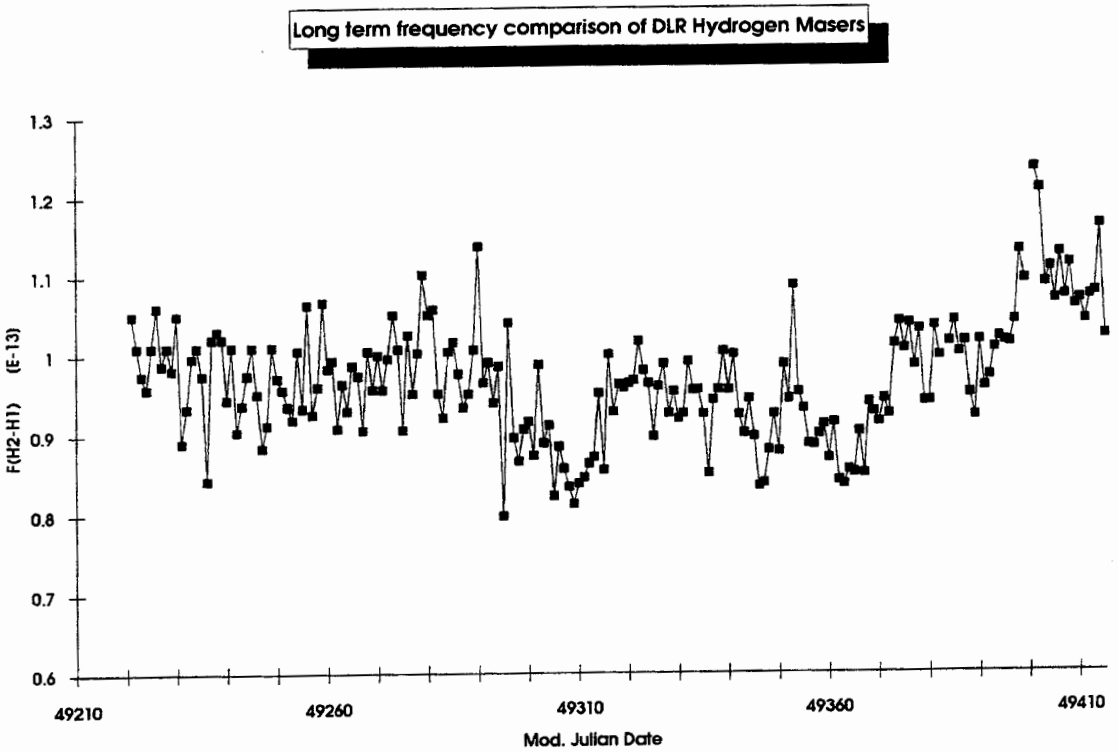


Fig. 2: Relative frequency difference between Maser H2 and Maser H1 - 1 day average values

The display gives an overview of about 5 days data. With this kind of display we can see very quickly any misbehavior of the clocks or the measurement system or the influence on clock frequencies and stabilities by changing operation modes or through external influences. All data are put on a local area network (LAN) so that at distant offices offline calculations and longterm graphs can be evaluated. In Fig. 2 the frequency difference between both Masers operating in the autotuned mode is demonstrated.

The average frequency difference of approximately $1 \cdot 10^{-13}$ corresponds to the manual frequency setting on both Masers about one year ago.

3. Comparison of frequency differences between the two H-Masers (H_1 - H_2) operated at Physikalisch-Technische Bundesanstalt (PTB) and at DLR

DLR and PTB use the same type of H-Masers from Kvarz, which can be operated in the cavity auto tuning mode (CAT). On both sides at midnight the frequency differences between the Masers (H_1 - H_2) have been determined during several months and exchanged via INTERNET, Fig. 3.

The frequency difference between the Masers (H_1 - H_2) on both sides is the result of the frequency setting at the beginning of operation; $1 \cdot 10^{-13}$ and $0.2 \cdot 10^{-13}$ respectively. The longterm stability looks very similar. In an 90 days interval a small drift $< 10^{-16}$ can be seen and demonstrates the good frequency stability.

4. Time Comparison to external time scales by means of GPS "common view" time transfer

4.1 DLR, PTB and the Russian time institute VNIIFTRI decided to exchange their GPS "common view" datafiles by E-mail

These data files contain the time differences between their reference clocks and GPS-time, collected during one week following the BIPM international tracking schedule by using an Allan Osborne time receiver.

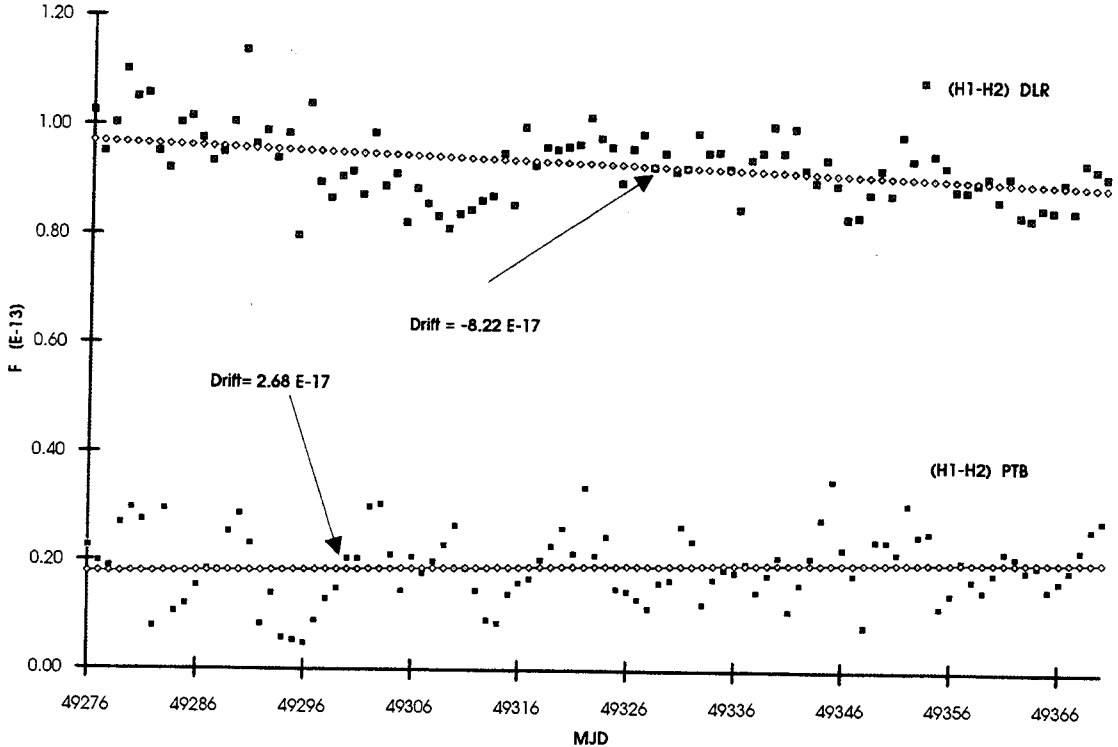


Fig. 3: Frequency Comparison of H-Masers from DLR and PTB

The following datafiles are available:

H_1(DLR) - GPS

H(SU) - GPS

UTC(PTB) - GPS; (UTC(PTB) is derived from the Cesium primary standard)

UTC(PTB) - H_1(PTB) (once per day at 0:00 h UTC)

The H-Maser clock at PTB is not the reference clock for the GPS time receiver but the cesium primary standard, therefore the H-Maser clock comparison via GPS could be evaluated only once per day. A one day average of about 20 measurements from the GPS data collected during one day had to be estimated.

4.2 Results

H-Maser intercomparisons from DLR to the time institutes PTB (Braunschweig) and VNIIFTRI (Moscow region) could be performed. First results are derived from a 7-weeks data set. The rates between the clocks and the standard deviation of the residuals s are (see Fig.4):

Rate [H_1(PTB) - H(SU)] = 21 ns/day; s = 5 ns

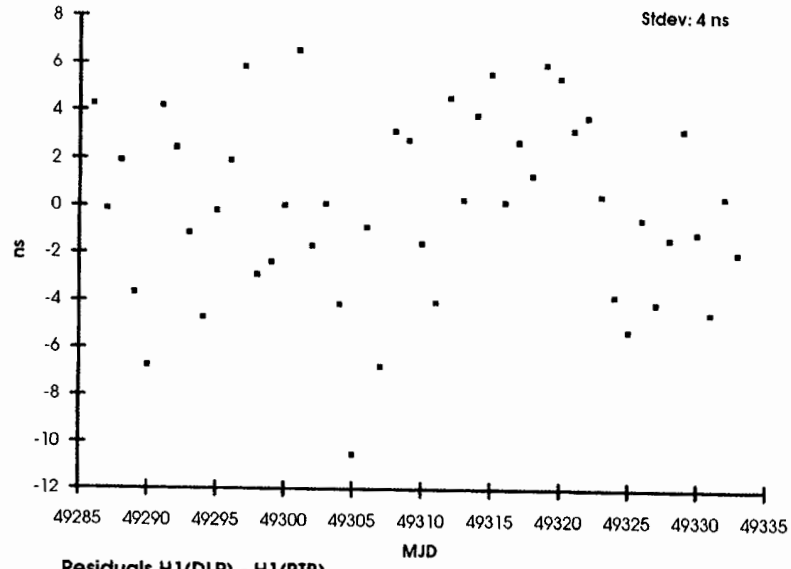
Rate [H_1(DLR) - H(SU)] = -12 ns/day; s = 5 ns

Rate [H_1(DLR) - H_1(PTB)] = 9 ns/day; s = 4 ns.

For the evaluation of the clock comparison between DLR and VNIIFTRI via the PTB's clock we received the same result as from the direct measurements from both sides.

The frequency instabilities between the clocks, during the above mentioned time period of 49 days, given by calculating the Allan variance are:

H_1(PTB) - H(SU)	$\sigma_y(\tau) = 8 \cdot 10^{-14}$	for τ = 1 day
	$\sigma_y(\tau) = 1 \cdot 10^{-14}$	τ = 10 days
H(DLR - H(SU)	$\sigma_y(\tau) = 7.6 \cdot 10^{-14}$	τ = 1 day
	$\sigma_y(\tau) = 1 \cdot 10^{-14}$	τ = 10 days
H_1(DLR) - H_1(PTB)	$\sigma_y(\tau) = 1 \cdot 10^{-13}$	τ = 1 day
	$\sigma_y(\tau) = 1 \cdot 10^{-14}$	τ = 10 days



Residuals H1(DLR) - H1(PTB)

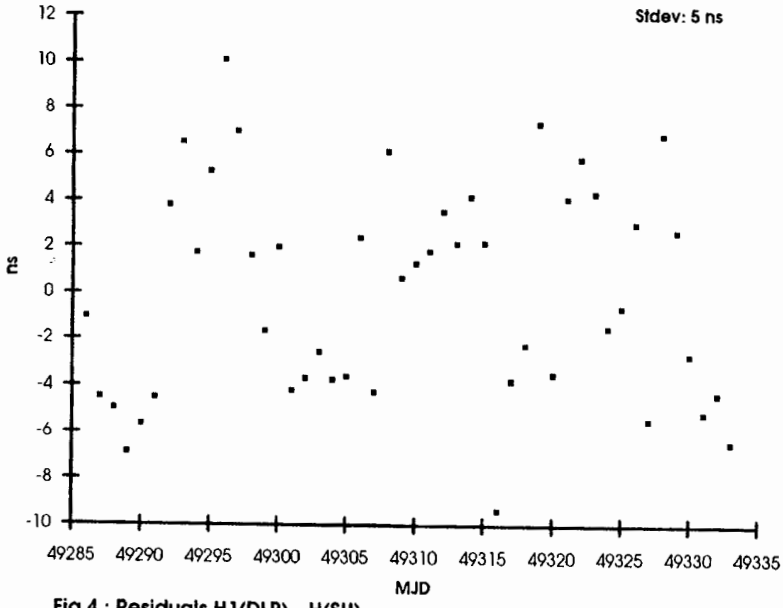


Fig.4 : Residuals H1(DLR) - H(SU)

5. Conclusion

The time and frequency system at DLR's tracking and telecommand ground station is equipped with two Hydrogen Masers from the Russian company KVARZ (Gorky). It could be seen that with these H-Maser clocks a very stable timescale can be generated. The frequency instability between the two Masers in the Cavity Auto Tuning mode was $5 \cdot 10^{-15}$ per day in a time intervall of 200 days with a frequency drift of $1 \cdot 10^{-17}$.

From the intercomparisons of the H-Maser at PTB, VNIIFTRI and DLR by means of GPS common view technique the instability between the timescales could be evaluated with the Allan variances: $\sigma_y(\tau) \cong 1 \cdot 10^{-13}$, τ = 1 day and $\sigma_y(\tau) \cong 1 \cdot 10^{-14}$, τ = 10 days as preliminary results. But these results are mainly influenced by the accuracy of the GPS common view method.

Oscillators (B 4)

INTERNATIONAL COMPARISON OF PHASE NOISE

Fred L. Walls,[1] Roland Barillet,[2] Raymond Besson,[3] Jacques Groslambert,[4] Peter Schumacher,[5] Jurg Rufenacht,[6] and Kurt Hilty[6]

ABSTRACT

During March of 1993 six different laboratories in three countries participated in phase modulation (PM) noise measurements of the PM noise standard that was developed at the National Institute of Standards and Technology (NIST). Measurements were made at 5 MHz, 10 MHz, and 100 MHz. The four different measurement systems that were used at the six laboratories differed widely in their approach to the measurements and in the degree of automation. The agreement of the PM noise at 5 and 10 MHz were typically better than ± 1 dB. The difference between the PM noise at the beginning and end of the trip was 0.2 dB at 5 and 10 MHz and 0.5 dB at 100 MHz. The results at 100 MHz were substantially worse, up to 3 dB, depending on details of the measurements. The source of the error was traced to an interference between harmonics of the 100 MHz with the reference signal. When the harmonic distortion was eliminated, the measurement error dropped to approximately 0.2 ± 0.5 dB. The initial error at 100 MHz was much larger than anticipated, indicating that the sensitivity in some measurement systems to harmonic distortion is substantially larger than previously documented. Based on the intercomparisons and earlier work at NIST, a detailed error model for PM noise measurements is presented.

1. NIST, Boulder, USA
2. LHA-CNRS, Orsay, France
3. ENSMM-LCEP,Besancon, France
4. LPMO-CNRS, Besancon, France
5. Observatoire de Neuchatel, Switzerland
6. Swiss Telecom PTT Research and Development, Bern, Switzerland

1. INTRODUCTION

In the past it has been difficult to evaluate the accuracy and noise floor of phase modulation (PM) noise measurement equipment because no artifact standards were available. Comparisons of measurement systems in different laboratories using commercially available oscillators as transfer standards were typically limited to a repeatability of roughly ± 3 dB due to the temporal variability of the oscillator noise. Recently a new portable secondary standard and associated measurement techniques for evaluating the noise floor and accuracy versus Fourier frequency of PM and amplitude modulation (AM) noise measurement systems were developed at NIST [1,2]. Evaluations of these new PM/AM noise standards at 5, 10, and 100 MHz yield an accuracy of better than 0.2 dB, a temperature coefficient of less than 0.02 dB/K, and a stability of better than 0.4 dB over 1 y. One of these new PM noise standards was used to compare the accuracy and noise floor of PM noise measurements at six different laboratories in three countries at the time of the 1993 European Frequency and Time Forum (EFTF). This paper reports on the results of these measurements and the insight into possible errors in PM noise

measurements. We present an error model which describes all of the commonly encountered errors in such measurements. These measurements in the six laboratories, which used the two-oscillator technique for measuring PM noise, demonstrate that by taking all of these possible error parameters into account, it is possible to make reliable accurate measurements to better than 1 dB.

2. NIST AM/PM NOISE STANDARD

The design of the new NIST AM/PM noise standard is described in detail in reference [1]. Figure 1 shows a simplified block diagram of the PM/AM noise standard model PMAM 115 used for these PM noise comparisons. A frequency source with very low PM and AM noise is regulated in amplitude and divided into the reference and signal outputs using a reactive power splitter. The amplitude and phase of these outputs track one another with great fidelity. The residual differential PM noise between the two outputs, when the noise source is off, is much smaller than the PM noise of the source. The ratio of the differential PM noise to the source PM noise can approach -100 dB at low Fourier frequencies, degrading to approximately -16 dB at $f = \nu/10$ [3], where ν is the carrier frequency. At 10 MHz, for example, the differential phase noise between the two channels, S_ϕ(10 kHz) = -194 dB relative to 1 rad^2/Hz. This feature is used to measure the noise floor of PM noise measurement systems.

A broad band power combiner can be used to add passband-limited Gaussian noise to a carrier. The added noise is roughly 40 dB above the noise floor of most measurement systems and 60 dB above the residual noise between the two signals. A switch can be used to change the noise power by 19.8 dB. This results in a spectral density of PM noise $S_\phi(f)$ given by

$$S_\phi(f) = \frac{PSDV_n(f)}{2V_O^2} \qquad (1)$$

where V_0 is the amplitude of the carrier, and $PSDV_n(f)$ is the power spectral density of voltage noise at Fourier frequency $\pm f$ from the carrier. $S_\phi(f)$ is constant from dc to approximately 10% of the carrier frequency. Within 100 kHz of the carrier $S_\phi(f)$ is flat to $\pm$ 0.05 dB. Since there is no phase coherence between the signal and the noise, the resulting modulated output has precisely equal AM and PM noise. These noise spectra can be made very nearly constant over a very wide temperature range by stabilizing the carrier and the noise separately using traditional approaches. If necessary, a phase shifter can be used to provide a 90° phase shift between the reference and modulated output. Measuring $PSDV_n(f)$ and V_0^2 separately provides, by Eq. (1), a primary calibration of $S_\phi(f)$. The accuracy of this calibration process is 0.2 dB [2]. This feature is used to evaluate the accuracy of PM/AM noise-measuring equipment as a function of Fourier frequency. The nominal PM noise of the source, the residual PM noise between the two outputs, and the calibrated PM levels versus Fourier frequency at 5, 10, and 100 MHz are listed in Table 1.

Implicit in these discussions is the assumption that the phase detector in the PM noise measurement system under test has sufficient discrimination against AM noise that the AM noise does not bias the PM noise measurements. These assumptions are not excessively restrictive since many devices to be calibrated also have similar PM and AM noise [3,4]. Discrimination of 15 dB reduces the unwanted effects below 0.14 dB. This discrimination level is easily met by virtually all AM and PM noise measurement techniques in use today [3,4] (typical levels of discrimination are 25 dB for PM and even higher for AM measurement systems.)

3. PM NOISE MEASUREMENTS AT 6 DIFFERENT LABORATORIES

The NIST PM/AM noise standard was evaluated using the method described above at NIST on 11 March, 1993 prior to the trip to Europe and on 1 April upon return. Several additional calibrations were made during the trip. The standard was delivered to each of the laboratories for determination of the PM noise for Fourier frequencies from roughly 10 Hz to 100 kHz. None of the laboratories participating in these measurements knew the expected PM noise within ± 5 dB until after they had completed their measurements. They were free to measure the PM noise at roughly -110 or -130 dBc/Hz. Each laboratory used their own set of parameters to evaluate possible bias terms.

3.1 PM NOISE MEASUREMENTS AT LABORATOIRE DE L'HORLOGE ATOMIQUE, ORSAY

The PM noise standard was delivered to the Laboratoire de l'Horloge Atomique-CNRS on 13 March and allowed to warm up for about 2 h. Measurements of noise floor and calibrated PM noise were made at 5, 10, and 100 MHz using their custom PM noise measurement system and a Scientific Atlanta 380 Fast Fourier Transform (FFT) spectrum analyzer.[1] A simplified block diagram of the measurement configuration is shown in Figure 2. All control of this system is manual. An external substitution source at approximately the same level as the reference frequency output was used to obtain a beat frequency to calibrate the mixer sensitivity k_d multiplied by the post amplifier gain G(f) at the zero crossing [3,4]. Figure 3 shows a typical wave form and the method of calculating $k_dG(f)$ using an oscilloscope or other recording device. The conversion from radians to volts at the output of the amplifier was about 650 V/rad at 5 and 10 MHz and about 250 V/rad at 100 MHz. Generally 800 averages were taken for the PM noise data at all three carrier frequencies.

The results obtained for the calibrated PM noise and the noise floor are summarized in Tables 2 to 4. Figure 4 shows the raw data for the measurement of the NIST PM noise standard at 5 MHz versus Fourier frequency. At Fourier frequencies of less than 20 kHz, the difference between the measured PM noise and the NIST PM standard is under 1 dB at carrier frequencies of 5 and 10 MHz. The 3 dB differences at 100 MHz is unexpectedly high. See section 3.5 for a

[1]Certain measurement equipment is identified to properly document the measurement configuration and does not imply endorsement by any author or laboratory.

discussion. The 95% confidence interval for 800 averages is approximately ± 0.26 dB. The noise floor of the measurement system is low enough that it does not bias these results.

3.2 MEASUREMENTS AT ECOLE NATIONALE SUPERIEURE DE MECANIQUE ET DES MICROTECHNIQUES, BESANCON

The PM noise standard was delivered to the ENSMM-LCEP and allowed to warm up for about 1 h. Measurements of both noise floor and the calibrated PM noise level were made at 5, and 10 MHz using a FemtoSecond FSS600 PM noise measurement system and an HP3561A FFT spectrum analyzer. All control of this system is manual. The measurement configuration was similar to that shown in Figure 2. An external substitution source at approximately the same level as the reference frequency output was used to obtain a beat frequency to calibrate $k_dG(f)$ at the zero crossing. An oscilloscope was used to measure the phase slope. The results for the calibrated PM noise and the noise floor are summarized in Table 2. The IF response of this system is extremely flat with Fourier frequency. The 95% statistical confidence interval for the 5000 averages recorded for f = 1 kHz, 10 kHz, and 100 kHz is approximately ± 0.12 dB. The 95% statistical confidence interval for the 1000 averages recorded for f = 10 and 100 Hz is approximately 0.27 dB. The difference between these PM measurements at 5 and 10 MHz and the NIST PM standard are less than 0.5 dB. The noise floor of the measurement system is low enough that it does not bias these results. No measurements were taken at 100 MHz.

3.3 MEASUREMENTS AT THE OBSERVATOIRE DE NEUCHATEL, NEUCHATEL

The PM noise standard was delivered to the Observatoire de Neuchatel and allowed to warm up for about 1 h. Measurements were made at 5, 10, and 100 MHz using a HP3048A PM noise measurement system. The measurement configuration was similar to that shown in Figure 2. All measurements on this system are controlled by a computer once the initial setup is completed. The maximum number of averages possible with the software was 200. The 95% confidence interval for 200 averages is approximately ± 0.5 dB. An external substitution source at approximately the same level as the reference frequency output was used to obtain a beat frequency to calibrate $k_dG(f)$ at the zero crossing. The rf power was adjusted so that the beat signal was a sine wave. The peak value of the sine wave is then equal to k_d. The phase noise data were then taken with this reduced power for the mixer. The disadvantage of using this simple method for determining $k_dG(f)$ is that the noise floor is somewhat higher than can be obtained by driving the mixer closer to its maximum ratings. This can be seen by comparing the noise floor of this configuration with that obtained in the next two tests. The results for the calibrated PM noise and the noise floor are given in Tables 2 and 3. Figure 5 shows the measurement of the calibrated PM noise at 10 MHz. The effect of a small number of averages on the measurement uncertainty is clearly visible. The IF response of this and the other two HP3048A systems was very nearly flat with Fourier frequency. The small number of averages that was possible with the automated system precluded looking at this effect with high resolution. A modified

software package is now available for these systems that greatly increases the number of averages that can be taken [5]. The difference between these measurements and the calibrated PM noise of the NIST standard at 5 and 10 MHz are less than 1 dB. Measurements made at 100 MHz as described above for 5 and 10 MHz, yielded values that were about 3 dB low. The probable cause of this error at 100 MHz is discussed in Section 3.5. The noise floor of the measurement system is low enough that it does not bias these results.

3.4 MEASUREMENTS AT LABORATOIRE DE PHYSIQUE ET METROLOGIE DES OSCILLATEURS, BESANCON

The PM noise standard was delivered to the LPMO-CNRS and allowed to warm up for about 1 h. Measurements were made at 5, 10, and 100 MHz using a HP3048A PM noise measurement system. The measurement configuration was similar to that shown in Fig. 2. All measurements on this system were controlled by a computer once the initial setup was completed. The maximum number of averages possible with the software was 200 [5]. This results in a 95% confidence interval of approximately $\pm$ 0.5 dB. A substitution source at the same level of the reference frequency output was used to obtain a beat frequency to calibrate $k_dG(f)$ at the zero crossing. The measurement of the PM of the NIST standard at 10 MHz is shown in Fig. 6. The results for the calibrated PM and the noise floor are given in Tables 2 and 3. The difference between these measurements at 5, 10, and 100 MHz and the calibrated PM noise of the NIST standard are 1 dB or less. The large biases seen at LHA and Observatoire de Neuchatel are absent; see Section 3.5 for a discussion. The noise floor of the measurement system is low enough that it does not bias these results.

3.5 MEASUREMENTS AT SWISS TELECOM PTT, BERN

The PM noise standard was delivered to the laboratory of the Swiss PTT Telecom and allowed to warm up for about 1 h. Measurements were made at 5, 10, and 100 MHz using a HP3048A PM noise measurement system. The measurement configuration was similar to that shown in Figure 2. All measurements on this system were controlled by a computer once the initial setup was completed. The maximum number of averages possible with the software was 200 [5]. This results in an approximate 95% confidence interval of $\pm$ 0.5 dB. As with the other HP3048A systems, a substitution source at approximately the same level of the reference frequency output was used to obtain a beat frequency to calibrate $k_dG(f)$ at the zero crossing. The results for the calibrated PM level and the noise floor are given in Tables 2-4. The results at 100 MHz are shown in Fig. 7. The difference between the measurements at 5, 10, and 100 MHz are less than 1 dB when measuring a PM noise at about -128 dBc/Hz. Measurements at a level 20 dB higher indicated significant errors similar to those seen at LHA and Observatoire de Neuchatel. The noise floor of the measurement system is low enough that it does not bias these results.

The occasional errors for the PM noise measurements at 100 MHz of approximately 0-4 dB were quite perplexing. Figure 8 shows the spectrum analysis of the 100 MHz signal. The distortion at 200 and 300 MHz, while relatively large,

seemed too small to explain the discrepancies. To explore the cause of the 100 MHz errors, the 100 MHz carrier was reduced 7 dB prior to the power amplifier, which reduced the harmonic content by more than 10 dB. The PM noise standard was recalibrated using the technique of [2] and the PM noise remeasured with the HP3048A. These measurements are shown in Table 4. The agreement with the primary calibration is excellent. To further test the role of the harmonics, a 135 MHz low-pass filter was used to eliminate the higher harmonics. A new primary calibration was carried out and the measurements with the HP3048A repeated. The results are given in Table 4. Again with the harmonics removed, the agreement with the primary calibration is excellent.

The errors in the 100 MHz PM noise calibrations was explored by Bob Temple of HP. He noted that the spectrum in Fig. 8 showed the PM noise on only the 100 MHz signal and not on the higher harmonics. If the higher harmonics were phased just right they could increase the value of k_d leading to an underestimation of the PM noise because the high harmonics contributed to k_d but not to the measured PM noise. He was able to show qualitatively that the presence of the second, third, and higher harmonics could cause the problem and that the error was to indicate a smaller PM noise than was actually present. If the second and third harmonics are more than 35 dB below the fundamental, this effect is negligible. This effect is about an order of magnitude higher than previously reported [3] and changes from one phase detector family to another and as a function of input drive. It may also depend on the IF termination. It is quite likely that this effect has caused errors in many previous measurements.

4. ERROR MODEL

Table 5 shows the various parameters which appear to be the most important contributors to errors and uncertainties in the measurement of PM noise. The first term is the bias and uncertainty in the determination of k_d. Errors in this parameter originate from errors in measuring the slope of the zero crossing (see Fig.3). It is most important that both the positive and negative going slopes be measured. It they differ by more than about 10% it indicates excessive injection locking between the two sources or possibly a damaged phase detector. Another bias occurs when a substitution oscillator is used, as in the measurements described here, which has an output power level and/or impedance that is different from the original source. Another common error is the use of a different length of coaxial cable, when determining k_d, than is used to take PM noise data. The cable transforms the apparent impedance of the mixer as seen by the source. If the voltage reference of the oscilloscope or other recording device used to determine k_d does not agree with the spectrum analyzer to measure $PSDV_n$, there will be additional errors. k_d may depend on frequency if the IF port is not terminated in 50 Ω for the rf signals [4].

The second term comes from the bias and uncertainty in measuring the amplifier gain G(f) versus Fourier frequency. The mixer output impedance is a function of frequency and power [4] making it difficult to always achieve a good

match to the low-pass filter and the post-amplifier. As f increases, G(f) may depend on the cable and the impedance of the spectrum analyzer.

The third term comes from the bias and uncertainty in determining the effect of phase-locked-loops (PLL) on the phase noise at Fourier frequencies near or below the bandwidth of the loop [4]. We can estimate the effect of the PLL on the phase response from measurements of the mixer sensitivity, oscillator tuning rate, and PLL gain. Often this approach is not satisfactory because the oscillator under test has a low pass filter in the control path that affects the loop response. The only secure way to account for this effect is to measure the loop response as a function of f. In most of the measurements discussed above, no PLL was needed, so this problem was not an issue. Several PM noise determinations at LPMO and at PTT did, however, measure the PM noise of the NIST standard against an external oscillator and PLL effects had to be taken into account by the computer software.

The fourth term comes from conversion of AM noise, in one or both of the signals under test, to apparent PM noise by the mixer and any other nonlinear element. The rejection of AM noise by most PM noise measurement systems is of order 15 to 25 dB. This is sufficient only when the AM noise is equal to or smaller than the PM noise to be measured. Therefore no PM noise measurement is complete until the AM noise and the AM to PM conversion factor have been measured. In many synthesizers and complex sources where the output power is leveled, the AM noise is actually larger than the PM noise. This is usually not a serious issue in simple oscillators [6,7].

The fifth term comes from the error in measuring both k_d and PSDVn in the presence of harmonic distortion in either the source under test or the reference oscillator. We have seen above that this can lead to errors of order 4 dB even when the distortion is less than -16 dBc. A good rule of thumb is to keep the harmonic content below -35 dBc.

The sixth term is due to the contribution of the system noise floor to the measured noise. Since the noise floor depends on k_d through the rf and LO drive power, these should be recorded for all measurements along with k_d. No measurement is complete without a determination of system noise floor relative to the PM noise to be measured. In some cases cross-correlation techniques can be used to reduce the contribution of system noise floor [8].

The seventh term is due to noise in the reference signal biasing the measurements of the device under test. By measuring the PM noise between three oscillators of comparable PM noise, output power, and impedance, we can generally reduce the contribution of the reference oscillators to the determination by about 6-10 dB [4]. The noise floor of the measurement system still contributes. Better cancellation of both the reference and system noise can be obtained by using cross-correlation because the measurements are made simultaneously. Both reference and system noise average towards zero as $1/\sqrt{N}$ where N is the number of averages

[4,8]. It is not uncommon to obtain a factor of 15-20 dB improvement in measurement noise floor using this approach [6,7].

Confidence intervals (term 8) of the FFT spectral density data are shown as a function of the number of averages in Table 6. The confidence intervals are independent of the noise type as long as the measurement bandwidth is small compared to the Fourier frequency, f. This requirement is satisfied if f is larger than the frequency span/75 for the Hanning window and f/23 for the flat top window [3,4].

The accuracy and linearity of modern spectrum analyzers (terms 9 and 10) limit the accuracy of these measurements to about ± 0.2 dB unless special calibrations are performed on the instruments.

5. CONCLUSION

PM noise measurements of a NIST PM noise standard made in six different laboratories using four different measurement configurations demonstrate that with careful attention to detail and the use of the error model given in Table 6, it is possible to make PM noise measurements that are accurate to $\pm$ 1dB. An accuracy of approximately ± 0.5 dB appears possible if more attention is paid to the harmonic content of the signals. We have used these results to make the first precise international comparison of PM noise between laboratories.

6. ACKNOWLEDGMENTS

We thank Bob Temple for his help in analyzing the errors in the PM measurements at 100 MHz. F. L. Walls thanks Tom Parker and John Lowe for helpful comments and all the participating laboratories for their hospitality and enthusiasm in making these comparisons.

7. REFERENCES

[1] F. L. Walls, U. S. Patent 5,172,064, "Calibration System for Determining the Accuracy of Phase Modulation and Amplitude Modulation Noise Measurement Apparatus."

[2] F. L. Walls, "Secondary Standard for PM and AM noise at 5, 10, and 100 MHz," IEEE Trans on I&M, 42, pp.136-143, 1993. See also F. L. Walls, "Practical Standards for PM and AM Noise at 5, 10, and 100 MHz", Proc. 1993 EFTF.

[3] F. L. Walls, C. M. Felton, A. J. D. Clements, and T. D. Martin, "Accuracy Model for Phase Noise Measurements," Proc. of 21st Annual Precise Time and Time Interval Planning Meeting, Redondo Beach, CA, pp. 295-310, Nov. 30-Dec. 1, 1990; F. L. Walls, A. J. D. Clements, C. M. Felton, and T. D. Martin, "Precision Phase Noise Metrology," Proc. National Conference of Standards Laboratories, pp. 257-275, August, 1991.

[4] "Characterization of Clocks and Oscillators," Eds. D. B. Sullivan, D. W. Allan, D. A. Howe, and F. L. Walls, National Institute of Standards & Technology Tech. Note 1337, 1990.

[5] As a result of this international comparison, updated software that permit up to 10^5 averages for the HP3048A systems is available from HP distributors.

[6] Lisa M. Nelson and F. L. Walls, "Relationship of AM to PM Noise in Selected rf Oscillators," Proc. IEEE Freq. Control Symp.,1993, pp 298-302.

[7] F. G. Ascarrunz, E. S. Ferre, and F. L. Walls, "Investigations of AM and PM Noise in X-Band Devices," Proc. IEEE Freq. Control Symp., 1993, pp.303-310.

[8] W. F. Walls, Cross Correlation Phase Noise Measurements," Proc. 46 Ann. Frequency Control Symp., 1992, p.257-261.

[9] D. B. Percival and A.T. Walden, Spectral Analysis for Physical Application," Cambridge Univ. Press, 1993.

[10] B. N. Taylor and C. E. Kuyatt, NIST Technical Note TN1297, 1993.

TABLE 1. System noise floor and phase noise of sources

NOMINAL SOURCE PHASE NOISE/CHANNEL, ±3 dBc/Hz								
	FOURIER FREQUENCY							
SOURCE FREQUENCY	1 Hz	10 Hz	100 Hz	1 kHz	10 kHz	100 kHz	1 MHz	10 MHz
5 MHz	-121	-151	-163	-171	-174	-174	-174	
10 MHz	-115	-145	-157	-165	-168	-168	-168	
100 MHz	-70	-100	-130	-156	-170	-170	-173	-173
MAXIMUM RESIDUAL NOISE BETWEEN CHANNELS, dBc/Hz								
5 MHz	-162	-172	-182	-190	-194	≤-175	≤-175	
10 MHz	-161	-176	-183	-191	-197	≤-175	≤-175	
100 MHz	-152	-162	-172	-182	-193	≤-175	≤-175	≤-175
DIFFERENTIAL PM/AM NOISE, ±0.2 dBc/Hz								
Attenuation 19.8 dB, 11 March 1993								
5 MHz	-127.3	-127.3	-127.3	-127.3	-127.3	-127.3		
10 MHz	-128.4	-128.4	-128.4	-128.4	-128.4	-128.4	-128.4	
100 MHz	-129.5	-129.5	-129.5	-129.5	-129.5	-129.5	-129.5	-129.8

TABLE 2. Summary of PM noise calibration by participating Frequency Laboratories

FREQUENCY	LABORATORY									
	NIST 11 March	LHA 13 March	ENSMM 15 March	NIST 15 March	OBS NEU 15 March	LPNO 19 March	NIST 20 March	PTT 20 March	NIST 1 April	AVG NIST
5 MHZ	-127.3	-128.0	-127.0	-127.6	-127.2	-126.3	-127.5	-127.5	-127.5	-127.48
10 MHZ	-128.4	-128.5	-128.6	-128.2	-127.4	-128.5	-128.3	-128.5	-128.5	-128.35
100 MHZ	-129.5	-126.3			-132.5	-129.3	-129.8	-129.6	-129.0[1]	-129.43[1]

[1] value may be biased because standard was modified at PTT for experiments discussed in Sectors 3.5 and Table 4 and then reassembled.

TABLE 3. Noise floor at 5, 10, AND 100 MHz
at participating laboratories vs Fourier frequency

LABORATORY FOURIER FREQUENCY (Hz)

Carrier Frequency	10	100	1K	10K	100K
LHA - CNRS					
5 MHZ	-153	-162	-171	-176	-177
10 MHZ	-153	-163	-172	-179	-179
100 MHZ	-146	-156	-166	-170	-171
ENSMM-LCEP					
5 MHz	-151	-160	-166	-175	-178
10 MHZ	-156	-165	-172	-179	-179
OBSERVATOIRE DE NEUCHATEL					
5 MHZ	-149	-158	-168	-174	-177
10 MHZ	-149	-160	-171	-176	-177
100 MHZ	-153	-163	-172	-175	-176
LPMO - CNRS					
5 MHZ					
10 MHZ	-158	-167	-174	-177	-178
100 MHZ	-156	-167	-176	-178	-179
SWISS TELECOM PTT					
5 MHZ	-153	-163	-172	-176	-178
10 MHZ	-159	-169	-176	-179	-180
100 MHZ	-157	-167	-177	-181	-181

TABLE 4. PM noise comparisons at 100 MHz for various configurations

Carrier (MHz)	NIST 11 March	NIST 20 March	NIST 1 April	PTT 20 March	Notes
100	129.5	-129.8	-129.0	-129.6	Fig. 2 without PLL
100	-109.7	-110.0	109.2[3]	-111.3	vs HP8663
100	-109.7	-110.0	109.2[3]	-112.3	vs HP8663
100		-106.2		-106.3	1
100		-110.2		-110.0	2

1. 100 MHz signal reduced 7 dB before output amplifier at 100 MHz.
2. Note 1 plus 135 MHz low pass filter yielding harmonics <-70 dBc.
3. Measured after NIST system reassembled following notes 1 & 2.

TABLE 5. Error model for PM noise measurements

1. Determination of k_d
2. Determination of Amplifier G(f)
3. PLL Effects (if any)
4. Contribution of AM Noise
5. Harmonic Distortion
6. Contribution of system Noise Floor
7. Contribution of Reference Noise
8. Statistical Confidence of Data
9. Linearity of Spectrum Analyzers
10. Accuracy of PSD Function

TABLE 6. Statistical uncertainty of FFT spectral density measurements as a function of N, the number of averages, where $S_m(f)$ is the measured spectral density, S(f) is the true spectral density, and k controls the confidence interval [9,10]

Number of Samples	k ≃ 1 (approx. 68%) $S_m = S[1\pm\delta]$, $S_m {}^{-\gamma}_{+\beta}$ dB			k ≃ 1.9 (approx. 95%) $S_m = S[1\pm\delta]$, $S_m {}^{-\gamma}_{+\beta}$ dB		
	δ	γ	β	δ	γ	β
4	0.54	-2 ,	+3.3	2.5	-3 ,	+6
6	0.42	-1.5,	+2.3	1.4	-2.5,	+5
10	0.32	-1.2,	+1.7	0.61	-2.1,	+4
30	0.18	-0.72,	+ .86	0.35	-1.3,	+1.8
100	0.1	-0.41,	+0.46	0.19	-0.76,	+0.92
200	0.058	-0.24,	+0.25	0.14	-0.46,	+0.51
1000	0.032	-0.13,	+0.13	0.06	-0.26,	+0.28
3000	0.018	-0.08,	+0.08	0.035	-0.15,	+0.15
10000	0.01	-0.04,	+0.04	0.019	-0.08,	+0.08

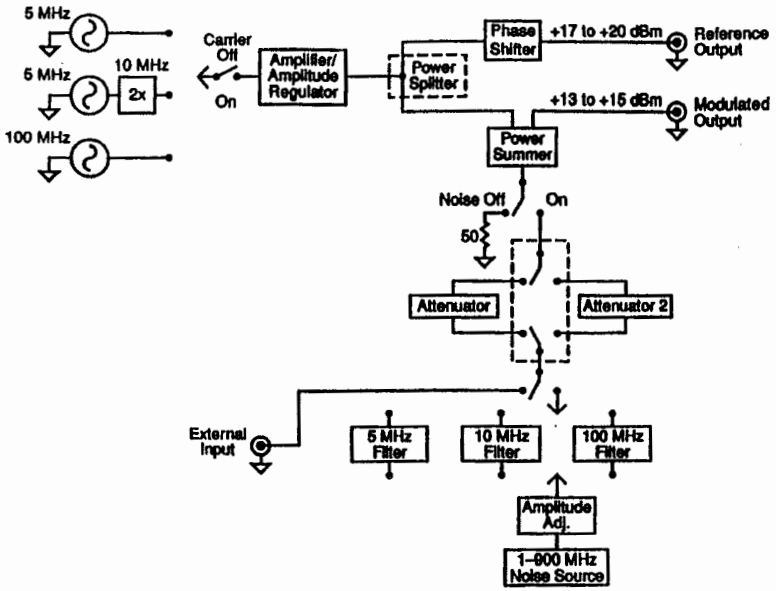


Fig. 1. Block diagram of Calibration Standard PMAM 115 for PM and AM noise.

Calibration of Noise Floor and Accuracy of Two Oscillator Phase Noise Measurement System

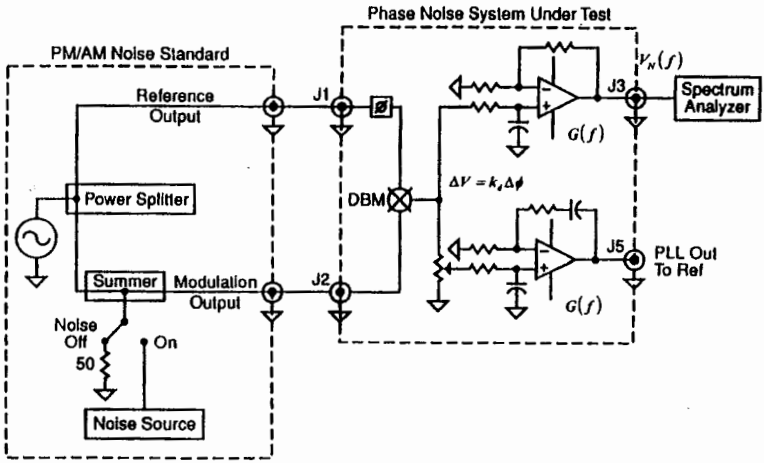


Fig. 2. Block diagram of the PM noise measurement system.

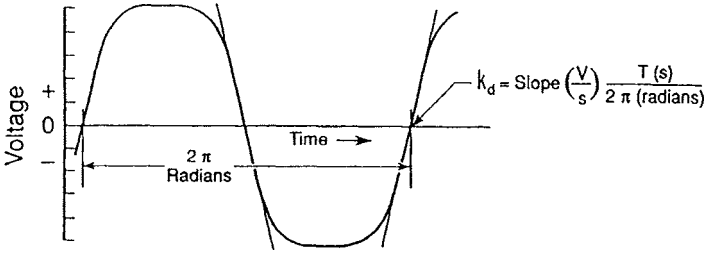


Fig. 3. One method of determining the phase to voltage conversion coefficient k_d G(f) from the beat signal.

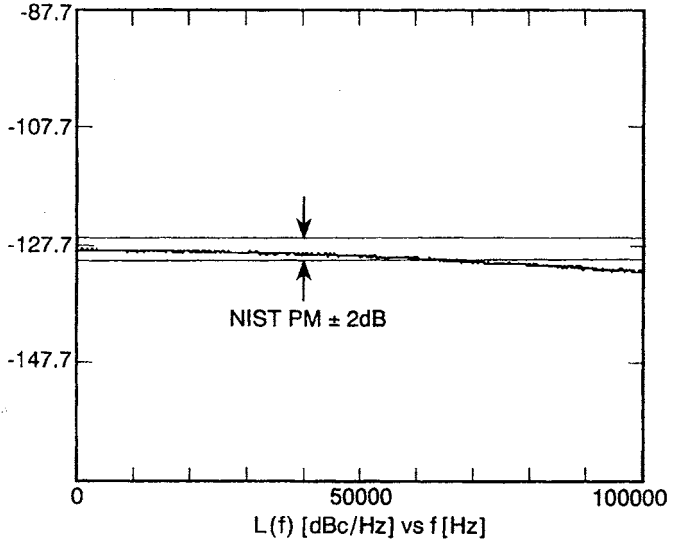


Fig. 4. Raw data for the measurement of the calibrated 5 MHz PM noise of the NIST standard at LHA-CNRS with 800 averages.

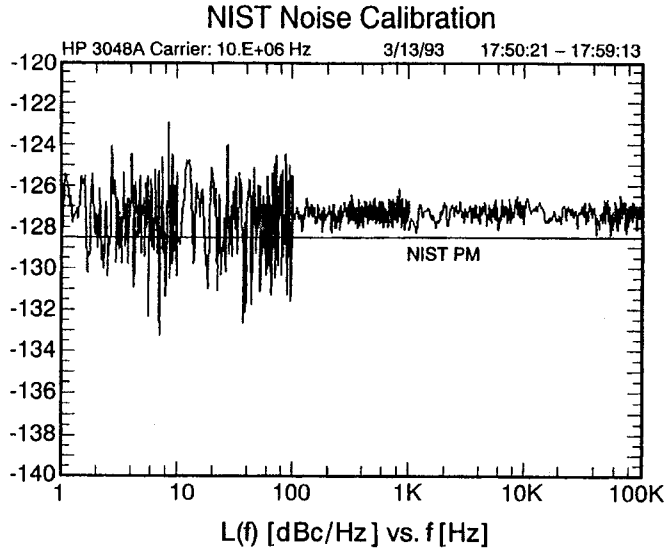


Fig. 5. Raw data for the measurement of the 10 MHz calibrated PM noise of the NIST standard system at Observatoire de Neuchatel. The fluctuations in the measured level is determined by the number of averages.

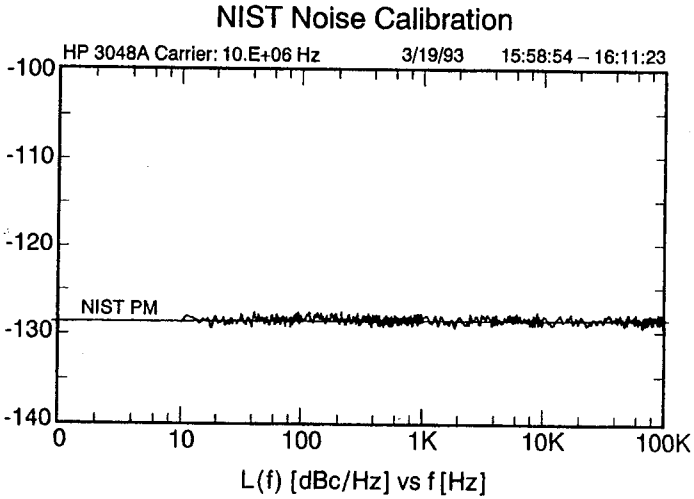


Fig. 6. Raw data for the measurement of the 10 MHz calibrated PM noise of the NIST standard system at LPMO-CNRS with 200 averages.

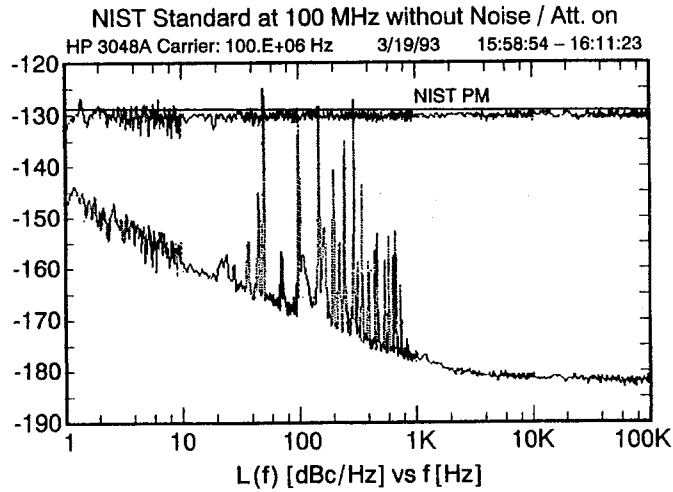


Fig. 7. Measurement of the calibrated 100 MHz PM noise of the NIST standard at Swiss PTT Telecom Research Laboratory without PLL The noise floor of the system is also shown.

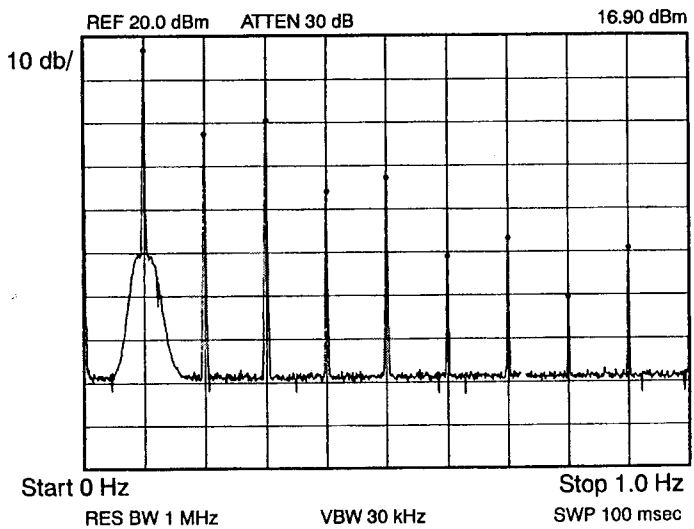


Fig. 8. Spectral analysis of the 100 MHz PM signal from the NIST standard.

PERFORMANCE OF ULTRASTABLE QUARTZ OSCILLATORS USING BVA RESONATORS

Jerry R. Norton
The Johns Hopkins University Applied Physics Laboratory
Johns Hopkins Road
Laurel, Maryland 20723
Telephone 410-792-5000

Summary

This paper addresses the frequency stability, and the effects of environmental stress on output frequency, of three ultrastable oscillators using BVA quartz resonators. Aging rates of $< 2 \times 10^{-12}$/24 h and Allan variances $< 4 \times 10^{-14}$ (10 s) have been achieved. One oscillator was designed and qualified for the TOPEX/POSEIDON spacecraft and has manifested excellent performance in orbit for 19 months. Performance data for the TOPEX/POSEIDON oscillator during flight qualification and in orbit are presented. An oscillator with a similar electronic design was used for a ground-based application. Both of these oscillator designs use 5-MHz resonators. A third oscillator being designed will have greatly reduced size and mass (≈19% of the TOPEX oscillator) and will use a 10-MHz resonator. The BVA resonators utilized in these oscillators are 3rd overtone, SC (stress compensated) cut devices.

BVA Resonators

A BVA resonator differs from a conventionally mounted resonator in two major ways: (1) A BVA resonator is fabricated by machining the mounting ring, connecting bridges, and critical vibrating center disk, which determines frequency, from a solid piece of quartz. (2) Metal electrodes for applying electrical signals to the resonator are not deposited directly on the face of the critical frequency-determining disk of a BVA resonator. These design features result in less stress between dissimilar materials (quartz and gold) than when the electrodes are deposited on the resonant area of the conventional resonator. A more detailed discussion of BVA resonator design is available in the literature /1,2/.

TOPEX/POSEIDON Oscillator

A new ultrastable oscillator using a 5-MHz BVA resonator was developed for the TOPEX/POSEIDON frequency reference unit (FRU). This oscillator demonstrated outstanding frequency stability (10-s Allan variance of 7×10^{-14}) and low aging rates ($< 2 \times 10^{-11}$ /24 h) during flight qualification /3/.

The TOPEX/POSEIDON spacecraft was launched into orbit in August 1992, and the oscillator has performed well. After increasing steadily for the first several months following launch, the aging rate has consistently decreased after going through a direction reversal and was -5.5×10^{-12}/24 h at the end of January 1994, as shown in Figure 1. This pattern of output frequency change has been observed in other spacecraft /4/. Spacecraft temperature and output frequency data are available, and the data correlate part of the time (note the flattening of frequency data around 7 April 1993). The data from orbit are averaged daily, and the spacecraft temperature could therefore vary with a period less than 24 h, which would adversely affect output frequency.

There is also a question of the effect of ionizing radiation on oscillator output frequency. The TOPEX spacecraft is in an orbit (1440 km) that is especially vulnerable to radiation, since it passes through the South Atlantic Anomaly (SAA), which is an area located roughly between the East Coast of South America and the West Coast of Africa that has a high concentration of charged particles. High-energy protons (> 80 MeV) in the SAA are the major cause of oscillator output frequency change due to radiation. The

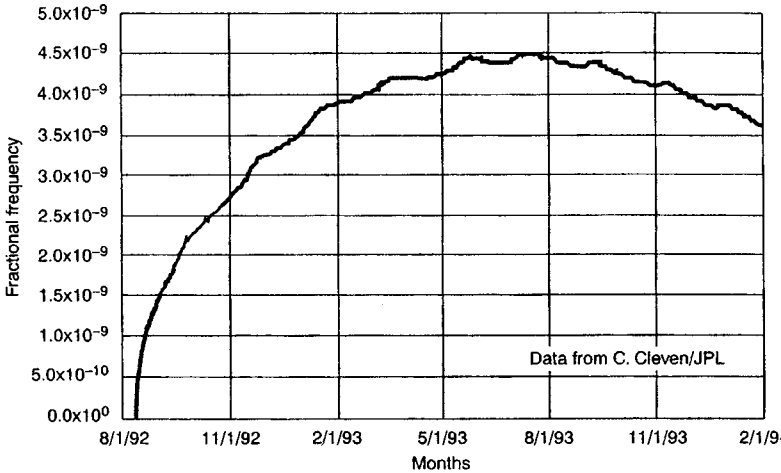


Figure 1. TOPEX ultrastable quartz oscillator in-orbit aging rate.

TOPEX spacecraft passes through the SAA at least once each day, and it could pass through parts of the SAA as often as three times a day. The predicted maximum accumulated radiation dose on each passage through the SAA is 0.6 rad(Si). Data presented in Figure 2 are from a ground test that simulates passage through the SAA and the resulting frequency change in response to the radiation. Frequency changed between 4.4×10^{-11} and 5.2×10^{-11} per exposure, which is very significant compared with an aging rate of 5.5×10^{-12}/24 h. Measurements of the actual radiation dose at the oscillator in orbit are not available; therefore, a definitive conclusion on the cause of frequency change cannot be reached; only the net frequency change can be measured.

Additional performance data for this oscillator during qualification testing are presented in Table 1. Allan variance and phase noise data are shown in Figures 3 and 4.

Ground-Based Oscillator

An oscillator with an electronic design similar to that of the TOPEX oscillator was repackaged for a ground-based application. This oscillator had the best 10-s Allan variance ($< 4 \times 10^{-14}$) and the best phase noise (−137 dBc/Hz @1 Hz from the carrier) of all the oscillators built at APL (see Figures 3 and 4). Additional data for this oscillator are shown in Table 1. The entire oscillator was placed in a sealed, thick-walled enclosure to provide isolation from atmospheric pressure and humidity changes. Unfortunately, the performance of this oscillator began to deteriorate approximately 4 months after it was built. The Allan variance deteriorated to $\sim 9 \times 10^{-13}$ (10 s) and remained at this value for almost 7 months before declining again to $\sim 1 \times 10^{-11}$ (10 s). There was little change in aging rate, which remained at $\sim 2 \times 10^{-10}$/24 h until the oscillator was disassembled to determine the cause for the decreasing Allan variance. The initial analysis indicates an unexplained problem with the resonator. Frequency jumps and changing frequency stability were previously reported /3/.

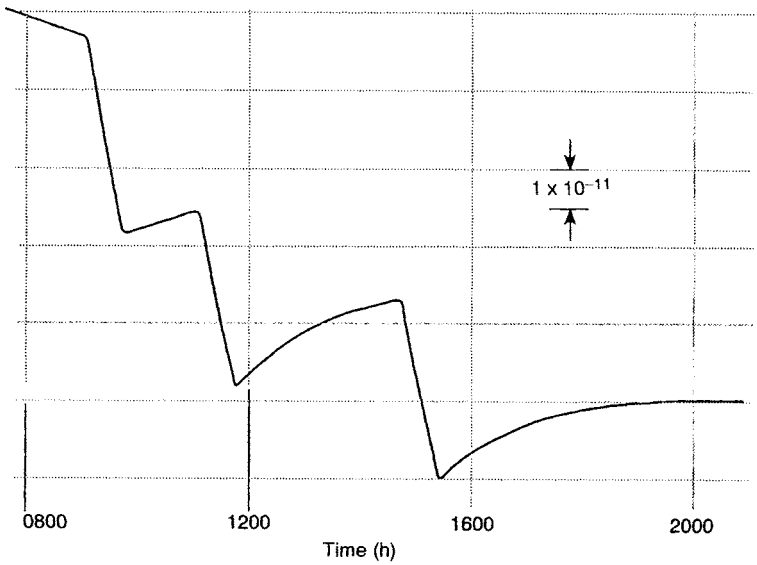


Figure 2. Frequency change in response to a simulated passage over the South Atlantic Anomaly.

Pluto Flyby Oscillator

A spacecraft mission to fly past Pluto is being planned. The trajectory of a spacecraft traveling to Pluto is difficult and a very large rocket is consequently required; therefore, spacecraft size, mass, and power consumption are critical. Design of an oscillator for this mission is under way with the goal of reducing the size and mass by at least 80% of that of the TOPEX oscillator while retaining a 100-s Allan variance of 1×10^{-13}. Good isolation from environmental effects such as temperature, magnetic fields, EMI, etc., must also be maintained. A 10-MHz tactical BVA resonator is the prime candidate for this design because of its smaller size and rugged internal construction. A test oscillator using a tactical BVA resonator has achieved an Allan variance of $< 2 \times 10^{-13}$ (100 s) with some periods having an Allan variance of $<1 \times 10^{-13}$ (100 s). Additional Allan variance and phase noise data from a test oscillator are shown in Figures 3 and 4.

This is an extremely challenging design; however, we believe the design goals can be achieved through circuit optimization and extensive mechanical changes. The projected mass of this oscillator is 0.32 kg, and the volume is 353 cm^3. Power requirements are projected to be < 0.8 W.

Radiation Sensitivity

The Applied Physics Laboratory has conducted comprehensive studies on the effects of both high and low levels of ionizing radiation on quartz resonators and oscillators /5,6/. These studies have disclosed that the magnitude and character of the radiation-induced frequency change depend on many parameters, including energy level, radiation dose, and dose rate. The radiation source was either omnidirectional, or

Table 1. Performance of APL ultrastable BVA oscillators.

	TOPEX Oscillator	Ground Oscillator	Pluto Oscillator
Output frequency	5-MHz	5-MHz	10-MHz
Aging rate/24 h	2.7×10^{-12}	8.10×10^{-11}	2.00×10^{-11}*
Allan Variance			
τ(s)			
0.01	3.78×10^{-12}**	NA	NA
0.1	5.39×10^{-13}	5.71×10^{-13}**	2.00×10^{-12}**
1	1.22×10^{-13}	8.38×10^{-14}	3.75×10^{-13}
10	7.97×10^{-14}	3.74×10^{-14}	1.27×10^{-13}
100	6.91×10^{-14}	6.09×10^{-14}	1.31×10^{-13}
1000	9.61×10^{-14}	NA	1.78×10^{-3}
Phase Noise			
Frequency offset (Hz)			
1	−131 dBc/Hz	−137 dBc/Hz	−118 dBc/Hz
10	−146 dBc/Hz	−149 dBc/Hz	−132 dBc/Hz
100	−156 dBc/Hz	−155 dBc/Hz	−142 dBc/Hz
1000	−158 dBc/Hz	−157 dBc/Hz	−148 dBc/Hz
10000	−158 dBc/Hz	−156 dBc/Hz	−150 dBc/Hz
100000	−160 dBc/Hz	NA	−150 dBc/Hz
Frequency as function of			
Temperature	-4.3×10^{-13}	NA	1.00×10^{-12}*
Acceleration [g]	1.4×10^{-10}	1.42×10^{-10}	1.50×10^{-9}*
Radiation [rad]	-4×10^{-11}	NA	1.00×10^{-10}*
Magnetics [G]	2×10^{-12}	7.00×10^{-12}	2.00×10^{-12}*
Load	2×10^{-12}	1.00×10^{-12}	2.00×10^{-12}*
Voltage	2×10^{-12}	3.00×10^{-12}	1.00×10^{-12}*
Output Characteristics			
Power level	6.99 dBm	6.7 dBm	0 dBm
Harmonic	−58 dBc	−59 dBc	−58 dBc
Spurious	−88 dBc	−85 dBc	−88 dBc
Power at 25°C (W)	1.37	1.48	0.8*
Weight (kg)	1.76	1.91	0.32*
Volume (cm^3)	1979	1744.7	353.2*

*Proposed values.

**Equal noise sources NOT assumed.

the radiation was applied perpendicularly to the face of the quartz resonator disk on the assumption that the resonator would be most sensitive when the large surface area of the disk face was exposed. During recent tests on 10-MHz SC-cut tactical BVA resonators, the direction of the radiation source was varied with surprising results. When the radiation source was applied parallel to the resonator disk face, we anticipated the radiation sensitivity would be smaller than when the source was applied perpendicularly. However, test results shown in Figure 5 are very similar in both magnitude and response shape. Tests were conducted on two oscillators. Mechanically the oscillators were made from different materials, and the configurations

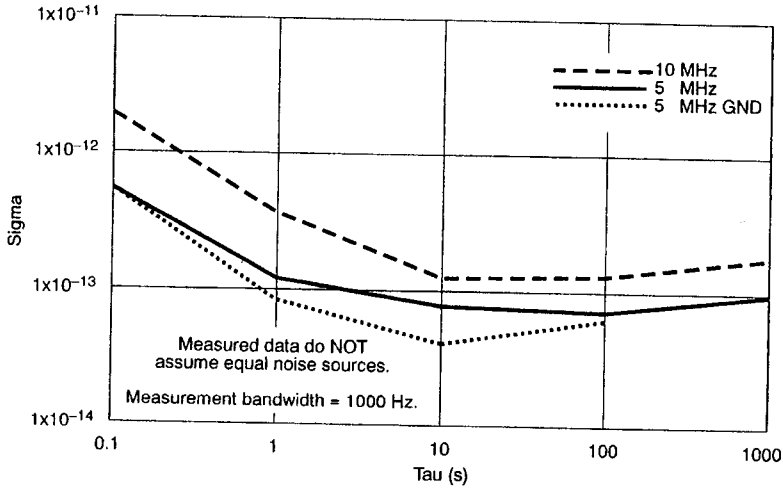


Figure 3. Allan variance of BVA oscillators.

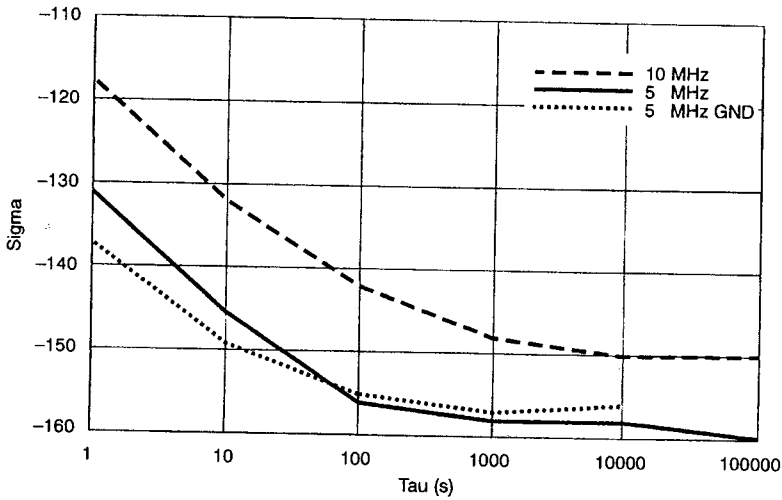


Figure 4. Phase noise £ $_{(f)}$ of BVA oscillators.

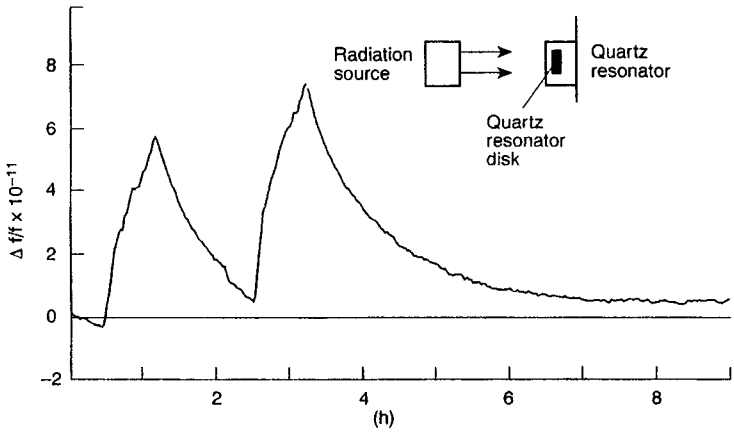


Figure 5A. Frequency response of BVA resonator SN1408 to radiation. Radiation source perpendicular to resonator.

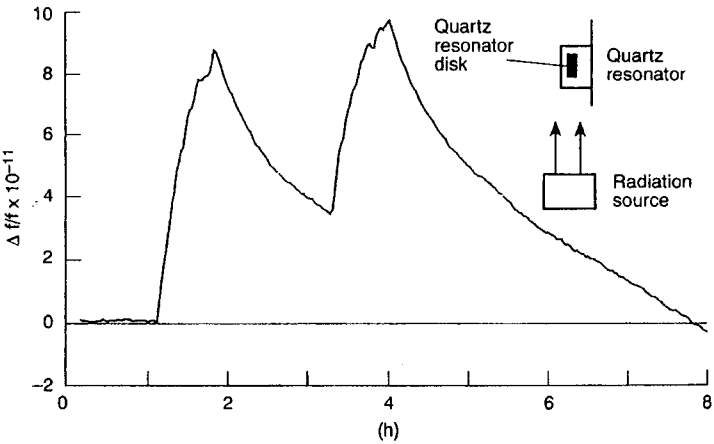


Figure 5B. Frequency response of BVA resonator SN1408 to radiation. Radiation source parallel to resonator.

were different. Schematically the electrical design was similar, but conventional components were used in one oscillator and chip components in the other. Even with these differences in oscillator configuration, the results from the radiation tests were quite similar.

Pressure/Humidity Sensitivity

Frequency changes induced by pressure and humidity variation in oscillators using conventionally mounted resonators have been reported /7/. The effects that weather-related pressure/humidity changes have on output frequency are shown in Figures 6 and 7 for BVA resonators. The output frequencies of two oscillators with the same design are shown in Figure 6. Oscillator 1 was exposed to atmospheric changes, but Oscillator 2 was in a sealed enclosure. The Oscillator 1 frequency has an obvious inverse correlation to pressure with a time lag of 8 to 10 hours, whereas the Oscillator 2 output frequency is unaffected by pressure. Daily output frequency of an oscillator with resonator SN430406.011 is presented in Figure 7. During the first 28 days, this oscillator was operated at atmospheric pressure, which resulted in a peak-to-peak frequency change of -7×10^{-10} to $+3 \times 10^{-10}$. The aging rate of the oscillator is masked by the large daily fluctuation. On day 29, the oscillator was placed in a vacuum chamber and the pressure was reduced to $<4 \times 10^{-4}$ torr. Output frequency decreased until about day 65, when a positive aging rate of 5×10^{-12}/24 h was established and continued for approximately 60 days. Again, the effect of pressure change on output frequency is obvious. Both the oscillator and the resonator were exposed to the same change in pressure; therefore, no definitive conclusion can be drawn regarding the individual effect of each

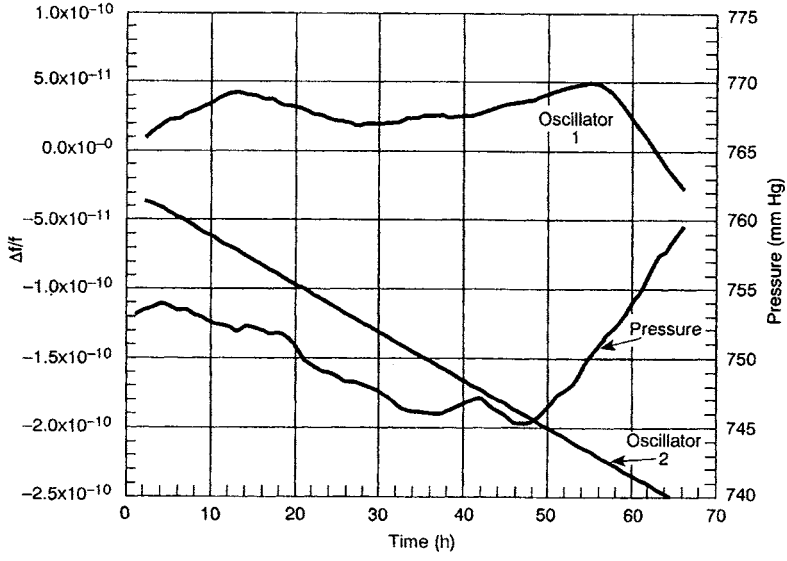


Figure 6. Atmospheric-pressure-induced frequency change.

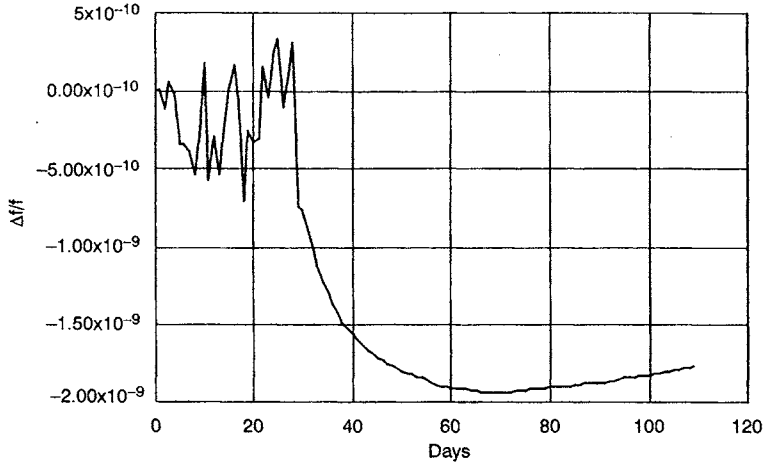


Figure 7. Daily frequency change of BVA quartz resonator SN430406.011.

component. Two test oscillators have been used to evaluate over 25 5-MHz BVA resonators. Some resonators are without doubt less affected by pressure changes than others. Only pressure data were available. Although humidity changes frequently occur with weather-front movements, these test were conducted in an area with the temperature controlled within a tolerance of ±2°C by a two-stage air conditioner, thus reducing the effects of humidity.

Conclusions

APL-designed ultrastable quartz oscillators using BVA resonators have demonstrated frequency stability and immunity to frequency changes caused by environmental stress second to none. Radiation-induced frequency change is becoming a major factor in overall performance of spacecraft oscillators, and progress in solving this problem is slow. The choice between a 5- or 10-MHz oscillator depends on the application. If size, weight, and input power are critical, a possible trade-off is to accept some degradation in frequency stability for a reduction in physical parameters that can be realized by using the smaller 10-MHz resonator. However, if the best frequency stability and phase noise are the design goals, a 5-MHz resonator should be used. Even with continuing improvements in quartz resonator design, the resonator is still the dominant player in oscillator performance.

References

/1/ Besson, R. J., "A New Electrodeless Resonator Design," in *Proceedings of the 31st Annual Symposium on Frequency Control*, June 1977, pp. 147/152.

/2/ Delaite, R., "Special Minaturised SC Cut Quartz Resonator for Severe Environments Applications," in *Proceedings of the 3rd European Time and Frequency Forum*, March 1989, pp. 296/300.

/3/ Norton, J. R., "BVA-Type Quartz Oscillators for Spacecraft," in *Proceedings of the 45th Annual Symposium on Frequency Control*, May 1991, pp. 426/430.
/4/ Rueger, L. J., Norton, J. R., and Lasewicz, P. T., "Long-Term Performance of Precision Crystal Oscillators in a Near-Earth Orbital Environment," in *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, Vol. 40 (1993) No. 5, pp. 528/531.
/5/ Norton, J. R., Cloeren, J. M., and Suter, J. J., "Results from Gamma Ray and Proton Beam Radiation Testing on Quartz Resonators," in *Proceedings of the 38th Annual Symposium on Frequency Control*, May 1983, pp. 63/72.
/6/ Suter, J. J., Bates, A. G., Cloeren, J. M., Norton, J. R., Schlueter, B., and Peier, U., "The Susceptibility of BVA-SC Resonators to Proton Ionization Effects," in *Proceedings of the Third European Frequency and Time Forum*, March 1989, pp. 11/21.
/7/ Norton, J. R., "Ultrastable Quartz Oscillator for Spacecraft," in *Proceedings 21st Annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting*, 1989, pp. 509/518.

Optimization of quartz crystal design for TCXO- and VCXO-applications

Eberhard Seydel , Ulrich Priess

KVG GmbH, R&D crystals, D-74922 Neckarbischofsheim , P.O.Box 61

1 Introduction

Activity dips influence changes in frequency and resistance behaviour versus temperature or versus load-capacitance variations. The basis for this discussion is the theoretical calculation of mode-charts and amplitude distributions for the vibrations of quartz-crystals with variable blank-geometry and variable electrode (material and geometry). In this paper causes of these effects will be described and possibilities of design variations to avoid activity dips will be shown.

2 Theoretical background and calculation

The calculation of vibration-modes of quartz crystal resonators with different dimensions and surface contours is based on the publication of Beaver /1/.
The starting point are the differential equations

$$h(\Psi_1+\Psi_{2,1}) - (\rho h^3\omega^2/D_{60})\,\Psi_1=0 \qquad (1)$$

$$[h(\Psi_1+\Psi_{2,1})]_{,1} - (\rho kh\omega^2/D_{60})\,\Psi_2=0 \qquad (2)$$

and the boundary conditions

$$\Psi_{1,1}=0 \qquad \text{at } x_1=\pm\, d/2 \qquad (3)$$

$$\Psi_1+\Psi_{2,1}=0 \qquad \text{at } x_1=\pm\, d/2 \qquad (4)$$

The differential quotients can be substituted by difference quotients.
i.e.

$$\Psi_{1,1}=({}_1\Psi_{i+1}-{}_1\Psi_{i-1})/h_i \qquad (5)$$

$$\Psi_{2,1}=({}_2\Psi_{i+1}-{}_2\Psi_{i-1})/h_i \qquad (6)$$

The function ${}_1\Psi_n$ describes the amplitude parallel to the surface (direction X1) at point number n and the function ${}_2\Psi_n$ describes the amplitude perpendicular to the surface (direction X2) at point number n (see fig.1)
These calculations can be done for different directions across the crystal blank.
In this way it is possible to calculate the vibrations of crystal resonators of various shapes (plano-parallel, bevelled, plano-convex, bi-convex, wedge-shaped), of different cut angles, with different electrode dimensions and at different temperatures.
The devised difference equations can be described as a band matrix with low occupation out of the band (coupling elements). The eigenvalues of the matrix represent the frequency of the vibrations and the eigenvectors represent the vibration amplitude.
The size of the matrix depends upon the resolution of the calculation (resolution n -> $(2n+2)^2$ - matrix elements) /3/, so for small or medium resolution calculations can be performed on PC's.

The program used for these calculations was self developed. It is written in FORTRAN and can run on any computer. Only the graphics output has to be adapted to the computer system.
The advantage of this approach over a finite element calculation is the simpler handling and the higher resolution in the considered direction /2//4/-/9/.
Therefore it is possible to find coupling modes of higher order than it would be with the FEM.
With this program variations in the amplitude distribution of each mode of vibration can be simulated as a function of the blank and electrode geometry, temperature, and cut angle. In this way predictions of crystal behaviour can be made.
The error in the calculations is approximately 5% and is accurate enough to perform a design trend analysis.
These calculations give us the possibility to present the vibration modes in different graphs :

mode chart versus diameter/thickness ratio
mode chart versus temperature
mode chart versus diameter of plane surface
mode chart versus diameter of surface contour
mode chart versus thickness of electrode
mode chart versus diameter of electrode

The mode charts show the coupling point of different modes and the conditions to avoid them.

amplitude distribution for calculated modes

The distribution of amplitude over the resonator allows us to determine the effects of the edge and the damping of vibration modes (e.g. inharmonic modes).

frequency shift of modes
ratio of amplitudes

The frequency shift of modes and the ratio of amplitudes shows dependencies on the geometrical parameters. This is an excellent instrument to optimize filters and oscillators and shows us e.g. the energy trapping conditions as a function of geometrical parameters.

Finally, with these calculations, the coupling points and amplitude-distribution can be determined as a function of the blank geometry and the electrodes. It is possible to determine the points of activity dips and the energy trapping conditions for blanks with given geometry.
With these calculations the quartz crystals can be optimized for different TCXO- or VCXO-applications and also for filter applications.
In the following part some of these possibilities of optimizing crystal design will be shown with some examples.

3 The influence of the X-tal-properties in TCXO's and VCXO's

Activity dips influence changes in frequency and resistance behaviour versus temperature or versus load-capacitance variations. It is very difficult to handle such behaviour for the temperature-compensation of TCXO's and the linearity of the pulling curve of VCXO's. The magnitude of these effects is a function of the measuring conditions, like drive level, load-capacitance and temperature.
We will consider these effects on the design of a 4 MHz crystal. This crystal was manufactured with different blank diameters in order to confirm the theoretically calculated values.
Fig. 2 , 3 and 4 show the mode charts vs. blank diameter for different bevel widths and constant electrode thickness.

The standard frequency is in this case the frequency of an infinite plate with electrode. From the diagramms one can see that the width of the bevels do not have a significant influence on the coupling in the considered region.
A greater effect can be achieved by varying the blank diameter.
By changing the blank diameter from 15 to 14 mm with a bevel of 2 mm (fig. 4) the region of the main resonance was uncoupled from the unwanted modes. The corresponding amplitude distributions (fig. 5) also show this effect clearly.
The main mode (fig.5 ❹) is coupled strongly to mode (fig.5 ❺) with a blank diameter of 15 mm. This coupling is much weaker for a blank with 14 mm diameter.
The measured frequency/temperature curves (fig. 6) of the sample crystals at different drive levels confirm the results of the calculations. The coupling effects at high drive level (1 mW) lead to a complete failure at certain temperatures for the 15 mm blank.

Summerizing it can be seen that :

the lower the order of the unwanted mode, the higher the degree of coupling.
the higher the amplitude of vibration (drive level), the higher the degree of coupling to unwanted modes and to the edge

With the 15 mm diameter crystals, the behaviour of frequency and resistance as a function of load capacitance at set temperatures was investigated. It could be seen that resistance changes were much larger than the frequency changes.
The largest influence on the thickness-shear mode did not originate from the frequency change but from the energy loss at the edge of the blank to the mounting system.

In fig. 9 the influence of temperature on the load capacity at which coupling occurs is shown. Here it can be seen how the occurence of coupling in VCXO's or TCXO's can change with temperature. Especially in VCXO's in which the crystals are driven over wide temperature and load capacitance ranges these coupling effects can lead to the failure of the oscillator.
The load capacitance - frequency curve can not only be deformed at the coupling points but can also show changes in the sign of the gradient. This can lead to oscillations of phase loops and failure of the system.
In cases where the main mode couples with the edge (fig. 6/14 mm) the resonance resistance increases over a wide temperature range resulting in a small distortion of the f/t - curve. The vibration energy flows due to poor energy trapping to the edge (mounting system).
On the other hand coupling with unwanted modes (e.g. fig. 6/15 mm) leads to an increase of resistance by many times within a temperature range of 5 - 10 degrees. In the same range the f/t curve is distorted and pertubations of several ppm are possible. The vibration energy is diverted through the unwanted mode. The behaviour of the frequency corresponds to that of a coupled pendulum.
The magnitude of the frequency dip depends on the order of the unwanted mode and therefore corresponds to the coupling degree to the main mode of the thickness-shear vibration.
The degree of coupling in turn depends on the difference quotient of this vibration distribution and therefore upon the vibration amplitude. The cause of drive-level-dependancy also lies partly in these effects.
These spurious modes lie very close to the main mode and have very small vibration amplitudes when they are not coupled with the main mode. For these kind of spurious modes it is not possible to apply the simple equivalent network model with uncoupled resonance circuits for the individual modes.

In order to describe the large distortions of frequency and resistance some degree of coupling between the resonances has to be assumed.Its magnitude evidently depends upon the amplitude of vibration i.e. on drive level.
By appropriate choice of blank diameter, surface contour, electrode diameter and thickness, it is possible to change the conditions of coupling.
In the case of high fundamental frequencies i.e. blanks with large diameter to thickness ratios, the designs for freedom of coupling can mean having to meet very tight manufacturing tolerances. It will then become necessary to make an exact prediction of regions free from coupling.
The here presented program is an appropriate tool for these predictions.

4 X-tals in crystal filters

The calculation of spurious modes (inharmonic and other modes) for filter crystals is another field of application. For example the frequency and amplitude behaviour of the inharmonic modes can be calculated for different electrode diameters and thicknesses (fig 10 to 13). In this way it is possible to find the best crystal design without tedious test series. From these diagramms it is very easy to find the best electrode diameters and thicknesses for optimal spurious suppression or frequency offset.
Fig. 14 and 15 show clearly the influence of the electrode thickness on the behaviour of the spurious modes for different electrode diameters. Fig. 16 shows the amplitude distribution for different modes and different electrode diameters. From these one can see the reflection effects of the edges and the influence of the electrodes.
It is also possible to determine the points for mode suppression.

5 Conclusions

The here presented program is a very powerful tool for developement and design of quartz crystals for various applications. Due to the simple handling and short calculation times compared to FEM it is possible to use this tool for each new design. By correct interpretation of the results, it is capable, especially for critical applications, to save valuable developement time.

Literature

/1/ W.D. Beaver
Design equations for bi- and plano-convex AT-cut resonators
27nd Annual Frequency Control Symposium 1973 p.11

/2/ R.Lerch
Simulation piezoelektrischer Körper mit der Finite-Element-Methode
Fortschritte der Akustik -DAGA '86 p.797

/3/ J.Stoer
Einführung in die Numerische Mathematik
Springer-Verlag 1979

/4/ K.Trümpy
Störungsrechnung mit finiten Elementen angewendet auf Schwingquarze
Congrès de chronométrie '86

/5/ Yook Kong Yong, J.R.Vig
Resonator surface contamination
IEEE transactions on ultrasonics, ferroelectrics and frequency control 36,4(1989)452

/6/ R.C.Peach
A normal mode expansion for piezoelectric plates and certain of its applications
IEEE transactions on ultrasonics, ferroelectrics and frequency control 35,5(1988)593

/7/ M.Naillon, R.H.Coursant, F.Besnier
Analysis of piezoelectric structures by a finite element method
Acta electronica 25,4 (1983) 341

/8/ D.C.L. Vangheluve
Finite element analysis of AT-cut crystals
32th AFCS 1978 p.134

/8/ D.C.L. Vangheluve
The frequency and motional capacitance of partial contoured crystal resonators
AFCS 1980 p.412

/9/ B.Dulmet
A general model for quasi thickness vibrations in contoured quartz plates
5th EFTF pp 164

$_1\Psi_n$ ⇔ parallel to surface at point number n

$_2\Psi_n$ ↑↓ perpendicular to surface at point number n

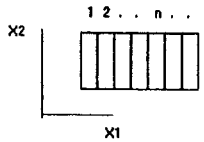


figure 1

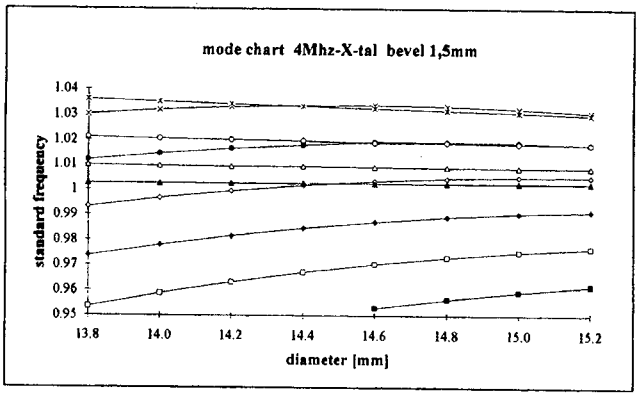


figure 2

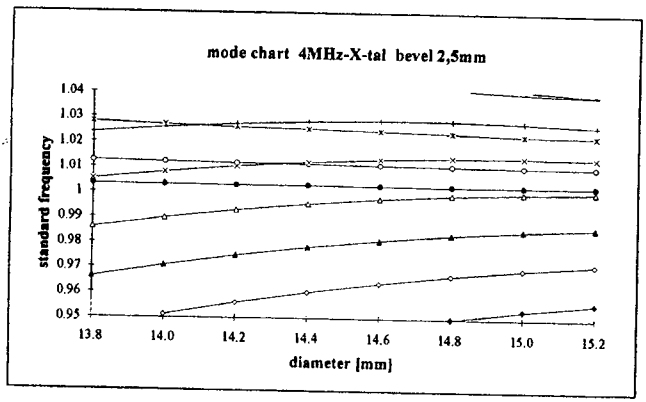


figure 3

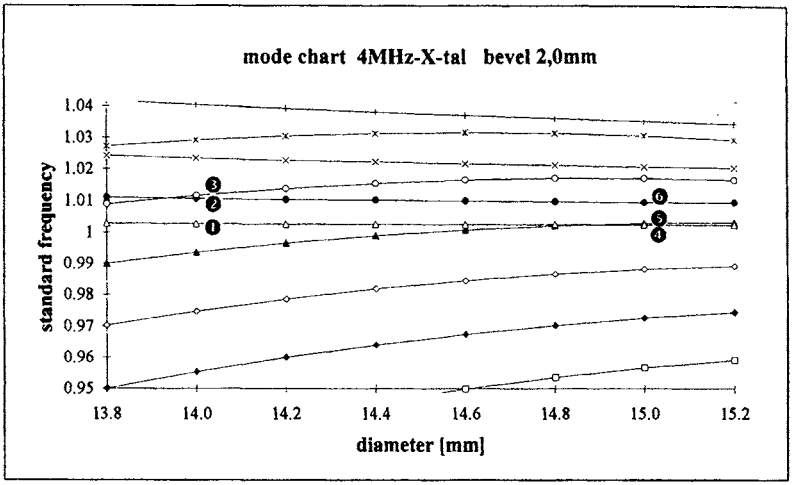


figure 4

amplitude distribution of diverse modes in figure 4

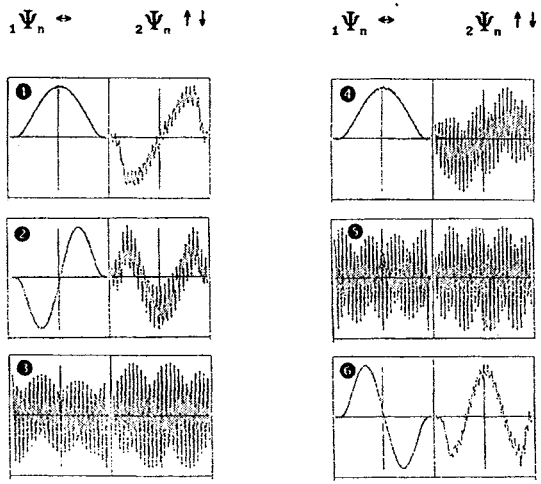


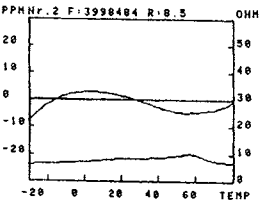
figure 5

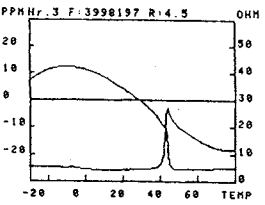
ft-curve at different drive level

diameter 14 mm

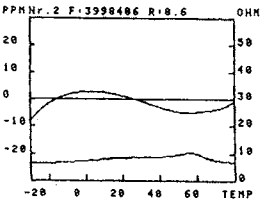
diameter 15 mm

0.01 mW





1 mW



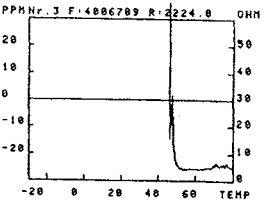


figure 6

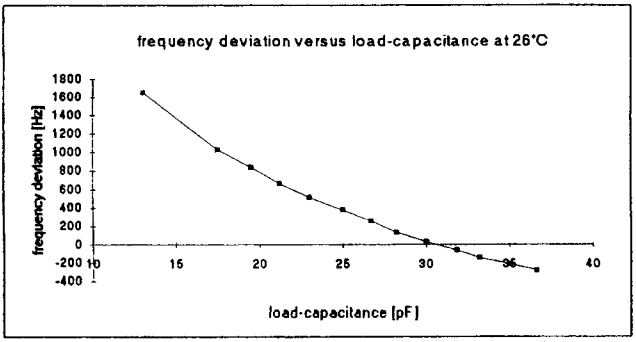


figure 7

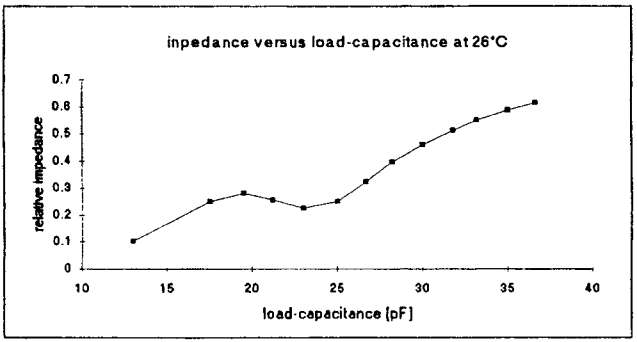


figure 8

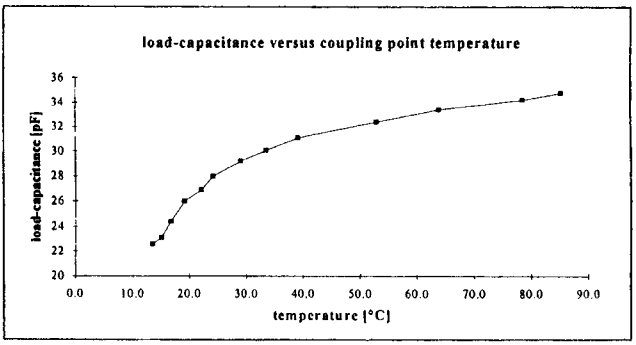


figure 9

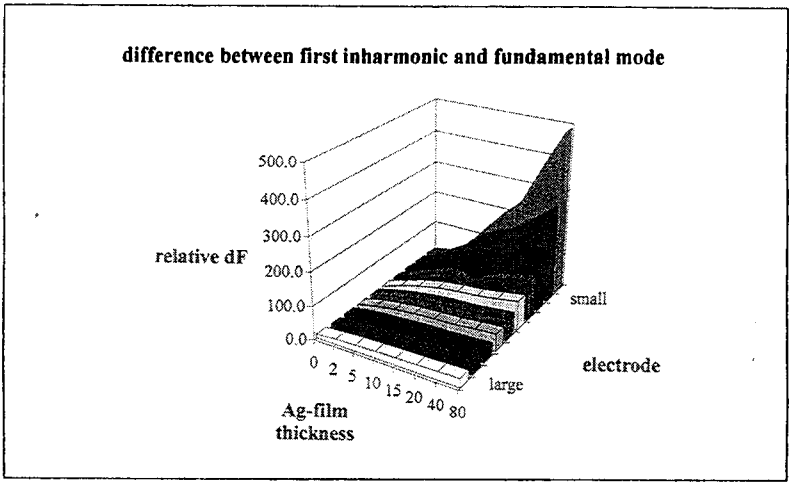


figure 10

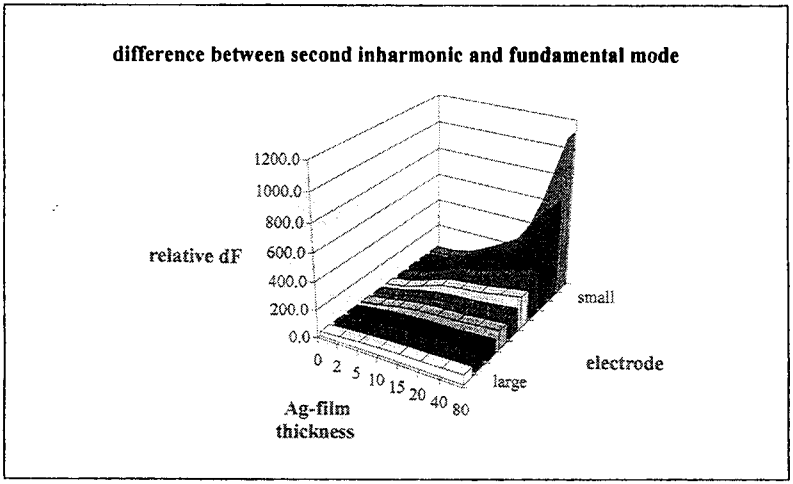


figure 11

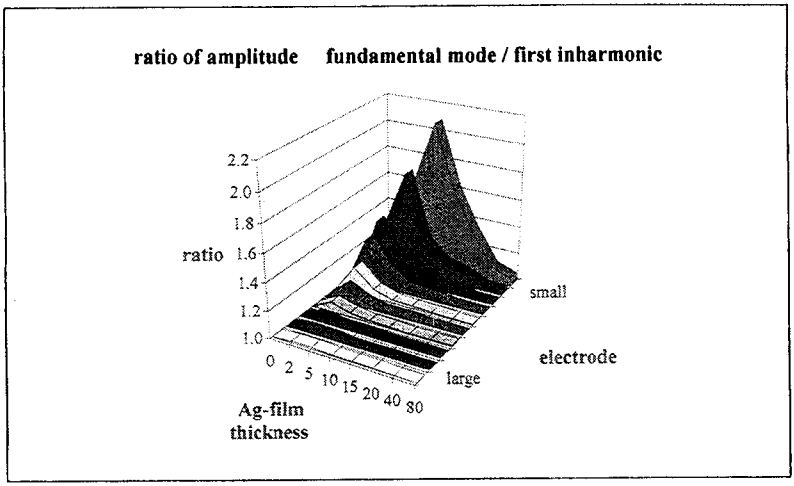


figure 12

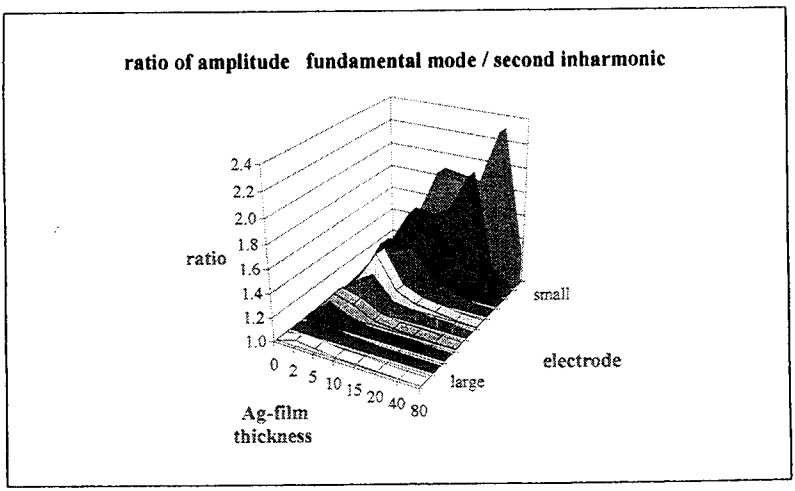


figure 13

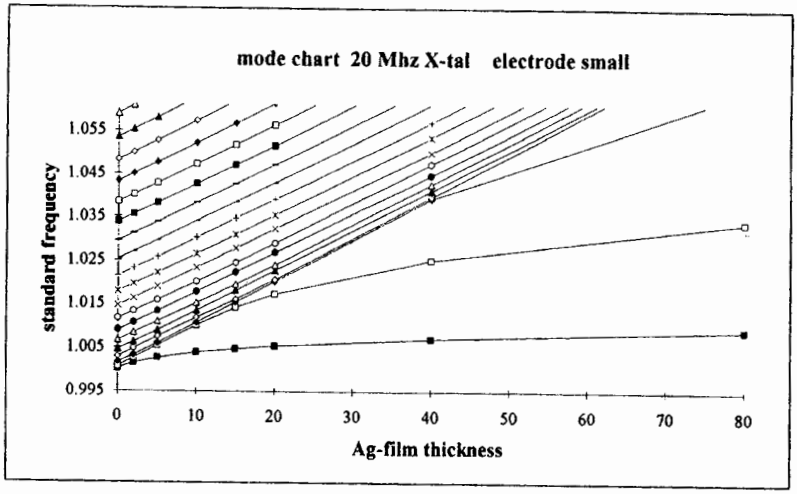


figure 14

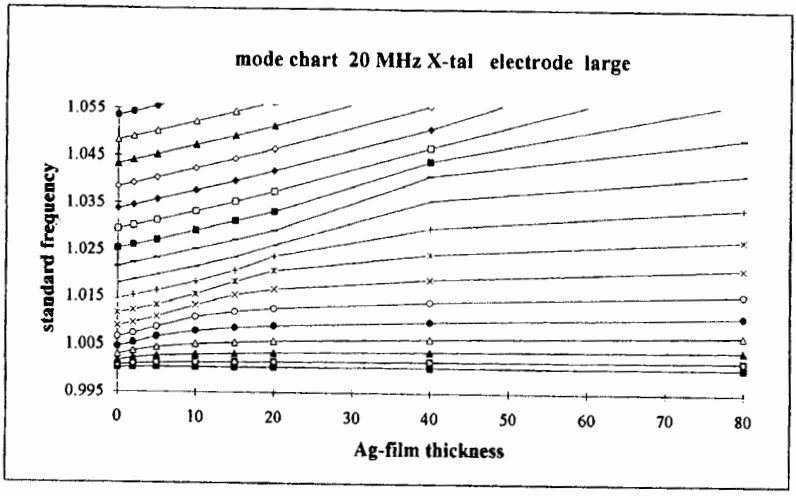


figure 15

amplitude distribution of harmonic , first and second inharmonic mode

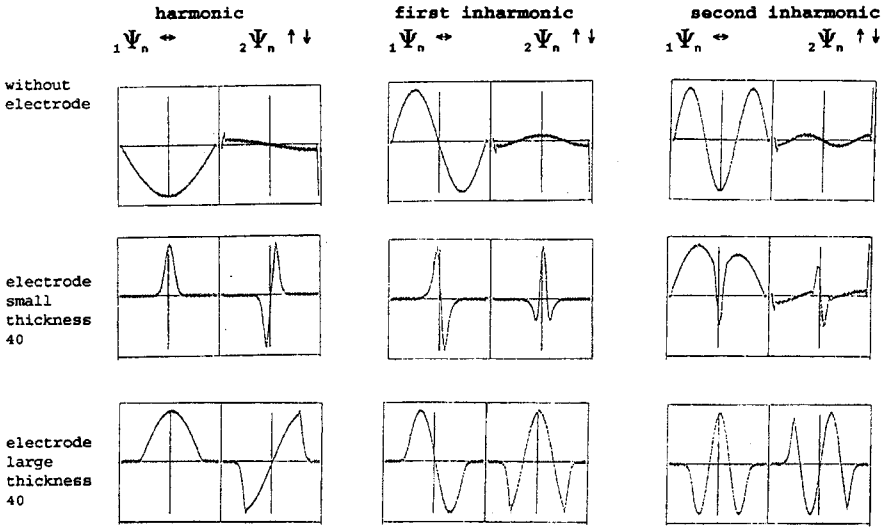


figure 16

SHORT LEAD TIME FOR THE DESIGN OF HIGH PERFORMANCE AIRBORNE OSCILLATORS

Ph. CATHELIN, B. NASREDDINE, P. CANZIAN, V. CANDELIER

THOMSON / CEPE
44, Avenue de la Glacière - BP 165
95105 ARGENTEUIL - FRANCE

1. ABSTRACT

Even in airborne applications, the customer requests a short lead time for the delivery of the first prototype. As the time to procure mechanical parts of a complex oscillator is quite long, no iteration is possible, the first prototype must be full-spec. For a new oscillator, every aspect must be under control, we have to watch over electrical, mechanical and thermal problems and interraction. The only way to achieve this challenge is to evaluate all the design choices with CAD tools.

The purpose of this paper is to describe a methodology, from the analysis of the customer's need to the delivery of the prototypes and to see which tool can be helpful, at each step, to avoid surprises on the finished oscillator.

The methodology followed at CEPE can be described on the diagram number 1. We will use this diagram as a guide within this article.

2. PRE-STUDY

The first part of the work is the most strategic one, it consists of the review of technical specifications. The problem is to announce the best performance to satisfy the customer and to insure the manufacturability of the product.

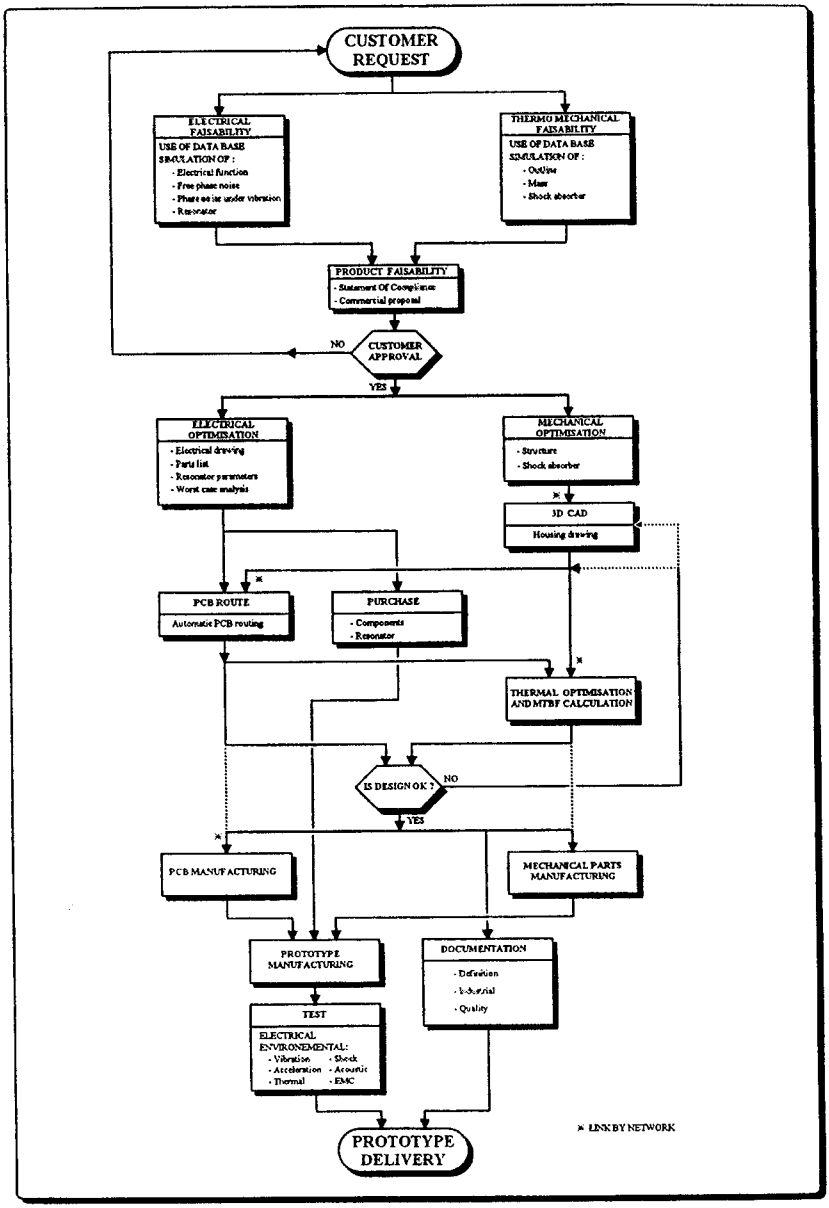
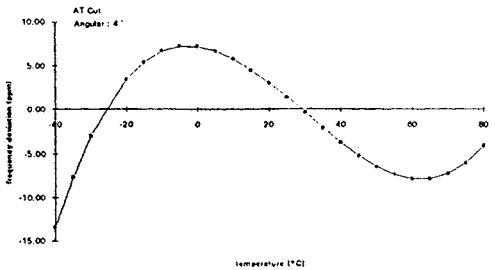

CUSTOMER REQUEST
ELECTRICAL FAISABILITY
USE OF DATA BASE
SIMULATION OF :
- Electrical function
- Free phase noise
- Phase noise under vibration
- Resonator
THERMO MECHANICAL FAISABILITY
USE OF DATA BASE
SIMULATION OF :
- Outline
- Mass
- Shock absorber
PRODUCT FAISABILITY
- Statement Of Compliance
- Commercial proposal
NO
CUSTOMER APPROVAL
YES
ELECTRICAL OPTIMISATION
- Electrical drawing
- Parts list
- Resonator parameters
- Worst case analysis
MECHANICAL OPTIMISATION
- Structure
- Shock absorber
3D CAD
Housing drawing
PCB ROUTE
Automatic PCB routing
PURCHASE
- Components
- Resonator
THERMAL OPTIMISATION AND MTBF CALCULATION
IS DESIGN OK ?
NO
YES
PCB MANUFACTURING
MECHANICAL PARTS MANUFACTURING
PROTOTYPE MANUFACTURING
DOCUMENTATION
- Definition
- Industrial
- Quality
TEST
ELECTRICAL
ENVIRONEMENTAL:
- Vibration
- Shock
- Acceleration
- Acoustic
- Thermal
- EMC
PROTOTYPE DELIVERY
* LINK BY NETWORK

DIAGRAM #1

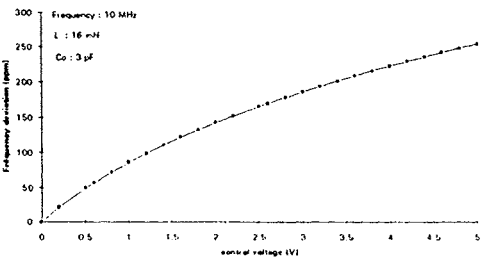
First, we have to take into account our experience. Is the oscillator we are about to design close to an existing one ? In this case, all results of the existing design can be taken as inputs for the study. All differences have to be confirmed.

2.1. Resonator and electrical circuit visualisation

We will begin with the visualisation of the resonator's parameters. We use specific tools, based in house specific made software, to give the correct cut angle for the thermal stability, and the motional parameters for the required frequency control characteristics (see the simulation's results on a VCXO on curves 1 and 2).



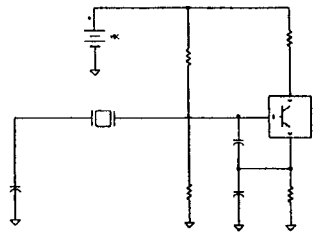
Curve # 1 : Frequency stability between -40 and 80°C



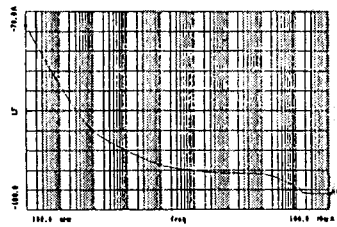
Curve # 2 : Frequency versus control voltage

We will now evaluate the phase noise of the free running oscillator. For this purpose, we use an electrical simulation software based on the harmonics balance principle. With some adjustments, this software is able to give the phase noise of the oscillator. The phase noise is calculated using LEESSON's formula, taking into account electronic parts

and crystal noise. The result of the simulation is iven on schematic number one, on curve number 3.

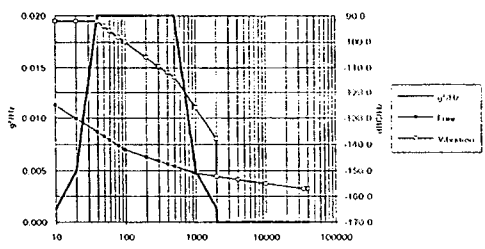


schematic # 1 : circuit description



Curve # 3 : Phase noise results

In addition to phase noise, it is possible to predict the behaviour of the oscillator under vibration. Using FILLER laws, software has been developed to give a view of the result.Curve number 4 gives the result of such a simulation of a 10 MHz oscillator mounted with a $1.10^{-9}/g$ crystal. The bold curve, relative to left scale, represents the random vibrations level. The other curves (right scale) are the free phase noise and the phase noise under vibration.



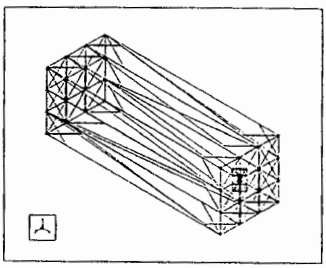
curve # 4 : phase noise under vibration

This software gives the required resonator's g sensitivity and, eventually, the need of absorbers. The software is able to combine the effects of the absorbers and the effects of the crystal. The cut-off frequency of the absorbers to meet the customer's need can be determined this way.

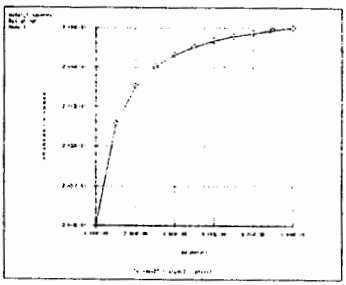
To complete the study, additional data is necessary. For standard components, you can consult manufacturer specifications. For the crystal especially for the g sensitivity and the noise, data bases have to be created.

2.2.Mechanical characteristics determination

At this stage, we know the resonator, the electrical schematics and the wanted mechanical frequency response. We now have to build the mechanics. At this step, we take the customer's outline drawing and make a basic representation of the oscillator, including the absorbers as shown on drawing number 1.



drawing # 1 : basic model : a box and absorbers



curve # 5 : calculation result : absorbers cutt off frequency versus absorbers rigidity

We can run a structure calculation of this model using finite element software until getting the expected mechanical response. At this stage the approximate shape of the oscillator, it's weight and the characteristics of the absorbers are determined. With curve number 5, for example, it is possible do determine the absorber's rigidity.
At this point we review the technical specification and compare it to the customer request ... and wait for his approval or the change of the technical need ; a change of the technical need may lead to a new design.

3. PROTOTYPE REALISATION

3.1. Study

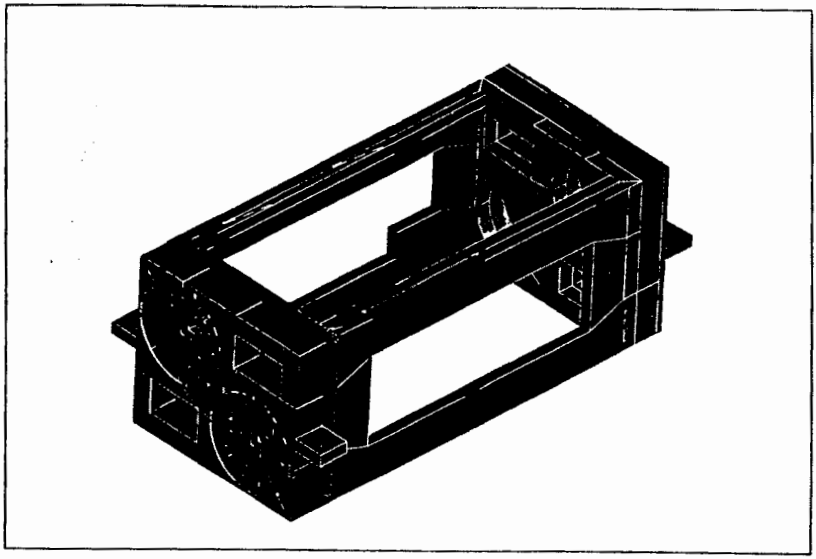
When the prototypes are ordered, the second phase, "study phase" can begin. It will be the continuation of the first phase : if the first phase has been well managed, this one will not cause any problem.

3.1.1. Electrical optimisation

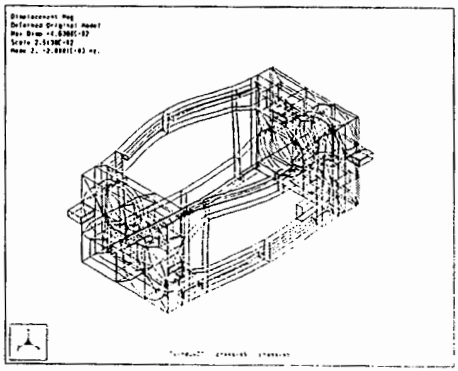
The electrical part will not be time consuming. The simulation software achieves sensitivity calculation, worst case analysis and wave form determination. These simulations yied relatively quickly to the final schematics and component values. At this point, it is possible to place orders for parts (components and resonator).

3.1.2. Mechanical optimisation

The mechanical part is important and must begin as soon as possible. During the first step, we simulated a box, we now have to built a complex mechanical. For this model, we include the external case, the cover(s), the PCB, and the significant weight components. We now run a calculation on this model. In avionics products, we have to avoid mechanical resonant mode in the vibration spectrum : the goal is generally to have a first mechanical resonant mode over 2000 Hz. One can see on drawing number 2 an example of housing and the result of the simulation, drawing number 3. The first resonant mode of this structure is at 2 800 Hz.



drawing # 2 : complex model of the external housing



drawing # 3 : visualisation of the first resonant mode

When the housing is correct, it can be sent, via network, to the 3D CAD facility in order to draw detailed plans (including screws, etc...). From these plans, it is possible to extract the outline of the PCB and give it to the electrical CAD facility were an automated circuit diagram can be constructed.

3.1.3. Thermo-mechanical optimisation

A special software can be used to determine the temperature of the different parts of the PCB. This software gives also the reliability (MTBF) of the circuit. The result of the temperature distribution is used to perform a thermo-mechanical calculation, using finite element calculation software. In the case of oven controlled oscillators, this calculation enables the optimisation of the dimensions of the oven to obtain the minimum consumption with the required temperature stability. Thermal behavior of ovenised oscillators can be determined from two analisies

- finite element
- electrical equivallent

The thermal behavior of absorberes, the role the different farts play in the thermal flow, ... have to be determined to get the complete oscillator thermal behavior in final configuration.

3.2. Definition review

All the calculation results are compiled and presented in the "definition review". The goal of this review is to validate the results and eventually decide to rework one or more step of the study phase. One has to keep in mind that since definition review is accepted, no iteration is permissible.

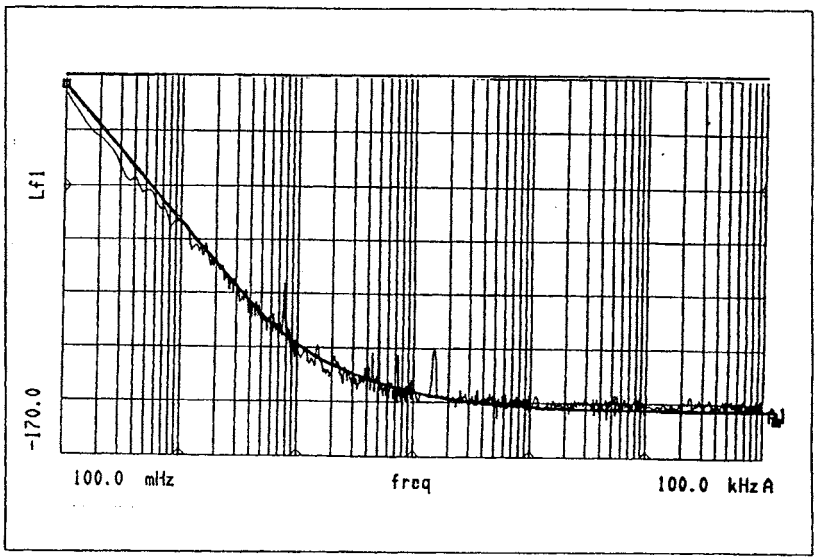
3.3. Prototype manufacturing

The third phase consists in the realisation of the prototypes. The mechanical parts and the PCB's are ordered and the.study results are gathered in three files (Definition, Industrial and Quality).

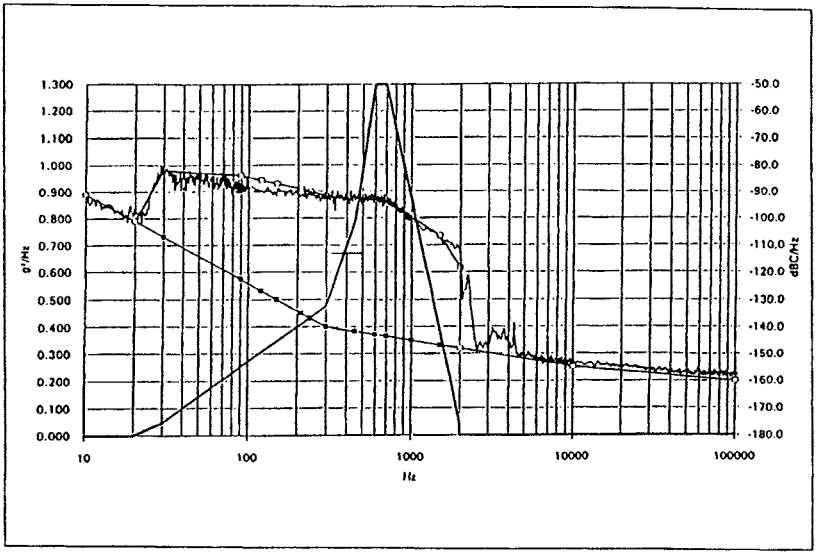
The next stages are the assembly and the adjustment of the oscillators. The measurements, electrical and environmental measurements ; can be performed, for example, measurement under vibrations, under vacuum, under moisture ...

4. Results analysis

It is now necessary to compare the measurements of the prototypes with the expected values from the simulations. The simulations are generally very close to reality. Yon can see, for example, on curve number 6 the comparison of the simulated phase noise of a 10 MHz OCXO, straight line, and the measured phase noise.



curve # 6 : Comparison between simulated and measured phase noise



Curve # 7 : Comparison between simulated and measured phase noise under vibrations

5- Conclusion

Airborne oscillators need an accurate optimization of electrical, mechanical and thermal design, in order to ensure the characteristics under the specified environmemental conditions, the reliability of the product and a competitive cost.

All of these design parameters being tightly connected a full spec oscillator will require a simultaneous optimization of all of them.

This optimization can be performed in a short lead time only at the expense of an efficient set of CAD step to determine the major characteristics of product behaviour and prevent design errors.

The paper has drawn the required CAD architecture. It demonstrated that the correct CAD tool gives reliable informations concerning the the prediction of the behaviour of

The paper has drawn the required CAD architecture. It demonstrated that the correct CAD tool gives reliable informations concerning the the prediction of the behaviour of the prototype making possible a one shot design fully compliant regarding the specification.

This methodology has been used to design a complex VCXO working in a severe environment. The result was good. The prototype has been delivered four months afer the order. This lead time includes two months for the manufacture of the mechanical parts. The prototypes were fully compliant at the first iteration.

Atomic Frequency Standards II (A 5)

PROGRESS ON THE LPTF OPTICALLY PUMPED PRIMARY FREQUENCY STANDARD.

E. de Clercq, G. D. Rovera, and A. Clairon
Laboratoire Primaire du Temps et des Fréquences
BNM-Observatoire de Paris
61, avenue de l'Observatoire, 75014 PARIS, France

Abstract

The results of the first evaluation of our optically pumped Cesium beam primary frequency standard are reported here. The short term frequency stability, measured with respect to an Hydrogen Maser, is characterized by $\sigma_y(\tau) = 6.7 \cdot 10^{-13} \tau^{-1/2}$, with a minimum level of $2.3 \cdot 10^{-15}$ for an averaging time of $10^5 s$. The first accuracy evaluation has resulted in an uncertainty of $1.1 \cdot 10^{-13}$ (1σ).

1 Apparatus

An optically pumped, thermal Caesium beam frequency standard is under development in our laboratory since a few years. The Cs tube has been described in more details elsewhere [1] and only main features are recalled here.

The Ramsey cavity has a drift length of 1 m. The atomic beam ($3 \times 8\,\mathrm{mm}^2$) goes through the cavity arms at a distance $\lambda/2$ from the terminal short circuits. The static magnetic field is directed vertically, transverse to the atomic beam direction, and covers the cavity length and the optical interaction regions.

The $4-4\,\sigma$ polarized transition of the Cs D_2 line is used for optical pumping. Detection is performed with the circularly polarized $4-5$ cycling transition. The use of a cycling transition weights the velocity distribution of the atoms by $1/v$ and yields to a better signal-to-noise ratio [2]. We use a single extended-cavity semiconductor laser locked to the 4-5 transition using the saturated absorption spectrum of an external Cs cell. A part of the laser beam is frequency shifted on the 4-4 transition by means of an acousto-optic modulator.

A block diagram of the servo system is shown in Fig. 1. The 5 MHz Quartz oscillator of the PTB multiplier chain [3] is locked to the 100 MHz coming from the maser with a second order loose loop (1 Hz). The $7.36\cdots$Mhz is generated by a commercial HP synthesizer with 1 mHz resolution connected to a personal computer via a IEEE bus. The computer also drives a microwave attenuator in order to set the microwave power at the desired level or to servo the power at the optimum level. The clock fluorescence signal, collected on a large area Si photodiode is then measured via an analog I/O board.

The modulation used is a slow square wave (1-2 Hz) with blanking time and it is performed in a totally asynchronous way by software. The error signal, given by the difference of the fluorescence signal in the two sides of the Ramsey central fringe, is digitally integrated by simple addition. In this way it is possible to realize a true integration that avoids all the problems of an analog integrator with finite gain. The value of the integrator

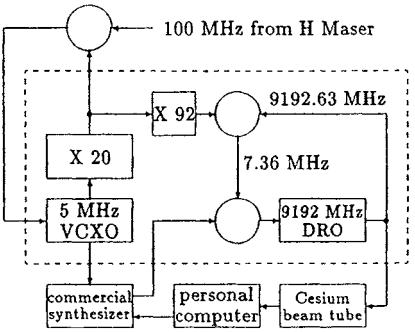


Figure 1: Block diagram of servo loop system. The dashed line encloses the PTB frequency multiplication chain.

register is then used to steer the central value of the synthesizer which is stored on the hard disk of the PC. By analyzing these data a posteriori, we can calculate the average frequency and variance of the measuring session. With this system we have to take care of the quantization noise of the I/O board and of the resolution of the synthesizer, but in both cases the quantization error is lower than other noise.

2 Stability

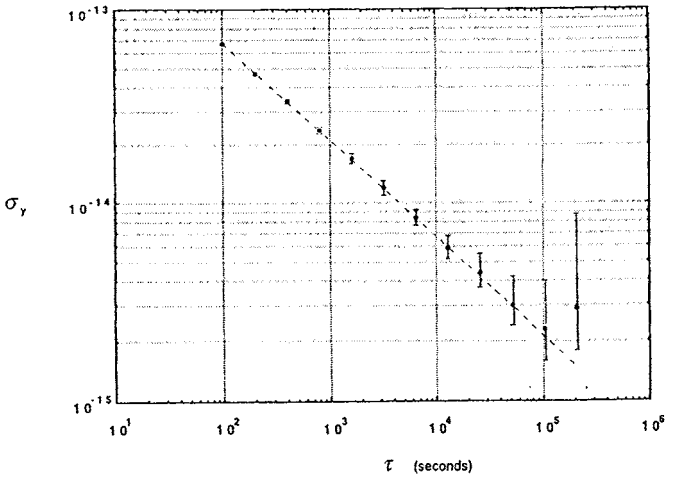


Figure 2: Overlapping Allan standard deviation.

The overlapping Allan standard deviation of our standard versus the Hydrogen maser is reported in Fig. 2. For sampling times of up to 10^5 s the frequency instability shows the

expected behaviour of a white frequency noise process. The measured data fit very well the line $\sigma_y(\tau) = 6.7 \cdot 10^{-13} \tau^{-1/2}$. The previously published [4] value $\sigma_y(\tau) = 5.8 \cdot 10^{-13} \tau^{-1/2}$ holds for the other direction of the atomic beam wich exhibits a higher Signal-to-Noise ratio. In these experiments the ovens temperature is $85°C$. To our knowledge, this short term stability is the best reported for a Cs standard and confirms the possibilities of optical pumping. The stability reaches a floor of $2.3 \cdot 10^{-15}$ for an averaging time of 10^5 s. This outstanding stability allows us to perform accurate frequency measurements in a very short time.

3 Frequency shifts

The first evaluation of our standard has been reported in [4] and is briefly reported here. The frequency shifts and their uncertainties are given in relative values. The quadratic Zeeman shift is calculated [5] from the linear Zeeman splitting between microwave lines (50.7 KHz) with an uncertainty of $3 \cdot 10^{-15}$ estimated from the C field inhomogeneity.

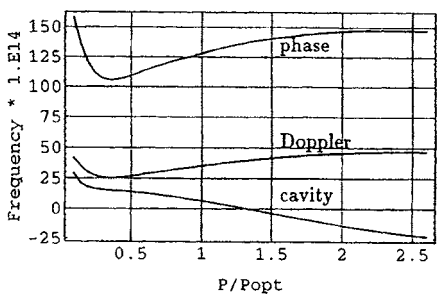


Figure 3: Microwave Power dependent shifts. Second order Doppler and cavity shift are represented with inverted sign.

The second order Doppler shift is calculated taking in account the microwave power, the modulation parameters and a Maxwellian weighted by $1/v$ velocity distribution, confirmed by fitting of experimental and calculated Ramsey patterns. As the real velocity distribution has not been measured, the 2^{nd} order Doppler uncertainty is conservatively estimated to $3 \cdot 10^{-14}$; The calculated shift versus the microwave power is shown in Fig. 3.

Considering the high Q (4500)and the large detuning (-1.2 MHz) of our cavity the calculation of the cavity pulling effect using the formula given in [6], making the approximation of small detuning, is no longer valid. The formula we actually use is:

$$\left(\frac{\delta_\nu}{\nu}\right)_c = \Lambda_c \frac{\nu}{2Q^2} \frac{1}{p} \frac{dp}{df} \tag{1}$$

which is one of the intermediate stages leading at the formula (9) given in [6]. In (1) p is the microwave power, Q the quality factor of the line and Λ_c a correction coefficient that takes into account modulation parameters. The relative variation of the microwave power versus frequency, around the operating point, is evaluated to be $-1 \cdot 10^{-6}$ per Hz, with an uncertainty of 5 %. Cavity pulling, as function of power is given in Fig. 3. Taking into account the present uncertainty in power measurement the cavity pulling effect is $-6.6 \cdot 10^{-14} \pm 1.5 \cdot 10^{-14}$.

The end-to-end phase shift is at present the biggest cause of uncertainty in our standard for the following reasons: firstly from our preliminary experiments on distributed phase shift we deduce a frequency shift equal to $71 \cdot 10^{-14}$ per mm of displacement of the center of gravity of the atomic beam. As the ovens are aligned with an accuracy better than 0.1 mm the uncertainty we have to account for this effect is $7 \cdot 10^{-14}$. The experiment is performed blocking a part of the atomic beam by a movable shutter in the vicinity of the oven exit, and then measuring frequency versus the shutter position. Secondly, the shift itself is as big as $128 \cdot 10^{-14}$ and therefore a 5 % uncertainty in determining microwave power can give a $2 \cdot 10^{-14}$ error. Finally, as the frequency measurements where performed with a $3 \cdot 10^{-14}$ uncertainty, the claimed total uncertainty for the phase shift is $8 \cdot 10^{-14}$. The end to end phase shift measured by beam reversal is then $128 \cdot 10^{-14} \pm 8 \cdot 10^{-14}$ at optimum microwave power.

Use of laser beams in a Cs clock can be the origin of new frequency biases via the light shift, the modification of the atomic trajectories, or via unexpected effects. The light shift due to the fluorescence light of the beam is calculated in our case to be less than $1 \cdot 10^{-15}$ using the results of Oshima et al [7]. However the light shift due to stray light is not a priori known, because the amount and direction of stray light propagating in the Ramsey cavity region is not known. In order to check for a possible stray light effect and other light related effects we have undertaken a set of frequency measurements varying the power of the pumping or detection beam, laser frequency, polarization of the detection beam, angle between laser beams and atomic beam (of a few milliradians), and using travelling waves instead of standing waves for the pumping or the detection beams. At present we have not found a simple correlation between optical parameters and the measured frequency, so we use, as an evaluation of uncertainty, the standard deviation of the ensemble of all measurements in both beam directions, that is: $6 \cdot 10^{-14}$.

The relative asymmetry of the neighbouring lines is always smaller than 1 % in our case. The Rabi pulling calculated [8] with the operating conditions for a such asymmetry, with the realistic assumption that the velocity distribution of the atoms in the m ± 1 sublevels is also Maxwellian multiplied by $1/v$, is $0.3 \cdot 10^{-14}$. We take this value as the uncertainty in order to be cautious.

Shifts due to the electronic system have been tested using two different servo systems, one is totally analogic and was borrowed from PTB, the other one is the digital system presented in this paper. The difference between the two systems is found to be $3 \cdot 10^{-14}$ and this value is taken as the uncertainty.

The shift due to the black-body radiation, equal to $-1.8 \cdot 10^{-14}$, is calculated from the temperature measured on the Ramsey cavity inside the vacuum vessel. The associated uncertainty is overestimated to be cautious. All other shifts are expected to be smaller than 10^{-14} and are not taken into account for this evaluation.

4 Total accuracy evaluation

Nearly all the frequency shifts depends on the microwave power actually experienced by the atoms. Studying the frequency shift versus the microwave power is a way of checking if the different frequency biases are correctly evaluated. Fig. 4 shows the measured relative frequency versus the normalized microwave power. The solid lines are calculated taking into account the second order Doppler effect, cavity pulling, and a 0.30 mrad end to end phase difference. The upper (lower) curve is for W-E (E-W) beam direction, middle curve is obtained excluding the phase shift. The theoretical curve fits very well the experimental data for the W-E direction. The agreement in not so good for the E-W direction. This

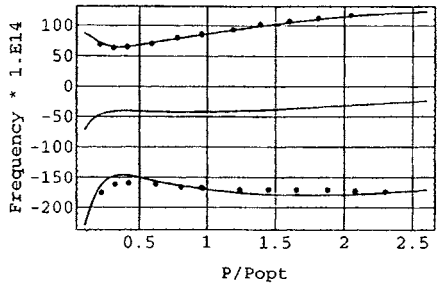


Figure 4: Microwave Power shift

Type of Frequency Bias	Fractional Frequency Bias $\times 10^{-14}$	Fractional Uncertainty $\times 10^{-14}$
C Field	24020	0.3
Second order Doppler	−35	3
Cavity phase shift	128	8
Cavity pulling	−6.6	1.5
Rabi pulling	±0.3	.3
Unexplained Shifts	–	6
Electronics	–	3
Black body	−1.8	0.1
$\sqrt{\Sigma x_i^2}$	–	11

Table 1: Systematic major frequency shifts in the LPTF OP primary frequency standard and uncertainties in parts in 10^{-14} at optimum power.

discrepancy is not explained at present time.

Finally the different considered biases are summarized in Table 1 with their uncertainty. The final accuracy (1σ) is given by the quadratic sum of the uncertainties: $\sigma = 1.1 \cdot 10^{-13}$. Our standard was compared with TAI [9] at the mid of may 93. The difference between the relative frequencies of the LPTF standard, taking into account the black body shift, and of TAI time scale was:

$$\nu_{cs} - \nu_{TAI} = 11 \cdot 10^{-14} \pm 13 \cdot 10^{-14} (1\sigma) \qquad (2)$$

5 Conclusion

This paper describes the present characteristics of our standard. The measured short term frequency stability is less than $7 \cdot 10^{-13}$ The first evaluation of the accuracy of our standard is reported here, it is estimated to $1.1 \cdot 10^{-13}$. The limiting biases could be reduced after modification of some hardware parts and further investigations, pushing then the accuracy to few parts in 10^{14}.

6 Acknowledgment

Helpful discussions with A. De Marchi are gratefully acknowledged. We would also thank G. Freon (LPTF) and C. Thomas (BIPM) for the comparison with TAI.

References

[1] E. de Clercq, G. D. Rovera, S. Bouzid, and A. Clairon, "The LPTF optically pumped primary frequency standard," *IEEE Trans. Instrum. Meas.*, vol. 42, pp. 457-461, April 1993.

[2] N. Dimarcq, V. Giordano, P. Cerez and G. Theobald, "Analysis of the noise sources in an optically pumped cesium beam resonator," *IEEE Trans. Instrum. Meas.*, vol. 42, pp. 115-120, April 1993.

[3] R. Schröder, "Frequency synthesis in primary cesium clocks," *Proc. 5th European Frequency and Time Forum*, 1991, pp. 194-200.

[4] G.D. Rovera, E. de Clercq, A. Clairon: "An Analysis of major frequency shifts in the LPTF optically pumped primary frequency standard", in *Proc. 1993 IEEE Int. Frequency Control Symposium*, 2-4 June 1993, pp 75-80.

[5] Y. Koga, "An improved metod of measuring the Zeeman shift in a Cesium beam frequency standard," *Jpn. J. App. Phys.*, vol. 23, pp. 97-100, January 1984.

[6] A. de Marchi, G.D. Rovera and A. Premoli, "Effects of servo loop modulation in atomic beam frequency standards employing a Ramsey cavity," *IEEE Trans. Ultrason. Ferroelec. Freq. Contr.*, vol. 34, pp. 582-591, November 1987.

[7] S. Oshima, Y. Nakadan, T. Ikegami and Y. Koga, "Light Shift in an optically pumped Cs beam frequency standard," *IEEE Trans. Instrum. Meas.*, vol. 40, pp. 1003-1007, Dec 1991.

[8] A. De Marchi, G.D. Rovera, A. Premoli: "Pulling by neighbouring transitions and its effects on the performance of Caesium beam frequency standards" *Metrologia*, vol.20, pp.37-47, 1984.

[9] BIPM, Circular T 66, 26 July 1993.

The PTB Primary Atomic Clocks CS3 and CS4: Determination of Further Uncertainty Contributions

A. Bauch, T. Heindorff
Physikalisch-Technische Bundesanstalt
Bundesallee 100, D-38116 Braunschweig

Summary

At the PTB two primary cesium atomic clocks with a vertical atomic beam have been constructed and put into operation in the last years. The uncertainty contributions traditionally considered as being of relevance have been determined. Instead of touching several effects in short, the method of determining the second-order Doppler effect, Rabi pulling and power shift is laid down in some detail in this report.

1 Background information

All PTB primary clocks follow a common design principle which facilitates the determination of the systematic shifts of the clock transition. It is worth mentioning here the axially symmetric atomic beam, 3 mm in diameter, the longitudinal static quantization field and, in particular, the use of magnetic lenses for the state selection and velocity filtering. In the clocks, CS3 and CS4, atoms with velocities as low as 70 m/s contribute to the clock transition signal. The vertical direction of the atomic beam helps to avoid the deflection of the atoms by gravitational forces and to retain axial symmetry [1, 2]. The amplitude of the microwave magnetic field in the two end sections of the microwave cavity varies along the beam direction. Thus the Rabi frequency b(t) (as defined in [3]) which describes the atomic interaction with the microwave field has the time dependence

$$b(t) = b_c \begin{cases} \sin(\pi t / \tau) & 0 \le t \le \tau \\ 0 & t < 0,\ \tau < t < \tau + T,\ t > 2\tau + T \\ \sin(\pi(t - T - \tau)/\tau) & \tau + T \le t \le 2\tau + T \end{cases} \qquad (1)$$

while the atoms traverse the cavity structure.

Here b_c is the maximum of b(t) in each interaction region, $\tau = \ell/v$ and $T = L/v$ are the flight time of the atoms through the interaction region and the drift region, respectively, and v is the atomic velocity. It is $L = 0{,}7677$ m and $\ell = 23{,}26$ mm.

2 Second-order Doppler effect

2.1 Simple calculation

For a monokinetic atomic beam of velocity v_0, the line centre of the Ramsey resonance would be shifted towards a lower frequency by

$$\delta f_D := -\left[v_0^2 / \left(2 \cdot c^2\right)\right] \cdot f_0 \qquad (2)$$
$$(f_0 = 9\,192\,631\,770\ \text{Hz})$$

The relative frequency correction to be applied is defined as $F_D = -\delta f_D / f_0$. In a simple approach we proceed as follows: The line width of the central fringe of the resonance is given by $W = v_m / (2 \cdot L)$, where v_m is the mean velocity of the atoms contributing to the clock transition signal. Replacing v_0 in (2) by v_m and using the experimental value, $W = 42{,}5$ Hz, we obtain $F_D = 236 \cdot 10^{-16}$. The numbers given are examples for one particular beam direction in the CS4 at a given epoch.

2.2 First refinement

P. Thomann [4] pointed out that the line width W is approximated with higher accuracy by

$$W = v/(2L_T) =: v/(2(L + 1{,}20 \cdot \ell)) \qquad (3)$$

Thus the result would be $F_D = 254 \cdot 10^{-16}$.

2.3 Experimental procedure

Fig. 1a shows a full Ramsey resonance curve of the clock transition, recorded with the CS4, Fig. 1b shows the central fringes. The line centre is determined by low square-wave frequency modulation with a modulation width $2 \cdot f_H$ = 44 Hz ≈ W. To simplify the experimental procedure, it is not W, but W', as defined in Fig. 1b, which is measured and used to calculate F_D. Many fringes are visible in Fig. 1a due to the small width of the velocity distribution. Even in this case, W' differs from W and, furthermore, W' depends on f_H as will be proved later.

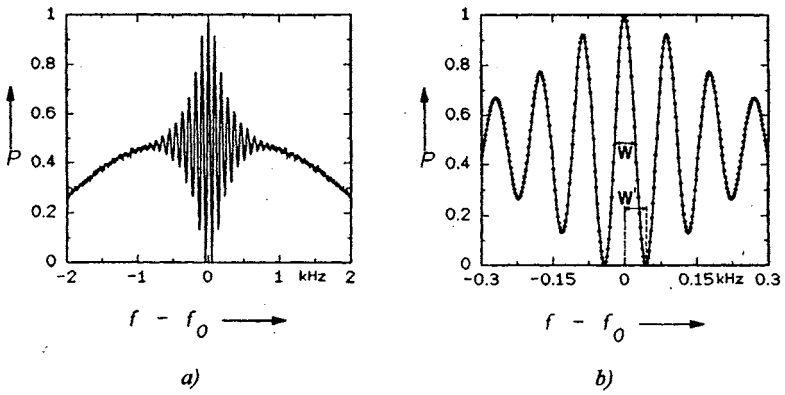


Fig.1: Ramsey resonance signal of the clock transition in the CS4; a) original data, plotted here as normalized transition probability P, b) central fringes, solid line: same data as in a), dots: calculated transition probability (see section 4).

2.4 The aim

F_D must be unambiguously expressed by the quantity W', measured with a precision of 0,1 Hz at a specified value of f_H. The uncertainty of F_D is to be

less than $1 \cdot 10^{-15}$. A simple analytical expression is needed for to facilitate the routine determination of F_D.

3. Determination of velocity distributions

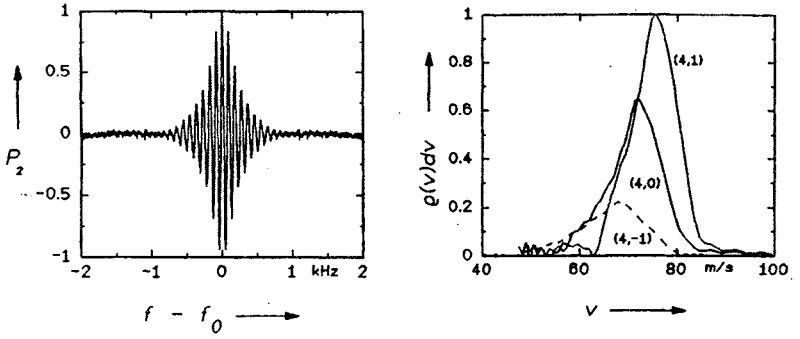


Fig.2: a) Data like in Fig. 1a after subtraction of the Rabi pedestal and normalization for use as input data for a Fourier inversion according to eq. 4.; b) normalized velocity distributions ρ(v)dv of atoms in the initial Zeeman states (4,0) and (4,±1). The area under the curves is proportional to the total number of atoms in the corresponding state.

3.1 ρ(T)dT

In case of a time-dependent microwave field amplitude b(t) (1), the transition probability $P(f - f_0)$ and thus the line shape of the Ramsey resonance (Fig. 1a) cannot be expressed analytically. The Rabi pedestal is approximated by eq. 5.2.43 in [3] (valid for constant microwave field amplitude) with an effective length $\ell' = \ell \cdot 7{,}03/9{,}56$ and a value for τ taken from a reasonable estimate of the mean velocity v_m. Finally the Rabi pedestal is subtracted. A pattern like the one in Fig. 2a is the result. It is used as an input data set (2 Hz step size, 2048 data points, 0 - 4096 Hz) for a Fourier inversion of the transition probability

$$P_2 = \frac{1}{2}\int_0^\infty \rho(T')\sin^2(b\tau)\left(1+\cos(\Omega_0 T')\right)dT' \qquad (4)$$

where $\rho(T)dT$ is the time-of-flight distribution of atoms moving through the section of length L_T and $\Omega_0 = 2\pi(f - f_0)$. For the Fourier inversion the term $\sin^2(b\tau)$ is approximated by 1. This is justified if the resonance curve is recorded at a microwave power which yields the maximum resonance signal.

3.2 $\rho(v)dv$

The distribution of atomic velocities, $\rho(v)dv$, follows straightforward from $\rho(T)dT$, if L_T is known. It has been found out, in agreement with [4], that L_T is slightly dependent on the shape of $\rho(v)dv$. Thus, L_T had to be determined by iteration. The adopted value of L_T is considered as a correct one, if the mean square deviation between the input Ramsey curve (Fig. 1a) and a calculated one (as described subsequently) is a minimum. In Fig 1b the dots indicate the agreement achieved, if for L_T

$$L_T = L + 1{,}15 \cdot \ell \qquad (5)$$

is valid (see also Fig. 3).

The steps mentioned above were carried out for the Ramsey resonance signals of the three neighbouring transitions connecting Zeeman substates with $m = 0, \pm 1$. Fig. 2b shows a plot of the three velocity distributions of atoms in the initial states (4,m), $m = 0, \pm 1$, which were obtained in this way. Concerning the total population of the states - expressed by the area under the curves - the differences are a consequence of the large velocity selectivity of the magnetic state selection.

4 Line shape calculation

4.1 Procedure

The method described in [3] is applied in the numerical calculation of the transition probability $P(f\text{-}f_0)$. The amplitude b(t) in each interaction zone is decomposed into p segments of constant amplitude. For each of these segments the integrated analytical solution is given by a 3 x 3 matrix. The final solution for one velocity value is the product of p x 1 x p matrices (the one for

the drift region). To obtain sufficient accuracy of the calculations p = 40 is a proper value. The velocity distributions are approximated by step functions of 100 steps and the sum of the individual contributions is formed. Figs. 1b and 3 show results of comparisons of input data (Fig. 1a) and calculated values of the transition probability P.

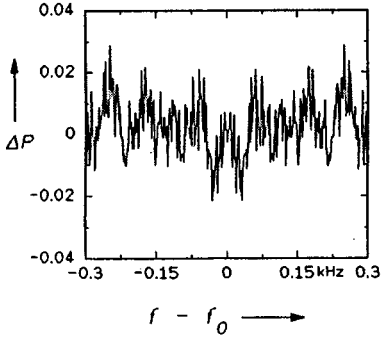


Fig. 3: Difference ΔP between input data and calculated values of P, shown in Fig. 1b. ΔP is mainly determined by the beam noise in the CS4, as the integration time of the beam signal in the analog-to-digital converter was only 128 ms.

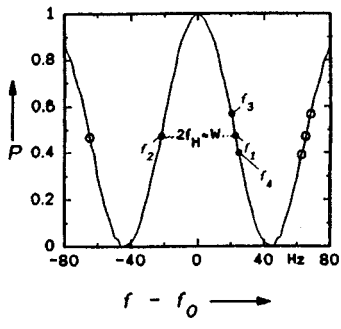


Fig. 4 Central fringes of a Ramsey pattern used to compute its line centre for $f_H = 22$ Hz (●) with the help of the P-values at $f = f_i$, i = 1,...,4. The line centre for $f_H = 66$ Hz is determined in the same way (O).

4.2 Determination of the line centre and F_D

To determine the line centre of a calculated resonance curve, $P(f - f_0)$ is calculated for four f-values which are indicated in Fig. 4. With these values it is possible to calculate the difference, $P(f_1 - f_0) - P(f_2 - f_0)$, and the steepness at $f = f_1$ and thus the shift of the line centre from $f = f_0$.

The first utilization of this procedure was to determine F_D. The result was $F_D = (267 \pm 5) \cdot 10^{-16}$ where the uncertainty was estimated from program runs with a different treatment of the input data file (integration limits in the Fourier transform, step size in ρ(v)dv, p). At the time of the recording of the resonance

curve with the CS4 clock W' was measured to be 43,5 Hz (2 f_H = 44 Hz). The result implies that

$$F_D = (1{,}410 \pm 0{,}026) \cdot 10^{-17} \cdot (W')^2 \qquad (7)$$

is valid. This is the relation asked for in section 2.4. W' has been found to be constant within ±0,1 Hz over many months.

4.3 Further studies

In order to avoid doubts that the adopted numerical procedure accounts for the physical reality in the clocks several checks have been done. It is easy to change f_H and the microwave field amplitude b_c in the calculations and, fortunately, in the flexible electronic system in operation in the CS4 [5], too. Fig. 5 shows a comparison of measurements and calculations of W' in dependence on f_H and on b_c. The small systematic offset seen in the data should not surprise. The measurements (open symbols) were done several months after ρ(v)dv had been calculated from the resonance curve as it was previously described.

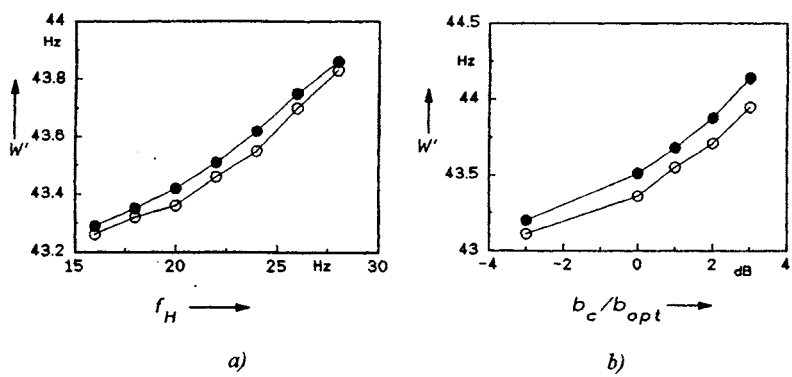


Fig. 5: Comparison of measured (open symbols) and calculated values of W' in dependence on a) the width of the square-wave frequency modulation, f_H, b) on the Rabi frequency (eq. 1). At $b_c = b_{opt}$ the maximum resonance signal is obtained.

Similar calculations have been done to determine the clock transition line centre assuming various amplitudes b_c and including the second-order-Doppler shift and an end-to-end cavity phase difference of 51µrad. As a result, the clock frequency should (in both beam directions) not exhibit excursions larger than $2 \cdot 10^{-15}$ / dB. It must be noted, however, that influences like spectral purity of the microwave signal, cavity pulling and spurious microwave excitation are not included in the calculation. Unfortunately, the short-term instability of the clocks is about $8 \cdot 10^{-12} / \sqrt{\tau}$ (τ in seconds) and an experimental verification of the prediction even around the optimum amplitude $b_c = b_{opt}$, is rather difficult. But experimental results are not at variance with the prediction.

5 Rabi pulling

The asymmetric velocity distributions shown in Fig. 2b reflect a highly asymmetric spectrum of the microwave σ transitions ($\Delta F = 1$, $\Delta m_F = 0$) in the clocks. The wings of the transitions between the $m_F = \pm 1$ Zeeman substates differ unavoidably in magnitude and slope at $f = f_0$. As Rabi pulling may cause serious problems the clocks should be operated at a sufficiently strong C-field /6/. The field chosen corresponds to $B_C = 7{,}714$ µT resulting in a Zeeman frequency, f_z, of about 54000 Hz which is equal to almost 12 times the width of the Rabi pedestal.

The magnitude of Rabi pulling was studied, again making use of the procedure described in section 4. The wings of the two neighbouring transitions were calculated using the far-off-resonance approximation (e.g. 5.2.88 in [3])

$$P_3 = \frac{15}{8}\pi^2\left[b_c \cdot \pi \cdot \cos(\Omega_R - \tau/2) / \left((\Omega_R \cdot \tau)^2 - \pi^2\right)\right]^2 \qquad (6)$$

with

$$\Omega_R = (f - f_0 + f_z) \cdot 2 \cdot \pi.$$

An average over the velocity distributions is done and the contributions are added to the values $P(f_i - f_0)$, $i = 1,\ldots, 4$. To illustrate the results, the wing P_3 of one neighbouring line is plotted in Fig. 6a together with the resulting shift, ΔF_R, of the clock transition frequency. It is noticeable that the extrema of the shift

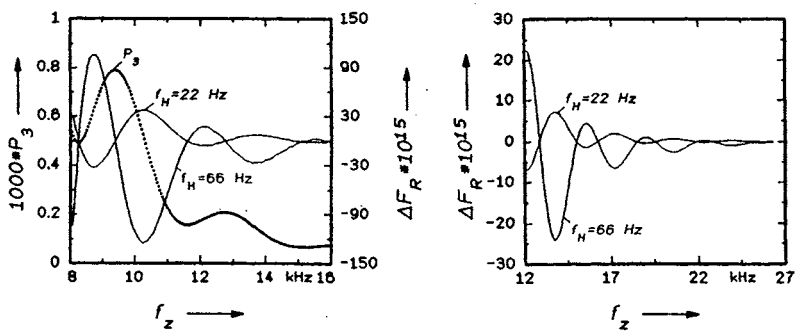


Fig. 6: Calculated Rabi pulling in the CS4. The velocity distributions shown in Fig. 2b were used as input data. a) Frequency shift ΔF_R of the clock transition ($f = f_0$) for two modulation widths, $2 f_H \approx W$ and $2 f_H \approx 3W$, caused by the wing of the (4,1) - (3,1) transition, centred at $f = f_0 + f_z$. P_3: amplitude of the wing at $f = f_0$;
b) ΔF_R including the contributions of both neighbouring transitions.

and those of the slope of P_3 coincide. By alternating the modulation width $2f_H$ between W and $3 \cdot W$, this type of frequency shift in a clock can clearly be distinguished from other cases because the shift behaves as indicated by the two solid lines in Fig. 6a. Some years ago experiments of this type were made with PTB's CSX and CS2 /7/ and showed excellent agreement between the experimental results and the predictions. Fig. 6b shows the predicted effect of both neighbouring transitions over a wider range of Zeeman frequencies f_z in the CS4. Unfortunately, for various reasons /1/ the clock must not be operated at low C-fields which would allow an experimental verification of the prediction. Yet, it can be safely claimed that at the operational C-field strength in CS3 and CS4 the Rabi pulling frequency shift is less than $0{,}3 \cdot 10^{-15}$, even at an increased microwave field amplitude b_c.

6. Conclusion

Some steps of a continuing evaluation of PTB's new primary clocks CS3 and CS4 have been presented. The results will help to make a careful statement of the uncertainty of the clocks which were designed to achieve a total value of one part in 10^{14}. Some recent results, concerning the long-term frequency instability, seem to call in question a success like this and have already caused a considerable delay in the completion of the clocks. Further efforts are necessary.

7. Bibliography

/1/ Bauch, A., Fischer, B., Heindorff, T., Schröder, R.: The new PTB primary cesium clocks. IEEE Trans. Instr. Meas. **IM-42** (1993), pp. 444-447.

/2/ Bauch, A., Fischer, B., Heindorff, T., Schröder, R.: The fourth primary cesium atomic clock of the PTB. Proc. 7th European Frequency and Time Forum EFTF (1993), pp. 531-535.

/3/ Vanier, J., Audoin, C.: The quantum physics of atomic frequency standards. Bristol and Philadelphia: Adam Hilger 1989.

/4/ Thomann, P.: On the length of the drift region in the Ramsey cavity. Proc. 22nd PTTI (1991), NASA Conference Publ. 3116, pp. 331-335.

/5/ Schröder, R.: Frequency synthesis in primary cesium clocks. Proc. 5th EFTF (1991), pp. 193-200.

/6/ De Marchi, A., Rovera, G.D., Premoli, A.: Pulling by neighbouring transitions and its effects on the performance of caesium-beam frequency standards. Metrologia **20** (1984), pp. 37-47.

/7/ 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), pp. 199-203.

Microwave Leakage as a Source of Frequency Error and Long-Term Instability in Cesium Atomic-Beam Frequency Standards

W.D. Lee, John P. Lowe, Jon H. Shirley and R.E. Drullinger
National Institute of Standards and Technology
Boulder, Colorado, USA

Abstract

We identify leakage from the microwave electronics in cesium atomic-beam frequency standards as a potential source of frequency error and instability in these standards. The problem arises when radiation that is phase coherent with the interrogating radiation finds its way to the atoms in regions outside the Ramsey cavity. We discuss techniques for evaluating and reducing the problem.

Introduction

Reduced to its most basic elements, an atomic clock is an electronic oscillator locked to an atomic resonance. In the case of cesium atomic-beam standards, the atoms are moving at thermal velocities, and the resonance is in the microwave region at 9.192 GHz. The atom motion could cause a first-order Doppler shift as great as a part in 10^6 if the interrogating radiation were counter- or co-propagating with the atomic beam. To obtain frequency accuracy or stability of a part in 10^{14}, this effect must be controlled to a part in 10^8. It is this extremely high sensitivity to the motion of the atoms that creates the sensitivity to microwave leakage.

The principal purpose of the microwave cavity is to provide the necessary control over the first-order Doppler effect. Inside the cavity, the microwave power flow in the direction of the atomic beam has been greatly reduced relative to that of a freely traveling wave. The remaining effects are interpreted as end-to-end and distributed-cavity phase shift[1]. In commercial standards, these residual effects are assumed to be small. In primary standards where accuracy is a consideration, these effects are measured. In either case, any frequency shift is assumed to be stable.

However, if radiation finds its way from the microwave generator to the atoms while they are outside the microwave cavity, the required control over the first-order Doppler shift may be lost. Only radiation that is phase coherent (i.e., radiation from the microwave generator operating the device in question) causes the most serious problems. Radiation that is nearly the same frequency (e.g., from a nearby clock) can also induce undesirable transition probability and, hence, frequency shifts. But the relative phase evolution should cause the shifts to average out. Phase coherent radiation, on the other hand, can cause a static frequency shift.

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

This is because it has different phase than the radiation in the cavity. The moving atoms see this spatial phase variation as a time varying phase, hence, a frequency shift. This is classical end-to-end phase shift and would be accounted for in a beam reversal evaluation if it is stable over the time of the evaluation[1]. However, frequency shifts caused in this way are often unstable because the leakage radiation travels over uncontrolled paths and arrives with uncontrolled phase. Changes in the paths cause changes in the phase and hence, changes in the frequency of the standard.

We have developed a theoretical model for the frequency shift caused by radiation outside the microwave cavity. The result includes power- and modulation-dependent factors multiplied by a spatial integral of the component of the leakage field in quadrature with the interrogating radiation, all averaged over atomic velocity. Since the spatial dependence of the leakage radiation is not known, only qualitative results have been obtained. Because the shift is linear in the radiation amplitude, leakage radiation that is 120 dB weaker than the interrogating radiation inside the microwave cavity can cause a significant frequency shift (parts in 10^{14}).

However, we do not need a quantitative model to deal with the problem. The fact that this effect is not widely recognized indicates that it is usually small. Most systems need only a little additional isolation to reduce the effect to acceptable limits. To achieve the desired isolation two areas should be investigated: leakage from the microwave electronics outside the beam tube and avenues by which the radiation gets into the beam tube.

Microwave leakage

All microwave equipment leaks at some level. Systems that have been assembled without special attention to this fact may leak orders of magnitude more than those that have been assembled with care. The flanges in X-band waveguides may leak badly if they are not extremely flat and/or located at nodes in standing-wave systems. Cavity-tuning screws are a special problem. Replacing waveguide structures with coaxial technology using the common SMA connectors is not better. SMA connectors require precise torque to achieve adequate shielding. Furthermore, many SMA-connectorized components (attenuators, circulators, isolators, directional couplers, couplers, etc.) are extremely leaky. Many of them actually have an epoxy-filled hole all the way through from the outer conductor to the inner conductor to hold them together. Solutions involve carefully selected components, rf gaskets, metal filled epoxy, and absorbing material placed over the leaking component.

Building a microwave leak detector and inspecting a system for leaks is relatively simple. We use a simple antenna, either a magnetic field loop or an electric field dipole, followed by a heterodyne receiver; Fig. 1. The microwave generator from a second set of clock electronics makes a good local oscillator (LO). Alternately, a microwave signal generator tuned near the cesium resonance works.

With the heterodyne beat signal thus translated to the audio region, an FFT spectrum analyzer is sensitive enough to detect most troubling radiation. Interchanging the LO and the device under test allows all of the equipment to be checked for leaks.

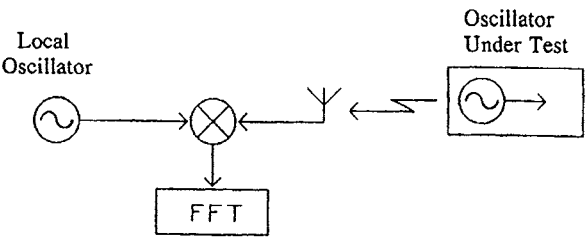


Fig. 1 Heterodyne detector used for rf leak detection.

The atomic beam tube

The mechanisms by which the radiation gets into the atomic beam tube are as important as the sources of radiation leakage. To find these paths, we split off some of the power from the microwave generator and applied it to a microwave horn. When this is directed at unprotected parts of the beam tube, large frequency shifts can be induced. One can then scale from the power radiated from the horn to the power leaked by the microwave plumbing to infer the size of the problem. We found that essentially all of the radiation is coupling into the beam tube through unfiltered, DC or low frequency feedthroughs such as C-field, Zeeman coils or degaussing leads. Unused pins on electrical feedthrough headers are often about $\lambda/4$ long and make effective antennas. Ferrite beads located right at the vacuum seal are effective filters. We have also used microwave-absorbing foam to shield unused pins.

Results

We have applied these techniques to NIST-7, the US primary frequency standard. Before the problem was identified and solved we had unexplained frequency shifts of several parts in 10^{14}. Once the problem was identified, we were able to force the frequency shifts by changing the phase delay between the microwave generator and the beam tube. Figure 2 shows a series of frequency measurements relative to a hydrogen maser where we systematically changed the phase delay between every measurement. The first four measurements were made before we closed the leaks, while the last four were made after. The improvement indicated by these measurements is at least 10 dB and is limited by the statistical

validity of the measurements. Our electrical measurements on the leaks indicate that we made an improvement of at least 50 dB.

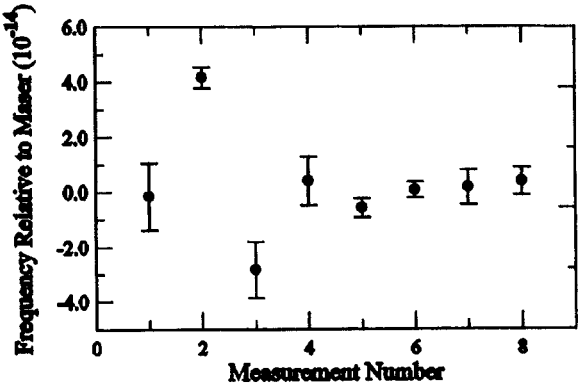


Fig. 2 Variation in the frequency offset measured between NIST-7 and a hydrogen maser while the phase delay from the microwave generator to the beam tube was systematically changed. The first four points are with leakage fields while the last four are after the leakage fields were reduced by the techniques outlined here.

Conclusion

Radiation leakage from the microwave electronics in a cesium atomic-beam frequency standard that finds its way into the beam tube can cause substantial frequency shifts. These shifts may be unstable with time because they are dependant on the phase of the radiation which arrives at the atoms via uncontrolled pathways. The effect may lead to frequency errors in primary standards and to long-term instability in commercial standards. The fact that this effect has gone largely unrecognized suggests that it is usually small. Hence, small improvements in the leakage from the microwave electronics and attention to the pathways by which this radiation enters the beam tube are sufficient to reduce the problem to acceptable levels.

References

1. Jacques Vanier and Claude Audoin, The Quantum Physics of Atomic Frequency Standards, Adam Hilger 1989, Section 5.6.4(A).

PERFORMANCE OF A 2dm^3 OPTICALLY PUMPED CESIUM BEAM TUBE : A PROGRESS REPORT

P. Petit, V. Giordano*, P. Cérez, B. Boussert, C. Audoin, N. Dimarcq

Laboratoire de l'Horloge Atomique
Unité Propre de Recherche du CNRS associée à l'Université Paris-Sud
Bât. 221 - Université Paris-Sud - 91405 ORSAY Cedex - France

Summary.

The main characteristics of 2dm^3 volume optically pumped cesium beam tube are briefly described. The resonator operates with a single commercially available laser diode ; its microwave cavity is designed to provide a phase difference of π between the two oscillatory fields. Some advantages are expected from this structure.

A suitable static magnetic field configuration is chosen in order to eliminate the detrimental non-linear Hanle effect which is associated with the one laser operation. In a preliminary test, the measured frequency stability is $4 \times 10^{-12} \tau^{-1/2}$ for $1 < \tau < 10^4$ and 4×10^{-14} for $\tau = 1$ to 2 days.

We also observe a strong dependence of the frequency of our device to the microwave power injected in the Ramsey cavity. This frequency shift seems to be highly connected to microwave leakages of the cavity.

1. Introduction.

The wide success of time and frequency systems relies mainly on the availability of small size, field operable, cesium beam frequency standards, whose frequency is very close to the one delivered by large size laboratory primary devices.

The method of optical pumping is able to improve the performance of small size cesium beam frequency standards as well as that of laboratory set-ups. In this respect, we report briefly the first demonstration that an optically pumped cesium beam tube provides the same level of medium term frequency stability as state of the art, high performance, manufactured tubes, but in a smaller volume.

---

* Present address : Laboratoire de Physique et de Métrologie des Oscillateurs, Besançon, France

2. Description of the small size optically pumped cesium beam tube.

The atomic resonator is a laboratory device plugged to an external 25 ℓ s^{-1} vacuum pump (Fig. 1). The overall volume of the tube is $2dm^3$, when equipped with an external magnetic shield.

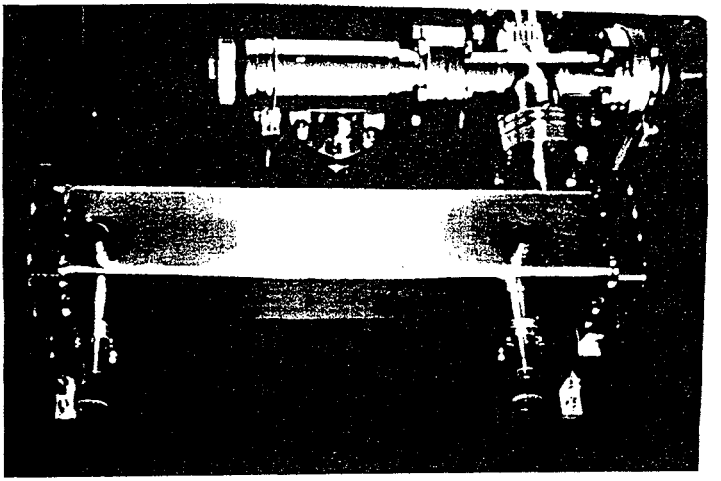


Figure 1 - Picture of the miniature resonator.

The cesium oven is of a classical design. The emissive area is 2 mm x 4 mm and the operating temperature is $\cong 90°$ C.

The microwave cavity is of the E-bent type. The length of the drift space between the two interaction regions is 16 cm. We have chosen to provide a phase difference of π between the two microwave oscillatory fields, which brings some advantages [1,2]. A coaxial feedthrough enables to supply the cavity at one extremity of the tube. This avoids the annoying effect of a misplaced hole in the magnetic shields.

A single laser is used for optical pumping and for optical detection. This is a Hitachi HLP 1400 diode laser with a linewidth of $\cong$ 25 MHz, tuned to the 3-3 σ transition of the D_2 cesium line. Two windows are implemented to enter the light in the tube. The incoming light is trapped internally. Two other windows on the top of the tube enable to collect the fluorescence light. The signal collected in the pumping region is used to lock the laser frequency. In the second optical region, a large area photodiode delivers the clock signal.

The magnetic field is transverse to the atomic beam direction. Its magnitude is 5 x 10^{-5} T (500 mG) in the two optical interaction regions and 7,65 x 10^{-6} T (76,5 mG)

in the microwave interaction region. This configuration eliminates the detrimental non-linear Hanle effect which is associated with the efficient 3-3 σ transition [3]. A record of the seven microwave transitions of the cesium atom around 9.2 GHz is shown in Fig. 2 and shows an excellent static field uniformity. Two magnetic shields are set inside the vacuum enclosure and a third one at the outside.

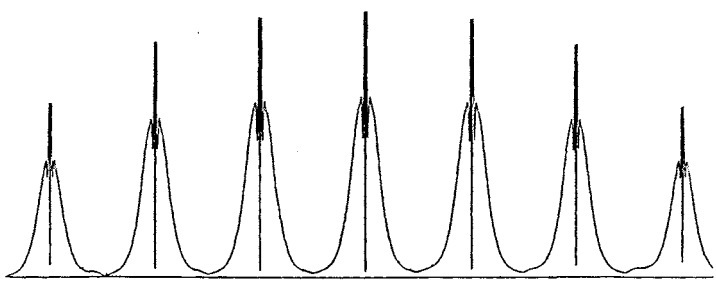


Figure 2 - Microwave spectrum delivered by the resonator.

The electronic system associated with this tube comes from a 5061 A Hewlett-Packard cesium beam frequency standard, with some minor modifications. A microwave amplifier boosts the signal coming from the Harmonic Generator. A New AC pre-amplifier at 137 Hz has also been designed.

3. Experimental results.

With $\phi = \pi$ between the two oscillatory fields, the Ramsey pattern shows a minimum at resonance. The background level of the stray light detected is 10% of the level of the useful fluorescence light, measured between the peak and the valley of the central Ramsey fringe (Fig. 3). The signal to noise ratio is 20000 in a 1 Hz noise bandwidth. The atomic linewidth is 660 Hz.

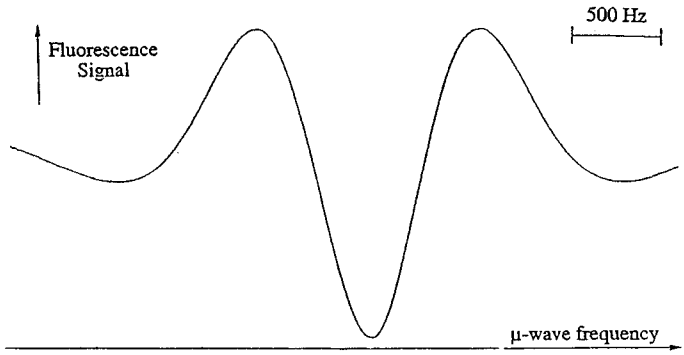


Figure 3 - Peak to valley signal of the central Ramsey fringe.

After a rough adjustment of the operating parameters (microwave level, modulation depth), a three weeks run has been completed. The frequency stability is $\sigma_y(\tau) = 4 \times 10^{-12}\ \tau^{-1/2}$ for $10\ s < \tau < 10^4\ s$ and $\cong 4 \times 10^{-14}$ for $\tau = 1$ to 2 days, as shown in Fig. 4. A hydrogen maser, whose cavity is auto-tuned [4] was used as the frequency reference in this measurement.

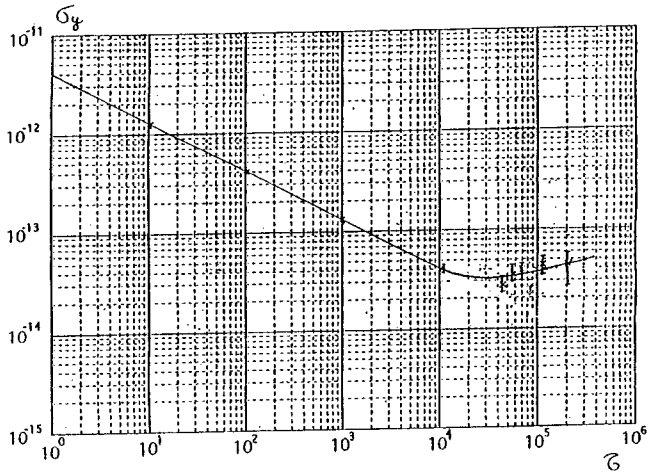


Figure 4 - Frequency stability graph of the small size optically pumped cesium beam tube.

The output frequency of the frequency standard has been corrected for the second order Zeeman frequency offset (2.49 Hz). Unfortunately, the output frequency of our device does not match with the expected one. Moreover, we noted a strong dependence of this frequency with the microwave power injected in the cavity.

In order to evaluate this dependence, we have measured the frequency shift Δf versus the microwave power variation ΔP around the optimal value P_0. P_0 is the power which maximizes the peak valley amplitude of the fringe. The measured sensitivity $\Delta f/\Delta P$ (10^{-11}/dB) is uncommonly high compared to the one expected by the theory.

Almost the frequency shifts are microwave power dependent. The major ones are connected with the imperfections of the cavity. We have verified that the good electrical symmetry of the two arms and the low detuning (≤ 2 MHz) of our cavity can not explain such a huge sensitivity. We are now convinced that the observed effect is linked to microwave leakages of the Ramsey cavity. As a matter of fact, we have pointed out a high level of microwave leakages at the short circuits and in the vicinity of the coaxial to waveguide coupling device. These leakages are known to induce uncontrollable frequency shifts [5] which can be large for a short cavity.

More precise investigations are undertaken in order to estimate the influence of these leakages. However, a precise calculation can not be easily performed due to the non-trivial spatial structure of the stray field outside the Ramsey cavity.

Let us also mention that a preliminary measurement of the sensitivity of the frequency to an external magnetic field shows up an effect of 10^{-13} per Gauss in any direction.

4. Conclusion.

We have designed and operated a miniature optically pumped cesium beam clock. It exhibits a good short term and medium term frequency stability but the delivered frequency does not match with the definition and suffers from a high sensitivity to the microwave power. The origin of this drawback seems to the highly connected with microwave leakages of our cavity. We are looking for Laboratory methods to get over with this detrimental effect.

Acknowledgements.

We gratefully acknowledge the support of STEI.

References.

[1] P. Petit, V. Giordano, N. Dimarcq, P. Cérez, C. Audoin and G. Théobald. "Miniature optically pumped cesium beam resonator". Proc. 6th European Frequency and Time Forum, Noordwijk, The Nederlands, 1992, pp. 83-86.

[2] C. Audoin, V. Giordano, N. Dimarcq, P. Cérez, P. Petit and G. Théobald. "Properties of an optically pumped cesium beam frequency standard with $\phi = \pi$ between the two oscillatory fields". Proc.7th European Frequency and Time Forum, Neuchâtel, Switzerland, 1993, pp. 537-540.

[3] V. Giordano, A. Hamel, P. Petit, G. Théobald, N. Dimarcq, P. Cérez and C. Audoin. "New design for a high performance optically pumped cesium beam tube". IEEE Trans. on UFFC, 38, 1991, pp. 350-357.

[4] R. Barillet, J. Viennet, P. Petit and C. Audoin. "Fast cavity auto-tuning for H Masers ; experimental results". Proc.5th European Frequency and Time Forum, Besançon, France, 1991, pp. 407-411.

[5] A. Bauch. "Majorana-übergänge in caesium-Atomstrahl-Frequenznormalen". PTB-Bericht Opt-20 february 86, ISSN 0341-6712, pp. 20-21.

ATOMIC BEAM TUBE FOR PROMISSING FREQUENCY STANDARD INSTALLED GLONASS SATELLITES ONBOARD

A.B.Bassevich, Y.G.Gouzhva, A.G.Gevorkyan, I.P.Gorlova
Russian Institute of Radionavigation and Time,
2 Rastrelli square, St.Petersburg, 193124, Russia

Used GLONASS satellites onboard are clocks on the basis of cesium beam frquency standards. Accuracy parameters of system, as well as its ability to operate in autonomous mode, depend on daily frequency instability of frequency standard. Therefore it is a very topical task to achieve an instability less than $5 \cdot 10^{-14}$ over 1 day compared to the present at the level of $(2-3) \cdot 10^{-13}$ without ageing.

The results of development of the new atomic beam tube for a frequency standard with the mentioned instability are considered. The design specific features of the tube are given, taking into account a decrease of destabilizing factors to the level being no more than $1 \cdot 10^{-14}$ as applied to operation conditions taken place GLONASS satellites onboard. The results of metrological parameters' measurement are presented for the frequency standard breadboard.

1 INTRODUCTION

An accuracy of atomic beam frequency standard is determined, along with other factors, by shifts of (3,0) - (4,0) transition resonant frequency in a cesium beam tube. Temporal changes of these shifts, under the effect of external destabilizing factors such as temperature and magnetic fields or as a result of changes of atomic beam tube operational electric parameters, i.e. beam source

temperature, magnetic field strength H_o within a region to be shielded, excitation microwave signal power, are transformed into the long-term frequency instability of frequency standard.

For compact industrial atomic beam tubes, the daily frequency instability at the level of $(1\text{-}5)\cdot 10^{-14}$ is determined by shifts connected with physical perturbations of atoms (sampling by H_o-field) as well as by transition excitation technique (microwave cavity pulling, phase shifts within a cavity, pulling by adjacent σ- and π-transitions). The values of these shifts define the limiting parameters' deviations for medium and operational mode allowable in view of providing desirable frequency stability.

While developing the new cesium beam tube USPEKH (translated as "SUCCESS"), such tasks were solved as to decrease the magnitudes of systematic shifts and to bring down its sensitivities to destabilizing factors.

Compared to basic UTRO atomic beam tubes developed jointly by RIRT and ISTOK Scientific Production Association in 1982-1989 with σ_{day} equal to $(2\text{-}3)\cdot 10^{-13}$, the USPEKH tube is equipped with such modernized units as cylindric magnetic shields' system, H_o-field coil, cesium vapour absorber, electric-discharge pump, as well as with newly designed microwave cavity and external magnetic shield.

The joint works carried out enabled to reduce systematic frequency shifts and beam tube's sensitivity to external destabilizing factors up to the values giving $\sigma_{day} = (2\text{-}5)\cdot 10^{-14}$.

2 DESTABILISING FACTORS

Considered is a number of destabilizing factors, as well as techniques to decrease its effects which restricted the long-term frequency stability of preceded UTRO cesium beam tubes to the level of $(2\text{-}3)\cdot 10^{-13}$.

The greatest systematic shift of 0-0 - transition frequency is caused by the H_o-field whose uniformity and stability determine to a large extent a long-term frequency stability of frequency standard.

In USPEKH cesium beam tube, the coil design is modernized and the magnetic guard system is used comprising four cylindric shields and six disk ones, four pairs of magnetic plates (shims) placed in transit zones between selecting magnets and cavity and four compensating coils. Such a design enabled to increase a shielding coefficient up to 10^5-10^6, to improve significantly H_o-field uniformity and to reduce its operational value to 60 mOe. Using thermal ceramic isolations within a region to be shielded make it possible, to bring down to a great extent H_o- field distortions caused by thermal currents.

Problem of reducing frequency shifts caused by cavity pulling and phase difference of oscillating fields within regions of interaction with atomic beam was posed while developing a new version of microwave cavity. Reducing shifts caused by cavity pulling while changing the cesium beam tube temperature mode up to $\sim 10^{-14}/^oC$ is achieved due to diminishing cavity loaded Q up to the values Q_{load} = 400-600 with its own $Q_{own} = 5 \cdot 10^3$. For the initial cavity tuning error of ±1 MHz, the cavity-pulling shift is of $1 \cdot 10^{-13}$.

An increase of waveguide duct uniformity along the cavity length, VSWR diminishing due to the use of smoothed radius turns instead of corner sections, as well as introducing operation of making cavity arms symmetric into the cesium beam tube manufacturing process, enabled to reduce a level of arms' asymmetry below 100 mu providing phase difference no more than $(1\text{-}5) \cdot 10^{-5}$ rad and a corresponding frequency shift $n \cdot 10^{-13}$. Dependance of shifts caused by cavity pulling and its phase differences on microwave excitation signal power is close to the linear law within the region of its optimal value /1/. Consequently, a change of the total shift does not exceed

$n \cdot 10^{-15}$ per 1% of power change.

The pulling of basic transition frequency by adjacent ($\Delta m_f = 0$) transitions (Rabi pulling) is a result of these transitions' asimmetry. One of reasons for such an asymetry is non-adiabatic transitions induced within a transit region between selecting magnets and cavity and giving rise to intensity changes for all the microwave transitions. As it was shown in experimental investigations, a significant decrease of non-adiabatic transitions' probability, as well as diminishing Rabi-pulling, compared to basic atomic beam tubes, is realized as a result of correction of magnetic fields topography within a transit region by means of magnetic plates and correcting coils. An oscillating nature of Rabi-pulling frequency shift dependence on H_o-field shows, thus validiting a theoretical prediction /2/, that there must be such field strength values which result in negligibly small shifts and minimum cesium beam tube sensitivity to microwave signal power.

Shifts connected with adjacent ($\Delta m_f = \pm 1$) transitions' frequency pulling (Ramsey-pulling) occur with oscillating magnetic field components normal to H_o-field and inducing these transitions. According to the theoretical estimate /3/, for USPEKH cesium beam tube, where a deviation of atomic trajectories from the wide wall's center of U-type cavity does not exceed 1.25 mm, the shifts caused by Ramsey-pulling are no more than $n \cdot 10^{-15}$.

3 EXPERIMENTAL INVESTIGATIONS

The new cesium beam tube has a figure of merit $F \geq 10$ for spectral line width $W \leq 500$ Hz and device total length of 500 mm. A photograph of the USPEKH cesium beam tube is shown in Fig.1.

The goal of experimental investigation was to find out a dependence between systematic frequency shifts and

H_o-field and microwave signal power , as well as to check a possible improvement of standard long-term frequency stability by means of choosing an optimal (in view of H_o-field) operational mode of atomic beam tube.

UTRO and USPEKH cesium beam tubes were investigated as a part of frequency standard breadboard. Frequency measurements were carried out with the measurement system of the RIRT's Secondary Time/Frequency Reference Standard. The CH1-70 hydrogen frequency standard was used as reference standard.

Experimental curves presented in Fig.2 and describing an effect of H_o-field on frequency of frequency standard with USPEKH (curves 1,2) and UTRO (curves 3,4) cesium beam tubes (a frequency of standard was updated for square Zeeman shift) correspond to two values of microwave signal power $P=P_{opt}$ and $P=2P_{opt}$ for each tube. The field strength corresponding to the minimal USPEKH cesium beam tube sensitivity to a microwave power (intersection of curves 1 and 2) was found to be equal to 61 mOe.

Further, a daily frequency σ-instability of the frequency standard was measured at H_o specified value of 90 mOe and at H_o of 61 mOe.

Following results took place:

for H_o= 90mOe $\sigma_{day}= 4.0\cdot10^{-14}$over observation interval of 10 days (from 3.05.93 through 12.05.93);

for H_o= 61mOe $\sigma_{day}= 1.5\cdot10^{-14}$over observation interval of 25 days (from 15.03.93 through 8.06.93);

$\sigma_{day}= 2.1\cdot10^{-14}$over observation interval of 25 days (from 15.05.93 through 8.06.93).

Allan variance of frequency standard with USPEKH cesium beam tube as a function of measurement time interval is presented in Fig.3.

CONCLUSION

In this way, investigations and measurements carried out confirm an opportunity to improve a frequency

stability of the frequency standard up to the level less than $5 \cdot 10^{-14}$ under operational conditions taking place GLONASS satellites onboard.

ACKNOWLEDGEMENTS

The authors are pleased to express their gratitude to I.I.Samartsev and his collegues in ISTOK Scientific Production Association for participation in designing of atomic beam tubes.

REFERENCES

/1/ Effects of servo loop modulation in atomic beam frequency standards employing a Ramsey cavity/ De Marchi A., Rovera G.D. and Premoli A.// IEEE Trans. Ultrason., Ferroelec., and Freq.Contr.- 1987, 34, No. 6, 582-591.

/2/ Frequency pulling by hyperfine σ-transitions in cesium beam atomic frequency standards/ Cutler L.S., Flory C.A., Giffard R.P. and De Marchi A.// Journal of Appl. Phys.-1991, 2780-2792.

/3/ Сдвиги частоты квантового перехода в атомно-лучевом спектроскопе, связанные со структурой возбуждающего поля. / Горлова И.П., Геворкян А.Г.// Вопросы радиоэлектроники, сер. ОТ. - 1980, вып. 9, 95-102.

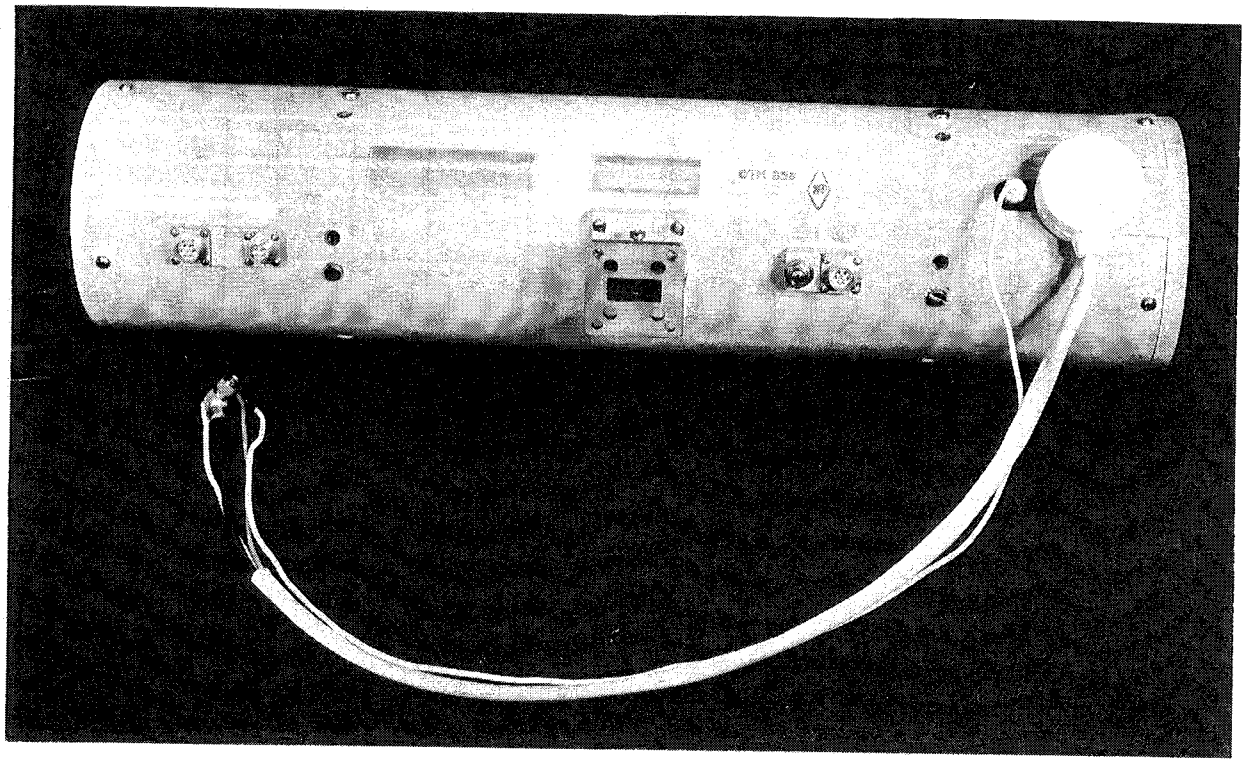


Fig.1. USPEKH cesium beam tube.

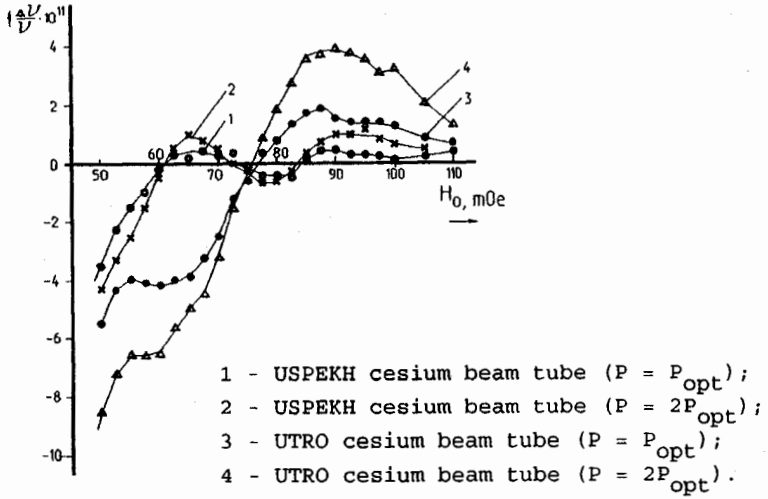


Fig.2. Frequency shifts in cesium beam tubes vs H_o-field strength.

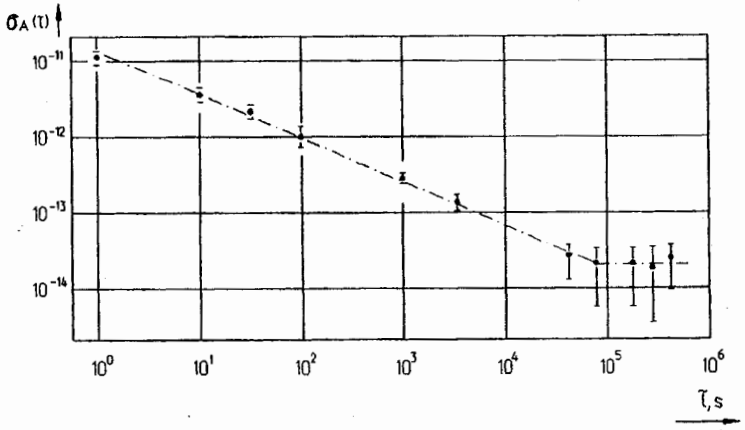


Fig.3. Allan variance of frequency standard breadboard vs measurement time interval.

SAW Devices (B 5)

Piezoelectric Frequency-Output Sensors

E. BENES, M. GRÖSCHL, M. SCHMID, W. BURGER

Technische Universität Wien, Institut für Allgemeine Physik, Wiedner Hauptstr. 8/134, A-1040 Wien, Austria

A review about sensors based on piezoelectric crystal resonators is presented. The survey focuses on the fundamental resonator modes more than on the variety of surrounding support configurations in the special sensor applications. First, the general properties of vibrating crystal sensors and their inherent superiority are described. The sensor concepts utilising either homogeneous resonators with temperature and pressure (stress) as primary measurands or composite resonators with areal mass density and viscoelastic properties of the "foreign" layer as primary measurands are discriminated. A comparison between Bulk Acoustic Wave (BAW) and Surface Acoustic Wave (SAW) resonators with respect to their primary sensitivity functions and principal capabilities for sensor applications is given and the importance of most recent investigations on Lamb Wave and Horizontal Polarised Shear Wave (HPLW) Inter-Digital Transducer (IDT) resonators is acknowledged. The importance of mode-purity for high dynamic range sensors based on resonators and some aspects of the demand on specialised electronics are emphasised. The present state of the established sensors based on primary sensitivities, e. g. quartz crystal thermometers, pressure transducers, thin film thickness and deposition rate monitors, viscoelastic layer analysers (crystal/liquid composite resonators) is reviewed. A selection of most promising, recently investigated vibrating crystal sensors utilising indirect sensitivities is described, like the wide field of analyst selective coatings and resonator based immunosensors or -assays. Finally, the potential of alternative piezoelectric materials for future sensor developments is briefly discussed.

Keywords: Frequency sensors, acoustic sensors, piezoelectric resonators

1. INTRODUCTION

A sensor is a transducer that transforms a measurand quantity into an electrical measurement value, usually voltage, current or frequency. In the case of frequency output the term "digital sensor" is sometimes used, although we have still an analogue change of the output with the measurand quantity. However, the frequency can be easily digitised by use of a counter with an appropriate time base.

The use of a counter as Analogue to Digital Converter (ADC) has an important advantage: The costs increase only slightly with the number of digits, the obtainable dynamic range, that is the ratio of measurement range over resolution, is not limited by the analogue to

digital conversion. In the case of a frequency sensor the dynamic range is only limited by the sensor principle itself, e. g. values of up to 10^7:1 are reached with the quartz crystal thin film thickness monitor.

Furthermore, a projection to frequency is equivalent to transferring the information into a time interval value. Time is invariant to a change from the mechanical to the electrical domain. That is, there is no loss of accuracy by the transfer from a mechanical vibration to the piezoelectrically coupled corresponding electrical signal. Finally, time is - with a gap of three orders of magnitude - the most accurately determinable measurement quantity at all.

The most wide spread piezoelectric single crystal material used in the frequency-sensor field as well as in the frequency/time-control field is quartz. Thus, in these fields the oxide form of Silicon plays the same dominating role as pure Silicon in the semiconductor industry.

Many vibrational modes and structures used in the sensor field have counterparts in the frequency-control field and there are many cross benefits from joining the research and development results of both scientific communities.

Piezoelectric crystal resonators for frequency control are commonly known, because of their wide spread use in watches, TV-apparatus, Video-recorders, CD-players, computers, and handy phones. In these applications an estimated number of more than 1.3 billion resonator units are currently world-wide in use. The main reasons for this amazing market penetration are the high acoustic resonator quality factor (Q-value) and the insensitivity of the resonance frequency to potentially perturbing parameters, like temperature, pressure, humidity, etc.. Therefore, at the first glance, it is very surprising that such stable resonator concepts can be used as well for different sensors with high sensitivity to various measur- and quantities. The explanation is twofold:

- Firstly, the anisotropy of the piezoelectric single crystals allows different angles of cuts (different orientations of the resonator shape with respect to the crystallographic axis), with very different properties. While, e. g., in the frequency control field temperature compensated cuts like the AT-cut and an additionally stress compensated cut like the SC-cut were developed, for a temperature or pressure sensor a respective high sensitivity of the resonance frequency is welcomed.

- Secondly, in the frequency control case the resonator is usually vacuum encapsulated and decoupled from mechanical stress transfer from the environment, while in the sensor application the resonator is kept in contact with the medium that transports the information about the interesting parameter quantity.

In the interest of conciseness, the paper emphasises on the principal crystal resonator types more than on the variety of surrounding supports and housings characteristic for the specific measurement task. Readers interested in detailed information about specialised sensor data are referred to the comprehensive cited literature list. In the following two chapters the fundamental resonator concepts are reviewed with special emphasis upon the parameters contributing to the value of the resonance frequency in order to show the basis for the derivation of sensor sensitivity functions. Then examples of the special demands on the measurement electronics, characteristic for the sensor application, are given. An overview follows about the already established primary and the wide field of secondary measurands. Finally, possible alternative piezoelectric materials for future sensor developments are discussed.

2. HOMOGENEOUS RESONATORS

An acoustic resonator based on one homogeneous material (substrate) is here called "homogeneous resonator" in order to discriminate it from a resonator that is in contact with a second medium. The latter case is treated as "composite resonator". The homogeneous resonator - at least within the limits of an one-dimensional treatment - has a trivial mechanical resonance frequency spectrum: harmonic overtone-frequencies are integer multiples of the corresponding fundamental resonance frequency.

2.1. BAW-Resonators

2.1.1. Principal vibration modes

Piezoelectric plates can be excited to fundamental or overtone flexural modes, face shear modes and thickness shear and twist modes as well as to several extensional modes. Long thin rods can be used in the bending, extensional or torsional modes. In order to obtain resonators with high Q-values and with properties determined only by the single crystal and not by the holder material and construction, the clamping of the resonator must be located at vibration nodes or other regions of vanishing amplitude. One of the advantageous resonator types is the tuning fork shape, where the angular momentums of both single-ended flexural mode beams cancel out and therefore no energy is lost into the support. The lower possible resonance frequency of the tuning fork design allows low power consumption of CMOS electronics. Therefore this design, miniaturised by use of photolithography and etching technology, revolutionised the watch industry. The miniaturisation and low cost production technology of these resonators are also very attractive properties for sensor applications.

2.1.2. Resonance spectrum of the one-dimensional resonator

For sensor applications, a clean resonance spectrum free of spurious modes is obligatory. This is because the resonance utilised for the sensor function shifts - corresponding to the

measurand quantity - within a certain frequency range. If there are any spurious modes or any other modes with different dependencies upon the measurand quantity, resonance frequency crossings and mode coupling may occur within the measurement frequency range. Such mode couplings produce not only significant deviations from linearity, but sometimes produce even frequency jumps and activity dips in the frequency versus measurand quantity curve. To avoid this, there are basically two possibilities:

a) Utilising the lowest mechanical resonance frequency at all. (E. g. the lowest bending mode of a tuning fork, but not a torsional mode).

b) Keeping the resonator geometries as close to the one-dimensional theoretical description as possible.

The latter resonators exhibit a simple resonance spectrum which is described by the following fundamental relation:

$$f_n = \frac{n}{2l}\sqrt{\frac{c}{\rho}}\,, \tag{1}$$

where f_n is the regarded natural mechanical frequency, n the harmonic number, l the resonance determining dimension (e. g. the thickness of a relatively large thin plate or the length of a thin long rod), c is the effective elastic stiffness constant (e. g. the shear stiffness constant in thickness direction of a plate or Young's modulus in the case of a thin rod), and ρ is the density of the crystal material. This fundamental equation shows that a change of the resonance frequency can take place only if the measurand quantity is altering the effective elastic stiffness constant, the resonance determining dimension, and/or the density of the crystal.

2.1.3. Thickness shear resonator

The boundary conditions of the infinite plate allow in the general case three sets of harmonic resonances corresponding to one longitudinal and two transversal eigen waves with different propagation velocities in the thickness direction of the plate. In the case of a singly rotated Y-cut quartz crystal, e. g. the AT-cut, only one of the three sets of eigenfrequencies is piezoelectrically excitable, that corresponds to the transversal wave with the displacement direction in x-direction. Fig.1 shows the displacement curves of an infinite plate thickness shear resonator for constant d.c.-voltage and for the different resonant a.c.-voltages supplied to the electrodes. Note that in the case of the homogeneous resonator only odd harmonics can be piezoelectrically excited.

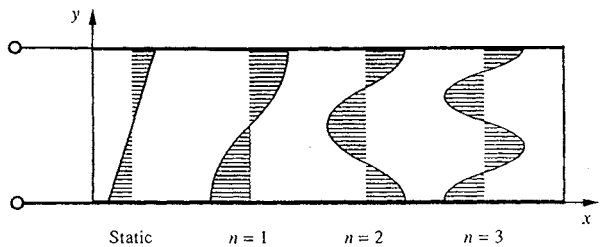


Fig.1. Infinite plate thickness shear resonator. Displacement curves for constant voltage and for excitation by an alternating voltage with frequencies corresponding to odd harmonic mechanical resonances. The acoustically defined but piezoelectrically not excitable even harmonic ($n = 2$) is also shown.

2.1.4. Consideration of finite active area

Because of the clamping problem, the thickness mode vibrations of a thin plate are kept in the centre of the plate either by the so called energy trapping effect of electrodes with a sufficient thickness and an area smaller than the plate area or by using lens-like convex disk shapes. In both cases the spatial vibration amplitude variation exhibits a maximum value in the centre of the disk and diminishes towards the edge, therefore a non-critical clamping at the disk edge is possible. It should be emphasised that - at least for quartz or other piezoelectric crystals with low electromechanical coupling factor - not the piezoelectric excitation area defined by the electrode area is essential for the restriction of the mechanically vibrating area, but the mechanical effect of the electrode (in the case of the plane-parallel plate) or the focusing effect of the convex curvature. It is useful to define an *effective active area* of the finite plate as the area that has to be regarded of the infinite plate model in order to obtain the same dynamic electrical admittance in the vicinity of a resonance /1/.

Fig.2 shows the spatial vibration amplitude distribution of a plano-convex AT-cut quartz crystal excited at the first, third and fifth harmonic thickness shear resonance frequency. The measurements were made utilising a non-interferometric in-plane sensitive laser speckle method /2/. In the left column the shear vibration amplitude $u(x,y)$ is shown on the vertical-axis in arbitrary units. In the right column an iso-amplitude representation of the same measurements is shown. The maximum amplitude of the fundamental mode is approximately 40 nm. (n, p, q) is the mode classification after Tiersten /3/. It can be seen clearly that the higher harmonics are more concentrated in the centre, the effective active area decreases with the harmonic overtone number n..

2.1.5. Frequency spectrum and mode-purity

The harmonic frequencies shown in Fig.2 are characterised in that at any moment every point of the surface is moving - with different amplitude but in-phase - in the same direction. Besides this harmonic frequencies, so-called anharmonic frequencies are always appearing. These are characterised in that some regions of the surface are moving in anti-phase. While in the frequency control application of piezoelectric crystal oscillators only the one utilised (harmonic) resonance frequency is of interest, in sensor applications usually the entire frequency spectrum has to be considered. This is because the influence

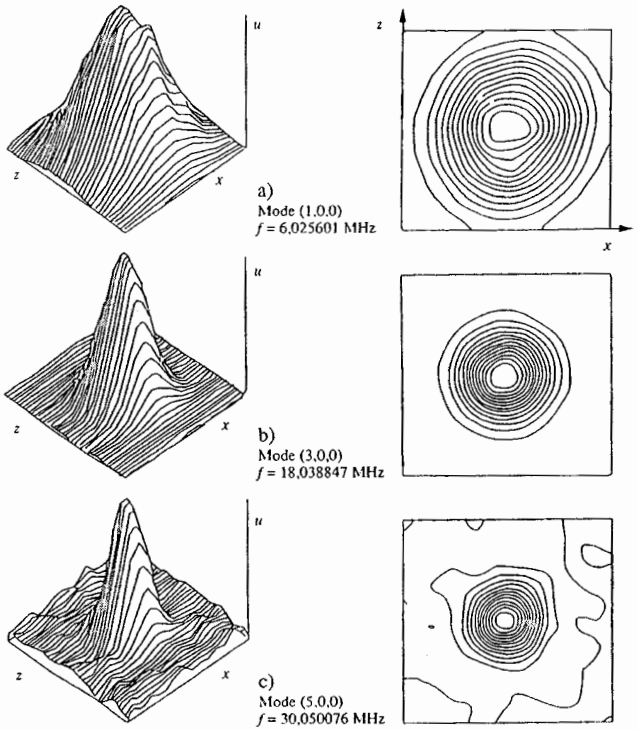


Fig.2. Spatial in-plane amplitude distribution $u(x,z)$ of the first, third and fifth harmonic thickness shear mode of a 14 mm plano-convex AT-cut quartz crystal. The square scan area was 8x8 mm. Left column: 3-dimensional representation, the z-axis shows the shear vibration amplitude in arbitrary units. Right column: iso-amplitude line representation.

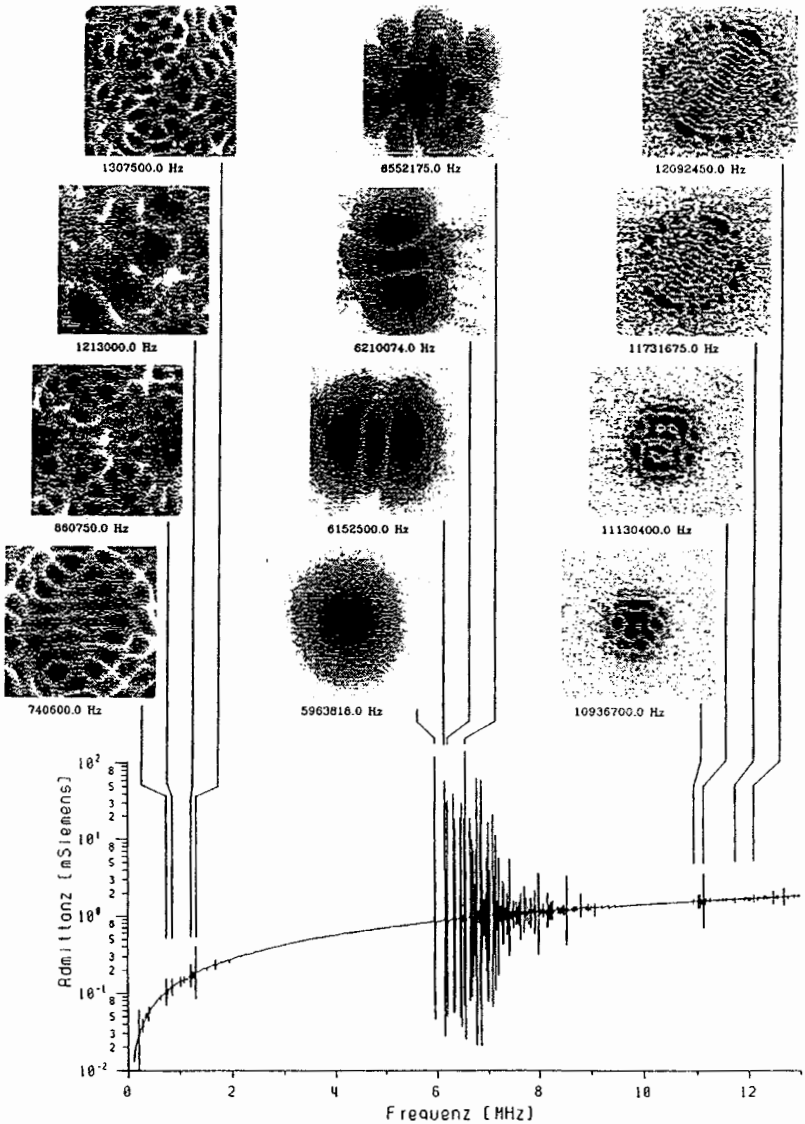


Fig.3. Vibration modes of a plano-convex SC-cut quartz crystal. The spatial vibration amplitude distribution within an area of 9x9 mm^2 of the 14 mm diameter circular disk is shown for any of the indicated resonances of the spectrum. Higher amplitudes correspond to darker areas in the grey scale representation.

of the measurand quantity on different frequencies is often very different. As a consequence, neighboured resonance frequencies may cross each other with changing measurement quantity. This temporarily causes mode coupling and, as a consequence, frequency jumps and activity dips in the sensor transfer function (See section 8.2.5 for example). Since the utilised resonance frequency is sweeping through a certain frequency range by varying the measurand quantity through the entire measurement range, this working frequency range should be free of other resonances. Therefore mode-purity is of premium importance in the sensor application field of piezoelectric resonators. To illustrate the problem, Fig.3 shows the frequency spectrum of a plano-convex SC-cut quartz crystal and the corresponding vibration amplitude pattern for a selection of frequencies of interest. Because in the case of the SC-cut all three acoustically defined eigen-waves are electrically excited, this cut exhibits a more crowded spectrum as the AT-cut. By use of a logarithmic scale in the admittance versus frequency plot, the much weaker excitable anharmonic and spurious modes are relatively emphasised. The measurements shown in Fig.3 were performed with the same laser speckle in-plane vibration amplitude measurement set-up /2/ as used for Fig.2, but this time a grey-scale representation of the amplitude values was chosen. Coupled bending and face shear oscillations (left column), harmonic and anharmonic thickness shear vibrations (centre column), and thickness extensional modes coupled with thickness shear modes (right column) can be identified.

Sometimes simultaneous use of two vibration modes with different sensitivities to two measurement quantities (e. g. force and temperature) is suggested /4/. Since in this case two equations for the two unknown measurand quantities are available and, for example, the temperature influence on the force/frequency sensitivity can be eliminated. But in practice the problem of mode coupling restricts these concepts significantly.

2.1.6. Tuning Fork Resonators

A tuning fork has high mechanical Q because the two tines move opposite of each other, the result being that the moments and forces from one tine cancels those of the second tine and the clamping at the tuning fork junction is in a zero amplitude region. The use of a tuning-fork shape for time keeping revolutionised the watch industry for two reasons. Firstly, the low resonance frequency of the flexural mode allows a low power consumption of the CMOS-electronics which is of advantage for battery powered watches. Secondly, the photolithography and etching process enables low cost mass production and matching of the two tines sufficient for a clean tuning-fork behaviour. Nevertheless, the sensor applications of this attractive device are fewer, only a few important ones have emerged. The main reason seems to be that the degrees of freedom for choosing appropriate cut angles are significantly restricted by the etching technology.

The double ended flexural mode resonator is of exclusive significance for the force or stress sensor and is, therefore, acknowledged in the respective section 7.

2.2. SAW-Delay Lines and Resonators

A SAW-device is characterised in that a Surface Acoustic Wave is electrically excited in a piezoelectric single crystal plate substrate (e. g. quartz, lithiumniobate, lithiumtantalate) by use of an metallic (e. g. Aluminium) Interdigital Transducer (IDT) structure /5/. The covered frequency range starts at about 50 MHz and ends up at a few GHz, corresponding to a state of the art interdigital finger linewidth below 0.5 µm. Two basic design concepts for the sensitive element can be discriminated, the SAW delay line and the SAW resonator. In both cases the influence of the measurand quantity on the acoustic wave phase propagation speed v is utilised for the sensor effect. Depending on where the interaction between the environment and the SAW-device takes place, two categories of sensors may be discriminated. In the first case, that is here treated as the homogeneous resonator case, the entire substrate is effected (e. g. SAW sensors for temperature and stress). In the second case, that is here treated as composite resonator case, the surface only is affected by an overlaying layer (see section 3.2). Already in the first case, the sensor transfer function between the measurand quantity and the measurement value is more complex than in the bulk case, because more than one elastic constant is contributing to the SAW propagation speed.

2.2.1. Two-port delay line

A two port SAW delay line is characterised by a propagation path between two separate IDTs, the first is working as transmitting transducer and the second as receiving transducer for the acoustic wave (see Fig.4). The edges at both ends are tilted or loaded with an absorbing resin to suppress reflections in the propagation direction of the primary wave. The delay time t_d between the electrical input and output signal is determined by

$$t_d = l / v \,, \tag{2}$$

whereby l is the mean spacing between the two IDTs, and v is the SAW phase propagation speed. The maximum acoustoelectric IDT interaction occurs at a characteristic frequency f_0 defined by the following equation

$$f_0 = \frac{v}{\lambda_0} = \frac{v}{p} \,, \tag{3}$$

whereby λ_0 is the corresponding characteristic wavelength, and p is the IDT's period length or pitch. Using a simple IDT as sending transducer, the insertion loss is significant, because the first IDT is radiating in both directions. Unidirectional IDT structures have been developed more recently /6/, but are more sophisticated. If the delay line is included as frequency-determining element in the feedback loop of an electronic oscillator circuit, the losses can be electrically compensated and a signal with the characteristic frequency f_0 can be maintained. Thus, the delay line together with the electric amplifier performs like a

resonator utilising a *resonant travelling wave* with electrically compensated damping. Unfortunately, there are many closely spaced resonances f_n

$$n\frac{\lambda_n}{2} = n_0\frac{\lambda_0}{2} = l = n\frac{v}{2f_n} = n_0\frac{v}{2f_0}\,, \tag{4}$$

excitable, whereby the excitability decreases with increasing distance of f_n from f_0.

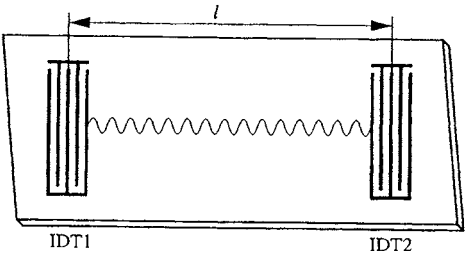


Fig.4. Two-port delay line SAW device

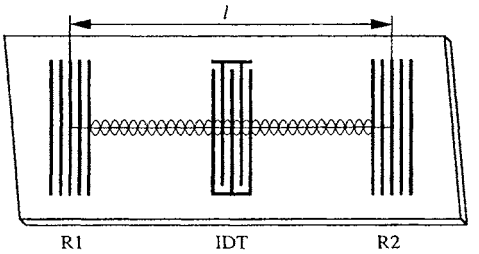


Fig.5. One-port SAW device with the IDT in the centre between two, at the desired resonance frequency totally reflecting, gratings

2.2.2. Single-port resonator

A single-port SAW resonator uses one IDT structure in the centre between two, at the characteristic frequency f_0 totally reflecting, gratings (see Fig.5). Between the reflecting gratings a *standing wave* can be electrically excited by the centre IDT. Thus, a high overtone acoustic resonator is formed. The resonance frequency f_n is given by

$$f_n = \frac{nv}{2l}\,, \tag{5}$$

whereby l is the effective lengths of the resonator, n is the overtone number and v the SAW propagation velocity. As in the case of the BAW-resonator, depending on the piezoelectric material properties and on the angle of cut, the sound speed is sensitive to changes in the elastic properties and the density. However, the Q value of SAW resonators is in the low range between 6000 and 12000 /7/.

3. COMPOSITE RESONATORS

In many sensor applications an additional layer is acoustically coupled with the basic monocrystal resonator. Since these composite resonator structures require a more sophisticated derivation of the sensor transfer function, they are treated here separately.

3.1. Layered BAW-Resonators

3.1.1. One-dimensional model

Harmonically excited thin plate thickness resonators can be described one-dimensionally (see Fig.6). That is, all physical quantities like the displacement, stress, and electrical potential depend exclusively upon one - the thickness - dimension. For the description of a layered piezoelectric BAW-resonator with two electrodes as well as with an arbitrary number of electrodes a rigorous one-dimensional mathematical model is available. /8,9/ In this model every layer is described by an 8x8 matrix with the input parameters elastic, piezoelectric, dielectric constants, density and thickness of the regarded layer. The total resonator is represented by multiplication of these matrices. If the influence of the measurement quantity on one ore more of these input parameters is known, the sensor transfer function can be obtained easily. In this model the very complex coupling between the three sets of eigen-modes at the layer interface planes is fully included. In practice, we have also an amplitude variation in lateral direction, but as long as this variation is very flat compared with the variation in thickness direction, the one-dimensional description accurately models the reality for harmonic resonance modes. The link between the one-dimensional description and the real finite resonator (e. g. plano-convex shape) is given by the effective vibration area introduced already in section 2.1.4. The influence on the effective area caused by loading the piezoelectric crystal with an additional passive layer was investigated by SCHMID /1/. In the case where only one set of harmonic eigen-frequencies (e. g. single rotated Y-cut quartz crystals like the AT-cut) is piezoelectrically excited and where the frequencies of the three sets are not coming to close to each other, the one-dimensional description becomes very simple. If we neglect the acoustic influence of the electrode layer, we obtain for the harmonic parallel (open electrodes) resonance frequencies f_n of a two layer composite resonator /10/

$$z_Q \tan Q_n + z_F \tan F_n = 0 \ , \quad Q_n = 2\pi f_n \frac{l_Q}{v_Q} \ , \quad F_n = 2\pi f_n \frac{l_F}{v_F} \ , \tag{6}$$

$$z_Q = \sqrt{\rho_Q c_Q} \ , \quad z_F = \sqrt{\rho_F c_F} \ , \quad v_Q = \sqrt{\frac{c_Q}{\rho_Q}} \ , \quad v_F = \sqrt{\frac{c_F}{\rho_F}} \ . \tag{7}$$

For the AT-cut the stiffness constant c_Q effective becomes

$$c_Q = \bar{c}_{66} = c_{66} + \frac{e_{26}^2}{\varepsilon_{22}} \ , \tag{8}$$

whereby c_{66} is the shear stiffness constant, e_{26} the piezoelectric constant, and ε_{22} the dielectric constant involved.

For the harmonic series resonance frequencies (short-circuited electrodes) relation (6) in principle remains valid, only the effective acoustic impedance of the crystal and the effective shear wave velocity has to be slightly modified, because the short-circuited electrodes produce a partial unstiffening of the piezoelectrically stiffened elastic constant $\bar{c}_{66}$ ($z_Q \rightarrow z_{Qs}$, $v_Q \rightarrow v_{Qs}$) /8/.

The inneracoustic losses in each layer can be considered by complex interpretation of the respective elastic constants

$$c = c' + jc'' \ , \tag{9}$$

whereby j is the imaginary unit. Introducing this complex interpretation of the stiffness constants, Eqs. (6) and (7) describe the general case of two *viscoelastic* layers forming a composite resonator. Vanishing imaginary part $c = c'$ describes the *elastic limit case*, vanishing real part $c = jc''$ the *viscous limit case*.

3.1.3. The crystal/liquid interaction

It should be emphasised that Eqs. (6,7,9) describe solid as well as liquid layers rigorously within the limits of the one-dimensional treatment. In the case of a solid layer in contact with the crystal, both layers are closer to the elastic limit and Fig.6 visualises the situation schematically. In the case of a solid layer in contact with a liquid, the piezoelectric layer is closer to the elastic limit and the liquid layer is closer to the viscous limit. The latter case has been investigated rigorously first by KANAZAWA and REED /11,12/. The situation is visualised by Fig.7. For a Newtonian liquid with $c_L'' = \omega_n \eta_L$, whereby ω_n is the regarded harmonic angular frequency and η_L is the viscosity of the liquid, we obtain from Eqs. (6,7,9) for the relative frequency change due to the liquid immersion of one side of the crystal

$$\frac{\Delta f_n}{n f_Q} = -\sqrt{\frac{f_Q \, \rho_L \eta_L}{n\pi \, \rho_Q c_Q}} \ , \quad \Delta f_n = f_n - n f_Q \ . \tag{10}$$

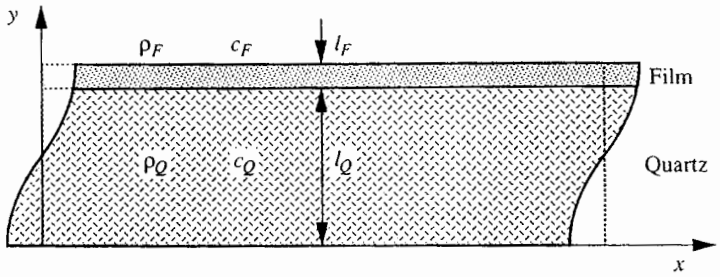


Fig.6. One-dimensional model (section of an infinite plate) of a thickness shear resonator crystal (e. g. quartz AT-cut) loaded with a foreign film and excited at the fundamental resonance. ρ_Q , ρ_F are the densities, c_Q , c_F the effective elastic constants, and l_Q , l_F the thicknesses of the quartz and film, respectively.

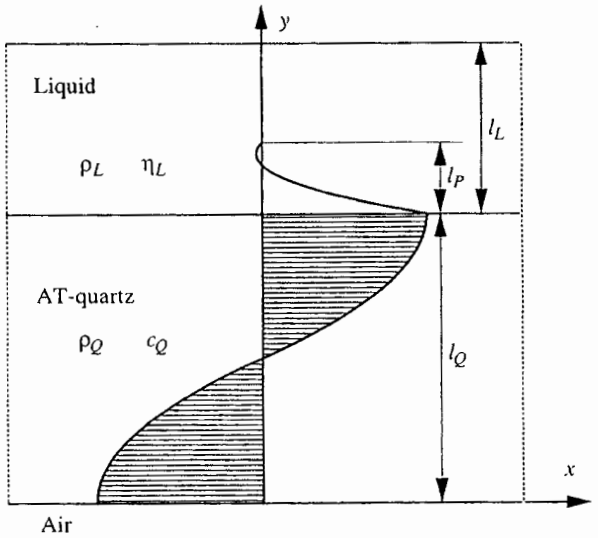


Fig.7. One-dimensional model of the crystal/liquid composite resonator. ρ_L , η_L , and l_L are the density, viscosity, and thickness of the liquid layer, respectively. l_P is the penetration depth of the shear wave into the liquid.

The most comprehensive and useful treatment of the crystal/liquid resonator was given by MARTIN ET AL. /13/. As long as the penetration depth of the shear wave into the liquid is sufficiently smaller than the thickness of the liquid layer, the liquid layer thickness is of no influence, that is, the liquid space can be regarded as semi-infinite. This is mostly fulfilled,

in the MHz frequency range and for low viscosities (e. g. water) the penetration depth, defined as the depth where the amplitude is decreased to a certain fraction of the value at the interface plane between the crystal and the liquid, is only a few μm.

3.2. Coated SAW-Resonators

In the sensor application field, mostly the propagation path between the two IDTs of a two port SAW delay line is used as sensitive area. This area is usually coated with an analyst sensitive coating. The composite resonator analysis is much more difficult than in the BAW resonator case, there are no rigorous analytical sensor functions available. The difference between the SAW and the BAW sensor is best illustrated for the simplest case of the microbalance application. If a foreign layer is deposited upon the BAW-resonator, the relative frequency decrease is exactly proportional to the relative mass load (see section 8.1). If the layer thickness increases, in addition *one* elastic constant of the film becomes some influence on the composite resonance frequency. If a foreign layer is deposited upon the SAW-resonator (or upon the back feeded delay line device), from the very beginning the frequency change is depending on the mass load as well as on several elastic constants. In addition, there is also an influence of the electric conductivity and the dielectric constant of the film. If the layer thickness increases, the influence of additional mass loading on the composite resonance frequency decreases towards zero.

3.2.1. SAW sensor utilising the difference principle

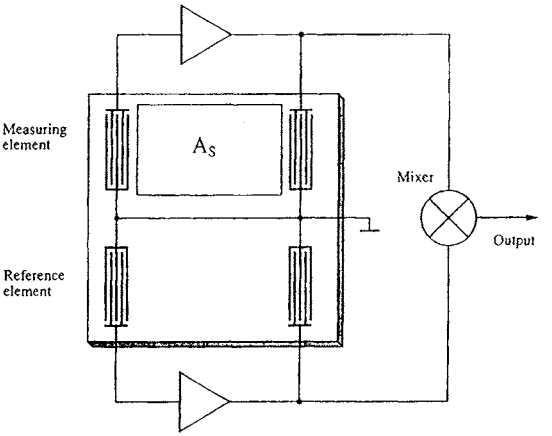


Fig.8. SAW sensor using the difference principle

Because many parameters may contribute to a frequency change of the SAW device, the reverse determination of a parameter change from a measured frequency change is ambiguous. The most universal solution of the problem is the use of two equal devices, whereby only one of the devices is coated with a layer that produces a known exclusive sensitivity to the measurand quantity (e. g. analyst concentration), while the other device is used as reference element (see Fig.8) /14/. The difference of the frequency changes of both devices is only sensitive to the measurand quantity, the frequency changes caused by all other influences are compensated by this difference principle.

3.2.1. Effect of thin layer coating

The strength of the SAW-sensor is the extraordinary sensitivity of the propagation velocity of the SAW on any changes of the surface material properties. Already an ultra thin foreign layer at the surface causes a detectable change of the SAW-controlled frequency, as long as the film has sufficiently different mechanical or electrical properties compared to the substrate. The SAW-sensor is more a sensitive monitor than an accurate measuring device.

3.2.2. Performance in liquids

The use of Rayleigh type SAW devices in liquids is not recommended, because there is an extreme damping of the SAW by the liquid. The problem is more serious than in the case of the BAW resonator, since the pronounced out-of-plane component of the Rayleigh-wave generates longitudinal waves in the liquid that carry a high energy flow density apart from the SAW. Promising alternatives are indicated in the following two sections.

3.3. Lamb Wave Resonators

A new sensor concept similar to SAW-sensors but employing Lamb waves was first published in 1988 /15/. The lamb wave is guided in a very thin plate whose thickness is small compared with the ultrasonic wavelength. The major benefit is the fact that the liquid is in contact with the substrate at the side opposite to the IDTs side and the quality factor is not much lowered, although the substrate is immersed at one side in a liquid. The low operating frequency in the MHz range is also convenient. The thin plate has a low heat capacity and hence a rapid thermal response.

3.4. Surface Localised HPSW-Resonators

With respect to the damping effect caused by the liquid immersion, the surface localised Horizontal Polarised Shear Wave (HPSW) resonator is directly comparable with the BAW thickness-shear resonator, only the operational frequency is much higher /16/. With a 148 MHz sample the frequency response reaches saturation at a layer thickness of about 1 µm.

4. DEMAND ON SPECIALISED ELECTRONICS

4.1. Passive measurements

The most information about the sensing resonator can be obtained by measurement of the locus of electrical admittance, that is the real and imaginary part of the admittance with the frequency as parameter, or the entire resonance spectrum, that is the absolute value of the admittance versus the frequency (see Fig.3), by use of a network analyser or a vector voltmeter. While such a measurement set-up /17/ is indispensable for scientific analysis of the sensing resonator, the measurement procedure and the resonance frequency evaluation is too time consuming for immediate utilisation in sensor applications.

The term passive sensor is sometimes used with a perfectly different meaning. A passive SAW sensors means that the two IDTs are directly connected to antennas, one is receiving a burst signal and the other is remitting the signal after the measurement value dependent delay time. This allows a very attractive wireless remote use of the sensor. The problem is, of course, the significant signal reduction by the return transmission path.

4.2. Oscillator circuits

The characteristic use of a sensing piezoelectric resonator is as frequency determining element in a feedback loop of an oscillator circuit. This is sometimes called an active measurement, because the resonator is actively controlling the frequency. The oscillator circuit keeps the sensor at one of its mechanical resonances by electrical excitation and outputs an electrical signal with the frequency of the mechanical resonance.

With respect to three properties the demands on the oscillator electronics for sensor applications of piezoelectric resonators differ from the oscillator electronics for frequency/time applications:

- Because the utilised resonance frequency changes with varying measurement quantity, instead of a fixed frequency a frequency range (in the thin film thickness sensor quartz example 20%) has to be allowed by the oscillator circuit.

- The Q-value of composite resonators often varies significantly with measurand quantity. As a consequence, the gain must vary over a wide range (in the thin film thickness sensor quartz example 10:1) to maintain the loop gain of the oscillator circuit equal to one for maintaining the harmonic excitation.

- In the sensor application a distance between the sensing resonator and the electronics is usually wanted. For instance, in the case of a temperature sensor it is advantageous to locate the sensor remotely to do not restrict the temperature range by the respective limits of the electronics and to avoid a heating of the sensor by the electronic's dissipation.

4.2.1. The cable problem

In the high frequency range above about 25 MHz a coaxial transmission line with a cable-length of an integer multiple of a half wavelength ($n\lambda/2$) of the electrical signal in the cable can be used to allow a distance between the sensor and the electronic oscillator circuit. The transformation properties of such a cable transform the electrical admittance of the resonator 1:1 to the position of the oscillator electronics. Thus, the oscillator circuit can be designed as for the case where the resonator is directly connected to the electronics. For example, by use of a 50 MHz AC-cut quartz crystal as temperature sensor, a coaxial cable (Teflon® insulator, $\varepsilon = 2$) corresponding to a λ-length of 4,23 m has been successfully used /18/. In this application the frequency change in the covered temperature range from -50 to 200 °C is less than 1%, therefore the λ-condition remains fulfilled.

In the case of high frequency SAW, especially if delay lines according to the difference principle are used, the electronics is usually located closely to the sensor position.

In the lower frequency range, e. g. for the quartz crystal thin film thickness sensor, a cable with a length l sufficiently smaller than a quarter wavelength must be used ($l << \lambda/4$). A $\lambda/4$-wavelength cable would transform a low impedance load (e. g. in case of series resonance of the sensor) into a high impedance load at the oscillator circuit, hence the cable influence would be maximum.

4.2.1. Balanced bridge oscillator

In the case of a cable with a length l sufficiently smaller than a quarter wavelength ($l << \lambda/4$) the cable performs essentially as a capacity C_C which adds to the resonator's static capacity C_0. The cable capacity C_C causes in some cases an impedance which is that low that the oscillator circuit oscillates already without any sensor crystal connected. Even if this extreme condition is not reached, a high contribution of the cable capacity C_C to the oscillator frequency remains. Since this is undesirable, the cable capacity should be compensated. For one frequency this can be done in principle by connecting an inductor of the same admittance as C_C in parallel. But an inductor has different and less perfect properties (especially with regard to temperature behaviour and losses) than a capacity. Furthermore, in the case of the thin film thickness monitor sensor, where the sensor frequency changes typically from 6 to 5 MHz with increasing film load, a compensation that works over the entire measurement frequency range is desired. Such a perfect compensation can be reached with the balanced bridge oscillator, the simplified circuit diagram of which is shown in Fig.9. This circuit is characterised by two almost equal bridge branches, the first one contains the sensor crystal connected via the sensor cable, the second one an equal cable terminated by a capacity C_p equal to the resonator's static capacity C_0. The circuit scheme also shows the rectified signal voltage U_{AGC} that controls the feedback loop amplification to the value 1 (Automatic Gain Control AGC), thereby maintaining a harmonic oscillation over the full damping range of the sensor crystal.

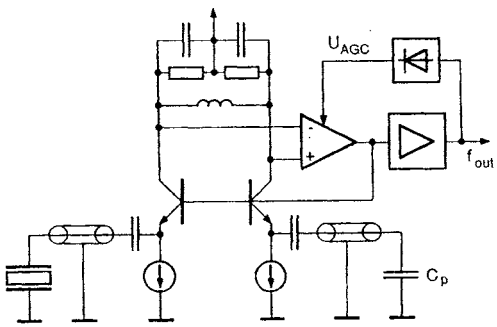


Fig.9. Simplified circuit diagram of the balanced bridge oscillator

4.3. Frequency versus (multiple) period-time measurement

It is often stated in literature that a higher sensor frequency (e. g. of SAW sensors) results in a higher sensitivity of the sensor. But this is only true in the case, where the resolution is determined by the frequency measurement itself. The resolution of the frequency measurement is directly dependent on the frequency f_x to be measured and the time base t_b of the counter (see Fig.10). The time base of the counter is identical with the measurement time and identical with the averaging period of the measurement value. A time base of 1 sec means that the counter content at the end of the measurement (= gate open) time is directly the frequency value f_x. In some cases a measurement time of 1 sec is too long and then the time base and thus the resolution has to be decreased. But in practice the resolution of a frequency sensor in the range above 100 MHz is always determined by other factors (signal noise and other instabilities of the frequency) than the frequency measurement method itself.

In the lower frequency range the (multiple) period time measurement method is more appropriate and allows a selection of the resolution and the measurement time in a very wide range, if the reference frequency is chosen sufficiently high (typically 100 MHz or higher). Hence in practice also below 100 MHz the resolution is always determined by other factors than the digitising measurement method itself.

In Fig.10 the frequency measurement and the (multiple) period measurement set up are schematically shown in one representation. It can be seen clearly that both methods are rather dual and differ only by the use of the inputs for the frequency f_x to be measured and for the reference frequency f_R. The change from one method to the other can be performed simply by exchanging these inputs. With both methods the measurement time is

determined by the time base counter capacity which can be fine-adjusted by the time base counter preset and the measurement zero (reference) point can be adjusted by the measurement counter preset.

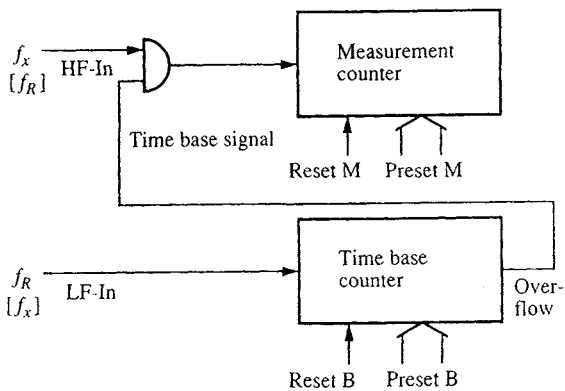


Fig.10. a) Frequency meter and b) multiple period measurement set-up

Having in mind the freedom of choosing the appropriate digitising method and the proper measurement parameters, it is misleading to compare sensitivities, e. g. of SAW and BAW sensors, by expressing them by absolute frequency over measurement quantity values. For an adequate characterisation of sensors, the reduced sensitivity S , that is the frequency change caused by unit measurand quantity referred to starting frequency, should be used exclusively. Higher sensor frequency (e. g. of SAW resonators) does not mean more *sable* sensitivity. For a fair comparison of sensors the performance figure SQ, defined as product of reduced sensitivity S (frequency change caused by unit measurand quantity referred to starting frequency) and resonator Q-value, is recommended:

$$SQ = S.Q$$

4.4. Multisensor bus connection

The following reasons indicate the need for a multisensor bus connection system:

- In measurement systems often many remote sensors are utilised.
- The length of the coaxial cables between the sensor resonator and the oscillator electronics is technically limited.
- Coaxial cables are expensive.

One solution is the so-called Interleaved Time Interval Sensor (ILTIS) bus /19/. It was

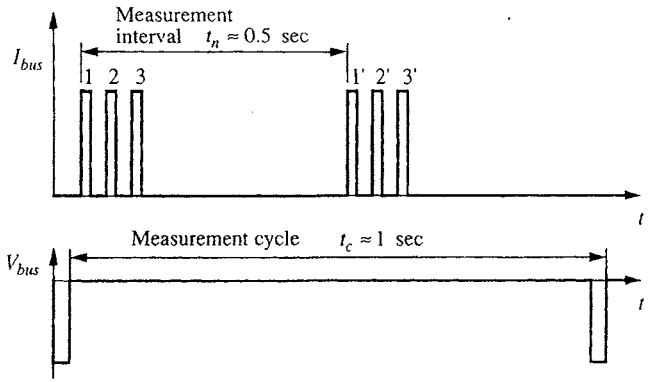


Fig.11. Time interval scheme of remote multisensor ILTIS bus. 1...First signal of sensor 1, 1'...Second signal of sensor 1, S...Synchronisation signal, I_{bus}...Current pulses, U_{bus}...Voltage pulses.

developed for a multiple quartz thermometer system using sensor crystals with a frequency of 16 MHz. These temperature sensors are connected via a short coaxial cable to the oscillator circuit and the output frequency of the oscillator is divided down to 2 Hz and a pulse-forming module generates short current pulses in time intervals of 0.5 sec. This time interval signal is transmittable by an inexpensive twisted pair cable of arbitrary length. Further, it is possible to connect up to 16 sensors at the same cable (bus system) by interleaving the time signals (see Fig.11). The synchronisation is performed by a voltage pulse on the same wire. A complete bus cycle takes about 1 sec.

5. SUMMARY OF PRIMARY AND SECONDARY MEASURANDS

In Table 1 an overview about the possible measurement quantities is given, whereby a discrimination between primary and secondary measurands is introduced. Primary measurands are quantities that appear directly in the resonance determining equations as parameters, secondary measurands influence one or more of those parameters and thus indirectly influence the resonance frequency. In the following sections a selection of established sensors based on piezoelectric crystal resonators is described.

Table 1. Summary of primary and secondary measurands

Primary sensitivities	Secondary sensitivities
Homogeneous resonator	
Temperature T [K]	Radiation intensity Anemometric flowmeter Dew point hygrometer Penning type vacuum pressure meter
Stress P [Pa]	Force Hydrostatic pressure Acceleration Mass (spring balance) Magnetic field
Composite resonator	
Areal mass density ρl [kg/m^3]	Film thickness Deposition or etching-rates Adsorption and absorption rates Analyte-selective coating Chemosensors in gas phase E. g. Hg-vapour / Au-electrodes
Specific acoustic impedance $z = \rho v = \sqrt{\rho c}$ [kgs/m^2]	
Viscoelastic case $c = c' + jc''$	Molecular weight of polymers Analyte-selective coating Chemosensors in liquid phase Antibody-Antigen selective coating Immunosensors
Elastic limit case $c = c'$	Effective elastic constant of solid films Auto-Z-Match thin film thickness monitors
Viscous limit case $c = jc''$ Newtonian liquid $c'' = \omega\eta$	Viscosity of liquids Liquid chromatography

6. QUARTZ CRYSTAL THERMOMETERS

Since the quartz crystal is an anisotropic medium, the resonance frequency of a plate according to Eq.(1) is highly dependent on the crystallographic orientation of the plate, the so-called angle of cut. This was already used in the early years of quartz crystal frequency control to produce crystal plates with vanishing temperature coefficient (AT- and BT-cut). According to the systematic investigation of BECHMANN, BALLATO and LUKASZEK /20/, the frequency versus temperature behaviour can always be sufficiently approximated by a third order polynom

$$\frac{\Delta f}{f_0} = a_0 + a_1 \Delta T + a_2 \Delta T^2 + a_3 \Delta T^3 \ , \quad \Delta f = f - f_0 \ , \quad \Delta T = T - T_0 \ , \tag{12}$$

where f and f_0 are the plate resonance frequencies at the temperatures T and T_0 (reference temperature), respectively, and the a_0, a_1, a_2, a_3 are the zero, 1st, 2nd, and 3rd order temperature coefficients (TCs), respectively. In 1962 the first systematic approach to promote the temperature sensitivity of the frequency from an unwanted effect in the frequency control application domain to an attractive temperature sensor concept was published by WADE and SLUTSKY /21/. With their non-rotated Y-cut crystal they obtained a resolution of 10^{-3} °C; one year later this was improved by SMITH and SPENCER to 10^{-6} °C /22/. But the crystal was enclosed in a rather clumsy glass vacuum bulb, that is the time constant and size was not really useful for practical temperature sensor applications. One of the compact quartz temperature sensor developments in the following decade used a biconvex AC-cut crystal with a temperature sensitivity of 20 ppm in Helium atmosphere of a few mb pressure in a copper TO-5 transistor case /18/. The crystal was driven on the third harmonic at 50 MHz, yielding a sensitivity of 1000 Hz/°C and a usable resolution of 10^{-4} °C for a 10 second time base frequency measurement.

A very successful development was the so-called LC- (Linear temperature Coefficient) cut by workers at Hewlett-Packard /23/. They systematically searched for a doubly rotated Y-cut with simultaneously vanishing second and third order temperature coefficients. The found solution exhibits a first order TC of 35 ppm. The sensor was sold over a period of more than 25 years in a general laboratory temperature measurement instrument for use from -80 °C to 230 °C with a calibration accuracy of 0.02 °C and a usable resolution of 10^{-4} °C for a 10 second counter gate time. The main disadvantage of that sensor is the significantly higher prize of doubly rotated compared to singly rotated Y-cuts. With the advent of the microprocessor the linearity of sensors became less important and consequently the last milestone in the development of shear mode plate resonator temperature sensors came back to the highest sensitive cut. ZIEGLER introduced in 1987 the HT-cut quartz temperature sensor based on a slightly singly rotated Y-cut (Θ = -4 °) with a frequency versus temperature sensitivity of 90 ppm/°C /24/. The nonlinearity is only a few percent over the full temperature range.

Tuning fork resonator based temperature sensors have been developed especially by UEDA /25/ (flexural mode, 45 ppm/°C, -270...250 °C) and EERNISSE /26/ (torsional mode, 30 ppm, -50...150 °C). The higher frequency torsional modes (fundamental torsional mode is typically at 200 kHz) tend to couple with high overtones of the lower flexural modes (fundamental flexural mode is typically at 2 kHz). The problem is serious, because the flexural mode frequencies decrease, while the torsional modes increase with increasing temperature (typically 0.3 to 0.5 % over a 100 °C range). Therefore the potential for frequency crossings and - as a consequence - frequency jumps and activity dips in the full temperature range is extremely high. There are only a few dimensional choices to avoid this problem associated with torsionally excited tuning fork temperature sensors.

The dissipation due to resonator damping of all mentioned bulk resonators is only in the order of 100 µW. Hence, selfheating of the sensor dispersed in flowing liquids is usually no problem.

The first SAW *delay line* based temperature sensor was described by HAUDEN et al. /27/. A frequency drift that corresponds to ±0.06 °C in 10 days and a non-linearity of 1 ‰ is typical /28/. The first SAW *resonator* based temperature sensor was published by VIENS and CHEEKE /29/, it features higher stability and resolution and a sensitivity of 80 ppm/°C. However, since the Q-factor is only 550, the performance of SAW temperature sensors remains almost two orders of magnitude lower than that of BAW temperature sensors.

7. FORCE, STRESS, AND PRESSURE TRANSDUCERS

The first investigations of the frequency versus stress sensitivity of quartz thickness shear resonator disks were reported by BALLATO and BECHMANN /30/, RATAJSKI /31/, and DAUWALTER /32/. Fig.12 shows a diametric force pair that squeezes a circular quartz plate. The frequency-force effect was thoroughly characterised by several workers /33,34/. The frequency shift is due to non-linear effects, that is the stiffness constant in Eq.(1) changes slightly with the stress applied, the effect of the stress on the dimension (strain) or the density is usually negligible. Zeros for the effect can occur when the squeezed diameter is aligned in certain directions for a given cut. While these directions are very recommended as preferred mounting points to improve the stability of the resonant frequency, in sensor applications the directions of maximal effect are of more interest, such as the x-direction on an AT-cut disk. The diametric force effect of Fig.12 has been used in a high-performance pressure transducer by KARRER and WARD /35/. The complete structure was formed from quartz to avoid unwanted forces from differences in thermal expansion. Another all-quartz structure has been developed by Hewlett-Packard workers /36/. In this construction hydrostatic pressure is radially applied via a quartz crystal cylindrical shell that forms one piece with the biconvex circular BT-cut resonator disk. It was realised that creep in a joint close to the resonator causes hysteresis and long-term zero drift. The incorporation of an integral shell moves the joint far enough away from the resonator to

minimise joint instability influence on frequency. With this construction excellent long term accuracy was obtained.

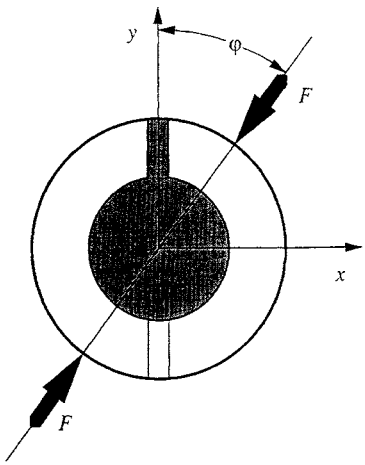


Fig.12. Force sensor using a disk thickness shear mode resonator with diametric force F pair applied.

The dynamic range (measurement range over resolution) of this outstanding device is 10^7, the overall accuracy is 0.01%. However, this transducer needs an oven-controlled temperature (the frequency/force-coefficient is slightly temperature dependent), it is very clumsy and expensive. Thus, its application was restricted to oceanography and oil and gas field exploration.

The key problem of pressure transducers based on crystal resonators are two counterbalancing demands: At one hand the resonator has the highest possible quality factor if it is fully embedded in vacuum as totally reflecting boundary condition. On the other hand, since pressure cannot be transferred via vacuum, the need for a pressure transfer path leads to a rather rigid support. An interesting approach to solve both demands without compromise used acoustic quarter wave length ($\lambda/4$) transformation layers to obtain total reflection by simultaneously squeezing the crystal plate homogeneously at its major surfaces /37,38/. In this case a singly rotated Y-cut crystal working on its 13^{th} harmonic thickness shear overtone mode at 100 MHz was used. The transformation layers were formed by sputtered films with alternating low and high acoustic impedance (Al /W). Because of the high frequency, the thicknesses of the layers were in the order of only 10 µm. The layers formed two sandwich packets that electrically took over the electrode function and acoustically the totally reflecting feature of the vacuum boundary condition by simultaneously allowing a very rigid pressure transfer to the vibrating crystal.

In all the above examples the total frequency range for full pressure range is only about 0.3%. A significant step forward to obtain a higher frequency-force effect was done in 1969 /39/ by the introduction of a string type resonator. This concept uses a long, slender, laterally (or flexurally) vibrating beam, the resonance frequency of which is sensitive to a longitudinal force roughly according to

$$f_n = \frac{n}{2l}\sqrt{\frac{T}{\rho}}\,, \tag{13}$$

where T is the stress applied to and l the length of the „string". Tension causes the frequency to go up, much like a guitar-string frequency goes up as the string is tightened. Eq. (13) is identical with Eq. (1), except that the stiffness constant c is replaced by the stress T. Both quantities have the same dimension (Pa), but the stress can be changed over a much higher range than the elastic stiffness, which latter one is varying with the applied stress only due to a second order effect. As a consequence, the string type pressure transducer features for full pressure range a corresponding frequency range of about 10%. Since the beam actually shows a bending stiffness, it is far away from a mathematical string and Eq. (13) is only a rough approximation. Therefore the resonance frequency does not go to zero with vanishing stress. Only the increase in frequency comes from the restoring force caused by the longitudinal tension adding to the restoring force caused by the bending rigidity of the beam. As can be already expected from Eq. (13) and the rather high frequency range, the string type pressure transducers exhibit a noticeable nonlinearity.

At the ends of the beam again the problem of simultaneous acoustic termination and force introduction via solid means occurs. There are two concepts used to circumvent the moments acting at the mounting attachments of the laterally vibrating beam in order to minimise vibrational energy losses. The first approach involves acoustical isolation by a spring-mass system incorporated into both ends of the laterally flexing beam (see Fig.13a). The effect can be compared with that of a $\lambda/4$ transformation line, that produces an infinite stiff support from a free end termination. The other approach uses the tuning fork moment compensation effect. Here two versions, the double- (Fig.13b) and triple-beam /40/ (Fig.13c) design is used. The bending stiffness is controlled in the double-beam device by the photolithographically defined tine width, while in the triple-beam device it is controlled by the much thinner thickness dimension. Thus the triple-beam device is much closer to the „string" concept, the force induced frequency shift can be made as large as 40%. The middle beam is twice the width of the two outside beams and it moves in anti-phase to the two outside beams. The crystallographic orientation is similar to that used in the single beam device, that is length along x-axis and thickness in y-direction. Thus, vibration can be piezoelectrically excited by producing shear deformations in thickness direction using appropriate electrode design.

The most important contribution to the success of the double ended tuning fork pressure transducers came from EERNISSE and PAROS /41/. The typical quality factor reached is 150000.

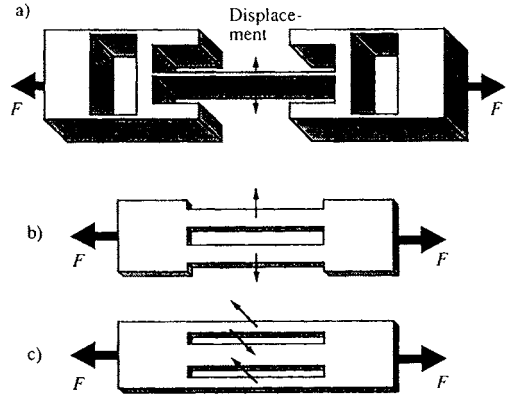


Fig.13. Double-ended flexural-mode "string tension type" resonator force sensor. Electrodes are not shown. F...Pulling force. a) Single beam device. b) Double-beam device, c) Triple-beam device.

Sandia Laboratories have developed a very small high-performance acceleration switch using a double-ended tuning fork which can be placed in a Dual-In-Line plastic package for mass production /42/. Single ended tuning fork resonators can be successfully used for monitoring vacuum pressure /43/, but here the primary effect is a damping increase with gas pressure, hence we leave the scope of this frequency-output sensor survey.

Several pressure sensor concepts based on SAW resonators have been reported, e. g. /44,45/, but no version is commercially available. Usually the substrate is single ended supported and bended or is part of a membrane which is deformed by the applied force. Interesting applications like strain gauge /46/ and electronic spirit level /47/ have been reported recently.

8. THIN FILM THICKNESS AND DEPOSITION RATE MONITORS

Quartz crystal thin-film thickness monitors are in wide use for control of deposition rate and for in situ film thickness measurement in coating plants of optics and semiconductor industry. The fundamental thickness shear resonance of an AT-cut quartz crystal is electrically excited while the crystal is coated by the same process as the substrate and the frequency change caused by the mass load on the crystal is used to calculate the current

thickness of the film. The sensor crystal is typically a plano-convex circular disk of 14 mm diameter and has a thickness of 330µm or 270µm corresponding to a fundamental resonance frequency of 5 MHz or 6 MHz, respectively.

8.1. Sensor transfer function

In the past, the major steps forward in the performance of the crystal thickness monitors were each correlated with improvements of the mass load versus frequency relation upon which the thickness monitors were based. The first generation of monitors, which sometimes are referred to as "frequency measurement technique" devices, assumed a direct identity between the reduced areal mass density m and the reduced frequency decrease $-\Delta f / f_Q$ caused by the deposited film, according to SAUERBREY'S /48/ original investigations,

$$m = \frac{m_F}{m_Q} = -\frac{f - f_Q}{f_Q} = -\frac{\Delta f}{f_Q} , \tag{11}$$

with

$$m_F = \rho_F l_F , \quad m_Q = \rho_Q l_Q . \tag{12}$$

Herein ρ_F and l_F are the density and the thickness of the deposited film on the sensing crystal, ρ_Q and l_Q are the density and the thickness of the crystal plate, respectively; f_Q and f are the fundamental harmonic resonance frequencies of the unloaded and of the loaded crystal, respectively; Δf is the frequency shift caused by the mass load. This simple Eq.(1) shows already that the quartz crystal thickness sensor head is primarily a mass sensitive device; in any case the density of the deposited material must be known and utilised by the monitor in order to provide a direct thin film thickness display, e. g., in nanometers.

Second generation devices assumed instead of Eq.(11) a direct identity between the reduced areal mass density m and the corresponding reduced period time increase $\Delta\tau/\tau_Q$ of the considered oscillation,

$$\frac{m_F}{m_Q} = \frac{\tau - \tau_Q}{\tau_Q} = -\frac{f - f_Q}{f} = -\frac{\Delta f}{f} , \tag{13}$$

according to BEHRNDT /49/. These so called "period measurement technique" devices offered an increased usable load range of the sensor crystal. That means, the crystal can be used over a longer evaporation period before the deviation of the assumed sensor transfer function from the real one becomes relevant. The most significant improvement of the utilised sensor function is based upon the fundamental analysis of MILLER and BOLEF /50/, as well as on the subsequent work of LU and LEWIS /10/, who for the first time treated the crystal and the film as a composite resonator structure. State of the art third

generation thin film thickness monitors utilise a sensor function which takes the influence of the different specific acoustic impedances of the deposition materials upon the resonance frequency into account. This sensor function is the explicite form of Eq.(6),

$$\frac{m_F}{m_Q} = F(\Delta f, f_Q, z)\,, \tag{14}$$

$$z = z_F/z_Q\,, \qquad z_F = \rho_F v_F = \sqrt{\rho_F c_F}\,, \qquad z_Q = \rho_Q v_Q = \sqrt{\rho_Q c_Q}\,, \tag{15}$$

where v_Q and c_Q are the phase velocity and the stiffness constant with respect to the piezoelectrically excited standing wave in the crystal, v_F and c_F are the corresponding parameters in the film; ρ_Q and ρ_F are the densities, z_Q and z_F the specific acoustic impedances, in the crystal and in the film, respectively. Eq.(14) indicates that the sensor function depends on the deposition material. In other words, the mass-sensitivity of the resonance frequency of the sensor crystal is different for various materials and this difference increases rapidly with increasing mass-load. Thickness monitors using Eq.(14) as sensor function are called "Z-Match" devices. The superior accuracy of these devices has been shown /51/ up to such extreme mass loads as $m_F/m_Q = 70\%$. However, these monitors need the z-value of the deposited film which can deviate from the tabulated bulk z-values of deposition elements. In the cases of non-standard alloys and of optical multilayer coatings the effective z-values are not known. For these reasons, the users often could not take advantage of the high accuracy potential of z-match devices.

In contrast to a homogeneous resonator, a composite resonator shows a non-trivial resonance frequency spectrum. In the general case of $z_F \neq z_Q$ the overtones are not exactly integer multiples of the fundamental. (Even multiples are also possible mechanical resonances, but they are usually not piezoelectrically excitable). Since the deviation from the trivial equally spaced spectrum increases with the deviation of the specific acoustic impedance of the film from that of the crystal, this characteristic frequency deviation can be used as a measure for the magnitude of the reduced specific acoustic impedance $z = z_F/z_Q$. Hence, the z-value can be calculated from two measured frequencies of the spectrum, e. g. from the fundamental and the third harmonic. This allows an automatic adjustment of the actual sensor function which strongly depends on the deposition material. The *effective* acoustic impedance of the film can be determined and therefore, in the frequent case of poorly or not known z-values, thickness determinations with higher accuracy are possible. With Auto-Z-Match the usable load range of the sensor crystal is only limited by the damping of the loaded crystal. The comprehensive analysis of the two-frequency concept has been described in previous works /52,53/ and is beyond the scope of this concised review paper.

It has been shown /54/ that an exchange of the sensor crystal is not influencing the accuracy of the monitor. This should be emphasised, since it evidences the extraordinary absolute accuracy of the quartz crystal as sensing element which means that a crystal

change causes no difference in the thickness display and there is not at least any need for individual calibrations of the mass sensitivity of the sensing element. The thickness display is only dependent on the resonance frequency f_Q of the virgin crystal, on the crystal preload m and on the z-value of the deposition material according to Fig.(14).

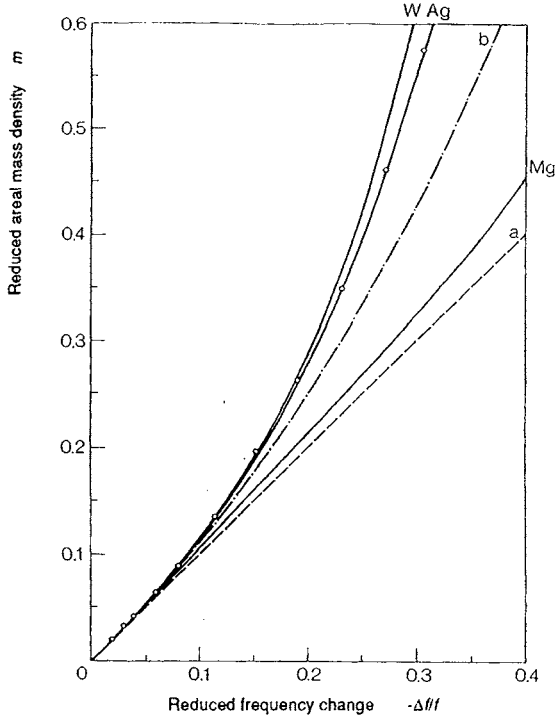


Fig.14. Dependency of the sensor transfer function of the sensor quartz crystal of a thin film thickness monitor on the deposition material (e. g. W, Ag, Mg) according to the two layer composite resonator model after Eqs. (6,7). For comparison the sensor functions according to the historical a) frequency measurement method after Eq. (11) and b) period measurement method Eq. (13) are also shown. o) Measurement points for Ag.

At that point it should be mentioned that a further two frequency concept had been described by WAHJDI /55/. This approach uses instead of two quasi-harmonic frequencies one quasi-harmonic frequency (the fundamental) and an anharmonic frequency. However, there are three drawbacks of this method:

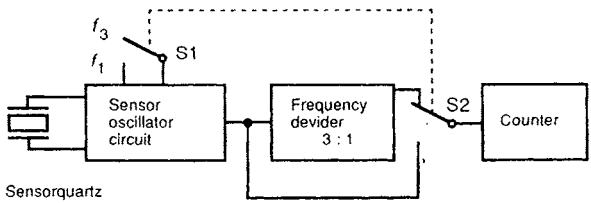


Fig.15. Two-frequency oscillator block diagram for state of the art Auto-Z-Match thin film thickness monitors

- There is no theoretical model available that supports the calculations based exclusively on empirical arguments.

- The two frequencies cannot be separated with a two frequency oscillator circuit concept. Therefore the anharmonic frequency has to be determined by passive measurement, which is much more sophisticated.

- The obtained accuracy is not at least as high as with the original "Auto-Z-Match" procedure.

Obviously the only reason for promoting this approach was the desire to avoid a patent infringement /56/.

8.2. Other factors influencing accuracy and usable load range

8.2.1. Film density

The film density is a parameter influencing the thickness display simply as a multiplication factor. Since the density of thin films varies with the deposition parameters like substrate temperature, deposition rate, and residual gas composition, this directly results in a corresponding error in terms of film thickness. In practice, however, the reproducibility of film ticknesses is much more important than the absolute thickness accuracy. Since the deposition parameters usually have to be controlled very carefully and kept at a certain constant value in any case in order to obtain well defined film properties, in practice the reproducibility of film-thicknesses is not diminished by the density uncertainty.

8.2.2. Relaxation

Especially at high deposition rates heavily preloaded crystals show an exponential frequency drift to higher frequencies after terminaton of the deposition process which corresponds to a virtually decreasing thickness display. This drift lasts over a period of several hours or days. The origin seems to be a recrystallisation process in the film. But

assuming a recrystallisation process, the mechanism influencing the resonance frequency of the crystal film composite is presently not really understood. The explanations offered in literature /57,58/ are contradicting and not convincing.

8.2.3. Intrinsic film stresses

The influence of film stresses on the resonance frequency of the sensor crystal is fully understood and the caused error in terms of film thickness can be quantitatively estimated /59/.

8.2.4. Temperature

The commonly used AT-cut sensor crystal exhibits a vanishing temperature sensitivity of the resonance frequency. But this compensation is only effective for the virgin crystal at about room temperature and as long as there are no temperature gradients appearing in the crystal. Moreover, the temperature sensitivity varies with mass load. Although the crystal holder is cooled, the temperature of the crystal plate rises after opening the shutter due to the high heat radiation from the evaporation source. For very small film thicknesses the error caused by the temperature influence becomes noticeable. The temperature influence

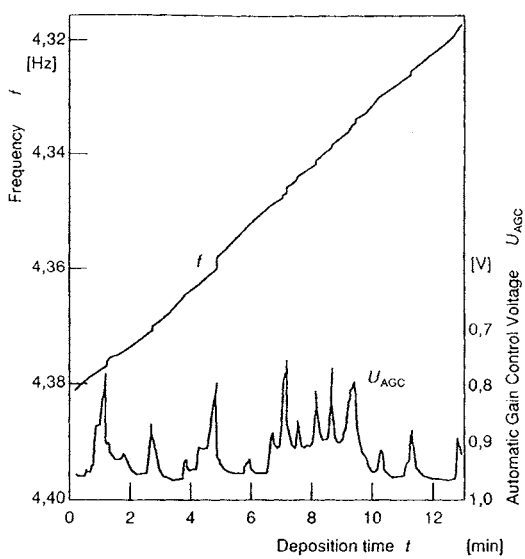


Fig.16. Frequency jumps and activity dips of a thin film thickness monitor sensor quartz with a deposition material splatter.

can be perfectly eliminated by using a chimney-like opening of the deposition shutter which exposes the sensing crystal to the heat radiation of the evaporation source already before starting the deposition process upon the substrates. Thus, constant temperature gradients remain in the crystal during the deposition process.

8.2.5. Frequency jumps (rate noise)

With increasing mass load conditions of temporary coupling with anharmonic thickness shear and twist modes or - more frequently - with high overtone bending modes arise (Compare Figs.(16,17)) /60/.

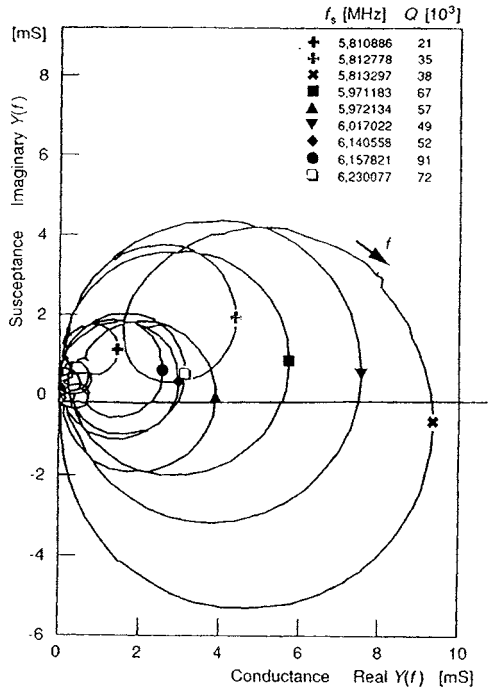


Fig.17. Locus of admittance curve of a thin film thickness monitor sensor quartz in the vicinity of a frequency jump. Mode coupling is indicated by the appearance of nooses in the locus curve.

These temporary mode couplings are responsible for so-called frequency jumps. Usually these jumps have small magnitudes and do not noticeable influence the thickness display

accuracy. Because the deposition rate is the time-derivative of the thickness increase, the frequency jumps appear in form of spikes in the pen trace of deposition rate strip chart records and are known as "rate noise". In the case of deposition material splatters on the crystal, however, these frequency jumps can become a serious problem. The tendency to such frequency jumps was greatly reduced by the development of a new crystal design with the code name "ppc-crystal". This "plane-plane-convex crystal" has the shape of a circular plate with a plane side and a convex side, whereby the convex side is flattened by a plane region in the centre (see Fig.) /61/. The new plane-plane-convex crystal shape improves the sensor crystal properties with respect to lower electrical impedance, higher signal to noise ratio, less rate noise and reduced temperature sensitivity /62/.

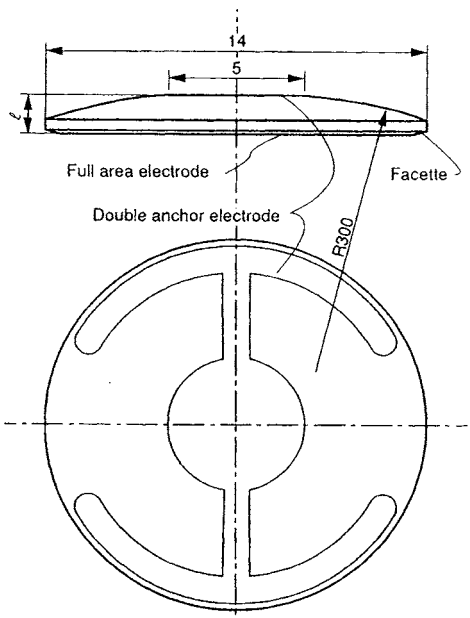


Fig.18. Plane-plane-convex (ppc-) quartz geometry.

8.2.6. Residual crystal life time indication

Presently used thin film thickness monitors only feature a simple "crystal fail" signal which informs the user about the end of the crystal life time. Actually the crystal does not suddenly loose its electrical excitability. As a matter of fact, the crystal damping increases rather continuously until the electronic oscillator circuit feedback loop gain is no longer

sufficient for sustaining the oscillation. Since the crystal damping is the parameter responsible for the determination of the usable crystal load range, this parameter can be utilised to evaluate a forecast of the remaining possible film thickness accumulation upon the crystal before the oscillation ceases. In practice this indication is very valuable since the usable load range of the sensor crystal is extremely material dependent. The usable crystal load range for the deposition material MgF_2 is, e.g., less than 1% of the range for the deposition material Al. The residual crystal life time indication based on the sensor crystal damping allows the safe use of the sensor crystals until the individual end of the usable load range. This is especially advantageous if the operator uses deposition materials for which no previous experience exists.

9. VISCOELASTIC LAYER PARAMETER MONITOR

The attractivity of the BAW composite thickness-shear resonator sensor principle for the characterisation of liquids is based upon the very good understood analytical model. But some fundamental problems must be considered:

Firstly, there is a dramatic decrease of the resonator's Q-value caused by the high damping of shear waves in liquids.

Secondly, the undesired excitation of spurious out-of-plane vibration amplitudes, that can never be avoided with finite shear resonator plates because of the angular momentum conservation /63/. Since longitudinal waves exhibit a comparable vanishing damping in the liquid, the probability that a high overtone longitudinal resonance coincides with the utilised shear resonance is rather high for within the covered viscosity range. A coupling between two different modes is not considered by the model and may cause serious irreproducibilities of the sensor's transfer function. One possible means to avoid unwanted longitudinal resonances is the use of a "spoiler" according to EGGERS /64/.

Thirdly, because of the small penetration depth of the shear wave into the liquid, the surface roughness of the crystal may have some influence on the accuracy. If the surface roughness is of the order of the penetration depth, an additional mass loading effect by the picked up liquid occurs /65/. Smooth-surface devices, which only viscously entrain a layer of contacting liquid, respond to the product of liquid density and viscosity. Textured-surface devices, which also trap liquid in surface features, exhibit an additional response that depends on liquid density alone. Combining smooth and textured resonators in a monolithic sensor allows separate measurement of liquid density and viscosity at the same time /66/.

10. CHEMOSENSORS

The use of the crystal sensor as an analytical device was first described by King /67/. Chemical sensors based on crystal BAW and SAW resonators are characterised by analyst

selective coatings and are reported both for the gas as for the liquid state /68,69,70,71/. Because the coating material is rarely exclusively selective for a single component, the resonating crystal sensor, as with most chemical sensors, is susceptible to interfering analytes. An alternate approach to the use of individual sensors for an analysis where interferents may be present is the multivariate approach with an array of sensors, each coated with a different partially selective coating material. If the number of sensors is equal to or larger than the number of components and all components possess different response patterns, quantitative analysis can be performed. This system is here referred to as *chemo assay* (Sometimes the term *smart sensor system* is used in literature). Such chemo assays have been investigated based both on BAW /72/ as well as on SAW /73/ systems.

One of the main problems of resonating gas sensors is the pronounced sensitivity to humidity /74/, the main problem of resonating sensors in liquids is the high damping. In both cases the SAW devices are more concerned. If the sensor is used for simple analysis rather than in situ monitoring of chemical reactions in liquids, only the interfacial reaction may be performed in the aqueous phase, but the final frequency measurement is made with the dry crystal.

11. BIOSENSORS

Biosensors are similar to chemosensors, but they utilise biological molecules as selective coating /75,76/. Extreme selectivity is obtained by the antibody-antigen reaction, such sensors are called *immunosensors*. In the case of a multiple sensor system the term *immuno assay* is recommended. Immunosensors work only in the liquid phase, most successful works are based on thickness shear AT-crystals. Some recent developments are of high relevance and seem to be close to commercialisation. E. g., an excellently performing blood group detection in the ABO system has been reported /77/. One side of the sensor quartz is coated with blood group antibodies in order to allow blood cells to interact with the quartz surface according to their group. Another example is a Human Immunodeficiency Virus (HIV) infection immunosensor /78/. With this analytical system antibodies specific against HIV have been detected in human serum samples within a few minutes and with a selectivity comparable to licensed Enzyme Linked Immunosorbent Assay (ELISA). One example for a SAW based biosensor is a Deoxyribo-Nucleic acid DNA hybridisation detector /79/. A major problem associated with biosensors is that they usually work non-reversible and the crystals can be reused only after repeating the sophisticated preparation procedure before each new measurement.

12. ALTERNATIVE PIEZOELECTRIC MATERIALS

Only a few of the many piezoelectric material structures have found wide spread technical application. The dominant monocrystal is quartz, the presently most used piezoelectric

ceramics is PZT (Lead-zirconate-titanate). In thin film technology ZnO is primarily used as piezoelectric layer material.

The decade of 1970...80 was characterised by Lithium-niobate and Lithium-tantalate substituting quartz in the field of SAW filter applications because of their higher elctromechanical coupling factors.

The decade of 1980...90 was characterised by the availability of PVDF (Polyvinylidenfluoride). PVDF foils are used in air for ultrasonic transducers, in water for hydrophones and in connection with solids as highly sensitive dynamic strain monitors.

To answer the question what might come next, one has to consider that a new material has only a chance if it offers significant advantages compared with already established ones. It is believed that the most promising candidate for the decade of 1990...2000 is the quartz-homeotype gallium-orthophosphate /80/. This material is of outstanding importance for sensor applications, since it features all the positive properties of quartz (e. g. high resonator quality factor) without the negative ones (e. g. usable temperature range, twinning, very low electromechanical coupling factor) /81,82,83/.

13. SUMMARY AND CONCLUSION

It should become apparent that BAW resonator sensors are primarily useful in the high-performance end of the measurement device market. This is particularly true for the thickness shear devices because of their size and costs. Present state of the art performance specifications are outstanding compared to most competing technologies. Short and medium term accuracies of some quartz crystal temperature and pressure sensors as well as of the quartz crystal areal mass density monitor are limited only by the traceability to primary standards. Further R&D efforts might take advantage of the availability of gallium-phosphate, which will allow smaller resonator sizes, broader full-scale ranges and better temperature capabilities.

The photolithographically produced single- and double-ended tuning fork designs are somehow between the thickness plate mode and the SAW device. In the cases of pressure or acceleration transducers they perform well enough to address the high performance instrument market. At the same time they allow miniaturisation and low cost mass production. The main restrictions of these devices are their fewer degrees of freedom for choosing an optimum cut angle.

The development of SAW sensors seems to be still more in its early stages. On the other hand, the SAW sensor exhibits a much higher potential for low-prize large-volume mass production than the BAW counterpart. This is primarily because a large number of SAW sensors can be produced at once, electrode (IDT) deposition is needed only on one side. SAW sensors are more attractive for monitoring purposes (in the extreme case yes or no

switches), rather than for physical measurand quantity determination with high absolute accuracy. Where this potential is needed, further R&D research is encouraged.

Higher sensor frequency (e. g. of SAW resonators) does not mean more *usable* sensitivity. To avoid continuing confusion in literature, for a fair comparison of sensors the introduced performance figure SQ, defined as product of reduced sensitivity S (frequency change caused by unit measurand quantity referred to starting frequency) and resonator Q-value, should be used. This figure is a general measure for the relevant stability and short term accuracy. For example, a BAW quartz thermometer has a performance figure SQ of one to two orders of magnitude higher than the SAW counterpart. Therefore other arguments must be found to launch a SAW temperature measurement device on the market.

14. ACKNOWLEDGEMENT

Part of the work was financially supported by "Österreichischer Fonds zur Förderung der wissenschaftlichen Forschung" (Austrian Science Foundation) under Project No. P 7198 and by Leybold AG Hanau, BRD, and Inficon/Leybold-AG, Syracuse, USA.

Literature

1 M. Schmid, E. Benes, W. Burger, V. Kravchenko, IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control, Vol.38, No.3, May 1991, 199-206, *Motional Capacitance of Layered Piezoelectric Thickness-Mode Resonators*

2 S. Hertl, E. Benes, L. Wimmer, M. Schmid, Proc. of the 39th Annual Frequency Control Symp. AFCS 85, Philadelphia, USA, (1985) 535-543, *Investigation of quartz crystal thickness shear and twist modes using a new noninterferometric laser speckle measurement method*

3 H. F. Tiersten, Proc. 31st Annual Frequency Control Symp. (1977) 44, *An analysis of overtone modes in contoured crystal resonators*

4 B. Dulmet, R. Bourquin, N. Shibanova, 7th European Frequency and Time Forum, Neuchâtel, (1993) 607-612, *Force-frequency transducer using a multimode doubly rotated quartz resonator*

5 R. M. White and F. W. Voltmer, Appl. Phys. Lett., 7 (1965) 314-316, *Direct coupling to surface acoustic waves*

6 G. Riha, K. Anemogiannis, R. Dill, C. Kappacher, Proc. Int. Symp. on SAW Devices for Mobile Communications, Japanese Phys. Soc., Sendai (1992) 113-120, *SAW low loss filters for mobile communication systems*

7 F. Seifert, W. Bulst and C. Ruppel, to be published in Sensors & Actuators, *Mechanical sensors based on surface acoustic waves*

8 H. Nowotny, E. Benes, J. Acoust. Soc. Am. 82 (2) (1987) 513-521, *General one-dimensional treatment of the layered piezoelectric resonator with two electrodes*

9 H. Nowotny, E. Benes, M. Schmid, J. Acoust. Soc. Am. 90 (3) (1991) 1238-1245, *Layered piezoelectric resonators with an arbitary number of electrodes (one-dimensional treatment)*

10 C. Lu and O. Lewis, J. Appl. Phys. 43 (1972) 4385, *Investigation of film-thickness determination by oscillating quartz resonators with large mass load*

11 K. K. Kanazawa and J. G. Gordon, Anal. Chem. Acta, 175 (1985) 1295-1300, *The oscillation frequency of a quartz resonator in contact with a liquid*

12 C. E. Reed, K. K. Kanazawa, and J. H. Kaufmann, J. Appl. Phys. 68 (1990) 1993-2001, *Physical description of a viscoelastically loaded AT-cut quartz resonator*

13 S. J. Martin, V. E. Granstaff, and G. C. Frye, Anal. Chem., 63 (1991) 2272-2281, *Characterization of a Quartz Crystal Microbalance with Simultaneous Mass and Liquid Loading*

14 D. Hauden et al., Proc. 34th Annual Frequ. Contr. Symp. (1980) 312-319, *Sensitivities of SAW-Oscillators to Temperature, Forces and Pressure: Application to Sensors*

15 S. W. Wenzel and R. M. White, IEEE Trans. Electron. Devices, 35 (1988) 735-743, *A multisensor employing an ultrasonic Lamb-wave oscillator*

16 H. Drobe, A. Leidl, M. Rost and I. Ruge, Sensors and Actuators A, 37-38 (1993) 141-148, *Acoustic sensors based on surface-localized HPSWs for measurements in liquids*

17 M. Schmid, E. Benes, R. Sedlaczek, Meas. Sci. Technol. 1 (1990) 970-975, *A computer-controlled system for the measurement of complete admittance spectra of piezoelectric resonators*

18 E. Benes, Elektronikschau 10 (1979) 31-32,*Verwendung eines AC-Schwingquarzes als Temperaturfühler*

19 H. Ziegler, H. J. Aulfes, H. Quint in Sensors: A comprehensive survey, Ed. W. Göpel,VCH NY (1990) Vol.4, Chapt.7: *Time/Frequency Thermometers*

20 R. Bechmann, A. D. Ballato, T. J. Lukaszek, Proc. of the IRE 50 (1962) 1812-1822, *Higher-order Temperature Coefficients of the Elastic Stiffness and Compliances of Alpha-Quartz*

21 W. H. Wade and L. J. Slutsky, Rev. Sci. Instr., vol. 33, 2 (1962) 212-213, *Quartz crystal thermometer*

22 W. L. Smith and W. J. Spencer, Rev. Sci. Instr., vol. 34, 3 (1963) 268-270, *Quartz crystal thermometer for measuring temperature deviations in the 10-3 to 10-6 °C range*

23 D. L. Hammond and A. Benjaminson, Instr. Control Systems, 38 (1962) 115, *Linear Quartz Thermometer*

24 H. Ziegler, Sens. Actuators 5 (1984) 169-178, *A Low-cost Digital Sensor System*

25 T. Ueda et al., Proc. 40th Annual Frequency Control Symp. (1986) 224-229, *Temperature sensor using quartz tuning fork resonator*

26 E. P. EerNisse and R. B. Wiggins, Proc. 40th Annual Frequency Control Symp. (1986) 216-223, *A resonator temperature transducer with no acivity dips*

27 D. Hauden, G. Jaillet and R. Coquerel, Proc. IEEE Ultrasonic Symp., Chicago, (1981) 148-151, *Temperature sensor using SAW delay line*

28 J. Neumeister, R. Thum and E. Lüder, Sens. Actuators, A21-A23 (1990) 670-672

29 M. Viens and J. N. Cheeke, Sens. Actuators A, 24 (1990) 209-211, *Highly Sensitive Temperature Sensor Using SAW Resonator Oscillator*

30 A. D. Ballato and R. Bechmann, Proc. IRE, 48 (1960) 261-262, *Effect of initial stress in vibrating quartz plates*

31 J. M. Ratajski, IBM J., 12 (21968) 92-99, *Force-frequency coefficient of singly rotated vibrating quartz plates*

32 C. R. Dauwalter, Proc. 26th Ann. Freq. Contr. Symp. (1972) 108-112, *The temperature dependence of the force sensitivity of AT-cut quartz crystals*

33 A. Ballato, E. P. EerNisse, and T. Lukaszek, Proc. 31st Annual Frequency Control Symp. (1977) 8-16, *The force-frequency effect in doubly rotated quartz resonators*

34 D. Janiaud, L. Nissim, and J. J. Gagnepain, Proc. 32nd Annual Frequency Control Symp. (1978) 169-179, *Analytical calculation of initial stress effects on anisotropic crystals: Application to quartz resonators*

35 E. Karrer and R. Ward, ISA Trans., 16 (1977) 90-98, *A low range quartz pressure transducer*

36 E. Karrer and J. Leach, IEEE Trans. Ind. Electron. Contr. Instr., IECI-16 (1969) 44-50, *A quartz resonator pressure transducer*

37 E. Benes and D. Hammer, U.S. Patent #4,166,967 (1979), *Piezoelectric resonator with acoustic reflectors*

38 E. Benes, R. Chabicovsky and K. C. Harms, Proc. 8th ICVM Int. Conf. Vacuum Metallurgy, Linz, Austria (1985) 83-101, *Acoustic layers*

39 L. Weisbord, U.S. Patent #3,470,400 (1969), *Single beam force transducer with integral mounting isolation*

40 R. G. Kirman et al., U.S. Patent #4,594,898 (1986), *Force sensors*

41 E. P. EerNisse and J. M. Paros, Proc. 37th Annual Frequency Control Symp. (1983) 255-260, *Practical considerations for miniature quartz resonator force transducers*

42 D. R. Koehler, U.S. Patent #4,479,385 (1984), *Double resonator cantilever accelerometer*

43 M. Ono et al., J. Vac. Sci. Technol., A4 (1986) 1728-1731, *Quartz friction vacuum gauge for pressure range from 0.001 to 1000 torr*

44 P. Das, C. Lanzl, and H. F. Tiersten, Proc. 1976 IEEE Ultrasonics Symp., NY (1976) 306-308

45 Y. N. Vlassov, A. S. Kozlov, N. S. Pashchin, I. B. Yakovkin, Proc. 1993 IEEE Int. Frequency Control Symp. (1993) 665-669

46 W. Baldauf, Dissertation, Technical University Munich (1993), *Drehmomentenmeßsystem mit Oberflächenwellen-Resonatoren*

47 F. Seifert, W. Bulst, and C. Ruppel, to be published in Sens. Actuators, *Mechanical sensors based on surface acoustic waves*

48 G.Z. Sauerbrey, Z. Phys. Verhandl. 155 (1959) 206, *Verwendung von Schwingquarzen dur Wägung dünner Schichten und zur Mikrowägung*

49 K.H. Behrndt, J. Vac. Sci. Technol. 8 (1971) 622, *Long term operation of crystal oscillators in thin-film deposition*

50 J.G. Miller and D.I. Bolef, J. Appl. Phys. 39 (1968) 5815, *Sensitivity enhancement by the use of acoustic resonators in CW ultrasonic spectroscopy*

51 C. Lu, Vac. Technol. 3 (1974) 45, *Improving the accuracy of quartz crystal monitors*

52 E. Benes, J. Appl. Phys. 56, 608 (1984) 608-626, *Improved quartz crystal microbalance technique*

53 E. Benes, K.C. Harms, and G. Thorn, 39th Annual Frequency Control Symp., Philadelphia, USA, (1985) 556-570, *Enhanced composite resonator analysis and its application to the quartz crystal microbalance*

54 E. Benes, M. Schmid, G. Thorn, Thin Solid Films 174 (1989) 307-314, *Progress in Monitoring Thin Film Thickness by Use of Quartz Crystals*

55 A. Wajid, Rev. Sci. Instrum. 62, 8 (1991) 2026-2033, *Improving the accuracy of a quartz crystal microbalance with automatic determination of acoustic impedance ratio*

56 K.C. Harms, E. Benes, Patent A #382963 (1987), Germany, France, GB, USA, *Verfahren zur Bestimmung der Dicke dünner Schichten*

57 D.R. Denison, J. Vac. Sci. Technol. 10 (1973) 126, *Linearity of heavily loaded quartz crystal microbalance*

58 E. Söllner, E. Benes, A. Biedermann, Vakuum-Technik 7 (1977) 197-201, *Einfluß der Schicht-Relaxation auf die Genauigkeit von Schwingquarz-Schichtdickenmeßgeräten*

59 E.P. Eer Nisse, 29th Annual Frequency Control Symposium, Philadelphia, USA (1975) 1

60 E.Benes, Proc. Optik-Symposium Moskau, USSR, May 25-27 (1987) *Auto-Z-Match® - Eine neue Generation von Schwingquarz-Schichtdickenmeßgeräten*

61 E. Benes, M. Schmid, H. Siegmund, Patent DE #3732594, USA, Japan, Europe, *Einrichtung zum Ermitteln der jeweiligen Dicke von sich verändernden Materialschichten auf einem Substrat (ppc-Quarz)*

62 M. Schmid, Dissertation, University of Technology Vienna, Institut für Allgemeine Physik (1989), *Möglichkeiten und Grenzen der Schwingquarz-Schichtdickenmessung*

63 Y. K. Yong et al., Proc. of the 45th AFCS Annual Frequency Control Symposium, Los Angeles, (1991) 137-147, *Thickness-Shear Mode Shapes and Mass-Frequency Influence Surface of a Circular and Electroded AT-Cut Quartz Resonator*

64 F. Eggers and T. Funck, J. Phys. E: Sci. Instum. 20 (1987) 524-530, *Method for measurement of shear-wave impedance on the MHz region for liquid samples of ≈1ml*

65 S. J. Martin et al., Anal. Chem. 65 (1993) 2910-2922, *Effect of Surface Roughness on the Response of Thickness-Shear Mode Resonators in Liquids*

66 S. J. Martin et al., Proc. of the 1993 IEEE International Frequency Control Symposium, Salt Lake City, USA, (1993) 603-608, *Measuring Liquid Properties with Smooth- and Textured-Surface Resonators*

67 W. H. King. Anal. Chem., 36 (1964) 1735-1739, *Piezoelectric sorption detector*

68 B. Henning et al., Sens. Actuators A, 25-27 (1991) 705-710, *Quartz Microbalance Sensors for Gas Detection*

69 H. Muramatsu et al., Analytica Chimica Acta, 215 (1988) 91-98, *Viscosity monitoring with a piezoelectric quartz crystal and its application to determination of endotoxin by gelation of limulus amebocyte lysate*

70 H. Muramatsu et al., Sens. Actuators, A21-A23 (1990) 362-368, *Piezoelectric Resonator as a Chemical and Biochemical Sensing Device*

71 M. S. Nieuwenhuizen and A. Venema, Sensors Mater., 1 (1989) 261-300

72 W. Wei, Z. Mo, and S. Yao, Analytica Chimica Acta, 251 (1991) 143-148, *Multi-component analysis in solution using piezoelectric quartz sensors*

73 A. J. Ricco and S. J. Martin, Sens. Actuators B, 10 (1993) 123-131, *Multiple-frequency SAW devices for chemical sensing and materials characterization*

74 W. Buff, Sensors and Actuators A, 30 (1992) 117-121, *SAW sensors*

75 M. Thompson et al., IEEE Trans. Ultrasonics, Ferroelectrics, Frequ. Contr., UFFC-34, 2 (1987) 127-135, *The Potential of the Bulk Acoustic Wave Device as a Liquid-Phase Immunosensor*

76 G. G. Guilbault, Anal. Chem., 55 (1983) 1682-1684, *Determination of formaldehyde with an enzyme-coated piezoelectric crystal detector*

77 L. Tessier et al., Proc. Ultrasonics International 93, Vienna (1993) 627-630, *AT-cut Quartz Crystal Biosensor for Blood Assay*

78 F. Aberl et al., to be published in Sens. Actuators, *HIV Serology using Piezoelectric Immunosensors*

79 J. C. Andle et al., Sens. Actuators B, 8 (1992) 191-198, *An acoustic plate mode biosensor*

80 P. Krempl, Fortschritte der Akustik DAGA'90, Proc. 16. Gemeinschaftstagung der Deutschen Arbeitsgemeinschaft für Akustik, Vienna (1990) 153-162, *Neue piezoelektrische Materialien*

81 P. Krempl, J. Stadler, W. Wallnöfer, E. Ellmeyer, R. Selic, Proc. 5th EFTF European Frequ. Time Forum (1991) 143-147

82 J. Detaint et al., Proc. 6th EFTF European Frequ. Time Forum (1992) 223-227, *Comparison of energy trapping resonators and filters using the AT-cut of quartz, berlinite and gallium-phosphate*

83 J. Schwartzel et al., Proc. 8th EFTF European Frequ. Time Forum (1994) (These Proc.), *Comparison of the Properties of the New Thermally Compensated Materials with Trigonal Symmetry*

SAW Devices for Frequency Control, Signal Processing and Sensor Application

WERNER BUFF
Department of Electrical Engineering and Information Technique,
Technical University of Ilmenau
D-98684 Ilmenau, P.O. Box 327, Germany

ABSTRACT

Surface acoustic waves have been employed in a large number of different devices for frequency control, special signal processing functions and for sensor application. After a short look at different types of surface acoustic waves the operational function of some devices are highlighted.

1 Introduction

Surface acoustic wave (SAW) devices have two important origins. The first one is the theory of waves in and along the surface of solid states beginning in 1885 with a fundamental paper of LORD RAYLEIGH /1/. For better understanding the nature of SAW has been under investigation up to now. The other origin is in the continuing advancement of the technological level of microelectronic integrated circuits. Starting with the invention of the interdigital transducer (IDT) for excitation of SAW in a piezoelectric solid in the middle of the sixties by WHITE and VOLTMER /2/ the SAW technique has followed the continuous miniaturization of dimensions in microelectronics. In terms of SAW devices this means a permanent progress towards higher frequencies.
The today's state of SAW devices can be characterized as a production of about 200 million pieces a year in different countries worldwide. There are no TV sets produced today without at least one SAW filter in the IF stage of the receiver. Moreover, there is a lot of other applications and the number of devices produced is increasing rapidly . The main reasons for that are their small sizes, low costs and high performance due to utilizing technological improvements of the well-developed microelectronic technique. Furthermore, they are the high reliability and reproducibility due to their solid-state properties and finally, the relatively good understanding of the basic physical processes because of the great efforts in R&D for about the last 30 years.

2 Surface acoustic waves

The common way for excitation of SAW is to use an interdigital transducer (IDT) overlaying on a piezoelectric substrate (fig. 1). An IDT is a metallic layered structure made by a photolithographic process and is a key element for all SAW devices. By using the inverse piezoelectric effect such a structure excites SAW with a velocity of about 10^5 times smaller than the velocity of electromagnetic waves in vacuum. For this reason it is possible to make SAW devices for a frequency range from 10 MHz up to some GHz.

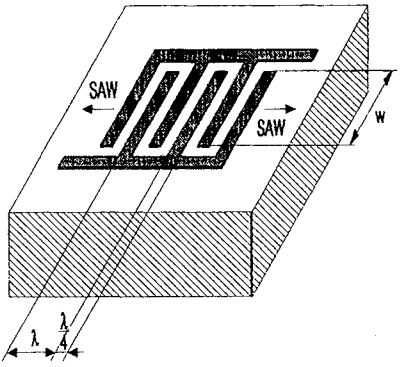


Fig. 1: Interdigital transducer (schematic)

Depending on the crystal cut and the direction of the traveling wave different coupling factors for different wave types do exist. Thus SAW can be distinguished depending on particle displacement (linear, elliptical or circular polarization), on the attenuation function from the surface into the volume of the substrate, on the coupling to an electric field or not and the material system where waves are traveling (monocrystaline, crystal with or without piezoelectric behavior, layered structure, membrane). If we introduce an orthogonal axial system x_1, x_2, x_3 and a particle displacement u_1, u_2 and u_3 in the axial directions we are able to describe the different types of waves. In addition we have to consider that in a piezoelectric crystal the wave is accompanied by a local difference of the electric potential, i.e., by an electric field.

The most common type of SAW is the so-called general SAW mode in an anisotropic, piezoelectric substrate. In this case all particle displacement directions and the electric field are coupled.

A general mechanical SAW is a mode in an anisotropic nonpiezoelectric material with displacement coupling in all three axial directions without coupling to any electric field. Those SAW modes, where at least one axial direction of particle displacement is uncoupled, are of practical importance. This means, the movement occurs in only one or in a plane built up of two axial directions. If the wave vector has the same direction as the energy flow then we have a so-called pure mode and that is what is used in realized SAW devices. /3, 4, 5/

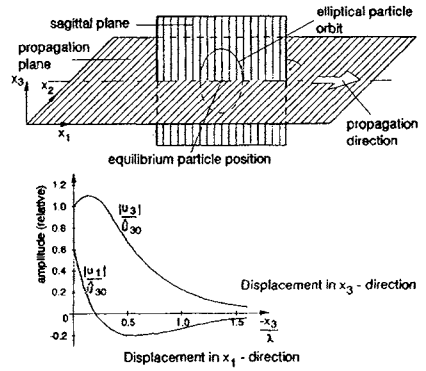


Fig. 2: Particle displacement of a Raleigh-type wave

A short look at the wave types mainly used gives an impression of the complexity of these waves. But it should be stated that the ratio between the wavelength and the amplitude is in the range of 10^5:1 (i.e. $LiNbO_3$, 100MHz). For frequency controlling devices there are almost always used Rayleigh-type waves with a particle displacement on elliptical orbits and a polarization in the x_1-x_3-plane (fig. 2). The displacement amplitude decreases exponentially from the substrate surface into the volume. In a strict sense a Rayleigh-mode exists only in an elastic isotropic nonpiezoelectric solid-state. But

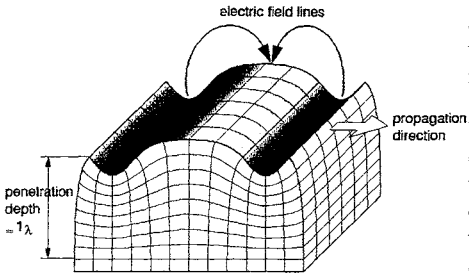


Fig. 3: Rayleigh-wave propagation (schematic)

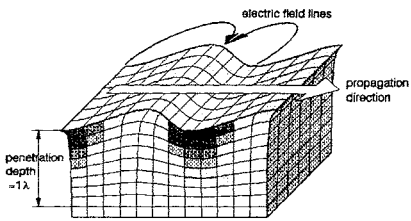


Fig. 4: Shear wave propagation

because of the exponential attenuation curve and polarization in the x_1-x_3-plane the most used wave mode in SAW devices on monocrystaline piezoelectric Lithiumniobate ($LiNbO_3$) or quartz is also called a Rayleigh-wave or piezoelectric Rayleigh-wave /6/. Waves with oscillating attenuation behavior into the depth of the substrate are known as generalized Rayleigh-waves and in piezoelectric crystals they are "stiffened" by the accompanying electric field (fig. 3).

Although the Rayleigh-type wave is the most used type the calculations show that there is a lot of other possibilities /6/ for wave modes. This concerns in the first order waves with a particle displacement only in one axial direction perpendicular to the propagation direction in the propagation plane (fig. 4), the so-called shear or horizontal polarized shear waves. They are practically used as shallow bulk acoustic waves (SBAW) or skimming bulk waves (SSBW) /7/, which are not pure surface acoustic waves but they propagate in a near surface layer with a very small deflection into the volume. The main advantage of such waves compared to Rayleigh-waves is the 1.6 times higher propagation velocity which allows to extend the application range for devices up to 2 GHz for resonator oscillators /8/ or 5 GHz for delay line oscillators /9/. In a special case, if the sagittal plane is normal to a twofold axis of a piezoelectric crystal or to an axis of higher even order, then a horizontal polarized shear mode with an associated electric field, the so-called Bleustein-Guljajev-wave, can exist there /10, 11/. The penetration depth is much higher than that of a Rayleigh-wave and the propagation velocity is less.

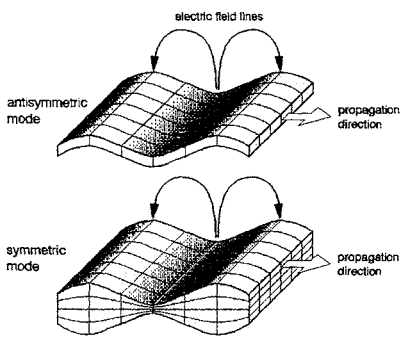


Fig. 5: Lamb-wave propagation

In structures with an overlaying piezoelectric or nonpiezoelectric layer the wave types are modified by the layer depending on the thickness-wave length ratio of the layer. The

main difference in layered structures is that the waves become dispersive, i.e., the phase velocity becomes dependend on the thickness of the layer in a certain frequency range. In layered systems, in addition to Rayleigh-type waves, a modified form of shear plane waves can propagate which are known as Love-waves /12/.
Especially for sensor applications they have become more and more important waves in membranes with a thickness which is higher or lower than the wave length (fig. 5) /13, 14/. There it is possible to excite plate waves in form of symmetric or antisymmetric modes and, furthermore, shear modes /15/. The latter type is reported to be suited to create high performance sensors for liquids /16/.
An overview of the different kinds of waves is given in table 1.

Table 1: Different types of surface acoustic waves

Wave Type	Coupled Particle Displacement	Coupled Potential	Material Features	Main Application
general SAW		ϕ	anisotropic piezoelectric	
general mech. SAW		-	anisotropic non-piezoelectric	
Rayleigh-wave		-	nonpiezoelectric with high symmetry, (piezoelectric)	
stiffened Rayleigh-wave		ϕ	piezoelectric	semiconductor coupled devices, air gap coupled devices, e.g. convolver, correlator
Rayleigh-type-wave		ϕ and -	piezoelectric and nonpiezoelectric layer	filter, resonators
shear wave (SH)		-	nonpiezoelectric and piezoelectric	highest frequency devices e.g. filter, resonator
stiffened SH (BG-wave)		ϕ	piezoelectric	sensors in liquids
Love-wave		ϕ and -	piezoelectric and nonpiezoelectric layer	sensors in liquids

general plate wave general mech. plate wave		ϕ -	anisotropic piezoelectric anisotropic non-piezoelectric	
Lamb-waves stiffened Lamb-waves		- ϕ	nonpiezoelectric with high symmetry (piezoelectric) piezoelectric	sensors sensors
SH plate waves stiffened SH waves		- ϕ	nonpiezoelectric and piezoelectric piezoelectric	sensors in liquids
SSBW (volume waves like)		ϕ	piezoelectric	resonators

3 Frequency control devices

For practical purposes the only way for exciting SAW and receiving them is to use the proper material and interdigital transducers. The most frequently used materials are the crystals of Lithiumniobate ($LiNbO_3$) and Quartz (SiO_2) in monocrystaline form. Another important material is Lithiumtantalate ($LiTaO_3$). Although one has investigated a large number of piezoelectric monocrystaline materials in the laboratories producers of SAW devices do not have a large choice.
However, there are many different ways and methods for the design of SAW devices using the great variety in constructions of interdigital transducers.

3.1 Delay line filters

The modeling and synthesis of SAW filters is quite different from those of LC filters. The physical reason is that there is no energy storage on reactive components as in LC filters but there an energy transfer occurs from the input interdigital transducer to the output interdigital transducer after the SAW traveled along the solid-state surface /17/.

This demands completely new approaches for SAW-filter design. The basic idea is to form a certain impulse response as weighted finger pairs as this is used in the design of transversal filters. This includes the design procedure of finite impulse response (FIR) modeling. So the filter design has concentrated on interdigital transducer design. Filter producers are nowadays able to construct filters with complicated bandpass characteristics using CAD systems /18/ after a very short design time. Today's design tools for filters include second-order effects like influences of finger reflections, conversion of SAW into bulk waves, diffraction, electromagnetic feedthrough, triple-transit-signals and more.

The limits of today's filters are approximately in a range from 10 MHz to 2 GHz, with fractional bandwidth up to 20%, sidelobe rejection better than 70 dB, passband flatness to within +0,2 dB, group delay ripples less than ±5 ns in the passband /17/ and insertion losses less than 1 dB /19/.

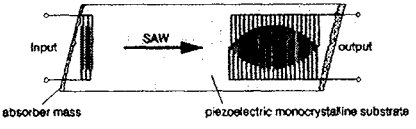


Fig. 6: SAW delay line filter

A typical structure for a delay line filter is shown in fig. 6. An uniform interdigital transducer is launching a SAW which is detected by a second transducer. The last one has a weighting of the finger pairs which is usually realized by overlapping (so-called apodization). The propagating time from one transducer to the other is the delay time of such a device.

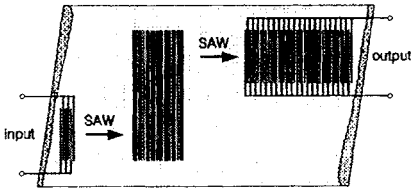


Fig. 7: Microstrip coupled filter

A further kind of weighting could be made by using different distances (withdrawal electrodes) between the fingers. The fingers themselves can be produced as full fingers with 2 fingers related to the wavelength (width of a finger: $\lambda/4$, distance between the centres of the fingers: $\lambda/2$) or as split fingers with 4 fingers per wavelength. This design concept allows a great variety of filter properties and has been successfully applied for a big number of filter devices. Microstrip coupler filters (fig. 7) avoid the drawback related to spurius signals due to bulk waves by transferring SAW with help of a metallic grating into another track directed to the receiving interdigital transducer. However, the chip area is much bigger than the area necessary for inline design.

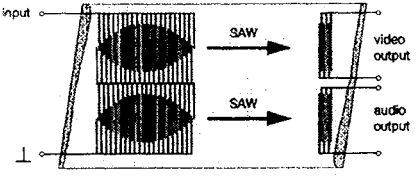


Fig. 8: Split sound filter

The SAW concept also allows the combination of different filter characteristics in one device. Examples are quasi-parallel or split sound filters (fig. 8) and switchable SAW filters (fig. 9). In the last case the actual filter characteristic depends on whether the input signal is connected

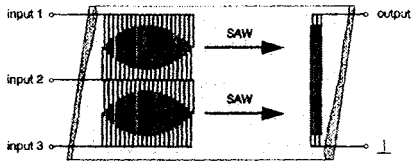


Fig. 9: Switchable SAW filter

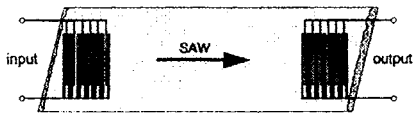


Fig. 10: Single phase unidirectional transducers

presently to the input electrodes 1,2 or 3. So it is possible to build devices with filter functions with either one of the filter characteristics or combinations of them as difference or sum. This idea allows a lot of possibilities including the combination of more than two SAW tracks on one device /19/.

In the simple case interdigital transducers used in delay line filters launch SAW in both directions in which one of them the energy transport can not be employed and is to be attenuated. Assuming identical behavior for the launching and receiving interdigital transducer we have an insertion loss for such devices of at least 6 dB. This is to much for several applications and therefore one has tried to find solutions for lower insertion losses. The concept of a single phase unidirectional transducer can meet the low-loss requirements especially for mobile telephone systems. This type of interdigital transducer (fig. 10) can be regarded as a split finger type transducer with a reflector grating between the fingers /20, 21, 22/. The centers of the SAW-exciting interdigital transducer and the reflecting grating are locally shifted in such a way that the waves in the desired direction are not reflected but the waves directed to the other side are totally reflected. This concentrates the energy flow almost exclusively in one direction. However, the design procedure is more complicated than in the case of using bi-directional transducers. Because of the great possibilities for variation of interdigital transducer layouts there are lots of configurations employed to implement low-loss filters.

3.2 SAW resonators

Bulk wave resonators such as quartz devices for frequency control and narrow band filter application have been well established for a long time ago. The quality factors reach values of 10^5 but they are difficult to produce in a frequency range above 50 MHz while working in the fundamental mode. When working on higher harmonics the electrical behavior of bulk wave resonators is disrupted. Nowadays, when the frequency range for signal processing is continuously shifting towards higher frequencies, the way

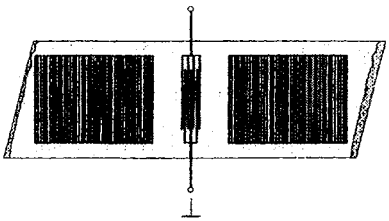


Fig. 11: 1-port resonator

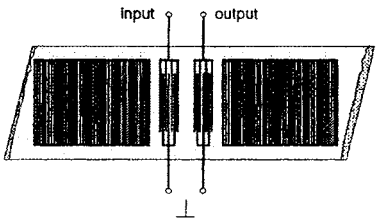


Fig. 12: 2-port resonator

out is to employ SAW resonators. They are available in a frequency range from 200 MHz to more than 1 GHz with typical insertion losses of about 7dB at 200 MHz and 15dB at 1 GHz with quality factors of 11000 (200 MHz) to 5000 (1 GHz) /19/ while operating on the fundamental mode.

SAW resonators are configurations where the excited SAW is reflected to the place of its origin and there it is exciting or amplifying a wave once more. The case of resonance means that only standing waves are existing and the wavelength (frequency) of them is strongly depending on the geometrical layout of the exciting interdigital transducer and the reflecting gratings. Resonators are commonly used as 1-port resonators (fig. 11) or as 2-port resonators (fig. 12). The reflecting gratings can be made in form of metallic stripes or for higher quality in form of etched grooves. The 1-port resonators are simple in layout and need a very small chip area. However, 2-port resonators are more complicated in layout, need more chip area but there is a higher quality factor available and they can better be matched in the electronic circuitry.

3.3 Resonator filters

In the middle of the 1970's TIERSTEN and SMYTHE first proposed a waveguide-coupled resonator filter as a counterpart to monolithic quartz filters /25/. The physical background is that metallic resonator gratings and the interdigital transducer can be considered as a waveguide where the SAW phase velocity is lower than in the completely metalized case because of an energy storing effect. The guided wave can be coupled into a second waveguide configuration similar to the coupling from one bulk wave resonator to an another in monolithic quartz filters. The design procedure for filters includes the superposition of the symmetric and antisymmetric coupling modes /26/ and also modes of higher order /27/.

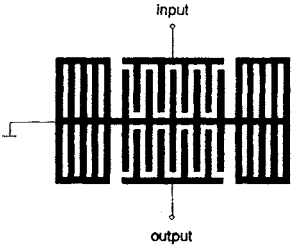


Fig.13: Transversely coupled resonator filter

In fig. 13 a possible layout of a waveguide resonator filter is shown. This concept is able to create narrowband filters and an available maximum bandwidth of 700 ppm together with an insertion loss of 3.4 dB are reported in /19/. Table 2 contains some of the most used filter types.

Table 2: Important SAW filters /19/ and /20/

Type	Layout	Principle	Properties	Application
delay line filter	1 track, 2 IDTs	1 IDT with overlap or withdrawal apodization	IL: 12-17dB	TV
microstrip coupled filter	2 tracks, 2IDTs, 1 MSC	1 IDT on each track, one or both apodized, change of tracks via MSC	IL: 25-35dB SR: 55dB	TV
quasi-parallel or split-sound filter	2 tracks, 2 coupled input IDTs and 2 separate output IDTs	splitting of one input signal to both a video channel (VC) and an audio channel (AC)	VC: IL: 12-17dB AC: IL: 16-20dB	TV, separation of video and audio channel
switchable SAW-filter	2 or more tracks, each with 1 input IDT, 1 output IDT	usage of either one or a combination of the transfer function of the tracks	IL: 12-17dB	TV, multi-standard receivers
SPUDT-filter	1 track, 2 SPUDTs	usage of internal reflections in the SPUDT for concentration of energy in one direction	IL: 3-6dB	mobile telephone, pagers
resonator filter	1 track, 2IDTs, reflectors on each end and between the IDTs	inner reflectors lead to the pass-band shape, outer reflectors and the IDTs influence the stopband rejection	IL: 1-3dB (tuned)	mobile telephone, pagers
transverse-mode coupled filter	2 or more tracks, each track with a resonator structure	the fundamental symmetric and first antisymmetric mode are of main importance	IL: 1-2dB	mobile telephone, pagers

IL: Insertion loss, SR: Sidelobe rejection, SPUDT: Single phase unidirectional transducer

4 Signal processing devices with special functions

The SAW technique offers the unique opportunity to realize very complicated signal processing functions in a simple way. Besides the mass production of filters and resonators this group of devices is becoming more and more important. Because the signal converted into a SAW is spread out at the surface of a solid-state during the propagating time it can be tapped at any place on the surface, for example by means of metallic stripes for sampling the electric field associated with SAW in a piezoelectric material. Furthermore, it is possible to use nonlinearities in elastic deformation or in the interaction of SAW with charge carriers in a semiconductor material. Because of the variety of such devices only a few of them will be considered which seem to have best chances for applications.

4.1 Chirp filter for pulse compression

Linear frequency modulated pulses are mainly used in radar configurations. In transmitter stages of radar equipments using correlation technique pulses are expanded into a frequency modulated wave train which will be emitted, reflected from an object and then compressed in the receiver circuitry. There are several possibilities to design chirp filters as SAW devices. The simplest case is to use interdigital transducers with nonequidistant fingers as dispersive configurations (fig. 14) /28/. The radar application was one of the basic applications of SAW devices and has contributed many ideas which make the SAW technique attractive.

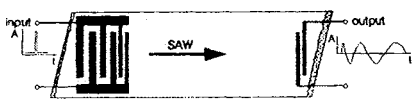


Fig. 14: Chirp filter configuration

Today the chirp concept is included in satellite communication systems and helps to meet very complicated specifications combined with digital signal processing /29/.

4.2 Tapped delay lines

If it is supposed that in usual receiving interdigital transducers each finger pair samples a position of the wave, then a local shift of finger pairs results in a phase shift of the received wave. This behavior is commonly used in matched filters where a phase shift of 180 degrees is employed. This results in biphase coded transducer structures, thus realizing a binary phase shift keyed (binary PSK) modulation (fig. 15) /30/. The coded signal is then translated via a spread spectrum system and after that detected by using the correlation in the receiving circuitry and so reconstructing the original input pulse. This leads to a certain correlation gain in the receiver which mainly

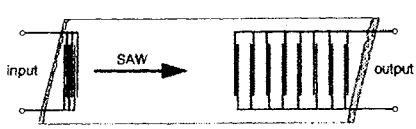


Fig. 15: PSK-filter configuration

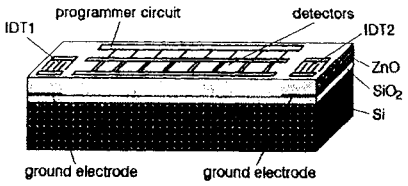


Fig. 16: Programmable matched filter

depends on the chosen code and the length in bit units of the correlating signals. The implementation of other SAW functions like weighting of finger pairs in form of overlapping or withdrawal electrodes results in further improvements of this concept. However, the simple SAW technology with one photolithographic step on piezoelectric monocrystaline material allows only a fixed code for each device and if it is necessary to change the code the only way out is to change the coding and decoding devices. A better way is to use semiconductor devices which contain piezoelectric layers for SAW excitation and in which SAW can be detected by using the piezoresistive effect in field effect transistor structures (fig. 16) /30-34/. Because these are semiconductor devices (e.g. in silicon) a program function varying with time can be implemented via integration of an electronic circuitry on the same chip. As one of the latest results a fully integrated SAW programmable transversal filter utilizing active piezoelectric junction FET detectors with 32 analog controlable active taps, a filter center frequency of 60 MHz and time-bandwidth product amounts to 16 has been reported on by van RHIJN et.al. /36/. On the same chip a low-noise readout amplifier and a bias control circuitry is also integrated.

4.3 Identification tags (ID tags)

The SAW concept allows to construct in a very simple way passive (e.g. without energy supply) identification tags. It has been predicted that such devices for identification systems promise an enormous potential for the future /37/. The function is very simple and this makes them small and inexpensive. The operation principle is shown in fig.17. From a basic station a pulse is transmitted to the input antenna of the ID tag which is connected to the input interdigital transducer. The input pulse will be converted in a wave train corresponding to the number of finger pairs. This SAW pocket can be coded in the coding area either by using a reflecting array or via a tapped delay line structure as it was considered above /38/. In the first case the output signal will be transmitted back to the basic station in form of a pulse sequence by the same antenna as it was used for receiving the input pulse with a shift in time. In the latter case the output signal can be a PSK signal transmitted from a second antenna. REINDL and RUILE have reported ID tags operating at 285 MHz and 2.45 GHz. The 8-bit 285 MHz tag has an insertion loss of 37 dB and a dynamic separation of about 30 dB while the 33-bit 2.45 MHz tag has an insertion loss of 55 dB and a dynamic separation of 55 dB /39/.

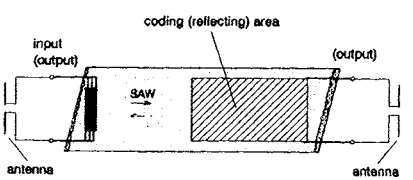


Fig. 17: Passive identification tag

4.4 Convolvers

Although there is only a weak nonlinearity in the elastic behavior of the mostly employed piezoelectric materials it is possible to use this for mixing two signals which are converted into SAW. Together with a signal inversion in time, a delay during the propagation time along the surface and an integration of the mixed signals, which is easy to perform in SAW technique, it is possible to realize the convolution of two signals /40/. But moreover such devices are able to carry out numerous other signal processing functions, for example the correlation of signals, signal compression and inversion, frequency selection by variable filter functions, as well as the coding and decoding of signals. The so-called elastic convolver (fig. 18) on $LiNbO_3$ or on quartz is under investigation since the begin of the 1970's and led, for example, to devices of small sizes, a center frequency of 350 MHz, an integration time of 3µs and a bandwidth of 50 MHz that results in a time-bandwidth product of 150 /41/.

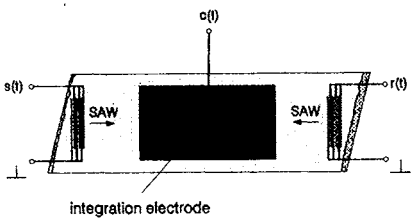


Fig. 18: Elastic SAW convolver

In contrast with the elastic convolver one can transfer the convolver function in MOS devices which utilize the nonlinear interaction of the electric field associated with acoustic waves in a piezoelectric medium. For example, ZnO-SiO_2-Si convolvers have the maximum output in the bias condition where the nonlinearity of the depletion layer capacitance occurs /42/. In addition, the drain current-gate bias relation of the MOSFET has a quadratic term in the saturation current region. So the MZOS - (metal-ZnO-oxide-semiconductor)-FET convolver combines the advantages of the SAW approach with strong quadratic terms of a MOSFET drain current characteristic. Recently some new constructions were proposed for MZOSFET convolvers utilizing a new mixing mechanism similar to that in a dual gate MOSFET /43/. The reported devices (fig. 19) avoid some drawbacks of former constructions and show a

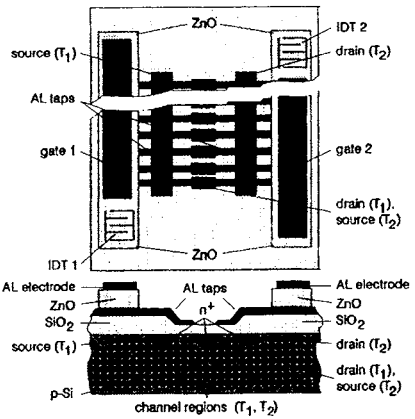


Fig. 19: Multiplicative FET-convolver

higher efficiency. However, a special problem in semiconductor SAW devices is the electromagnetic feedthrough /44/.

5 SAW sensors

Devices used for signal processing must have a high stability and have to be insensitive to variations of environmental condition. However, this is not possible. It is known that SAW devices react to changes in physical conditions in the propagation path by changes of both their amplitude and phase velocity, so it is possible to make use of such changes in the electrical characteristics for "feeling" changes in physical environmental conditions. In the last decade several kinds of SAW sensors were under investigation /45/ and, as reported in some papers, SAW sensors have the advantages of extremely high sensitivity and a signal in the form of a frequency shift. However, many prototypes reported in the literature were not fabricated. It seems that, on the one hand, there are lots of difficulties to overcome in the manufacture and in practical use of SAW sensors, but that, on the other hand, not all possibilities included in the SAW operation principle have been explored or exhausted.

In order to measure the changes in frequency caused by changes in the boundary conditions for propagating SAW's, a simple sine oscillator is commonly used. There is often a second generator which is identical to the first but protected against the environmental influence that is to be measured (fig.20). Especially the ever present cross-sensitivity against changes in temperature is disturbing often and is to be excluded.

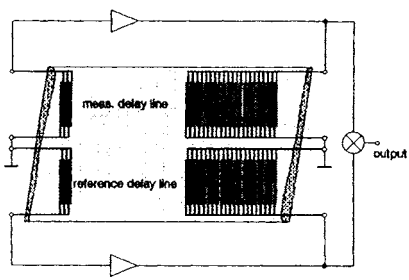


Fig. 20: Differential configuration with two delay lines

This particular disadvantage can be avoided by means of crossed delay lines construction (fig. 21) that is able, under consideration of certain design rules, to provide a constant difference frequency over a wide temperature range while one of the two oscillator frequencies alone is shifting with the temperature corresponding to the temperature coefficient in the applied crystal direction /46/. The concept of crossed delay lines was first considered for temperature sensors but it seems to be able to solve the temperature cross-selectivity problems also for other types of SAW sensors.

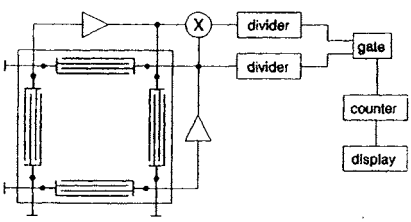


Fig. 21: SAW thermometer with crossed delay lines

Describing the different types of SAW sensors, we can distinguish between three main groups. Depending on where the interaction between the environment and the SAW occurs,

we have a group where the whole substrate, the surface only or an overlaying layer is affected.
If we distinguish between sensors for physical (temperature, pressure, acceleration, force etc.) and chemical (gas concentration, chemical components concentration etc.) magnitudes, then it is worth mentioning that commonly the chemical substance that is to be measured by means of absorption, adsorption or chemical reaction in a coating layer leads to a change in the physical properties and consequently to a change in boundary conditions for SAW propagation. Considering the advantages of SAW sensors compared to other sensor solutions one can expect a continuous development of this area and a broader application. In table 3 some sensor aspects are mentioned.

Table 3: Examples for the use of SAW-sensors

Measured parameter	**Main principle**	**SAW-construction**	**Sensitivity**
temperature	temperature coefficient of phase velocity	resonator	7 kHz/K
force	strain dependence	dual delay line	1800 Hz/N
pressure	strain dependence	resonator / membrane	2,5 kHz/psi
acceleration	flexion	dual delay line /beam	3279 Hz/G
voltage	electrical strain	lamb-wave delay line	200 Hz/V
chemical	surface layer conductivity, stiffness, thickness or mass	delay line	10 kHz/g
flow	temperature dependence/ thermic flow	delay line	142 kHz*min/l

6 Summary

SAW devices open the door for improving conventional and creating new solutions in TV circuitry, telecommunications, measurement techniques, security and identification tasks, sensor techniques, radar equipments etc. Besides the briefly discussed operational functions there were investigated SAW principles additional during the last decades. The research was with such devices like Fourier transformator, frequency discriminator, SAW amplifier, light deflector, phase shifter, different types of memories, memory correlator etc. The well-established SAW technology is able to serve a rapidly growing market. The next step is directed first of all at higher frequencies. Recently YAMANOUCHI et al. reported about 20 GHz SAW devices produced using a nanometer technological procedure /47/. Furthermore, it is predicted a higher complexity inside integrated devices in components of microsystems. In such components can be interconnected to promote interactions parts of microelectronics, micromechanics, microoptics and microacoustics.

Acknowledgments

The author wishes to acknowledge to the members of the acoustoelectronic group at the Technical University of Ilmenau Dr. Frank Möller, Olaf Schmeckebier, Kerstin Plath and Frank Plath for helpful discussions and assistance in completing this paper. For editorial help he is very grateful to his son-in-law, Joel Spencer-Buff.

References

/1/ Lord Rayleigh, *On Waves propagated along the plane surface of an elastic solid*, Proc. London Math. Soc., **Vol. 7**(Nov. 1885), pp. 4-11

/2/ R.M. White and F.M. Vollmer, Direct coupling to surface elastic waves, Appl. Phys. Lett. **17**(1965), pp. 314-316

/3/ C.K. Campbell and J.R. Jones, *A method estimating optimal crystal cuts and propagation directions for exciting of piezoelectric surface waves*, IEEE Trans. Sonics and Ultrasonics, **SU-15**(1968)No. 4, pp. 209-217

/4/ G.W. Farnell, *Symmetry considerations for elastic layer modes propagating in anisotropic crystals*, IEEE Trans. Sonics and Ultrasonics, **SU-17**(1970)No. 4, pp. 229-238

/5/ R.M. White, *Surface elastic waves*, Proc. IEEE, **Vol. 58**(1970)No. 8, pp. 1238-1276

/6/ D.P. Morgan, *Surface wave devices for signal processing*, Elsevier Amsterdam-Oxford-New York-Tokyo 1991

/7/ T.I. Browning and M.F. Lewis, *New family of bulk acoustic wave devices employing interdigital transducers*, Electronic Letters **Vol. 13**(1997)No. 5, pp. 128-130

/8/ I.D. Avramov, *Gigahertz range resonant devices for oscillator applications using shear horizontal acoustic waves*, IEEE Trans. on Ultrasonics, Ferroelectrics and Frequency Control, **UFFC-40**(1993)No. 5, pp.459-468

/9/ K.F. Lau, K.H. Yen, A.M. Mang, K.V. Rosseau, *SBAW Oscillators at 3 to 5GHz range*, 1980 IEEE Ultrasonics Symposium, pp. 263-266

/10/ J.D. Bleustein, *A new surface wave in piezoelectric materials*, Appl. Phys. Letters, **Vol. 15**(1968), pp. 412-413

/11/ Y.V. Guljajev, *Poverchnostnye elektrozvukovye volny v tverdych telach*, ZETF pisma, **Vol. 9**(1969)No. 37, pp. 63-65

/12/ A.E.H. Love, *Some problems of geodynamics*, Cambridge Univ. Press, 1911, Dover 1967

/13/ R.M. White, P.J. Wicher, S.W. Wenzel and E.T. Zellers, *Plate mode Ultrasonic Oscillator Sensors*, IEEE Trans. on Ultrasonics, Ferroelectrics and Frequency Control **UFFC-34**(1987)No. 2, pp. 162-171

/14/ M.J. Vellekoop, A.J. van Rhijn, G.W. Lubkin and A. Venema, *All-silicon plate wave Oscillator system for sensors*, Sensors and Actuators A, **25/27**(1991), pp. 699-703

/15/ E.L. Adler, *Electromechanical coupling to Lamb and shear horizontal modes in piezoelectric plates*, IEEE Trans. on Ultrasonics, Ferroelectrics and Frequency Control, **Vol. 36**(1989)No. 2. pp. 223-230

/16/ S.J. Martin, A.J. Ricco, T.M. Niemczyk, G.C. Frye), *Characterization of SH acoustic Plate mode liquid sensor*, Sensors and Actuators, **20**(1989), pp. 253-268

/17/ C. Campbell, *Surface Acoustic Wave Devices and Their Signal Processing Applications*, Academic Press, Inc., Hacourt Brace Janovich, Publishers Bosten, San Diego, New York, Berkeley, London, Sydney, Tokyo, Toronto 1989

/18/ S.M. Richie, B.P. Abbott, D.C. Malocha, *Description and Development of SAW Filter CAD System*, IEEE Trans. on Microwave Theory and Technique, **Vol.36** (1988), No. 2. pp. 456-466

/19/ C.C.W. Ruppel, R. Dill, A. Fischeraner, G. Fischeraner, W. Gawlik, J. Machni, F. Müller, L. Reindl, W. Ruile, G. Scholl, I. Schropp, K.Ch. Wagner, *SAW Devices for Consume Communication Application*, IEEE Trans. on Ultrasonics, Ferroelectrics, and Frequency Control, **Vol.40** (1993), No. 5. pp. 438-452

/20/ P. V. Wright, *A Rewiew of SAW Resonator Filter Technology*, IEEE Ultrasonics Symposium 1992, pp. 29-38

/21/ K. Hanma, B. Hunsinger, *A Triple Transit Suppression Technique*, IEEE Ultrasonics Symposium Proc. 1976, pp. 328-331

/22/ C.S. Hartmann, P.V. Wrigth, R.J. Kansy, E.M. Garber, *An Analysis of SAW Interdigital Transducers with Internal Reflections and the Application to the Design of Single Phase Unidirectional Transducers*, IEEE Ultrasonics Symposium Proc. 1982, pp. 40-45

/23/ P. Ventura, M. Solal, P. Dufilie, S. Chamaly, *A Global Optimization Procedure for SPUDT Filters*, IEEE Ultrasonics Symposium Proc. 1993, pp. 5-8

/24/ S. Jen, C.S. Hartmann, *Synthesis and Performance of Medium Bandwidth, Low Time-Spurious, High Out-of-Band Rejection SPUDT Filters*, IEEE Ultrasonics Symposium Proc. 1993, pp. 9-13

/25/ M.F. Tiersten, R.C. Smythe, *Guided Acoustic Surface Wave Filters*, IEEE Ultrasonics Symposium Proc. 1975, pp. 293-294

/26/ C.S. Hartmann, D.P. Chen, J. Heighway, *Modeling of SAW transversely coupled resonator filter using coupling-of-modes modeling technique*, IEEE Ultrasonics Symposium Proc. 1992, pp. 39-43

/27/ G. Martin, B. Wall, R. Kunze, M. Weihnacht, *Four Modes Waveguide Resonator Filter*, IEEE Ultrasonics Symposium Proc. 1993, pp. 35-39

/28/ J.D. Maines, *Surface Wave Devices for Radar Equipment*, In: H. Matthews (Ed.): Surface Wave Filters John Wiley & Sons, New York, London, Sydney, Toronto 1977

/29/ P.M. Bakken, A. Ronnekleiv, B.R. Andersen, *Frequency Multiplexers and Demultiplexers Based on the SAW Chirp Fourier Transformer Combined with Digital Processing*, IEEE Ultrasonics Symposium Proc. 1993, pp. 137-142

/30/ E.A. Ash, *Fundamentals of Signal Processing Devices*, In: A.A. Oliner (Ed.): Acoustic Surface Waves, Springer-Verlag Berlin, Heidelberg, New York 1978

/31/ E.W. Greeneich, R.S. Muller, *Acoustic-Wave Detection via a Piezoelectric Field-Effect-Transducer*, Appl. Phys. Lett. **Vol. 20** (1972), No.2, pp. 156-158

/32/ F.S. Hickernell, *The effect of the structure and composition of zinc oxide films on their elastic properties*, IEEE Ultrasonics Proc. 1977, pp. 309-312

/33/ F.S. Hickernell, *The microstructural characteristics of thin-film zinc oxide SAW transducers*, IEEE Ultrasonics Symposium Proc. 1984, pp. 239-242

/34/ F.S. Hickernell, M.D. Adamo, R.V. Delong, J.G. Minsdale, M.J. Bush, *SAW Programmable Matched Filter Signal Processor*, IEEE Ultrasonics Symposium Proc. 1980, pp. 104-108

/35/ J.Haartsen, A. Venema, *The junction FET used as a SAW detector in ZnO-SiO_2-Si structures*, IEEE Ultrasonics Symposium Proc. 1989, pp. 207-211

/36/ A.J. van Rhijn, G.W. Lubking, J.C. Haartsen, *A fully integrated SAW programmable transversal filter with programming readout circuitry*, IEEE Ultrasonics Symposium Proc. 1991, pp. 89-94

/37/ J.Heighway, D.P. Morgan, *1991 and beyond: The current status and future trends for SAW in Europe*, IEEE Ultrasonics Symposium Proc. 1991, pp. 217-223

/38/ A.S. Kozlov, I.P. Puzynya, N.I. Tolstoukhov, I.B. Yakovkin, *Surface Acoustic Wave Devices in a Radio Monitoring System*, Telecommun. Radio Eng. **Vol. 45** (1990) No. 5, pp. 56-57

/39/ L. Reindl, W. Ruile, *Programmable Reflectors for SAW-ID-Tags*, IEEE Ultrasonics Symposium Proc. 1993, pp. 125-130

/40/ R.B. Thompson, C.F. Quate, *Convolution and correlation in real time with nonlinear acoustics*, Appl. Phys. Lett. **Vol. 16**, pp. 494-496

/41/ A. Fauter, L. Reindl, R. Weigel, P. Russer, F. Seifert, *Miniaturized SAW Convolver for Indoor Mobile Communication*, IEEE Ultrasonics Symposium Proc. 1993, pp. 73-77

/42/ H. Gautier, G. S. Kino, *A detailed theory of the acoustic wave semiconductor convolver*, IEEE Trans. on Sonics and Ultrasonics, **Vol. 24** (1977), No. 1, pp. 23-33

/43/ R. Böttcher, W. Buff, *New Type of MZOS FET Convolver*, IEEE Ultrasonics Symposium Proc. 1991, pp. 259-262

/44/ F. Möller, W. Buff, *Electromagnetic Feedthrough in Si/ZnO-SAW-Devices*, IEEE Ultrasonics Symposium Proc. 1992, pp. 245-248

/45/ A. D' Amico, E. Verona, *SAW sensors*, Sensors and Actuators **17** (1989) 55-66

/46/ W. Buff, *SAW sensors*, Sensors and Actuators A, **30** (1992) pp. 117-121

/47/ K. Yamanouchi, H. Odagawa, T. Meguro, Y. Wagatsuma, K. Yamamoto, *Nano-Meter Electrode Fabrication Technology Using Anodic Oxidation Resist and Application to 20 GHz-Range SAW Devices*, IEEE Ultrasonics Symposium Proc. 1993, pp. 1263-1266

Generalized Elemental Solutions in Periodic SAW-Devices

H.P. Zidek and A.R. Baghai-Wadji

Vienna University of Technology, Department of Electrical Engineering and Electronics
Applied Electronics Group E359/2, Gußhausstraße 27–29, A-1040, Vienna, AUSTRIA

ABSTRACT — An easy-to-implement method is presented, allowing the efficient construction of elemental solutions in periodic arrays for 3D full-wave SAW-device modeling. The proposed elemental solutions represent manifold generalizations of Datta and Hunsinger's element factor, and consider *(i)* single and double periodic arrays of electrodes with finite extension, *(ii)* straight and curved electrodes, *(iii)* fundamental cells containing a finite number of electrodes with arbitrary driving potentials, *(iv)* electrically floating electrodes, and finally *(v)* rigorous electro-mechanic field-interaction. The wide-range applicability of the method is demonstrated by calculating various elemental charge and potential distributions.

I. Introduction

THE MAJORITY of existing models for surface acoustic wave (SAW) devices assume an infinite number of electrodes, and thus neglect longitudinal end-effects due to the finiteness of the electrode number. Furthermore, this models are not capable of describing the transversal end-effects which result from the finiteness of the aperture length of the electrodes. In a recent work [1] we presented a theory for rigorous 3D full-wave analysis of SAW-devices. This theory allows a frequency-dependent analysis, takes into account launching and scattering of surface and bulk waves, and accounts for both longitudinal and transversal end-effects. However, the application of this theory to real devices demands a considerable amount of computer resources; both time and memory requirements are considerable.

Many modern SAW devices involve a huge number of **short** electrodes, the analysis of which demands finite-aperture electrode models. Neglecting the longitudinal end-effects, this type of transducers can be analyzed by means of periodic models. One of the aims in this paper is the presentation of an efficient formulation to solve this kind of problem. Thus, while we neglect the longitudinal end-effects, we fully

account for the transversal end-effects. In structures like reflecting arrays [2], the substrate surface is covered by a matrix of metallic patches in the form of a double periodic array. A further goal in this paper is the analysis of this type of arrays.

As in antennae problems, the periodicity assumption in an array suggests the introduction of an elemental solution associated with this array, and an array-characteristic function. The array function in our problem describes the electrodes' potentials and thus its construction is trivial. In contrast to this the elemental solution is highly complicated. A significant part of this paper is devoted to a theory which enables us to construct a variety of useful elemental solutions for 3D SAW-device modeling.

Already existing elemental solutions: Based on Engan's electrostatic field solution of periodic interdigital-transducers on a dielectric semi-space [3], Datta and Hunsinger calculated an element factor in closed-form [4, 5]. However, this factor applies to a small subclass of **static** field problems in SAW-device modeling. In [6] we extended this theory to include frequency-dependent problems.

Generalization: In this paper we generalize the concept of element factor to cover the following:

(i) 3D field-problems,

(ii) arbitrarily-shaped electrodes (e.g. curved electrodes in beam compressors),

(iii) periodic sequences of groups of electrodes (e.g. a pair of electrodes in split-finger transducers, including the influence of the bus-bars),

(iv) isolated and/or interconnected floating electrodes,

(v) single- and double-periodic electrode-arrays,

(vi) full-wave analysis (rigorous calculation of SAW and bulk-wave interaction).

The reader will recognize that the proposed theory can be easily modified in order to include problems with even more complicated boundary conditions. We should mention the following examples: *(i)* substrates with a finite thickness (the backside of the substrate can be free or metallized), *(ii)* multilayered media as they arise in the package modeling and in SAW sensor applications, and *(iii)* SAW propagation in stacked media involving semiconducting layers.

The usage of the term "elemental solution" instead of "element factor," serves merely to emphasize the fact that our field problems may or may not be frequency dependent.

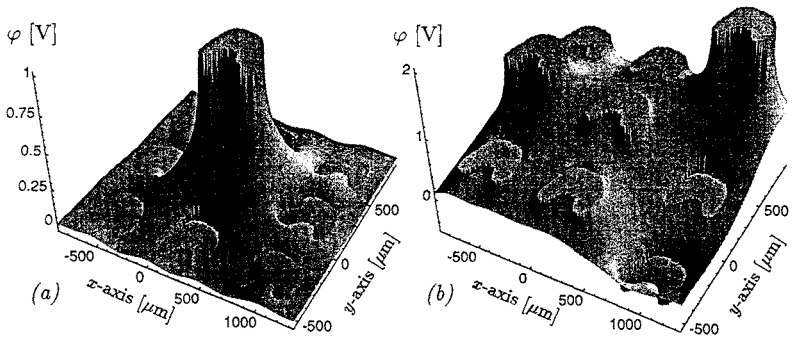


Figure 1: Potential distribution on an array of kidney-shaped electrodes. *(a)* Elemental solution. *(b)* Solution for a particular electric feeding. The potential takes on values between -0.5 and 2 V.

Elemental solutions: An elemental solution is defined as the field-reaction in a periodic array, if a group of driven electrodes is surrounded by an infinite set of neighboring grounded electrode-groups. Once an elemental solution related to a periodic array is known, the field variables in this array, for any potential feeding, can be constructed by shifting, scaling, and superposition processes. This computationally advantageous feature results from the geometric periodicity, and assumed linearity properties.

An example of an elemental potential distribution in an array of metallic patches with fairly complicated geometry is depicted in Fig. 1*a*. Only the nearest neighbors to the driven electrode are shown. The structure is driven at the 3^{th} harmonic on a 128° rot $LiNbO_3$ substrate. Constant potential areas indicate the location of the electrodes.

Because of the finiteness of real arrays, deviations from the real solution (longitudinal end-effects) will occur by using the above-defined elemental solution. However, the end-effects in real arrays with many electrodes can be neglected in the case of rapidly decaying electrostatic fields. Fig. 1*b* presents a result obtained by using the elemental potential distribution in Fig. 1*a*. The electrodes' potentials vary between -0.5 and 2 V. The surrounding electrodes—which are not shown here—are grounded.

In our approach the calculation of elemental solutions requires the field distribution in a **fundamental cell** only (Fig. 2). The geometry of FC is periodic in x, and possibly in y-direction. The method relies on the recently developed concept

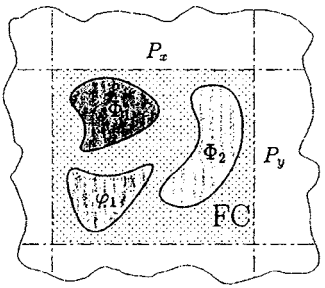


Figure 2: Fundamental cell: Φ_1 and Φ_2 are the given potentials of the driven electrodes, φ_1 denotes the unknown potential of an floating electrode.

of **periodic Green's function** [6]. The computations are fast and the resulting numerical data are extraordinarily noise-free. These advantages result from the fact that the construction of the involved periodic Green's functions needs only sampled data in the wave number domain, and that we employ the error minimizing method of **weighted residuals** [7].

II. Statement of the Problem

The aim is the efficient field calculation in geometrically single- and double periodic electrode-configurations for arbitrary potential feedings. A harmonic time-dependence according to $\exp(j\omega t)$ will be assumed. The arbitrarily shaped, ideal-conductive metallic electrodes with finite extension are taken to be deposited on the plane surface of a piezo-electric substrate with anisotropic material parameters $\underline{\underline{\varepsilon}}$, $\underline{\underline{e}}$ and $\underline{\underline{c}}$, [8]. The problem is the calculation of the array-characterizing elemental field-solutions.

III. Conventions

In Fig. 2 we have defined a fundamental cell (FC) as a rectangular area on the substrate surface ($z = 0$). Within FC we assume a finite number of arbitrarily shaped electrodes with given potential values, and possibly a number of electrically floating ones. A function $\Phi(\boldsymbol{r})$ defines the driven-electrodes' potentials in FC. $\Phi(\boldsymbol{r})$ may take on any real or complex number.

We consider now a **geometrically** single-periodic array which is defined in the following way: The m^{th} cell of this array is in x-direction a distance $m \times P_x$ apart from the FC and the electrode-configuration is identical with the one in FC. In addition, a function $C_m\Phi(x - mP_x, y)$ describes the driven-electrodes' potentials in this

cell. Analogously, we define a geometrically double-periodic array with a function $C_{mn}\Phi(x - mP_x, y - nP_y)$ defining the driven-electrodes' potentials in $(m,n)^{\text{th}}$ cell. The constants C_m and C_{mn} can be either real or complex.

Let us now concentrate on single periodic arrays. Among all the possible realizations of this type of arrays, the following special cases play a central role in our analysis: a **phased-periodic array** is an array with C_m being equal to $\exp(jk_{0x}mP_x)$, where the wavenumber k_{0x} can vary between $-\pi/P_x$ and π/P_x (K-interval). An **elemental array** is an array with C_m vanishing for $|m| \geq 1$ and C_0 being equal to unity. The field solution on an elemental array will be referred to as **elemental field-solution**.

Notice that any multiple of FC can also be regarded as a 'fundamental' cell. However, from a computational point of view the FC should be the one with the smallest geometrical extension.

IV. Governing Equations

In the free space Laplace's equation is valid. The governing equations in the substrate are the following coupled partial differential equations:

$$\begin{aligned} -\rho\,\ddot{u}_j + c^E_{ijkl}\,u_{k,li} + e_{kij}\,\varphi_{,ki} &= 0, \\ e_{ikl}\,u_{k,li} - \varepsilon^S_{ik}\,\varphi_{,ki} &= 0. \end{aligned} \tag{1}$$

φ is the electric potential, $\underline{u}$ is the elastic displacement vector, ρ is the mass density, $\underline{\underline{c}}^E$ is the elastic stiffness tensor measured at constant electric field, ε^S is the permittivity tensor measured at constant strain, and $\underline{\underline{e}}$ is the piezo-electric coupling tensor. A dot over a symbol denotes $\partial/\partial t$ and subscripts preceded by a comma denote partial spatial derivatives.

Boundary conditions: The interface plane ($z = 0$) is stress-free:

$$T_{i3} = 0. \tag{2}$$

On moving from the substrate (media I) to the upper half-space (media II), the electric potential is continuous across the interface plane

$$\varphi^{\text{II}}(\boldsymbol{r}, z = 0^+) = \varphi^{\text{I}}(\boldsymbol{r}, z = 0^-). \tag{3}$$

The potential of driven electrodes is known:

$$\varphi^{\text{II}}(\boldsymbol{r}, z = 0^+) = \varphi^{\text{I}}(\boldsymbol{r}, z = 0^-) = U(\boldsymbol{r}). \tag{4}$$

The *a-priori* unknown potential value of a floating electrode has to be determined in such a way, that the charge integral on this electrode vanishes. The jump discontinuity of the normal component of the displacement vector on the interface is equal to the charge density:

$$D_z^{\mathrm{II}}(\boldsymbol{r}, z = 0^+) - D_z^{\mathrm{I}}(\boldsymbol{r}, z = 0^-) \;=\; \sigma(\boldsymbol{r}, z = 0). \tag{5}$$

In addition, the fields are required to vanish at the limit $|z| \to \infty$. These conditions together with a solution Ansatz for the field-variables φ and u in the form

$$\exp(-jk_x x)\,\exp(-jk_y y)\,\exp(k_z(k_x, k_y)z)\,\exp(j\omega t) \tag{6}$$

lead to the equation

$$\varphi(k_x, k_y) \;=\; G(k_x, k_y)\,\sigma(k_x, k_y) \tag{7}$$

The involved Green's function $G(k_x, k_y)$ in wave number domain can be determined if we consider a point-charge excitation of the medium.

V. Outline of the Method

Assume that we are equipped with a convenient tool to calculate the field distributions on a given phased array specified by k_{0x}. Then we let k_{0x} in the associated phasor $\exp(jk_{0x}mP_x)$, as introduced above, vary continuously in K, and calculate the field distributions on the corresponding phased arrays. It is not difficult to show that an elemental solution can be obtained by integrating the fields on the phased arrays over k_{0x}. In practice, a sampling of K at N equally-spaced points $k_{0x}^{(n)}, n = 1 \ldots N$, leads to N phased arrays, and thus to an approximate elemental solution. Note that due to the discretization of K, the elemental solution is periodic in x-direction with the periodicity $N \times P_x$.

We now outline the main steps by the simple example shown in Fig. 3. We sample the K-interval ranging from $[-\pi/P_x, \pi/P_x]$ at the following three points: $-6\pi/6P_x$, $-2\pi/6P_x$, and $2\pi/6P_x$, as indicated in the figure. These points correspond to three phased arrays characterized by the phasors $\exp(-j6\pi/6)$, $\exp(-j2\pi/6)$, and $\exp(j2\pi/6)$. Let us concentrate on the potential distributions only. Fig. 3*a* shows the potential distribution of the $\exp(-j6\pi/6)$-phased array, which is a real function. The potential distributions of the $\exp(-j2\pi/6)$ and $\exp(j2\pi/6)$-phased arrays are conjugate complex. The real parts of these functions are equal and shown in Fig. 3*b*. Adding together the potential distributions of the above-mentioned three phased arrays we obtain the elemental potential distribution as shown in Fig. 3*c*. (From the above discussion the reader will recognize that the functions in Fig. 3*c* are real-valued.)

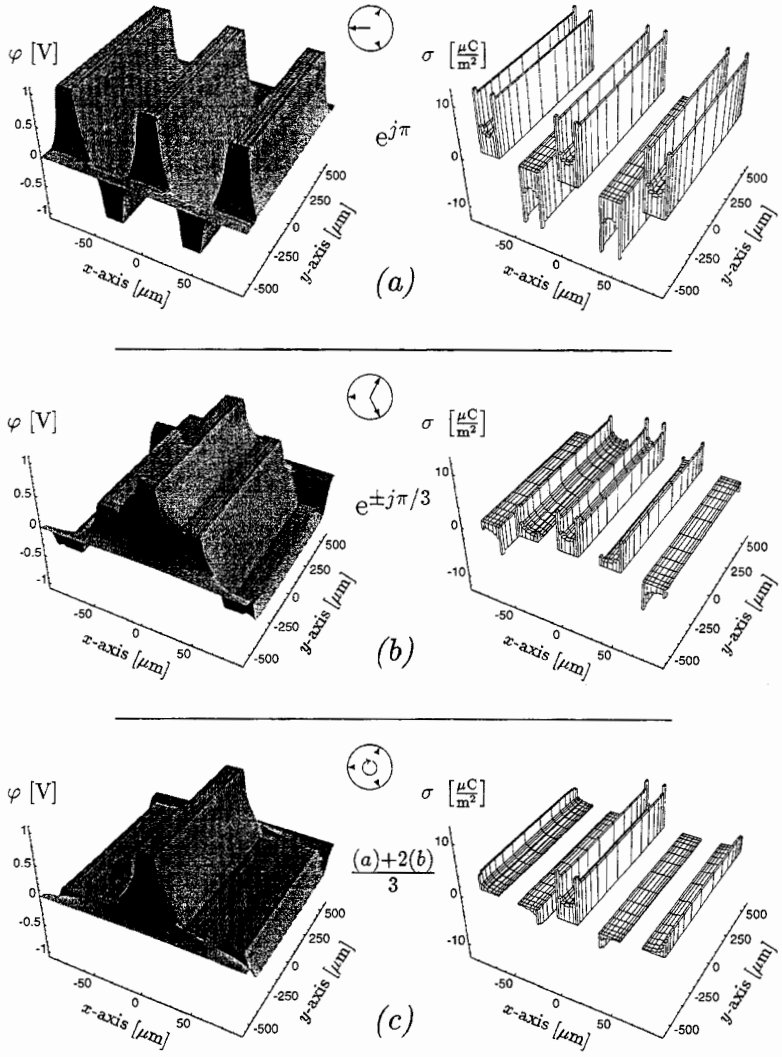


Figure 3: Construction of elemental solutions: The real part of the partial solutions for the potential (the left column) and charge (the right column) distribution. The elemental solutions *(c)* are the superposition of field-solutions for the feedings $\exp(j\pi)$, $\exp(j\pi/3)$, and $\exp(-j\pi/3)$. Note the different scale of the x- and y-axis.

Phased arrays: The analysis of the above-mentioned phased arrays is straightforward, and relies on the problem's **phased-periodic Green's function** $G^{\rm ph}(\boldsymbol{r}-\boldsymbol{r}';k_{0x})$. $\boldsymbol{r}'$ and $\boldsymbol{r}$ refer, respectively, to the position vectors of the source and observation point in FC. The function $G^{\rm ph}$ can be constructed from the Green's function $G(k_x,k_y)$ of an isolated point-source (7) in the following form:

$$\begin{aligned} G^{\rm ph}(\boldsymbol{r}-\boldsymbol{r}';k_{0x}) &= \sum_{n=-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{dk_y}{2\pi}\,G\left(k_{0x}+2\pi n/P_x,k_y\right) \\ &\quad\times\exp\left[j\left(k_{0x}+2\pi n/P_x\right)(x-x')\right]\exp\left[jk_y(y-y')\right] \\ &= \exp\left[jk_{0x}(x-x')\right]G^{\rm per}(\boldsymbol{r}-\boldsymbol{r}';k_{0x}). \end{aligned} \tag{8}$$

Evidently, $G^{\rm per}$ and $G^{\rm ph}$, respectively, possess the following properties:

$$\begin{aligned} G^{\rm per}(x-x'+P_x,y-y';k_{0x}) &= G^{\rm per}(\boldsymbol{r}-\boldsymbol{r}';k_{0x}), \qquad \text{and} \\ G^{\rm ph}(x-x'+P_x,y-y';k_{0x}) &= \exp\left(jk_{0x}P_x\right)G^{\rm ph}(\boldsymbol{r}-\boldsymbol{r}';k_{0x}). \end{aligned}$$

Using $G^{\rm ph}$ and applying the method of moments (MoM) to the electrodes in FC, the charge distribution on these electrodes, $\sigma_{\rm FC}(\boldsymbol{r},k_{0x})$, can be calculated. The charge distribution in $m^{\rm th}$ cell is then given by $\sigma_{\rm FC}(\boldsymbol{r},k_{0x})\exp\left(jk_{0x}mP_x\right)$. All other field variables can be conventionally calculated from $\sigma_{\rm FC}(\boldsymbol{r},k_{0x})$ in a post-processing step.

Below we summarize the main steps in constructing elemental field-solutions associated with single periodic arrays:

(i) Construct the Green's function $G(k_x,k_y)$. associated with an isolated point-charge residing on the surface of a piezo-electric semi-space, (7).

(ii) Set up $G^{\rm ph}(\boldsymbol{r}-\boldsymbol{r}';k_{0x})$ according to (8), using $G(k_x,k_y)$.

(iii) Divide $\mathbf{K}$ into N subintervals with midpoints $k_{0x}^{(1)},\ldots,k_{0x}^{(N)}$.

(iv) Construct the phased arrays associated with $k_{0x}^{(1)},\ldots,k_{0x}^{(N)}$.

(v) Calculate the charge distribution in FC of each phased array, employing MoM.

(vi) Calculate all other desired field variables.

(vii) Superpose the field distribution functions to obtain the elemental solution in question.

These steps are *mutatis-mutandis* valid for double periodic arrays.

It is worth mentioning that the information contained in an elemental solution can be stored in a laboratory PC, and can be readily used to construct the field distribution for any periodic or non-periodic feeding of electrodes in geometrically periodic arrays.

We conclude the paper with the following numerical results.

VI. Numerical results

Fig. 4 presents the elemental charge and potential distribution in a double periodic array of electrodes. Each electrode is circumscribed by a floating punched electrode. The structure is driven at the 3rd harmonic. The substrate material is YZ-$LiNbO_3$. Similar results are obtained for an array of electrodes embedded in a grounded mesh (Fig. 5).

In addition, we have calculated the electrostatic charge and potential distributions on the electrode structures described in Figs. 4 and 5. In order to facilitate a direct comparison of the results, the charge and potential distributions, respectively, are presented in combined form as given in Fig. 6. It can be seen that the grounded mesh shields more effectively than the floating punched electrodes. In the case of the floating electrodes ($x > 0$) the capacitance C_e between the active electrode in the centre and the surrounding grounded electrodes is 405 fF. The calculated capacitance between these electrodes is 60 fF in the case of the grounded mesh. However, the capacitance C_g between the active electrode and the mesh is 397 fF.

Our final example (Fig. 7) shows the charge and potential distribution on a finite-aperture transducer on YZ-$LiTaO_3$ at center frequency. The fundamental cell consists of two active electrodes, sandwiched between two floating ones. (The aperture length is equal to $1.5 \times P_x$.) The frequency is assumed to be the center-frequency of the structure. The elemental charge distribution in this structure possess discontinuities on the bus-bars. However, by shifting and superposing the elemental solutions, these discontinuities cancels in the final result as is obvious in Fig. 7.

This calculation has been carried out to check the accuracy of the numerical results. As the reader can see, the charge distribution in this structure can be obtained directly, because it can be regarded as a phased array with $k_{0x} = 0$. We compared this result with the numerical data we obtained by applying the above-described theory of elemental solutions. The results are equal within a few per cent.

Conclusion

A variety of novel elemental charge-distributions in SAW-devices has been constructed, taking into account the surface and bulk waves, as well as the finite exten-

sion of the electrodes. The construction relies on phased-periodic Green's functions, the method of moments, and the superposition-principle as applied to phased arrays. We have demonstrated the applicability of the formulation to 3D full-wave SAW-device modeling.

ACKNOWLEDGMENT

This work has been supported by Siemens AG, Corporate Research and Development, Munich, Germany. We gratefully acknowledge helpful discussions with Prof. F. Seifert and Dr. C. Ruppel.

REFERENCES

[1] H. P. Zidek, A. R. Baghai-Wadji, and O. Männer, "Full-wave 3D analysis of wave scattering on SAW-structures with finite aperture," in *IEEE Ultrason. Symp. Proc.*, pp. 149—152, 1993.

[2] F. Huang and E. G. S. Paige, "The scattering of surface acoustic waves by electrical effects in two-dimensional metal film structures," *IEEE Trans. on Sonics Ultrason.*, vol. SU-35, no. 6, pp. 723–735, 1988.

[3] H. Engan, "Excitation of elastic surface waves by spatial harmonics of interdigital transducers," *IEEE Trans. on Electron Dev.*, vol. ED-16, no. 12, pp. 1014–1017, 1969.

[4] S. Datta, B. J. Hunsinger, and D. C. Malocha, "A generalized model for periodic transducers with arbitrary voltages," *IEEE Trans. on Sonics Ultrason.*, vol. SU-26, no. 3, pp. 235–242, 1979.

[5] S. Datta and B. J. Hunsinger, "Element factor for periodic transducers," *IEEE Trans. on Sonics Ultrason.*, vol. SU-27, no. 1, pp. 42–44, 1980.

[6] A. R. Baghai-Wadji, C. Mecklenbräuker, and F. J. Seifert, "Convenient method for noise-free elastodynamic field calculations in general periodic SAW structures," *Electronics Letters*, vol. 28, no. 15, pp. 1466–1468, 1992.

[7] R. F. Harrington, *Field Computation by Moment Method.* New York: Macmillan, 1968.

[8] B. A. Auld, *Acoustic Fields and Waves in Solids.* New York: John Wiley & Sons, 1973.

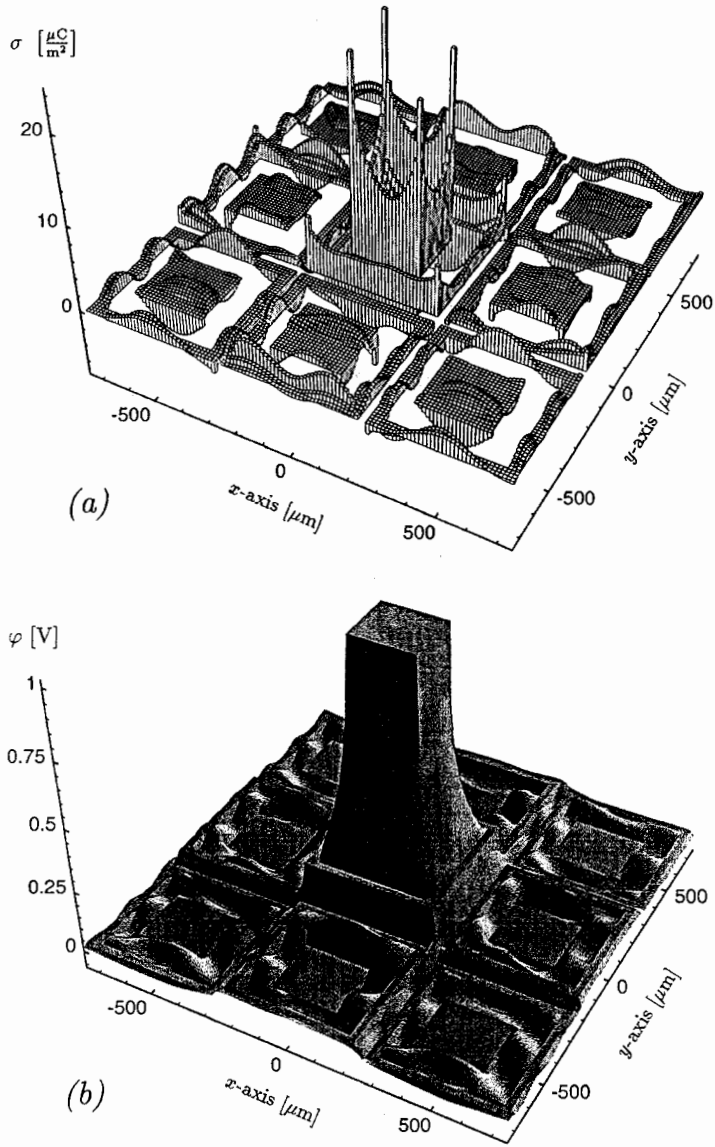


Figure 4: Charge and potential distribution on an array with electrodes, surrounded by floating punched electrodes at 3rd harmonic.

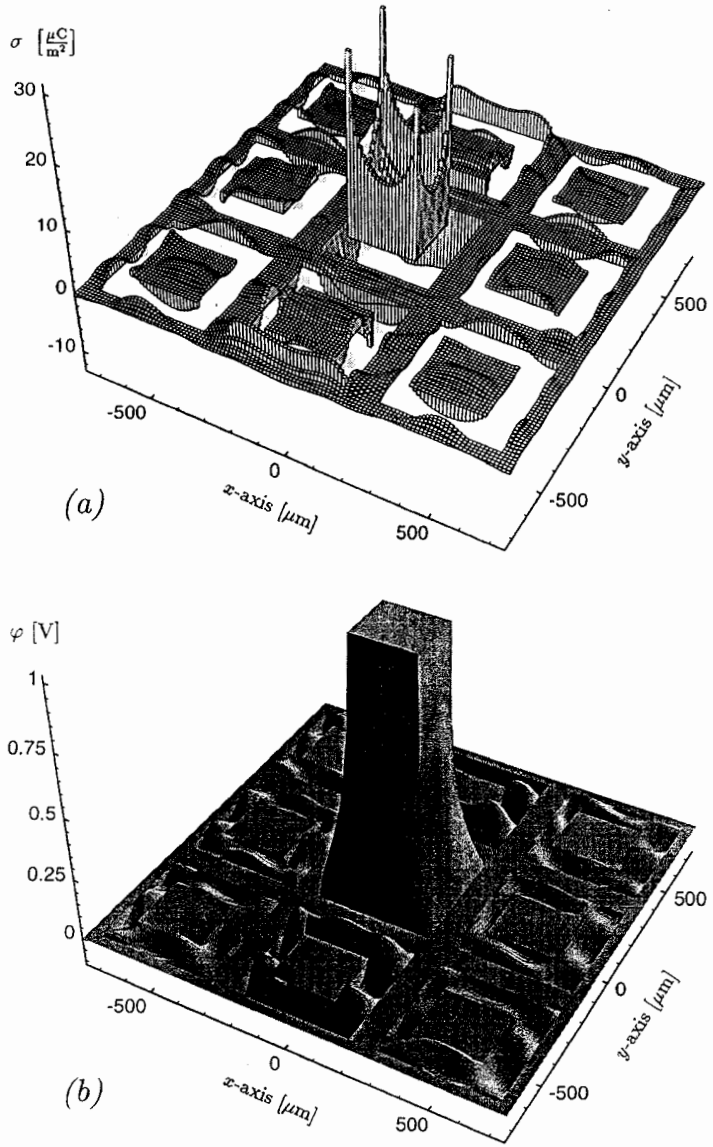


Figure 5: Charge and potential distribution on an array of electrodes embedded in a shielding mesh at 3rd harmonic.

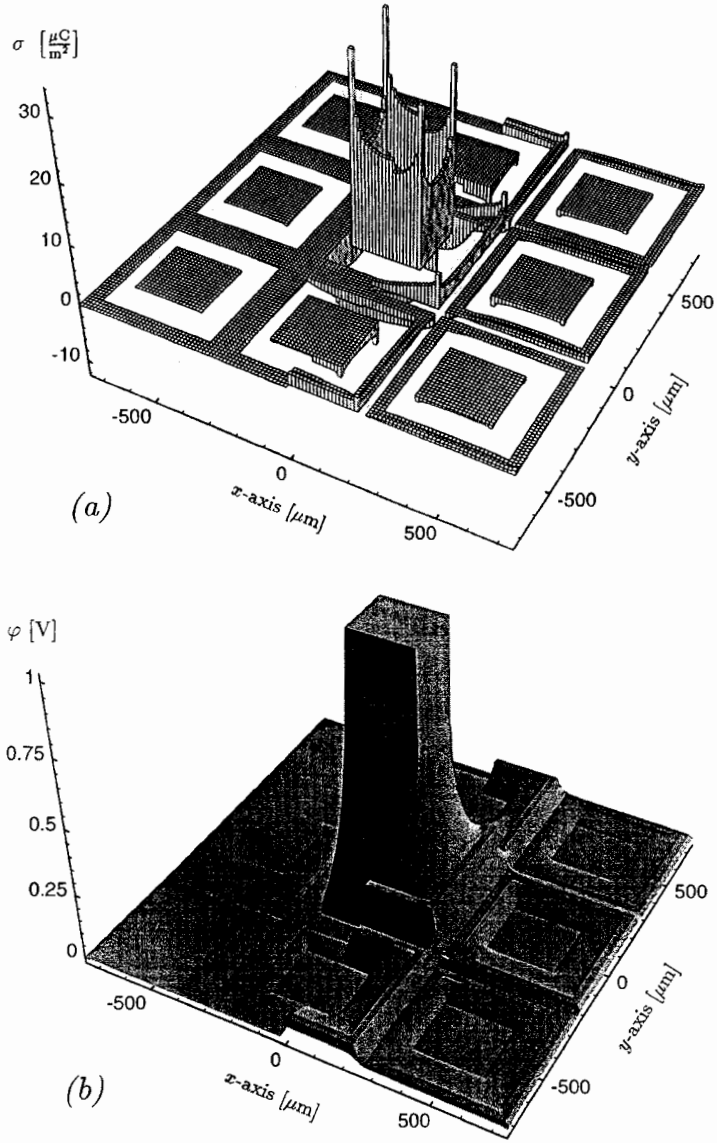


Figure 6: Combined results from calculations for the structures in Figs. 4 and 5. The charge and potential distributions are show (see text for details).

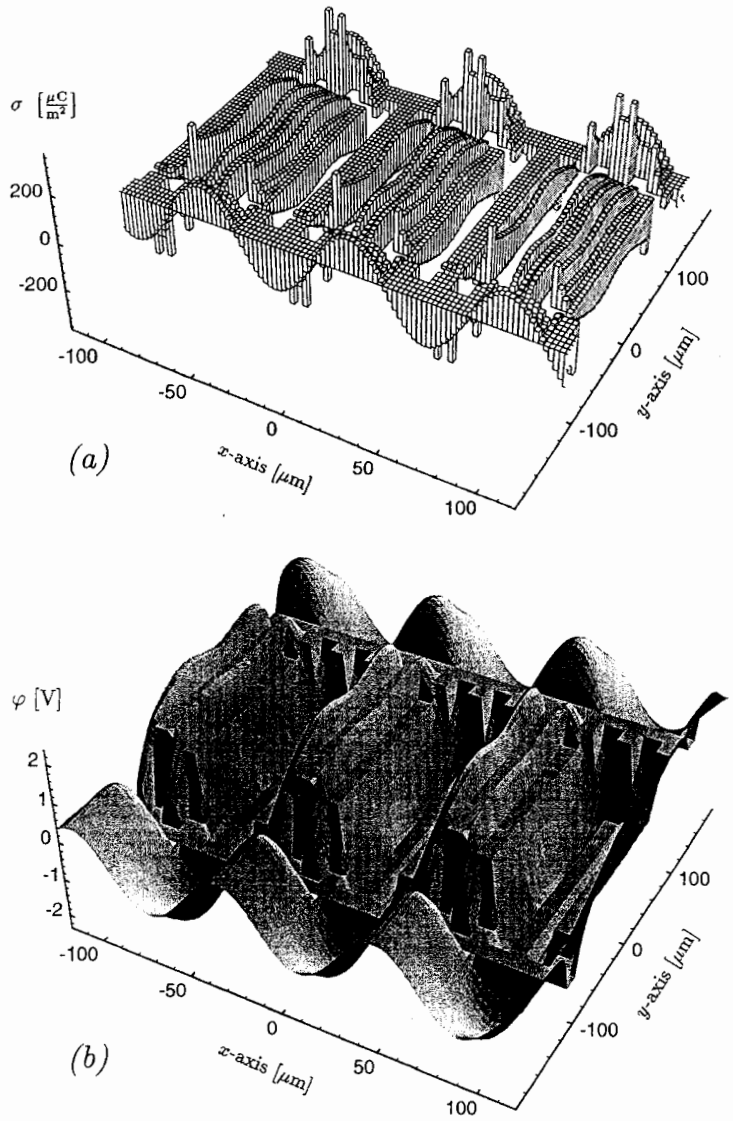


Figure 7: Charge and potential distribution on an array at center frequency.

GHz range STW resonators and narrowband filters.

Victor P.Plessky and Thor Thorvaldsson
Advanced SAW Products SA, CH-2022 Bevaix, Switzerland

ABSTRACT

This paper presents results of a development of a low loss fundamental mode high frequency surface transverse wave (STW) filter. A new type of balanced bridge impedance element (IE) structure was applied. The measured insertion loss at a center frequency close to 1.5 GHz was 3 dB. The sidelobe rejection was more than 35 dB and the 3 dB bandwith was about 1 MHz.

1. INTRODUCTION

During the past few years, interest has been growing in the use of surface transverse waves (STW) /1,2/ for high frequency resonators and narrow band filters. These waves have some obvious advantages over the Rayleigh type surface acoustic waves (SAW). In materials like quartz, the velocity of STW propagation is about 60% higher than that of SAW, which simplifies photolithography, and temperature stable cuts are also available. This and some other arguments make the choice of STW on quartz practically inevitable for narrow band devices operating at frequencies higher than 1.3 -1.5 GHz.

The theoretical study of STW propagation on periodic system of electrodes /3-6/ shows that its behavior is similar to that of Rayleigh waves: STW are reflected intensively at Bragg frequencies, stopband frequency range being dependent on the thickness of *Al* electrodes. But, the localization depth of the wave is not constant. The wave is strongly localized near the surface around the lower edge of the stopband and becomes non localized at frequencies higher than the upper edge of the stopband. At these frequencies the attenuation of the wave increases as a result of bulk wave scattering /5,6/. To avoid these additional losses, STW resonators are designed so that the resonance frequency is shifted towards the lower edge of the stopband. In practice this means that synchronous resonators are of interest.

2. ONE PORT RESONATOR

A one port synchronous resonator is a very suitable test structure for extraction of wave propagation parameters. The measurement of admittance dependence on frequency (see Fig.1) shows directly the frequency edges of the stopband. Other important parameters like strength of electromechanical coupling, attenuation of the wave, static capacitance and parasitic resistance can also be determined from this curve /5,6/. For the simulation of the admittance dependence on frequency, a standard coupling of modes (COM) model was used as well as a modified variant of the model which included 2 parameter description of the reflection coefficient of STW on periodic grating /7,8/. In this particular case, presented in Fig.1, the reflection coefficient is approximately 0.038 per wavelength, that is, per 2 electrodes.

This value is even slightly higher than the reflectivity of Rayleigh wave for the same relative thickness and metalization ratio. The magnitude of the reflection coefficient for the complete grating is shown in Fig.2. It is observed that the upper edge of the grating stopband is practically independent on electrode geometry and always occurs at about 1020 MHz which corresponds to a bulk surface skimming wave propagation.

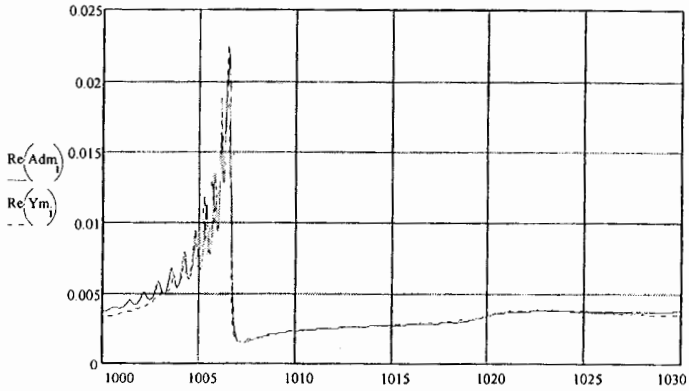


Fig.1 The real part of the admittance of one port synchronous resonator (dotted curve is measured, solid curve is simulated). There are 150 electrodes in each transducer, there are 1000 strips in each grating, the period of the structures is 2.5 μm, the aperture is 500μm, the Al thickness is 1100 Å and the metallization ratio is about 0.42

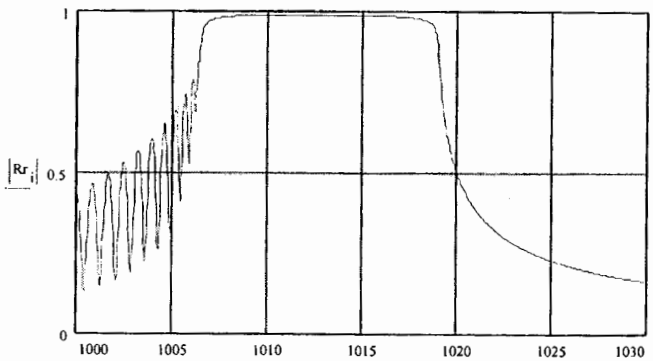


Fig.2 Magnitude of the grating reflection coefficient.

3. BALANCED BRIDGE NETWORK

By connecting together one port synchronous resonators (Fig.3) in a balanced bridge network as shown in Fig.4, narrowband filters can be designed.



Fig.3 A one port resonator building block.

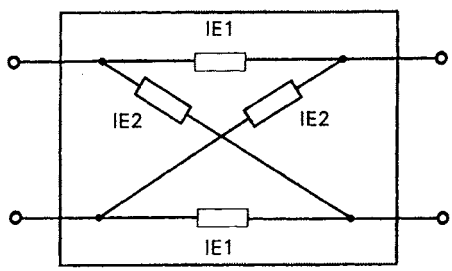


Fig.4 A balanced bridge network.

The resonators are used here as impedance network elements. They do not interact acoustically and only the change of impedance due to STW generation inside the resonators is of importance. In this scheme (Fig. 4) both arms of the bridge include the IDTs with the same static capacitance, but with different center frequencies (impedance elements IE1 and IE2 in Fig.4). The frequency shift is chosen so that the maximum admittance of one transducer roughly coincides with minimum admittance of the other. At this frequency the bridge is no more balanced and the input is connected to the output through low impedance giving low loss in the passband of the device. At other frequencies all resonators are just static capacitors of the same magnitude of capacitance, and therefore the bridge is balanced and the signals at the output are progressively decaying away from the passband frequency range.

4. RESULTS

A filter with the structure shown in Fig.4 and one port resonators as building blocks (Fig.3) was realized on a quartz substrate. In order to obtain the frequency shift mentioned in section 3 the resonators have a different periodicity.

The measured insertion loss at a center frequency close to 1.5 GHz was 3 dB. The sidelobe rejection was more than 35 dB and the 3 dB bandwith was about 1 MHz (see Fig.5 and Fig.6).

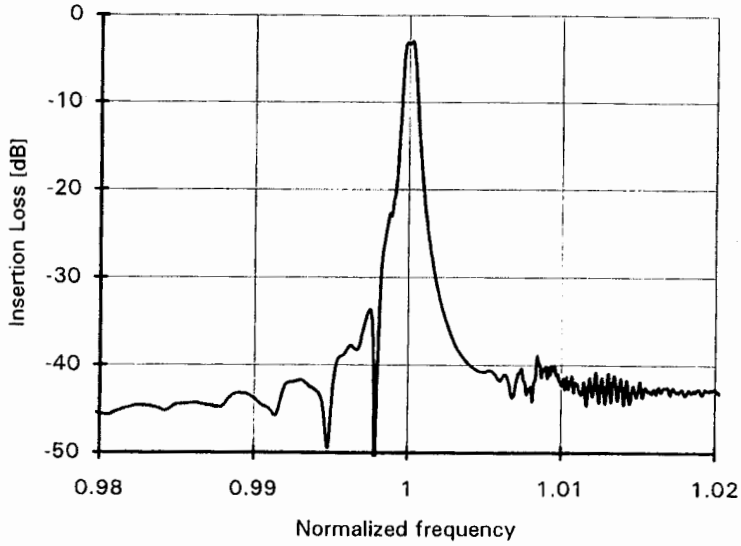


Fig.5 Measured frequency response of a balanced bridge STW filter

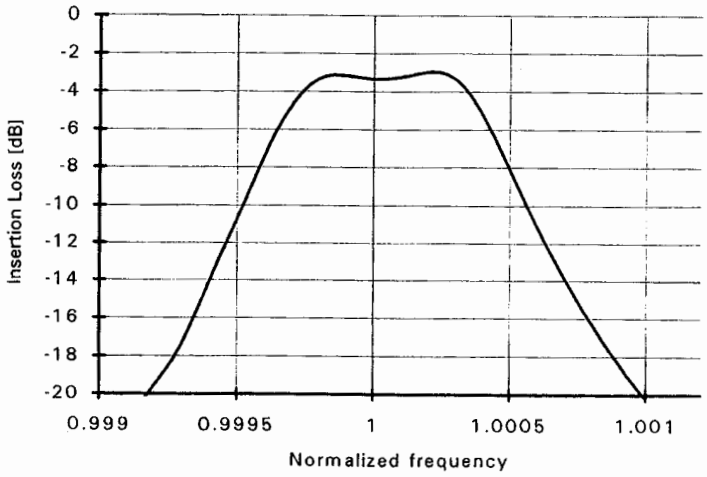


Fig.6 A close up view of the passband of the filter in Fig.5

5. CONCLUSION

A high frequency low loss STW filter using a balanced bridge network has been developed. Low loss was achieved at frequencies close to 1.5 GHz. Using STW in this configuration has the further advantages of high power capability and a good suppression of sidelobes.

6. ACKNOWLEDGMENTS

The authors would like to express gratitude to Mr. S.Muckenhirn for the manufacturing of the devices and to Dr.Choffat of CSEM for his support. Also thanks to Ascom Microelectronics for the access to the thickness measurement equipment and the Swiss PTT for financial support.

7. REFERENCES

/1/ B. A. Auld, J. J. Gagnepain, and M. Tan, "Horizontal shear waves on corrugated surfaces", *Electron. Lett.*, **12** (24), pp. 650-651.

/2/ Yu. V. Gulyaev and V. P. Plesskii, "Slow" acoustic surface waves in solids, *Sov. Tech. Phys. Lett.*, Vol.**3**, No. 3, March 1977,pp. 87-88.

/3/ Yu. V. Gulyaev, T. N. Kurach, V. P. Plesskii, and V.P.Tereshkov, *Akust. Zh.* **30**, 47 (1984), (in Russian), *Sov. Phys. Acoust.* **30**, 27 ,1984 (English translation).

/4/ E. Bigler, B. A. Auld, E. Ritz, and E. Sang, " An analysis of the influence of design parameters on the resonant frequency and Q-factor of surface transverse wave (STW) resonators", *Proc. 45th Ann. Freq. Cont. Symp.*, Los Angeles, June 1991, pp. 222-229, (see here also Ref. at earlier papers of B. A. Auld and co-workers).

/5/ V. P. Plessky and C. S. Hartmann, "Characteristics of leaky SAWs on 36°-LiTaO3 in periodic structures of heavy electrodes", *Proceedings of IEEE Ultrasonics Symposium*, 1991, pp. 1239-1246.

/6/ C. S. Hartmann and V. P. Plessky , "Experimental measurements of propagation, attenuation, reflection and scattering of leaky waves in AL electrode gratings on 41°-, 52°- and 64°-LiNbO3", *Proceedings of IEEE Ultrasonics Symposium*, pp. 1247-1250.

/7/ V. P. Plessky, "A Simple Closed Form Dispersion Equation for Shear Types of Surface Waves Propagating in Periodic Structures", *IEEE Trans. on Ultrasonics, Ferroelec. and Freq. Contr.*, Vol. **40**, No. 4, July 1993, pp. 421-423.

/8/ V. P. Plessky, "A two parameter Coupling-of-Modes model for shear horizontal type SAW propagation in periodic gratings", *Proceedings of IEEE Ultrasonics Symposium*, 1993, pp. 195-200.

EXACT MATERIAL PARAMETER EXTRACTION FOR NARROW BAND SAW FILTERS AND RESONATORS DESIGNED ON QUARTZ

Victor P. Plessky, Marc-André Schwab and Svetlana N. Krasnikova

Advanced SAW Products SA, CH-2022 Bevaix, Switzerland

1. ABSTRACT

Narrow band SAW filters and resonators have become a key element for mobile phones and other advanced communication systems. ST-quartz is normally used as a basic material for the device design due to its excellent temperature performance, ease of processing and stable mechanical characteristics. The Coupling-Of-Modes (COM) model gives the best simulation results when accurate parameters are known. Existing theoretical models allow the estimation of the reflection coefficient (coupling of opposite propagating waves, K -parameter), the strength of electro-mechanical interactions (parameter α), the reduction of the wave velocity (V) due to the mass-load of electrode fingers, and the static capacitance of the interdigital transducers (Cst). As a rule the accuracy of the theoretical data is not sufficient for the design of devices with absolutely correct frequency characteristics. Some other important parameters , the attenuation coefficient (γ) of the wave and the parasitic resistance of electrodes and bus-bars (Rp) are strongly technology dependent and cannot be modeled theoretically at the present stage with the necessary accuracy. Measurement remains the most accurate and reliable source of the parameters for the COM model. It was found that many of the desired model parameters could be determined based solely on the measured admittance of a one port synchronous resonator [1], [2]. The frequency dependence of the admittance clearly shows the edges of the stopband, which allow the determination of the SAW velocity and the reflectivity of the electrodes, the magnitude of the admittance is proportional to the electro-mechanical transduction coefficient and the wave attenuation is related to Q-factors of the resonance peaks. Also the static transducer capacitance and parasitic series resistance can be determined from these admittance measurements. The measured admittance versus frequency curves are remarkably free of spurious bulk, reflected and other parasitic responses. We use here this technique for characterization of Rayleigh -type SAW propagation and reflection in periodic structures on ST-quartz.

2. TEST STRUCTURE

We have used a test bar (Fig. 1) especially designed for extraction of COM parameters [1], [2]. This test bar can be considered as a simplification of a device used by C. S. Hartmann and co-authors [3] and includes only one interdigital transducer (IDT), and two reflectors synchronously placed on both sides of the transducer. For the measurements ofSAW characteristics on quartz the test structure included an IDT with typically 100 electrodes and two synchronously placed reflectors with 2 to 3 times greater number of electrodes in each of them. In particular in the case shown in Fig. 2, the test structure had a period close to $p=17.9\ \mu m$, the number of electrodes in the InterDigital Transducer (IDT) $Nt=101$, the number of electrodes in each of the reflectors $Nr=231$ and the aperture $W=9.5*2p$. This narrow aperture waveguide type structure was used for a Transverse Coupled Filter (TCF) design.

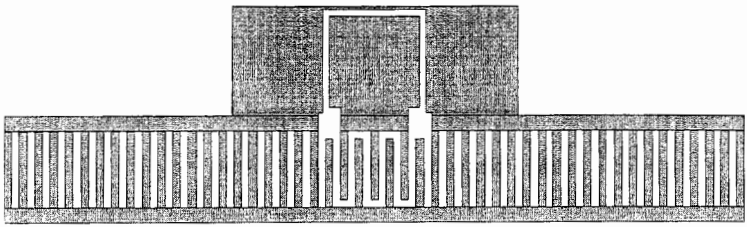


Fig. 1 Test structure geometry

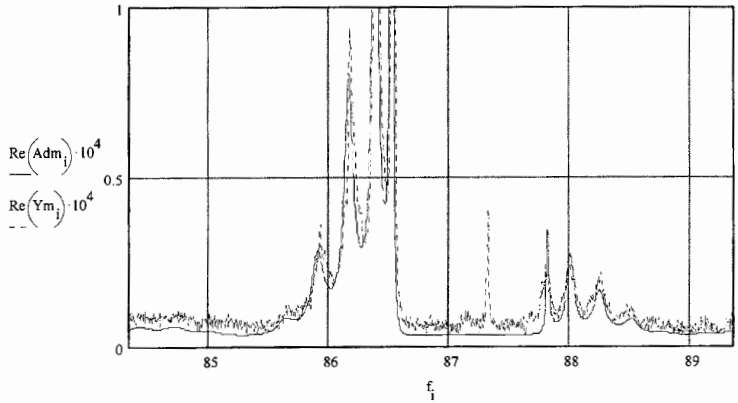


Fig. 2a Real part of admittance , (solid curve - simulation, dotted curve - measurement)

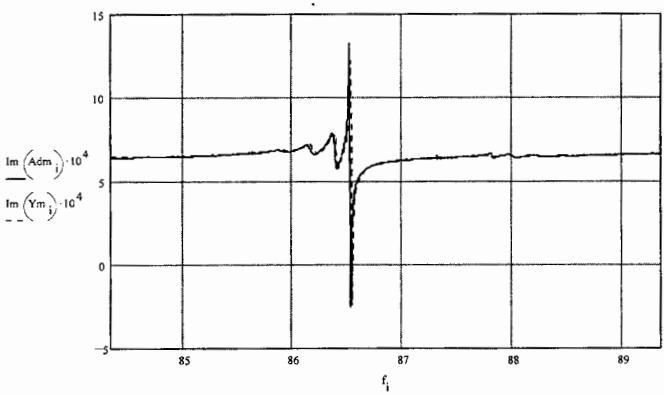


Fig. 2b Imaginary part of admittance (solid curve - simulation, dotted curve - measurement).

It is seen that admittance values although not matched to 50 Ohms, can be easily used for parameter extraction procedure.

3. ADMITTANCE CURVES AND COM PARAMETERS

The scattering parameter S11 or admittance of the device was measured and then fitted to a theoretical curve. The advantage of this device is that the resonance peaks (See Fig. 2a) appear on the admittance curve on the left and right side of the stopband, clearly showing the width and the center frequency of the stopband which is directly related to the coupling parameter K and the velocity *V*. In some cases additional transverse modes could be found. In the above example one such mode is clearly seen inside the stopband. The value of the admittance (which for wide apertures can be chosen around 50 Ohm) determines the parameter of interaction α and , finally , the relative height and sharpness of the resonances at the frequencies lower than the stopband gives the value of attenuation. The parasitic resistance *Rp* and static capacitance determine the pedestal levels and general behavior of the real and imaginary part of the admittance curves. These values can also be determined. So, using the measurement of only one frequency dependent characteristic of the device all necessary parameters for COM modeling of SAW devices can be obtained.
The admittance of a one port synchronous resonator was simulated using COM model:

$$Adm = P_{33} - 4 * P_{13}^{\ 2} * \frac{Rr}{1 - Rr * (Tt + Rt)} \qquad (1)$$

where *Pij* stands for *P* - matrix components of the IDT and *Rr, Tr* and *Rt, Tt* are reflection and transmission coefficients of the reflector and the transducer. These values are calculated using well known COM formulas.
In the COM model we have 4 *a priori* unknown parameters. They are the velocity *V* in the centre of the stopband , K (kappa) - the reflectivity per wavelength of the electrode structure, electo-mechanical transduction coefficient α (alpha), describing the strength of the piezoelectic coupling of the waves to electric fields of the IDT and the wave attenuation coefficient γ (gamma; the wave amplitude decays as $\exp(-\gamma x / \lambda)$). The accuracy of theoretical estimates of these parameters is usually not sufficient for practice. Two other parameters are the static capacitance and the parasitic resistance of the electrodes and the busbars. These parameters often also play very important roles as far as the passband width and attenuation is concerned.

For a systematic extraction of parameters for different central frequencies, apertures and electrode-to-period ratios software was developed which allows the processing of a great number of measured files using batch methods. In the fitting procedure, the measured values of both the real and imaginary parts of the admittance are compared with the simulated ones according formula (1) . The sum of squares of mismatches over typically 1600 frequency points is minimized by an optimization routine. The routine we used is quite slow but robust. It changes all 6 parameters with appropriate scaling of the steps and then moves to the point where the smallest value of the mismatch function is found. Then the cycle is repeated with gradually decreasing steps. Typically 5 to 10 minutes time is necessary for one fitting.

4. RESULTS AND DISCUSSION

Fig. 3 illustrates results obtained using the bove technique for the dependence of the electrode reflectivity on the cross section parameter (proportional to the product of the electrode thickness and m/p ratio) for Rayleigh SAW on 3λ- quartz. The velocity dependence on the same parameter is shown on Fig. 4. The results obtained with this technique are quite

accurate for all fitted parameters (except the attenuation) and display such fine effects as the dependence of the reflection coefficient on aperture or the dependence of velocity on the aperture (wave guide effect). Of course, stronger effects like α and *Cst* dependence on aperture , *m/p* ratio and thickness of *Al* electrodes are easily detected. For thick electrodes, second transverse modes can be seen on the admittance curves, but these modes were ignored in the fitting procedure.

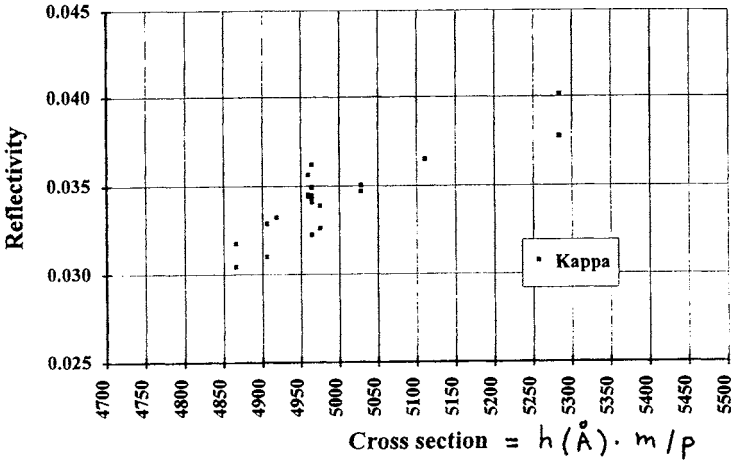


Fig. 3 . Kappa (reflectivity per wavelength) dependence on $h/\lambda * m/p$.

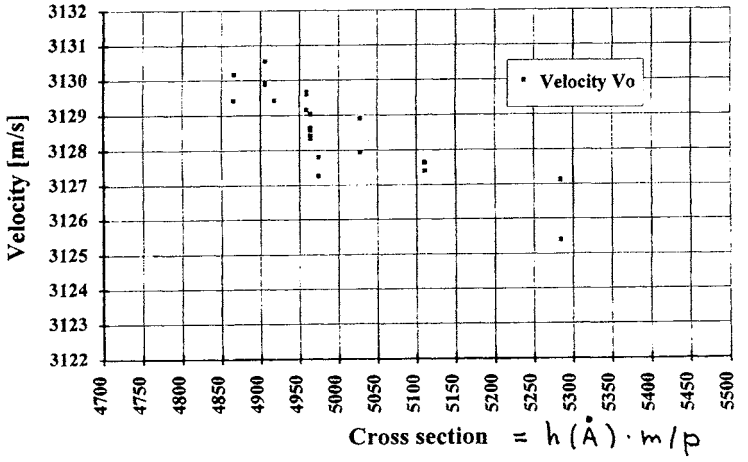


Fig. 4 Velocity of SAW on 34 degree rotated cut of quartz as a function of $h/\lambda * m/p$.

As to attenuation we get the value of about $\gamma \approx 10E\text{-}4$, which must be considered as an estimate, because the strong influence of the attenuation on the Q- factors of the resonance peaks is observed only if the length of the reflectors is of the order of the attenuation length. Other parameters (alpha, static capacitance and parasitic resistance) are strongly dependent on the aperture and layout details and are not presented here.
This test structure is often small compared to actual SAW devices and can be put on the same mask for routine checking of the parameters of the wave.

5. ACKNOWLEDGMENTS

The authors are grateful to C. S. Hartmann for many fruitful discussions during which the idea of the technique was generated and the design of the testbar was reviewed and to J. Heighway for help and attention.

6.REFERENCES

[1] C. S. Hartmann and V. P. Plessky, "Experimental measurements of propagation, attenuation, reflection and scattering of leaky waves in Al electrode gratings on 41- 52- and 64-LiNbO3", 1993 Ultrasonics Symposium Proc., pp. 1247-1250.

[2] V. P. Plessky and C. S. Hartmann, "Characteristics of leaky SAWs on 36-LiTaO3 in periodic structures of heavy electrodes", 1993 Ultrasonics Symposium Proc., pp. 1239-1242.

[3] C. S. Hartmann, D. P. Chen, and J. Heighway, Experimental determination of COM parameters for SAW transversely coupled resonator filters, 1992 IEEE Ultrasonics Symp. Proc., pp. 211-214.

ANALYSIS AND DESIGN OF A VERY LOW LOSS SAW RESONATOR FILTER ON 36° LITHIUM TANTALATE

M. SOLAL, P. VENTURA
THOMSON MICROSONICS, 399 rte des crêtes, 06903 Sophia Antipolis, FRANCE

ABSTRACT

36 ° lithium tantalate is a very widely used material for SAW filters. It provides a high coupling coefficient as well as a relatively low temperature dependence. Unfortunately, the wave propagation is not a true SAW but a P-SAW (Pseudo-SAW) . The propagation losses due to the leaky nature of the wave and the existence of close SSBW's have to be taken into account in the filter analysis.
In this paper, a model of P-SAW propagation based on harmonic admittance is described. This model was used to design a very low loss filter at 184 MHz using a ladder type structure. The result is a 2 dB insertion loss for a 1.8 MHz bandwidth and a 40 dB far out rejection. Analysis and experiment are in very good agreement on both the exact value of the IL and the shape of the transfer function.

1. INTRODUCTION

Classical SAW filters structures use the propagation between two transducers to generate the desired transfer function. When a low insertion loss is specified, this leads to unidirectional transducer structures or double mode resonator filters [3,4]. A new approach [1,2] consists of using SAW transducers only for their impedance characteristics. This technique is similar to the one used for bulk wave or ceramic devices and allows for very low insertion losses.
We applied it in the design of a 1% relative bandwidth low loss filters at 184 MHz on 36° rotated Lithium Tantalate. One key point for this kind of filter is the prediction of

insertion loss. To obtain accurate results, the leaky nature of the wave and the presence of close SSBWs have to be taken into account.

A harmonic admittance model [5] was used to determine the wave propagation characteristic parameters. These parameters were introduced in the filter analysis and allowed us to predict the insertion loss.

Two filters were designed. A 4 section 1.8 MHz bandwidth filter exhibiting a 2 dB insertion loss for a 40 dB far out rejection and an 8 section filter exhibiting a 3.2 dB insertion loss for a 80 dB far out rejection and a 45 dB close in rejection and a 1.5 MHz bandwidth.

2 .FILTER STRUCTURE

Figure 1 shows the SAW ladder filter structure. The basic filter basic is made by two resonators having different resonant frequencies. The parallel resonator is designed for zero admittance at the center frequency while the serial resonator has zero impedance at center frequency. Under these conditions and for infinite Q resonators, the filter is equivalent to a short circuit at center frequency so that there is no insertion loss. Furthermore, there is a zero in transmission for the frequencies where the parallel resonator has a zero impedance (respectively the serial resonator has zero admittance). Outside the stop band, the acoustical part of the resonators impedances are neglectible and only the transducers capacitances have to be considered. The rejection level is almost constant and depends only on the ratio of the two static capacitances.

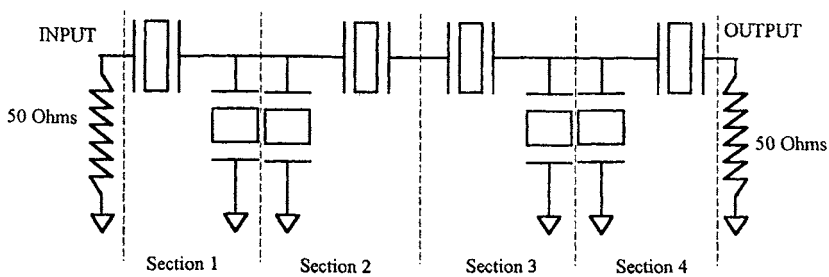


FIGURE 1 : FOUR SECTIONS FILTER STRUCTURE

3. RESONATOR ANALYSIS

To analyse a resonator using Pseudo SAW, it is necessary to take into account the leaky nature of the wave and the presence of losses due to the coupling with SSBW. We used the harmonic admittance model [5] to analyse the wave propagation under an infinite periodic grating.

This model allows us to compute the dispersion curve and the coupling between electrodes due to the P-SAW. We analyse electrical effects while mechanical effects due to the electrode thickness are deduced from experimental results. From the dispersion curve which gives the eigen mode phase variations with frequency, it is possible to obtain the propagation characteristics of the coupled modes which are used in our SAW analysis software. Althrough true SAWs can be described properly with a linear phase variation, this is not possible for P-SAW and we had to define a model which takes into account coupling with SSBWs. We define the phase φ for one period by the expression :

$$\varphi(f) = \varphi_r(f) - j\sum_{i=0}^{n}\varphi_i(f)$$

where $\varphi_r(f)$ is linear with frequency

$\varphi_i(f)$ is the attenuation due to the bulk wave i (i=0 corresponds to the attenuation due to the leaky nature of the P SAW)

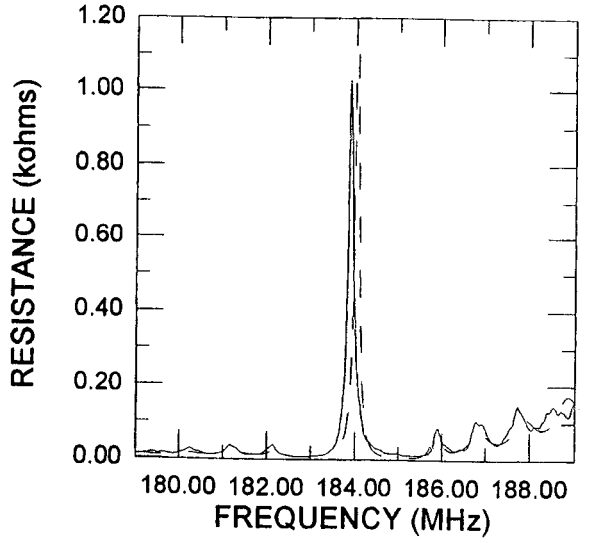
The attenuation frequency variation $\varphi_i(f)$ and the reflection coefficient for one period are determined by a polynomial type fit on the dispersion curve and are used in our analysis software. Figure 2 shows an example of comparison between the simulated and measured admittances of a 1 port SAW resonator. These results show a good agreement between measurement and simulations specially for the resonator Q. The value of Q was overestimated by our previous models.

4. FILTER DESIGN

We used a ladder structure to design a 184 MHz filter. The requirements were a 3 dB bandwidth larger than 1.5 MHz, a -20 dB bandwidth of 4 MHZ and a 40 dB far out rejection. The insertion loss had to be lower than 2 dB.

Since resonators characteristics have a direct impact on the filter transfer function, the main design issue is the resonators definition.

Figure 3 shows a 1 port SAW resonator. This resonator is designed by inserting a transducer between two reflective gratings.



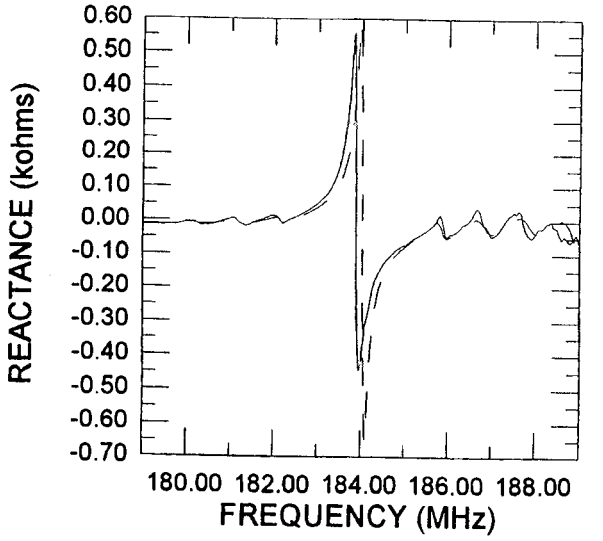


FIGURE 2 : 1 POLE RESONATOR IMPEDANCE
(solid line :measurement, dashed line simulation)

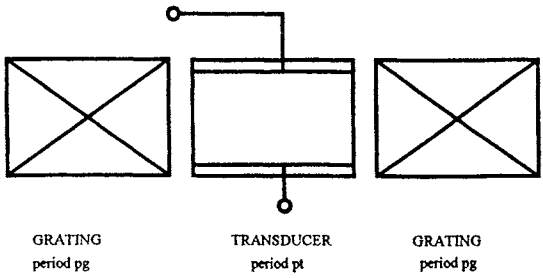
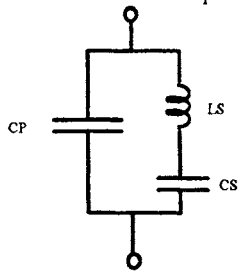


FIGURE 3 : STRUCTURE OF A SAW 1 POLE RESONATOR

The grating period pg is chosen for both resonators to have the filter center frequency at the stop band center frequency. The transducer period pt is used to adjust the resonant frequency (or the antiresonant frequency) to the filter center frequency.
The equivalent circuit of a 1 pole resonator in the stop band is shown below.



The parallel capacitance CP is the transducer capacitance and depends on the number of active periods and on the acoustic aperture. The inductance LS depends on the material coupling coefficient and the cavity length.

- the serial resonator resonance frequency and the parallel resonator antiresonance frequencies have to be adjusted to the filter center frequency.
- the serial resonator antiresonance frequency and the parallele resonator resonance frequencies will give two transmission zeros to the filter transfer function.

Figure 3 shows the impedances and admittance of the two resonators. This figure shows that the parallel resonator has a resonance at 182.5 MHz while the serial resonator has an antiresonance at 185.5 MHz. So, we will obtain zeros on the filter transfer function for these frequencies. The 3 MHz separation between zeros is coherent with the required 4 MHz -20 dB bandwidth.

Outside the stopband, the resonators are reduced to their parallel capacitances CP and the rejection level depends primarly on the ratio between the parallel capacitance of the parallel resonator on the parallel capacitance of the serial resonator. Unfortunately, as shown by [1], increasing the rejection is paid for with a degradation of the filter insertion loss and bandwidth and a trade off has to be made to define the resonator acoustic apertures.

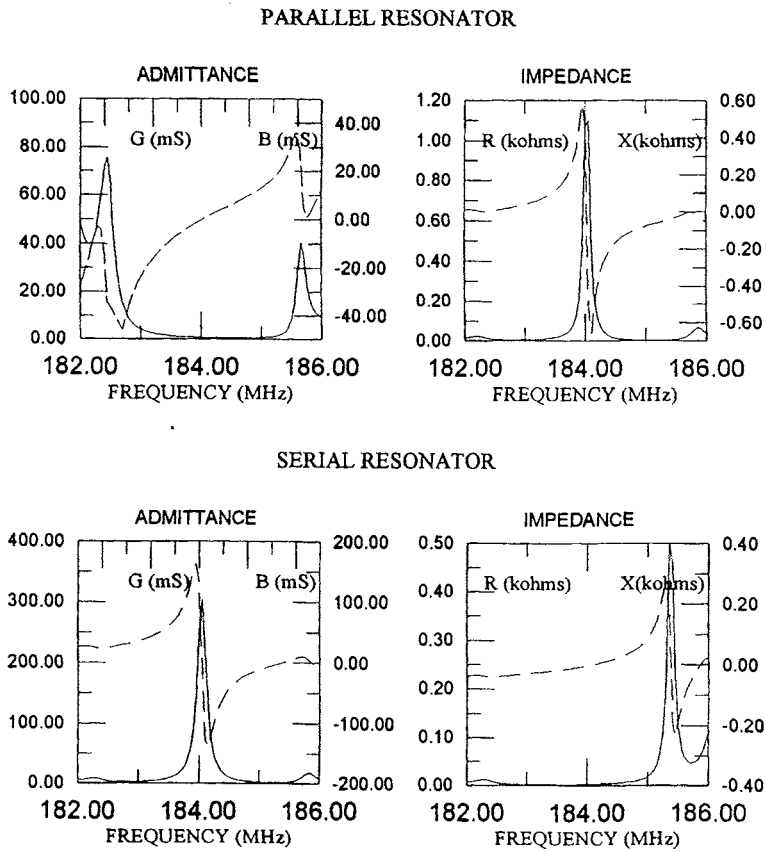


FIGURE 3 : PARALLEL AND SERIAL RESONATOR CHARACTERISTICS

For the required specifications, the bandwidth was not critical and, due to the low center frequency it was found possible to keep low insertion loss even with a high capacitance ratio. So we used a ratio between capacitances of 2 and a 4 sections structure was chosen to obtain a sufficent far out rejection.

5. EXPERIMENTAL RESULTS

To avoid the effect of stray elements on the transfer function, we decided in a first step to wafer probe the individual resonators. The measurements were used to simulate the complete filter. Figure 4 shows the comparison between the simulated transfer function and the expected transfer function obtained from resonators measured impedances. The fit is good as well for the bandshape as for the insertion which is predicted to within less than 0.1 dB. There is a frequency shift of about 50 kHz between the two transfer functions.

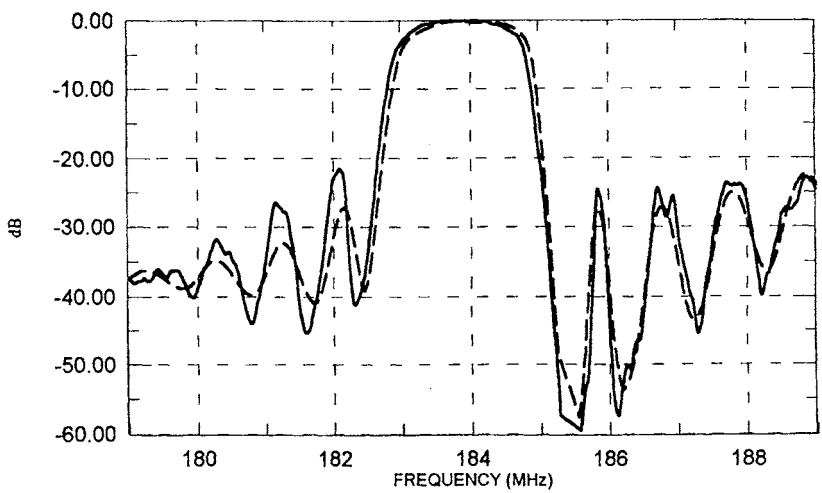


FIGURE 4 :EXPECTED FILTER TRANSFER FUNCTION

full line : from simulated resonator (reference loss = 1.853 dB)

dashed line : from resonator measurements (reference loss = 1.928 dB)

Figures 7 and 8 shows the transfer function of the mounted filter measured on 50 Ω. The 3 dB bandwidth is 1.9 Mhz, the insertion loss is 1.8 dB while the far out rejection is more than 40 dB as required. The differences between these results and those of fig. 4

are mainly due to errors in evaluation of stray capacitances and wire bonding inductances which has an impact on filter characteristics.
Figures 9 and 10 show the transfer function for an eight section filter designed using the same resonators. The 3 dB bandwidth is 1.5 MHz while the close in rejection is about 45 dB and the far out rejection is higher than 80 dB. The insertion remains no more than 3.2 dB.

6. CONCLUSION

Ladder structures are very attractive for designing SAW low loss filters. They provide very low losses without any tuning element. Althrough this kind of structure was primarly used for high frequency low rejection filters, it allows higher rejections for lower frequencies. A four section filter with 1.8 dB insertion loss and 40 dB far out rejection was designed. In addition, we demonstrated the feasibility of 8 section structures. A 80 dB rejection for 3 dB insertion loss was obtained.
We developed a P-SAW propagation model which allowed us to predict the insertion loss within 0.1 dB. Future work will tend to apply FEM-BEM results to filter analysis. This will improve the accuracy of metallization effects in the analysis.

6. REFERENCES

[1] O. Ikata et al., "Development of low-loss band pass filters using SAW resonators for portable telephones", 1992 IEEE U.S. proc., pp 111-115

[2] M. Hikita et al., "Design methodology and synthesis techniques for ladder-type SAW resonator coupled filters", 1993 IEEE U.S. proc., pp 15-24

[3] T. Morita et al., "Wideband low loss double mode SAW filters", 1992 IEEE U.S. proc., pp 95-104

[4] P. Ventura, M. Solal, P. Dufilie and S. Chamaly, "A Global Optimization procedure for SPUDT filters", 1993 IEEE U.S. proc., pp 5-8

[5] Y. Zhang, J. Desbois, L. Boyer, "Characteristic Parameters of Surface Acoustic Wave in periodic metal gratin on a piezoelectric substrate", IEEE trans. U.F.F.C.,vol. 40, May 1993

[6] P. Ventura, J. Desbois, L. Boyer, "A mixed FEM/analytical model of the electrode mechanical perturbations for SAW and PSAW propagation", 1993 IEEE U.S. proc., pp 205-208

[7] P. Ventura,"2D analysis of SAW and PSAW periodic transducers using FEM and BEM", 1994 EFTF proc.

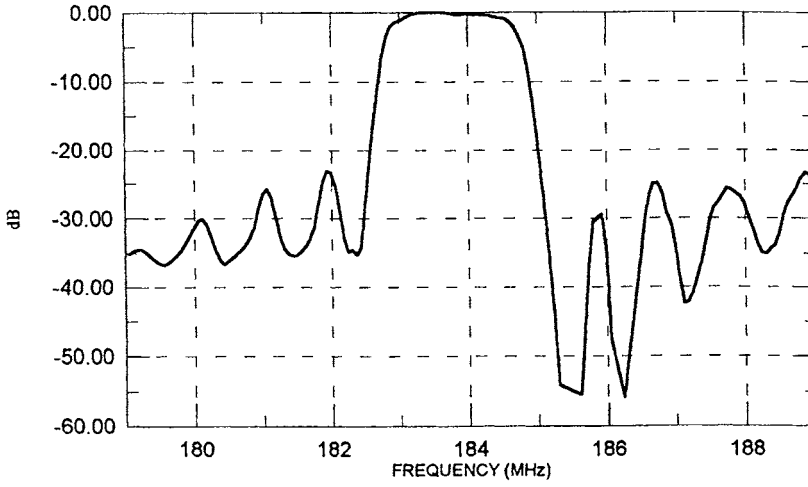


FIGURE 7 : FOUR SECTION FILTER MEASURED TRANSFER FUNCTION

(reference loss= 1.78 dB)

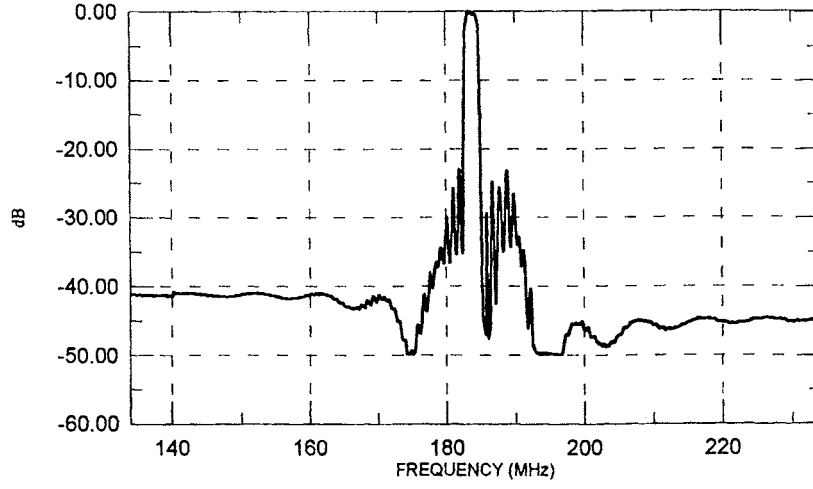


FIGURE 7 : FOUR SECTION FILTER MEASURED TRANSFER FUNCTION

(reference loss= 1.78 dB)

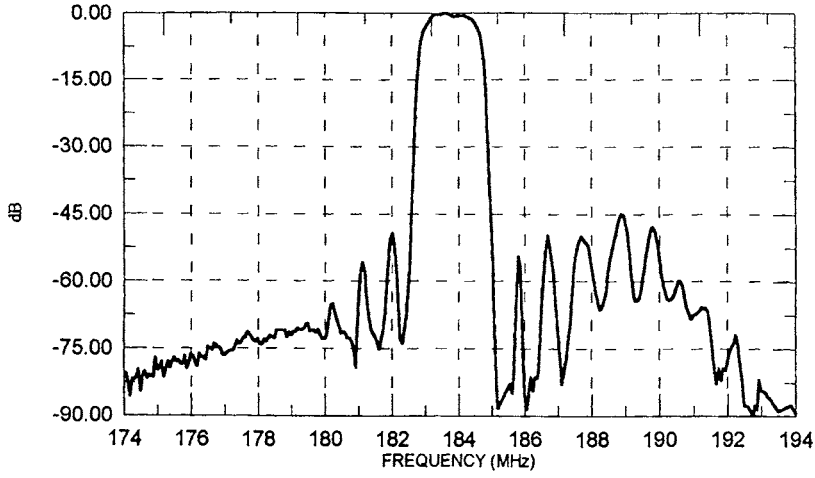


FIGURE 8 : EIGHT SECTION FILTER MEASURED TRANSFER FUNCTION

reference loss= 3.2 dB

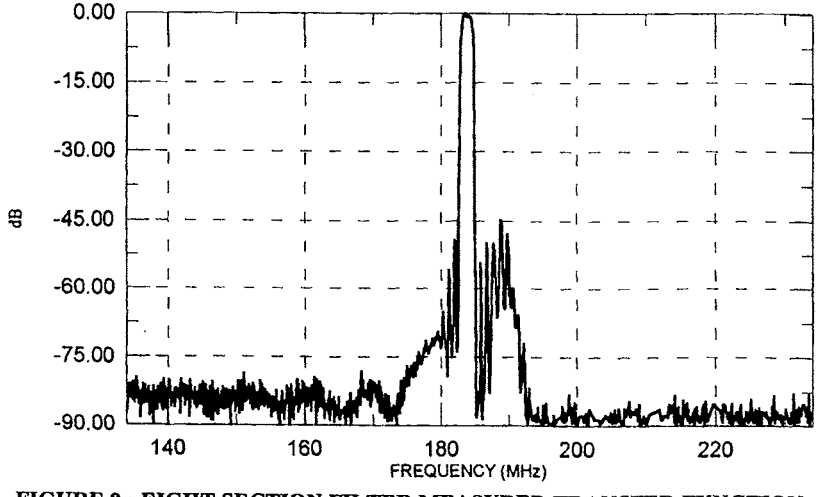


FIGURE 9 : EIGHT SECTION FILTER MEASURED TRANSFER FUNCTION

reference loss= 3.2 dB

Interrogation Signals in Atomic Frequency Standards (A6)

Intermodulation Signals in Direct Digital Frequency Synthesizers

by

Věnceslav F. Kroupa

Summary: Generally each DDFS exhibits a lot of spurious signals, in the output pass band, all of which are harmonically related to the smallest frequency step Δf which is equal to the clock frequency divided by the largest accumulator content, i.e. $f_c/2^R$. Some of these spurious signals are more prominent than the rest. To this particular group also belong intermodulation signals between the output frequency and the clock frequency. In the following we shall discuss the problem in more detail.

1.0 Introduction.

All frequency synthesizers are plagued with undesired signals accompanying the desired output signal. Very troublesome, particularly known from the theory of "Direct Analog Synthesizers" /1,2/, are the so called intermodulation signals.They are generally defined as multiples of input frequencies generated and mixed in a nonlinear circuit to a frequency which may pass the output filter,i.e.

$$r*f_1 + s*f_2 = f_p \qquad (1)$$

where r and s are integers (either positive or negative) and f_p is the spurious frequeny in the output filter pass band. Experience has taught us the higher the intermodulation product is (e.i. the larger the integers r and s are) the smaller the respective intermodulation signal is. The above equation can be solved in different ways.

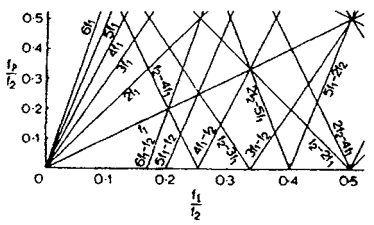


Fig.1 The "straight line" intermodulation diagram (adapted from /1/).

Very instructive is the assistance of the "straight line" intermodulation diagram reprinted in Fig.1 from /1/. Note that there also exist computer solutions /e.g. 3/. One very simple and particularly suited for investigation of **DDFS** is given in appendix **7.1.**

Finally we have to note that often there are more than two signals at the input of the mixer, however, the solution in these instances is not a simple one /4/.

Since a **DDFS** is a highly nonlinear device one is tempted to conclude that all spurious signals are intermodulation products of the type **(1)** /5,6/. However, pursuing this idea we arrive at very confusing results and as a consequence we have to start a more exact investigation of all major spurious signals.

2.0 Principles of DDFS.

The task of **DDFS** is to generate a desired output frequency f_x from the input standard or clock frequency f_i with the assistance of **Modulo-N** arithmetics. The simplified block diagram of the system is shown in Fig. 2. The governing equation of this system is

$$f_x = \frac{k}{N} * f_i \qquad (2)$$

where the denominator **N** is generally a power of two, i.e.

$$N = 2^R \qquad (3)$$

For examination of spurious signals we must start from the normalized frequency

$$\xi_x = \frac{f_x}{f_i} = \frac{k}{2^R} = \frac{X}{Y} \qquad (4)$$

where R is number of bits stored in the accumulator and X and Y are mutually prime integers, as a consequence X is always odd and Y a power of two as high as R (e.g. R=24, 32, etc.) or often even lower, some R_{eff}, in the following we shall refer to it as to R.

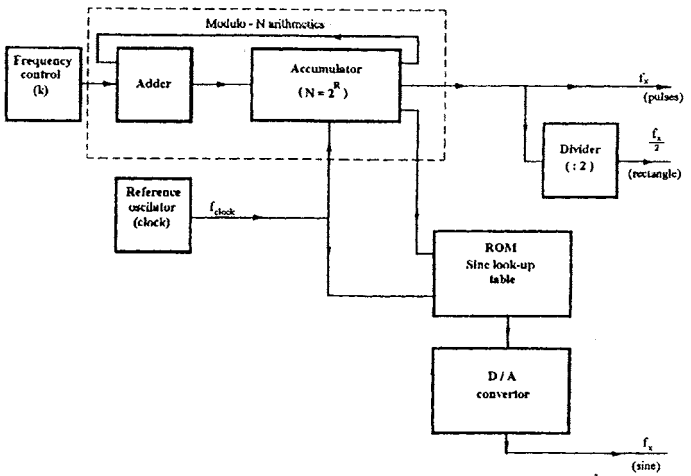


Fig.2 The simplified block diagram of DDFS based on Modulo-N arithmetics.

3.0 Spurious signals in DDFS.

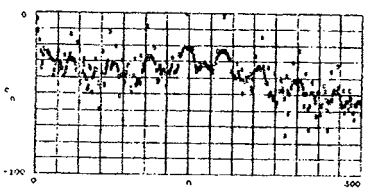


Fig.3 Spurious signals generated in PAL single frequency rectangular wave DDFS;it was used the 7th order approximation $\xi_{x,7}$ = 137/309 to the ideal ξ_o = 2216809/5*10^6.

Pulse and rectangular output waves are plagued with very large spurious signals /7,8,9/ and are used only in exceptional instances. See e.g. Fig. 3 for spurious signals generated in a PAL single frequency DDFS. However, generally we use sine wave outputs which exhibit much lower level of

spurious signals. Their sources are shown in the simplified block diagram in Fig. 4.

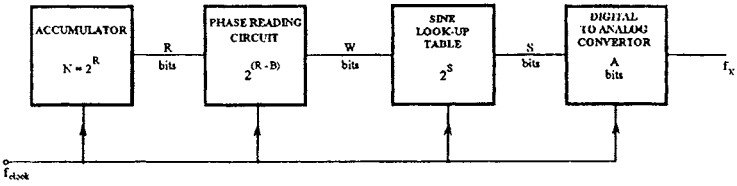


Fig.4 The simplified block diagram of sources of spurious signals in DDFS due to the loss of information in different building blocks.

3.1 Truncation of the phase bits - basic relations.

In accordance with Fig. 3 we use from the **R** phase bits stored in the accumulator only **W** bit for steering the sine look up table. The problem was first discussed by Mehrgardt /10/ and later solved by Nicholas and Samueli /11/. However, it was the author /8,9/ who presented a very simple closed form relation for the level of the spurious signals of this origin. We shall repeat,here the results in short. In instance where we neglect B phase bits, i.e.

$$B = R - W \tag{5}$$

the phase readings are still performed at each $\mathbf{mT_c}$ ($\mathbf{m} = 0,1,\ldots,Y-1$) . As a consequence amplitudes of individual samples (rectangles) in the output sine wave change into

$$\sin\left[2\pi\frac{2^B}{Y}integer\left(\frac{mX}{2^B}\right)\right] = \sin\left\{2\pi\frac{2^B}{Y}\left[\frac{mX}{2^B} - s(m)\right]\right\} \tag{6}$$

where

$$s(m) = \frac{mX}{2^B} - integer\left(\frac{mX}{2^B}\right) \leq 1 \tag{7}$$

evidently

$$\frac{2^B}{Y} s(m) \ll 1 \qquad (8)$$

consequently eq. (6) can be expanded into

$$\sin\frac{2\pi mX}{Y} - 2\pi\frac{2^B}{Y} s(m)\cos\frac{2\pi mX}{Y} \qquad (9)$$

For the evaluation of the noise spectra introduced by the phase error **s(m)** we shall use the theory of the "quantization noise spectra" /12/ namely that for the **m-th** error we can write the Fourier expansion as

$$e_m = \sum_l \frac{1}{2\pi jl} * e^{2\pi jl * s(m)} \qquad (l \neq 0) \qquad (10)$$

After introducing the remainder R_b for **s(m=1)**

$$s(m=1) = \frac{X}{2^B} - integer\left(\frac{X}{2^B}\right) = R_B \qquad (11)$$

and neglecting the DC part in (10) and further taking into account that **s(m)** is a periodic function we arrive with the assistance of (10) at the sum of the sine waves, i.e.

$$s(m) \doteq \frac{1}{\pi} \sum_{r=1,2,..}^{2^{(B-1)}} \frac{1}{r}\sin(2\pi rmR_b) \qquad (12)$$

Note that we can compute the remainder R_b from the normalized frequency as

$$R_b = \xi_x * 2^W - integer(\xi_x * 2^W) \qquad (13)$$

In instances where $R_b > 0.5$ we may refer without any loss of generality to $1-R_b$, i.e. to refer only to the range

$$0 \leq |R_b| = \frac{P}{Q} \leq \frac{1}{2} \qquad (Q = 2^B) \qquad (14)$$

where **P** and **Q** are integers. This simplifies location of spurious signals.

After introducing the r-th harmonic from (12) into (9) we can expand the 2-nd term into

$$\frac{2^B}{rY}\{\sin[2\pi m(\frac{rP}{Q}+\frac{X}{Y})]+\sin[2\pi m(\frac{rP}{Q}-\frac{X}{Y})]\} \qquad (15)$$

Evidently both terms of the above expression are staircase values of spurious sine waves around the carrier. Consequently we can compute their amplitudes with the assistance of FFT and we arrive at /8,9/

$$C_{sr}=\frac{2^B}{rY}\frac{\sin[\pi(\frac{X}{Y}\pm\frac{rP}{Q})]}{\pi(\frac{X}{Y}\pm\frac{rP}{Q})} \qquad (16)$$

After dividing the above relation by that for the carrier, i.e. C_X, we get for the r-th order spurious signal level

$$\frac{C_{sr}}{C_X}=\frac{2^B}{rY}\frac{X}{X/Y\pm rP/Q}\frac{\sin\pi(X/Y\pm rP/Q)}{\sin(\pi\frac{X}{Y})} \qquad (17)$$

$$\doteq\frac{2^B}{rY}=\frac{2^{-R+B}}{r}=\frac{2^{-W}}{r}$$

or in dB measure

$$\frac{C_{rs}}{C_X}\doteq -6W-20\log(r)\qquad [dB] \qquad (18)$$

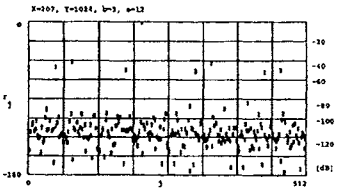


Fig.5 Spurious signals in a DDFS with the normalized frequency ξ_X=207/1024,W=7,S=12.

As an example we have computed with the assistance of FFT the spectrum of a normalized frequency $\xi_X = 207/10^{10}$ with 3 **LSB** neglected , i.e. **B=3** and **W=7**, and plotted it in Fig. 5.

The actual and estimated values of spurious signals in the above example are given in Tab. 1. Note, that the largest spurious signal, for r=1, is about -42 dB in accordance with the relation **(18)**.

Spurious spectral line numbers meet condition

$$n_r = X + (r\frac{PY}{2^B} + sY) < Y/2 \qquad (s = \ldots,-2,-1,0,1,2,\ldots) \qquad (19)$$

Investigation of the above inequality reveals different n_r only for $r < 2^B/2$ and a double one for $r=2^B$. After taking into account that there are two side bands around the carrier for each r (i.e. both positive and negative) their number is, inclusive the carrier, 2^B.

harmonics s(m)	spurious lines	spurious frequency [MHz]	spurious measured [dB]	spurious estimated [dB]
carrier	207	10.10	0	0
1st	335 79	16.36 3.86	-42.88 -41.48	-42.14
2nd	463 49	22.61 2.39	-49.78 -46.72	-48.16
3rd	433 177	21.14 8.64	-51.70 -49.43	-51.69
4th	305	14.89	-50.98	-54.19+3

Tab. 1. Spurious signals caused by the phase truncation of the normalized frequency $\xi_x \approx 207/10^{10}$ with 3 LSB neglected, i.e. B = 3 and W = 7, with f_c = 50 MHz.

3.2 General relation for computation of spurious frequencies due to the phase bit truncation.

Generally the denominators in the normalized frequencies in accordance with **(4)** are too large to compute spurious frequencies with the assistance of **(16)**. However, the problem can approximately be solved by introducing into relation **(18)** the remainder R_b, by replacing the normalized frequencies

with actual ones, and by assuming both signs for r and s

$$f_{sr} = \frac{n_r}{Y} f_c = f_x + (r*R_b + s) f_c \qquad (20)$$

For demonstration we shall use an actual published example /6/, where

f_x = 131.77 kHz, f_c = 1 MHz, R = 32, W = 14, D = 12

from **(13)** we get for the remainder

R_b = 0.91968

or in accordance with **(14)** we use the value

1 - (R_b) = 0.08032

With the assistance of relation **(20)** we get for the largest spurious signal

$f_{s,1}$ = 51.450 kHz and also $f_{s,-1}$ = 212.09 kHz

and for their level -6*W = - 84 dB

Actual measurement revealed for the $f_{s,1}$ the level -83 dB.

harmonics of **s(m)**	spurious frequency [kHz]	estimated level [dB]
carrier	131.77	0
1st	51.45 212.09	-84.29
2nd	28.87 292.41	-90.31
3rd	109.19 372.73	-93.83
4th	189.51 453.05	-96.33
$3*f_x$	395.31	-79.70
$f_c - 5*f_x$	341.15	-84.13

Tab. 2. Estimated levels of spurious signals for **DDFS** with f_c = 1 MHz, f_x = 131.77 kHz, R = 32, W = 14, and D =12.

Frequency (Hz)	Level (dBc)	Origin
263540	-72.5	$2Fs$
395310	-74.1	$3Fs$
472920	-78.2	Fc-$4Fs$
30710	-78.4	$23Fs$-$3Fc$
341150	-79.2	Fc-$5Fs$
449470	-79.7	$11Fs$-Fc
418760	-80.1	$2Fc$-$12Fs$
51450	-83.0	$(2^{14}+1)Fs$-$2159Fc$

Tab 3.

In Tab. 2 we have summarized few larger spurious signals with their estimated level. Comparison with the Tab. 3, reprinted from /6/, reveals other large spurious signals due to the intermodulation effects between f_x and f_c. We shall revert to the problem in the later section 5.

The level of the smallest spurious signals due to the phase truncation would be

$$-6*14 - 20*\log(2^{(32-14)}) = -182 \text{ dB}$$

evidently safely buried even in the background thermal noise.

4.0 Truncation of sine words in look-up tables.

Investigation of staircase sine wave with ideal amplitudes, i.e.

$$\sin(2\pi m\frac{X}{Y}) \qquad (21)$$

would reveal no spurious components in the output signal. However, the economy of actual DDFS requires that the information about the sine values in look-up tables is generally stored only as S-bits long words. The consequence is that besides those high level spurious signals caused by the phase truncation there is a large set of much smaller spurious signals, the level of which depends of the number of bits used in look-up table words. Consequently, truncation of stored sine values, to words of S bits long, results in changing the above equation into

$$\frac{1}{2^S} integer[2^S \sin(2\pi m\frac{X}{Y})] \qquad (22)$$

or even into

$$\frac{1}{2^S} integer\{2^S \sin[2\pi \frac{2^B}{Y} integer(\frac{mX}{2^B})]\} \qquad (23)$$

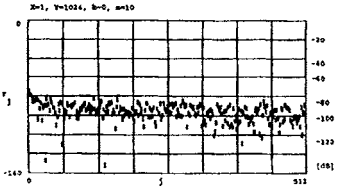


Fig.6 spurious signals in a DDFs with the normalized frequency, ξ_x=207/1024, W=0, S=10.

With the assistance of the FFT algorithm we can compute all major spurious signals caused by the phase truncation and also the "background" spurious signals due to the finite word length in the sine look-up tables. Some examples are shown in Figs. 5 and 6. However, in instances where Y is too large, what is generally the case, the above approach is not feasible. The only possibility is to estimate the level of spurious signals caused by sine word truncation.

4.1 Estimation of background spurious signal level - due to sine word truncation.

At the first sight at Figs. 5 and 6 one has a feeling that amplitudes of these background spurious signals are randomly distributed. However, this cannot be true since we have evaluated a periodic system. To get more insight we shall compute and plot the error signal

$$\epsilon(m) = \sin(2\pi m \frac{X}{Y}) - \frac{1}{2^S} integer\{2^S \sin[2\pi \frac{2^B}{Y} integer(\frac{mX}{Y})]\} \qquad (24)$$

From the above relation we easily find out that

$$|\epsilon(m)_{max}| < 1/2^S \qquad (25)$$

and a plot in Fig. 7 confirms the above inequality. Its close investigation reveals an odd symmetry and even other periodicities, i.e. the process $\epsilon(m)$ is periodically

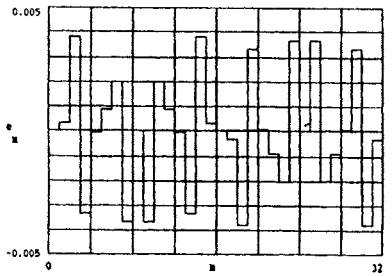


Fig.7 Plot of the eroor $\epsilon(m)$ $\xi_x = 7/32$, S= 8.

stationary with zero mean and a discrete spectrum. By computation of the power spectral density $S(n_r)$ we shall proceed in accordance with Papoulis /13/ (see Appendix 7.2) we and arrive at the mean power level of one sided spectral lines as

$$S(n_r) \doteq \frac{2}{Y}\sigma_\epsilon^2 \qquad (26)$$

By assuming that $\epsilon(m)$ has a uniform distribution in the interval from -1/2s to 1/2s we get for variance of $\epsilon(m)$

$$\sigma_\epsilon^2 \doteq (1/3)*2^{-2S} \qquad (27)$$

After computing actual variances

$$\sigma_\epsilon^2 = \sum_{k=1}^{Y} \frac{\epsilon_k^2 * T_i}{Y * T_i} \qquad (28)$$

and comparing the results with values arrived at with the assistance of (27) we have found quite small differences and concluded that we are justified to use relation (26) in combination with (27) in our pursuit of the estimation of the "background" spurious signal levels /14/.

4.2 Spurious signals due to the finite resolution of DAC convertors.

There are three major sources causing degradation of the output spectral purity by DAC namely, finite bit resolution, nonlinearity, and transient effects or glitches. Here, we shall discuss only the first one.

From the point of view of the present technology we are not able to build DAC's with any arbitrary resolution on one

hand. On the other hand the higher the resolution the more expensive **DAC's** are.

For discussion of the **DAC** bit resolution effect we must revert to the relation **(24)** and take into account that from the **S** bits long sine word we use only **A** bits for conversion the digital information about the sine wave amplitude into its analog version. Note that in instances where a D bit **DAC** is used in **2's** complement version i.e.

$$A = D - 1 \qquad (29)$$

the output spectrum can be arrived at if the sine word length **S** in **(24)** is replaced by the bit resolution **A** (cf. paper by Garvey et al./15/). The conclusion is that the extent of the sine look-up table and the bit resolution of **DAC** are closely related in properly designed **DDFS**.

4.3 Background level of spurious signals.

After combination of eqs. **(26)** and **(27)** with conclusions of the above section we get for the background level of spurious signals

$$S(n_r) \doteq -1.76 - 6.02A - 10\log(Y) \qquad (30)$$

However, in instances where the denominator **Y** is large, particularly if the smallest frequency step **Δf** is much smaller than **1 Hz**. i.e. if

$$\Delta f_x \ll \frac{f_c}{Y} \qquad (31)$$

then the number of spurious spectral lines in 1 Hz bandwidth is (note that all lines with **r** even are missing)

$$\frac{1}{2\Delta f_x} = \frac{Y}{2f_c} \qquad (32)$$

and the discrete background spectrum changes into effectively a continuous and white one

$$S(f) \doteq -6.02*A - 10*\log(f_c) \quad dB \qquad (33)$$

By reverting to the example in section 3.2 we get that

$$S(f)_{lim} \doteq -126 \text{ dB}$$

which is a much higher level than the thermal noise of the clock oscillator.

4.4 "Signal to noise ratio" in DDFS.

In technical data about **DDFS** we generally find the statement that a perfect **DAC** generating a full scale sine wave will have a rms signal to rms noise ratio (over the full Nyquist bandwidth by ignoring the Sin(x) rolloff), e.g. /16 and 17/

$$S/N = 1.76 + D6.02 \quad dB \qquad (34)$$

However, by assuming the sine wave amplitude to be 1 V we get with the assistance of the variance in **(27)** and after replacing the resolution **S** in accordance with **(29)** for S/N

$$S/N = 10*\log\left(\frac{1/2}{1/3*2^{-2A}}\right) = 1.76 + 6.02A \quad dB \qquad (35)$$

The difference of 6 dB is caused by application of the **ADC** theory where the error $\epsilon(m)_{max}$ is assumed to be only one half of the **LSB**.

5.0 Intermodulation signals in DDFS.

There are several sources of intermodulation signals in **DDFS** . Even an ideal sine look up table (or an ideal **DAC**) is an intermodulation generator. Further, we have to take into account the nonlinearity of the **DAC** transfer characteristic and often also the glitch effect.

5.1.0 Ideal DAC as harmonic generator.

Let us revert to the eq. **(24)** and Fig. 7 and 8 and compute the mean value of $\epsilon(m)$ both for the positive and negative half period of the output sine wave. We get a rectangular wave with the peak to peak amplitude 2^{-S}. Its Fourier series expansion reveals only odd harmonics with magnitudes

$$a_h = \frac{2*2^{-S}}{\pi} * \frac{1}{h} \quad (h = 3,5,..) \qquad (36)$$

From which the level of the 3rd harmonic of the output frequency in dB measure is

$$20*\log\left(\frac{C_{s,3x}}{C_x}\right) \approx -6S - 14 \quad [dB] \qquad (37)$$

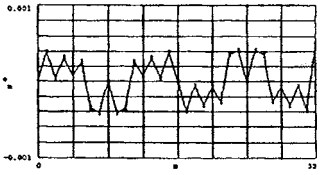


Fig.8 Plot of the error function $\epsilon(m)$ for the normalized frequency $\xi_x = 3/32$, A = 11.

Note that we have again neglected the **Sin(x)** effect of sampling on amplitudes. The above relation is in a good agreement both with many computer simulations (cf.Fig.5,6) and experimental measurements.

5.1.1 Aliasing of spurious harmonics.

In instances where the normalized frequency is very small we can detect several low order output frequency harmonics in accordance with **(36)** - c.f. Fig. 6 - till they merge with background spurious signals; note that some of them may be about 10 dB larger than mean values **(30)** or **(33)**.

However, generally the normalized frequencies are large and even low order harmonics of type **(36)** are returned as alias in agreement with the following equation

$$f_{s,h} = h*f_x - s*f_c \qquad (h,s = 0, \pm1, \pm2, \ldots) \qquad (38)$$

which recalls the basic intermodulation equation **(1)**.

However, amplitudes of these signals are governed by **(36)** and again only low order harmonics can be identified as stated above.

5.2 DAC as intermodulation generator.

Another possibility of generation of intermodulation signals of the type **(38)** is the leakage of the clock signal into **DAC** which acts as switching mixer /15/ controlled by the **MSB**. This behaviour explains the fact that actual intermodulation signals are often larger than those computed in accordance with the theory of higher harmonics considered in the preceding sections.

Investigations of real output spectra of **DDFS** reveal also even harmonics of the output signal and odd harmonics often larger than values expected from **(39)**.(c.f. Tab.2 and 3). The reason is the nonlinearity of the **DAC** transfer characteristics which are difficult to predict since they differ between individual **DAC**. However, they hardly exceed one **LSB** and as a consequence we can expect that the background spurious signal level will not deteriorate more than about 6 dB with respect to **(30)** or **(33)**.

5.3 DAC as glitch generator.

By switching a **DAC** converter we encounter some transient effects due to the charging or discharging of spurious capacities, to delays between individual switching commands, etc. Generally they are defined in **pAs** or **nVs**. It is difficult to estimate their influence, however, at lower output frequencies.They can be successfully removed with the assistance of sample-and-hold circuits /e.g. 6/. Otherwise they can generate unpredictable spurious or intermodulation signals.

6.0 Conclusions.

In the present paper we have shortly surveyed the sources of spurious signals generated in **DDFS's**. We have found that the most dangerous are side bands generated by the truncation of the phase bits stored in the device accumulator. The **-90 dB** level requires use of 15 **MSB** at lest. Though relations **(19)** and **(20)** resemble the basic intermodulation equation **(1)** the respective multiplication factors would be unrealistic high.

Next we have investigated the effect of the sine word and amplitude truncations which are generally performed in the **DAC**. It is usually stated that the spurious level is approximately **-6*D dB** where **D** is the resolution of the DA converter. However, relation **(30)** corrects the estimation to **-6*(D-1) dB** and the term dependent on the actual value of the denominator **Y**. Therefore designers take steps to rearrange the accumulator in such a way to keep $Y \doteq 2^R/6,11/$ or merely to change **X to X ± 1**.

Further we have shown that **DAC** is a harmonic generator and in connection with aliasing we arrive to a quasi intermodulation behaviour, however, of very low orders.

Actual intermodulation process is generated by switching properties of **DAC** and leaking clock signal. Again we encounter only low order products.

7.0 Appendices.

7.1 Programme for computation of intermodulation signals.

We reprint a programme written in MathCad for computation of intermodulation signals encountered in the example referred to in section **3.2**.

```
X := 131.77        Y := 1000

r := 0 ..11            s := 0 ..10

                Y
lim := -
                2
n     := r·X + (s - 5)·Y
 r,s

g     := if[|n     | ≤ lim,n     ,0]
      r,s   | r,s |        r,s
```

$$g = \begin{bmatrix}
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 131.77 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 263.54 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 395.31 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & -472.92 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & -341.15 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & -209.38 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & -77.61 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 54.16 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 185.93 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 317.7 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 449.47 & 0 & 0 & 0 & 0 & 0 & 0
\end{bmatrix}$$

7.2 Computation of the background power spectral density.

After replacing the variable **m** with t in relation **(24)** we can expand the random variable $\epsilon(t)$ as

$$\epsilon(t) = \sum_{-\infty}^{\infty} a_n e^{jn\omega_o T_i} \qquad \omega_o = 2\pi/YT_i \tag{39}$$

Evidently the n-th spectral line of the k-th rectangle $\epsilon_k(t)$ is

$$\begin{aligned} a_n &= \frac{1}{YT_i}\int_{kT_i-T_i/2}^{kT_i+T_i/2} \epsilon(t)\, e^{-jn\omega_o t}dt \\ &= \frac{1}{YT_i} * \epsilon(kT_i) * e^{-jn\omega_o kT_i} \frac{\sin(n\omega_o T_i/2)}{n\omega_o/2} \end{aligned} \tag{40}$$

from which the varience is

$$\begin{aligned} \langle a_n a_n^* \rangle &= \frac{1}{Y^2}\sum \epsilon^2(kT_i)\, Sin^2(n\omega_o T_i/2) \\ &= \frac{1}{Y}\sigma_\epsilon^2 Sin^2(n\omega_o T_i/2) = \alpha_n \end{aligned} \tag{41}$$

where we have introduced

$$\frac{\sin(x)}{x} = Sin(x) \tag{42}$$

Now we proceed with computation of the autocorrelation function **R(τ)**

$$R(\tau) = \sum_{-\infty}^{\infty} \alpha_n e^{jn\omega_o\tau} \qquad (43)$$

From it we aesily arrive at the power spectral density

$$\begin{aligned} S(\omega) &= \int_{-\infty}^{\infty} R(\tau)\, e^{-j\omega\tau} d\tau \\ &= \int_{-\infty}^{\infty} \sum \alpha_n e^{jn\omega_o\tau} e^{-j\omega\tau} d\tau \\ &= \sum \alpha_n \delta(\omega - n\omega_o) \end{aligned} \qquad (44)$$

and finally,after multiplication by 2 and with asistance o f **(27)**,we get for the mean value of one sided power spectral density

$$\begin{aligned} S(n\omega_o) &= \frac{2}{3Y} * 2^{-2S} Sin(n\omega_o T_i/2) \\ &\doteq \frac{2}{3Y} * 2^{-2S} \end{aligned} \qquad (45)$$

The approximation is justified since the transfer function **Sin(x)** is rather flat in the pass band from zero to **Y/2** (the maximum error is -4 dB).

References:

/1/ Kroupa, F. Věnceslav: Frequency Synthesis: Theory, Design et Applications. London: Ch. Griffin; New York: J. Wiley, 1973.

/2/ Manassewitch, Vadim: Frequency Synthesizers, Theory and Design. New York: J. Wiley, 1976,1980.

/3/ Rohde, L. Ulrich: Digital PLL Frequency Synthesizers: Theory and Design. Englewood Cliffs: Prentice-Hall, 1983.

/4/ Kroupa, F. Věnceslav: Amplitude of the general intermodulation product. Proc. IEEE vol. 58, pp. 851-2, May, 1970.

/5/ Zavrel, R.J. Jr.: Alias and spurious responses in DDS Systems, Application Note 102,in the DDS Handbook Stanford Telecom, 2nd ed. July 1990.

/6/ Giffard, P. Robin and Cutler, S. Leonard: A low

frequency, high resolution digital synthesizer", 1992 IEEE Frequency Control Symposium, pp. 188-192.

/7/ Kroupa, F. Věnceslav: Spectra of Pulse Rate Frequency Synthesizers. Proc. IEEE, vol. 67 (1979), No. 12, pp. 1980/82.

/8/ Kroupa, F. Věnceslav: Spectral Purity of Direct Digital Frequency Synthesizers. Proc. of 44th Annual Symposium on Frequency Control, (1990), pp. 498/510.

/9 / Kroupa, F. Věnceslav: Principles of Direct Digital Frequency Synthesizers. Kleinheubacher Berichte, vol. 36, pp. 663/678.

/10/ Mehrgardt, Soenke: Noise Spectra of Digital Sine-Generators Using the Table-Lookup Method. IEEE Trans. vol. ASSP-31, (1983), No. 4, pp. 1037/9.

/11/ Nicholas, T. Henry, III and Samueli Henry: An Analysis of the Output Spectrum of Direct Digital Frequency Synthesizers in the Presence of Phase-Accumulator Truncation. Proc. of the 41st Annual Frequency Control Symposium, (1987), pp. 495/502.

/12/ Gray, M. Robert: Quantization Noise Spectra. IEEE Trans. vol. IT-36, (1990), No. 6, pp. 1220/1244.

/13/ Papoulis, Athanasios: Probability, Random Variables and Stochastic Processes, New York: McGrow-Hill 1965.

/14/ Kroupa, F. Věnceslav, Discrete Spurious Signals and Background Noise in Direct Digital Frequency Synthesizers, 1993 IEEE Frequency Control Symposium, pp. 242/250

/15/ Garvey, F. Joseph and Babitch Daniel: An exact spectral analysis of a number controlled oscillator based synthesizer. Proc. of 44th Annual Symposium on Frequency Control, (1990), pp. 511/21.

/16/ Plassche van de, J. Rudy: High-speed and high-resolution analog-to digital and digital-to-analog converters: Ph.D. thesis, University of Technology Delft: 1989.

/17/ Krump, Gerhard: Halten Wandler-Board, was sie versprechen. Elektronik, vol. 26 (1993), pp. 95/106.

A NEW ARCHITECTURE FOR A SINEWAVE OUTPUT DDS WITH A HIGH SPECTRAL PURITY

Letizia LO PRESTI, Giuseppe CARDAMONE
Andrea DE MARCHI, Enrico RUBIOLA

Dipartimento di Elettronica
Politecnico di Torino
Corso Duca degli Abruzzi, 24 10129 Torino - Italy

Abstract

Direct Digital Frequency Synthesizers (DDFS's) are generally implemented by successively scanning through a look-up table stored in a Read-Only Memory (ROM) and converting the recalled sine samples to an analog waveform via a Digital to Analog Converter (DAC).

This paper describes a new architecture of a DDFS using a second order IIR resonator, whose impulse response is a sine wave; as this impulse response can be expressed in a second order recursive form, sample generation only requires few multiplications and additions. Therefore, the idea is to use a fast floating point DSP microprocessor that can implement the recursive equation in "real time".

This technique guarantees a larger set of output frequencies, a high spectral purity and a very simple, compact, and reliable hardware design.

1 Introduction

The applications of frequency synthesizers are numerous. They are used in areas like communication, control, music synthesis, and instrumentation. Designers must choose between using analog or digital synthesizers. Although stable synthesizers with low harmonic distortion can be built with analog circuits, automatic control of the parameters of the sinewave, namely amplitude, frequency and phase, is difficult with this approach. Direct Digital Frequency Synthesizers (DDFS's) have several advantages which include ease of control, flexibility, excellent temperature and ageing stability and low cost. They have the disadvantage of increased harmonic distortion caused by different numerical errors. Using a Digital Signal Processor (DSP) provides complete control and the speed required to generate a wide range of frequencies with the ability to perform distortion reduction calculations in "real time".

The most popular technique for Direct Digital Frequency Synthesis is the Sine Look-Up Table method first introduced by Tierney, Rader, and Gold [1]. This method synthesizes a sinewave by successively scanning through a look-up table stored in a Read-Only Memory (ROM) and converting the recalled sine samples to an analog waveform via a Digital to Analog Converter (DAC).

Using this architecture, both the frequency and phase resolution of the synthesizer are determined by the word length of the phase accumulator. This technique has some limitations:

- the set of output frequencies is limited,
- the spectral purity is strongly related to the word length of the phase accumulator,
- a large memory is required to achieve an acceptable frequency and phase resolution.

A theoretical analysis of the DDFS output spectrum is available [2], which also suggests a methods for improving the worst case performance. In more recent work [3], the phase noise is related to the resolution of the digital to analog converter (DAC) and to the intermodulation between output and clock frequencies.

The DDFS technique seems to be actractive for atomic clocks frequency synthesis. It has been studied for hydrogen masers [4], for rubidium cell standards [5] and for both laboratory [6] and commercial [7] cesium beam standards. This interest is due to the high resolution and to the fair spectral purity of the DDFS devices.

This paper describes a new architecture of a sinewave DDFS. The basic idea is to use a second order IIR resonator, whose impulse response is a sine wave; as this impulse response can be expressed in a second order recursive form, sample generation only requires few multiplications and additions. Therefore, the idea is to use a fast floating point DSP microprocessor that can implement the recursive equation in "real time" [8]. This technique guarantees a large set of output frequencies, a high spectral purity and a very simple, compact, and reliable hardware design.

The crucial point of the approach presented here is that the finite precision of calculations performed with the microprocessor does not guarantee that in the numerical implementation of the IIR filter the two complex conjugate poles in the z–plane be exactly on the unit circle. As a result the amplitude of the output sinewave may not be stable in time, and increase or decrease according to the random direction toward which truncation errors move the poles from the unit circle, that is outside or inside respectively. A number of techniques are described in this paper, which have been experimentally tested and shown to be adequate to keep the output sinewave stable. These techniques are referred to as *reset methods*.

Numerical and experimental results are presented in the following, which show that the spectral purity of the analog output is limited by the resolution of the digital-to-analog converter. A comparison with the results obtainable with the look-up table method is also given.

2 The sine look-up table method

In all DDFS's that use some variation of the Look-Up Table technique, one sample of the sinewave is recalled from a ROM at each clock pulse of a stable frequency reference f_s. L sinusoidal samples are stored in the ROM, and the sine frequency is assigned by means of a *Frequency control word* F_r, which is an integer number, expressed as a binary word, in the range $1 \leq F_r \leq 2^{L-1}$. The phase increment $\theta(n) = nF_r$ addressing the ROM is normally implemented by accumulating the binary frequency control word F_r, using a standard two's complement adder and a register (the *phase accumulator*). Therefore, in the ideal case of infinite precision memory and no phase quantization, the output sequence of the DDFS is given by

$$S(n) = \sin\left(2\pi\frac{F_r n}{2^L}\right) = \sin\left(2\pi\frac{\theta(n)}{2^L}\right) \tag{1}$$

The phase accumulator operates on the principle of overflow arithmetic using the modulo 2^L property of an L bit periodically overflowing accumulator register to simulate the modulo 2π property of the sine function by exploiting the relationship

$$\sin\left(2\pi\frac{\langle\theta(n)\rangle_{2^L}}{2^L}\right) = \sin\left(2\pi\frac{\theta(n)}{2^L}\right) \tag{2}$$

where $\langle\cdot\rangle_{2^L}$ stands for the integer residue of a number modulo 2^L.

2.1 Frequencies of the output sinewave

The frequency of the output sinewave depends on the DAC clock frequency $f_s = 1/T_s$ and on the ROM dimension L. The highest possible output frequency is

$$f_f = \frac{1}{2T_s} = \frac{f_s}{2}$$

while the lowest one is related to the ROM dimension L, that is

$$f_{sl} = \frac{1}{2^L T_s} = \frac{f_s}{2^L}$$

The other output frequencies, f_o, can be obtained by decimating the sinewave at frequency f_{sl}; the decimation step is related to the frequency control word F_r, that is

$$f_o = F_r \cdot f_{sl} = F_r \cdot \frac{f_s}{2^L}$$

2.2 Bit truncation in the phase accumulator

Using this architecture, both the frequency and phase resolution of the synthesizer are determined by the word length of the phase accumulator. For most applications, a large phase accumulator and, hence, a prohibitively large ROM are required to achieve an acceptable resolution. Therefore, in actual practice, all sine DDFS's, which require fine phase resolution, must truncate part of the phase accumulator output when addressing the ROM. In this way, smaller ROM's are needed.

Therefore if the output of the phase accumulator is truncated to W bits, only 2^W (instead of 2^L) ROM positions can be addressed. In this case, if $B = L - W$ is the number of truncated bits, the phase value addressing the ROM will be the integer part of $(F_r n)/2^B$. Hence, the output sequence of the DDFS becomes

$$S_t(n) = \sin\left(\frac{2\pi}{2^W}\left[\frac{F_r n}{2^B}\right]\right) = \sin\left(\frac{2\pi}{2^L}2^B\left[\frac{F_r n}{2^B}\right]\right) \tag{3}$$

where $[\cdot]$ stands for integer part; it follows that the sequence addressing the ROM can be written as

$$\langle\left[\frac{F_r n}{2^B}\right]\rangle_{2^W}$$

The phase accumulator truncation has the effect of producing undesired spurs in the line spectrum of $S_t(n)$, as described in [2].

In the next sections, we present an altogether new approach for the generation of a sinewave, which does not require any ROM. A greater frequency resolution is achieved, and the errors, due to the implementation in a finite precision microprocessor, can be efficiently controlled.

3 The IIR filter method

A fast and efficient method to generate a digital sinewave is to implement a digital resonator in a recursive form. The basic idea is to design the digital resonator as an IIR filter with two complex conjugate poles on the unit circle on the z plane, and to implement it by means of its input-output relationship, which is a simple recursive form. In fact, it is well known from the z-Transform theory that the impulse response of a filter with two poles on the unit circle is a sinewave, whose frequency depends on the pole locations. Such sinusoidal impulse response can be written in a recursive form, and generated by initializing the recursive equation with a digital impulse at time $n = 0$.

However, the theoretical impulse response cannot be perfectly achieved because of round-off errors. Therefore, the recursive algorithm that implements the IIR resonator must be modified in order to control the propagation of numerical errors during the sinewave generation and to guarantee the filter stability and a low spur level.

Both the idea of using an IIR resonator, and the methods implemented to reduce the effect of using a finite precision arithmetic, are totally innovative with respect to the most popular technique for Direct Digital Frequency Synthesis described in section 2.

3.1 IIR resonator

A digital resonator, implemented as an IIR filter with two complex conjugate poles ($e^{\pm j\varphi}$) on the unit circle, has a transfer function $H(z)$ of the form

$$H(z) = \frac{\sin\varphi z^{-1}}{1 - 2\cos\varphi z^{-1} + z^{-2}} \tag{4}$$

while its impulse response, $h(n)$, is an undamped sinusoid with unitary amplitude, which can be written in the recursive form

$$h(n) = c_0\delta(n-1) + c_1 h(n-1) - h(n-2) \tag{5}$$

where $\delta(n)$ is the Dirac delta sequence, defined as $\delta(n) = 1$ for $n = 0$ and $\delta(n) = 0$ for $n \neq 0$, $c_0 = \sin\varphi$ and $c_1 = 2\cos\varphi$. The implementation of Eq.(5) only requires few operations that can be executed using a fast DSP microprocessor.

The output frequency f_o of the sinewave is related to the pole locations by the equation

$$\varphi = 2\pi f_o T_s$$

where T_s is the DAC clock period. Therefore, f_o is related to the filter coefficients by the equations

$$\begin{aligned} c_0 &= \sin 2\pi f_o T_s \\ c_1 &= 2\cos 2\pi f_o T_s \end{aligned}$$

These two coefficients must be evaluated by the DSP microprocessor before starting the sinewave generation.

In our implementation of Eq.(5), we impose some constraints on the value of f_o, writing it in the form

$$f_o = \frac{M}{N} f_s \tag{6}$$

where $f_s = 1/T_s$ is the DAC clock frequency and M, N are integers. These two integer parameters are introduced to control the error propagation, as it will be explained in section 4. By using Eq.(6), the resonator coefficients can be written as

$$c_0 = \sin\left(2\pi\frac{M}{N}\right) \qquad c_1 = 2\cos\left(2\pi\frac{M}{N}\right) \tag{7}$$

Hence, different frequencies may be synthesized by changing the values of M and N.

3.2 Sinewave generation with a TMS320C30

The DSP microprocessor, used to implement Eq.(5), is a third generation Texas Instruments floating-point microprocessor (TMS320C30). The DDFS algorithm does not require a large memory; therefore, both instructions and data can be stored in the on-chip memory of the TMS320C30.

The procedure for the sinewave generation can be divided in three phases:

1. Microprocessor initialization. Prior to the execution of a digital signal processing algorithm, it is necessary to initialize the processor. Generally, initialization takes place any time the processor is reset. After reset, the processor should be initialized to meet the requirements of the system. Instructions should be executed that set up operational modes, memory pointers, interrupts, and the remaining functions needed to meet system requirements.

2. Algorithm initialization. In this phase, the IIR coefficients c_0, and c_1 are calculated. This calculation is performed as follows:

   - calculation of $\varphi = 2\pi(M/N)$,
   - calculation of $\cos\varphi$ and $\sin\varphi$.

   The calculation of φ is complicated by the restriction that division should be avoided since the TMS320C30 has no divide instructions; therefore, good approximations of φ can be obtained using the Newton-Raphson algorithm [9], while $\sin\varphi$ and $\cos\varphi$ are calculated using the CORDIC algorithm [10]. These algorithms are optimized for DSP applications.

3. The third phase is the dynamic one, where the sinusoid samples are evaluated and sent to the DAC every T_s seconds. If τ_m is the TMS320C30 clock period, the microprocessor will spend $\tau_D = k_D\tau_m$ seconds, where $\tau_D < T_s$, to produce one sine sample. During this time interval the arithmetic operations of the recursive equation and the instructions necessary to control the propagation error (see section 4) must be executed.

   In our implementation, we will try to reduce τ_D as much as possible. In fact, a low value of τ_D will give the microprocessor the possibility of spending some time $(\tau_R = T_s - \tau_D)$ to implement other functions.

4 Reset methods

In theory, Eq.(5) allows the digital generation of an infinite precision sinewave; but, in practice, its implementation, based on the TMS320C30, is affected by different numerical errors (round-off errors, propagation errors, drift errors, ...). These errors can produce damping or divergence of the signal amplitude. Therefore different techniques to control the amplitude have been conceived, which are described in the following sections.

4.1 IIR filter without stability control

This technique can be used if a short segment (few periods) of the sinewave has to be generated. In this case, no control of the numerical errors is generally necessary. In fact, the drifts towards an unstable or damped output are generally slow in a floating point implementation of Eq.(5), and they produce negligible errors for many periods of the sinewave. Therefore, in this case, a very simple software routine can be written in order to achieve a fast execution time and a small value of τ_D.

4.2 IIR filter with stability control

A simple method to control the error propagation, when generating the sinusoidal impulse response, is to reset the recursive equation (5) at each period of the generated signal. Notice that such period is generally different from the inverse of the sinewave frequency f_o. It is very easy to show that the period of the sequence is given by

$$R = \frac{N}{(M, N)} \tag{8}$$

where (M, N) is the greatest common divider of M and N. Therefore the sinusoidal sequence repeats itself every R samples. Hence, it is possible to make the DDFS stable by resetting the recursive equation every R output samples. The evaluation of R exploits the Euclidean algorithm.

This technique has the following characteristics:

- high stability and ease of control,
- τ_D greater than the one of the previous technique; in fact, as it is necessary to implement a reset routine, a higher number of software instructions are needed,
- the time necessary to initialize the DDFS is longer, but this is not a problem if a high switching frequency is not required.

4.3 IIR filter with stability control and a low value of τ_D

With a technique similar to the previous one, it is possible to reduce τ_D by storing the first R sinewave samples into the on-chip microprocessor memory.

After the calculation of the reset period R, the signal generation can be divided in the following steps:

1. R sine samples are calculated and stored in the on-chip microprocessor memory;
2. the R samples, previously stored in memory, are outputted.

This technique is limited by the on-chip memory dimension of the microprocessor.

4.4 DCSW method

The Dynamic Correction of Sine Waves method allows to get higher performances in the generation of a sinewave with an IIR resonator.

As each sinewave sample is calculated using a recursive equation (see Eq.(5)), the drift error is responsible for the performance degradation. In particular, the reset methods described previously do not guarantee a spur level of -135 dBc in their whole frequency range.

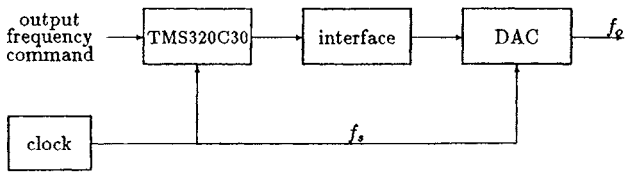


Figure 1: DDFS block diagram.

The basic idea of this method is to correct numerical errors by calculating some sinewave samples with a higher accuracy than the one obtained using Eq.(5). The number of higher accuracy samples calculated in a period of the output digital sequence of the DDFS can be chosen by the user.

The output samples of the DDFS can be divided in two sets:

1. sinewave samples calculated by means of the recursive equation,
2. sinewave samples calculated with an algorithm that guarantees high accuracy and fast generation.

Both sets are computed in "real time"; since the last group of samples requires a more complex process for their generation, and therefore a longer τ_D, at first sight it seems to be the main cause which limits the maximum output frequency, f_o. However, in an optimized implementation, the high accuracy sample calculation can be spanned over the dead time between the recursive equation samples.

This method guarantees an output signal with a high spectral purity and it makes possible to maintain the spur level below –135 dBc, indipendently of the output frequency.

4.5 IIR-ALT method

The IIR-Adaptive Look up Table method is based on the same idea of the DCSW method, but realizes it in a different way.

In the DCSW method, the correction is made in "real time", while in the IIR-ALT method higher accuracy samples are calculated during the DDFS initialization. Then these samples are stored in a memory that can be accessed by the microprocessor. In this way, the maximum output frequency is higher than the one of the DCSW method as only few microprocessor instructions are required to generate an output sample.

In the IIR-ALT method, the choice of the samples to be corrected is fundamental; it has been found that the best results are obtained if the correction is made before and after the waveform reaches its maximum values. Therefore, if the sinewave argument is $2\pi(M/N)n$, then the selected samples must satisfy the following inequalities

$$2\pi\frac{M}{N}n \leq \frac{\pi}{2} + 2k\pi \qquad 2\pi\frac{M}{N}(n+1) > \frac{\pi}{2} + 2k\pi$$

and

$$2\pi\frac{M}{N}n \leq \frac{3\pi}{2} + 2k\pi \qquad 2\pi\frac{M}{N}(n+1) > \frac{3\pi}{2} + 2k\pi$$

Hence, if the number of periods of the sinewave at frequency f_o contained in the period of the output digital sequence is:

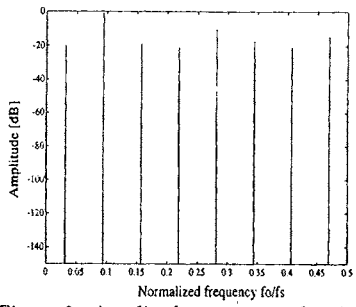


Figure 2: Amplitude spectrum of a sine wave at frequency $f_o/f_c = 3/32$ (ROM method) $L = 6$ and $B = 4$.

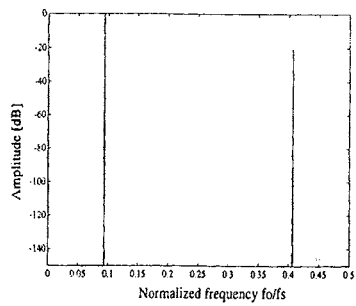


Figure 3: Amplitude spectrum of a sine wave at frequency $f_o/f_c = 3/32$ (ROM method) $L = 8$ and $B = 4$.

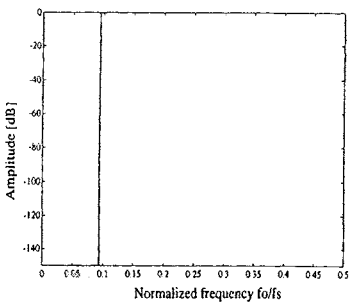


Figure 4: Amplitude spectrum of a sine wave at frequency $f_o/f_c = 3/32$ (IIR method).

$$L = \frac{M}{(M, N)}$$

then the number of samples, calculated and stored in memory during the microprocessor initialization, is $4L$.

The spectral purity of the sinewave generated with this technique is very high (the spur level is below −135 dBc).

5 Experimental results

In figure 1, a block diagram of a DDFS implemented with the IIR filter method is shown. The output frequency can be selected through a software program loaded in the on-chip memory of the TMS320C30. The digital samples generated by the microprocessor pass through a Digital to Analog Interface, and finally the DAC constructs the analog output waveform.

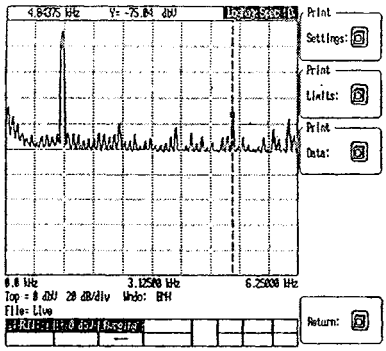


Figure 5: Amplitude spectrum of a 1.21 kHz sinewave.

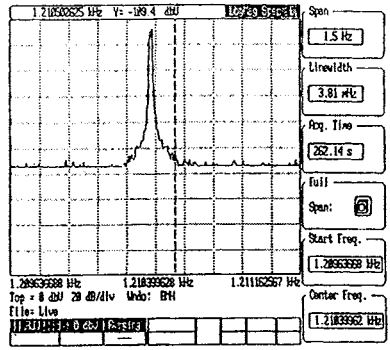


Figure 6: Amplitude spectrum of a 1.21 kHz sinewave.

In this section we will examine the spectral characteristics of the DDFS output signal referring to both the digital and the analog domain. In the first case, the spur level is due to the numerical errors of the algorithm, while in the second one, the spur level is determined by the DAC's number of bit.

5.1 Numerical output

In figures 2 and 3 the spectrum of a digital sinewave at frequency $f_o/f_s = 3/32$ is shown as obtained with the ROM method with two different ROM sizes. This frequency belongs to the set of frequencies that can be realized with both the IIR and the sine look-up table methods. This choice was determined by the need to compare the two methods.

In figure 4 the spectrum of this same frequency as obtained with the IIR method is shown. Calculated spectra have been found to be very similar in spectral purity for any frequency.

5.2 Analog output

Experimental results are given as obtained synthesizing a 1.21 kHz sinewave with a 16 bit DAC. This frequency was chosen becausde it is in the range needed by our final application, in the primary cesium standard. It is shown in figures 5 and 6 that the spur level and the spectral purity of the output signal are limited by the resolution of the DAC. If a smaller number of bit were used in the output DAC, for example for speed reasons, the spectral purity would turn out to be accordingly worse. Reported spectra were obtained with an FFT analyzer. It must pointed out that the latter is also a 16 bit machine.

6 Conclusion

In this paper a new architecture for a sinewave output DDFS has been described.

Its implementation is based on a very simple recursive equation representing the impulse response of an IIR resonator. Few instructions of the TMS320C30 allow the generation of a larger set of output frequencies than the one obtained with the Sine Look-Up Table method. Besides, other frequencies have been tested; the results validate the IIR

resonator as an interesting candidate for a new hardware architecture for the DDFS implementation.

References

[1] J. Tierney, C. M. Rader and B. Gold, "A digital frequency synthesizer", *IEEE Trans. on Audio Electroacoustics*, vol. AU-19, p. 48–57, 1971.

[2] H. T. Nicholas III and H. Samueli, "An analysis of the output spectrum of direct digital frequency synthesizers in the presence of phase-accumulator truncation", *Proc. 41st Annual Frequency Control Symposium* p. 495–502, IEEE, 1987.

[3] V. Kroupa, "Intermodulation signals in direct digital frequency synthesizers", *Proc. 8th European Frequency and Time Forum*, Weihnstephan, Germany, March 9–11, 1994.

[4] E. M. Mattison, L. M. Coyle, "Phase noise in direct digital synthesis", *Proc. 42nd Annual Frequency Control Symposium* p. 352–356, IEEE 1988.

[5] P. Schumacher, "A narrow band frequency synthesizer with a NCO circuit", *Proc. 5th European Frequency and Time Forum*, p. 123–128, Besançon, France, March 12–14, 1991.

[6] J. Viennet, "Synthetiseur digital de fréquence pour standards atomique de fréquence", *Proc. 3rd European Frequency and Time Forum*, p. 333–336, Besançon, France, March 21–23, 1989.

[7] R. P. Giffard, L. S. Cutler, "A low frwquency, high resolution digital synthesizer", *Proc. 46th Annual Frequency Control Symposium* p. 188–192, Philadelphia, USA, May 27–29, 1987.

[8] G. Cardamone and L. Lo Presti, "A direct digital frequency synthesizer using an IIR filter implemented with a DSP microprocessor", *Proc. 1994 International Conference on Acoustics, Speech & Signal Processing*, Adelaide (Australia) April 19–22, 1994.

[9] Texas Instruments, "*Theory, algorithms, and implementations*, ©Texas Instruments Inc., 1990.

[10] M. D. Ercegovac and T. Lang, "Fast cosine/sine implementation using on-line CORDIC", *Proc. 21st Annual Asilomar Conference on Signals, Systems and Computers*, 1987.

[11] Texas Instruments, "*Third-generation TMS320 user's guide*", ©Texas Instruments Inc., 1989.

On a Possibility of Reducing the Frequency Stability Limitation by the Interrogation Noise in Passive Frequency Standards

D. Venot, R. Barillet and C. Audoin

LABORATOIRE DE L'HORLOGE ATOMIQUE
Unité Propre de Recherche associée à l'Université Paris-Sud
CNRS - Bâtiment 221 - Université Paris-Sud - 91405 ORSAY-Cedex - France

Abstract - It is known that the local oscillator phase noise in passive frequency standards may limit the frequency stability because of an intermodulation effect [1,2]. The purpose of this paper is to describe a possible optimisation in order to reduce or eventually suppress the limiting effect. First suggested by Kramer[3], the likely optimisation is studied theoretically and significant results are found. The optimisation principle is to find the best demodulation function according to a given system. In fact, among all the possible functions it will be shown there are some waveforms able to improve the frequency stability by reducing the spurious effect. It will also be shown that the use of a notch filter in the frequency synthesis gives limited improvements but could as well give very interesting results if combined with the demodulation optimisation.

I. INTRODUCTION

In the field of passive frequency standards, large improvements are expected from optical pumping and laser cooling. But presently the frequency stability may stongly be limited by an intermodulation effect related to the modulation process and to the frequency noise present in the probe signal [1,2].

A first approach has been implemented at NIST in order to solve this problem. It consists in reducing the main noise contribution by using proper notch filters in the frequency synthesis [4]. In any case, a limit to this approach is given by the thermal noise, which is not so far for systems operated at ambient temperature. A practical limit, too, is related to the performance of the frequency multipliers in the synthesis chain. When one looks at the block-diagram of a typical passive frequency standard, it is obvious that improvements may be tried either in the frequency synthesis - as made at NIST, or after the output of the atomic resonator.

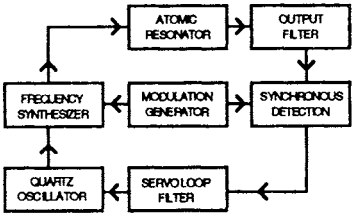


Figure 1
- Block-diagram of a typical passive frequency standard -

This last possibility has already been suggested [3]: with slow square wave modulation and demodulation, when the response of the atomic resonator is not filtered, the limiting effect is supposed to disappear. This remark led us to start a theoretical analysis, in order to determine whether it is possible to reduce or

suppress the limiting effect by means of a convenient signal processing at the output of the atomic resonator.

In this paper, we consider a passive frequency standard in which the probe signal is continuously applied. The calculation is made with the quasi-static approximation. In the Fourier frequency range of interest, the noise of the local oscillator is supposed to be a white phase noise, whilst the amplitude noise is not considered. After the general expression is introduced, we will shorten the field of investigations to the square wave frequency modulation.

II. THEORETICAL APPROACH

II.1. Interrogation Signal

The instantaneous angular frequency $\omega(t)$, at the output of the frequency synthesiser, should be locked at the resonance frequency ω_o. The expression of $\omega(t)$ involves the central angular frequency ω_o, the modulation function introduced by the system and two noise components $\delta\omega_s(t)$ and $\delta\omega_s(t)$.

$$\omega(t) = \omega_o + \omega_m g(t) + \delta\omega_s(t) + \delta\omega_f(t)$$

In this expression, $\omega_m g(t)$ represents the frequency modulation function and ω_m is the depth. The shape $g(t)$ is a periodic function of frequency f_M. It is assumed that $|g(t)|$ is equal to 1. The contribution of noise is divided into two terms $\delta\omega_s(t)$ and $\delta\omega_f(t)$ relatively to the cutoff frequency f_c of the low pass filter of the servo loop. The noise component $\delta\omega_s(t)$ represents the slow fluctuations of noise in the frequency range from 0 to f_c, $\delta\omega_f(t)$ is the fast fluctuation of noise in the frequency range beginning at f_c. The modulation depth is supposed to be much more important than any noise component.

$$|\delta\omega_s(t)| << \omega_m \text{ and } |\delta\omega_f(t)| << \omega_m$$

The noise involved in the limiting effect is in fact wholly included in narrow spectral windows of width $2f_c$ and centred around frequencies $2nf_M$, even multiples of f_M. Therefore, we will consider the Rice decomposition $\delta\omega_f^{f_M}(t)$ of the fast noise component $\delta\omega_f(t)$ around $2nf_M$.

$$\delta\omega_f^{f_M}(t) = \sum_{n=1}^{+\infty} p_{2n}(t)\sin(4\pi n f_M t) + q_{2n}(t)\cos(4\pi n f_M t)$$

$p_{2n}(t)$ and $q_{2n}(t)$ are random variables which are not correlated. The one-sided power spectral densities $S_{q_{2n}}(f)$ and $S_{p_{2n}}(f)$ will have the following properties :

$$S_{q_{2n}}(f) = S_{p_{2n}}(f) = 2.S_{\delta\omega_f}(2nf_M) \quad \text{if } f<f_c$$

II.2. The Atomic Resonator Response.

The response I(t), of the atomic resonator is an even function h of the variable $\omega(t)-\omega_o$. There is a maximum I_o at the angular frequency ω_o.

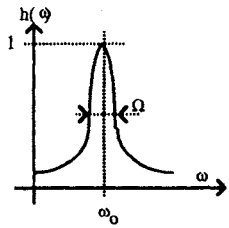


Figure 2
Atomic resonator response versus angular frequency

$$I(t) = I_o \times h\left(\frac{\omega(t)-\omega_o}{\Omega}\right)$$

Assuming the quasi-static approximation, it is possible to use a first order development of this expression according to the variable $\omega(t)$.

$$\frac{I(t)}{I_o} = h\left(\frac{\omega_m}{\Omega}.g(t)\right) + \left(\delta\omega_s(t)+\delta\omega_f(t)\right)\frac{\partial h\left(\frac{\omega_m g(t)}{\Omega}\right)}{\partial\omega}$$

The first term of this last expression has Fourier components only at frequencies $2nf_M$ whereas the second term has Fourier components only at frequencies $(2n+1)f_M$. Practically it is assumed that a filter is settled at the output of the atomic resonator to cut off the even multiples of f_M. The odd components $(2k+1)f_M$ are just affected by a filtering coefficient ϕ_{2k+1}. I(t) can be restricted to a response I'(t) with only the odd multiples of f_M, coming from the second term.

$$\frac{I'(t)}{I_o} = (\delta\omega_s(t)+\delta\omega_f(t))\sum_{n=1}^{+\infty} \phi_{2n+1}P_{2n+1}\sin(2(2n+1)\pi f_M t)$$

The coefficients P_{2n+1} are the Fourier coefficients involved in the decomposition of $\frac{\partial h\left(\frac{\omega_m g(t)}{\Omega}\right)}{\partial\omega}$.

With the Rice decomposition of $\delta\omega_f(t)$, it is possible to express I'(t) as a sum of sines and cosines. Because of the synchronous detection the only terms to take into account are factors of sines.

$$\frac{I''(t)}{I_o} = \sum_{k=0}^{+\infty} A_{2k+1}(t)\ \sin(2\pi(2k+1)f_M t)$$

with the following expression of $A_{2k+1}(t)$:

$$A_{2k+1}(t) = \phi_{2k+1}P_{2k+1}\delta\omega_s(t) + \frac{1}{2}\sum_{n=1}^{+k} \phi_{2n+1}(P_{2k+1-2n}+P_{2k+1+2n})q_{2n}(t)$$

$$+\frac{1}{2}\sum_{n=k+1}^{+\infty} \phi_{2n+1}(P_{2n+2k+1}-P_{2n-2k-1})q_{2n}(t)$$

The demodulation function $g_d(t)$ can be expressed by a Fourier sum.

$$g_d(t) = \sum_{k=1}^{+\infty} d_k \sin(2\pi k f_M t)$$

The signal I''(t) resulting of the demodulation is the product of I'(t) and $g_d(t)$.

$$\frac{I''(t)}{I_o} = \sum_{n=1}^{+\infty} A_{2n+1}\sin(2(2n+1)\pi f_M t) \times \sum_{k=1}^{+\infty} d_k\sin(2\pi k f_M t)$$

Eventually the low pass filter of the servo loop will give the D.C. component of this product. The cutoff frequency of this filter is obviously assumed to be far below the frequency f_M.

$$e(t) = \frac{1}{2}\alpha.\delta\omega_s(t) + \sum_{n=1}^{+\infty} \beta_n.q_{2n}(t)$$

with

$$\alpha = \sum_{k=0}^{+\infty} \phi_{2k+1}P_{2k+1}d_{2k+1}$$

and

$$\beta_n = \frac{1}{4}\sum_{k=0}^{n-1} \phi_{2k+1}d_{2k+1}(P_{2n+2k+1}-P_{2n-2k-1}) + \frac{1}{4}\sum_{k=n}^{+\infty} \phi_{2k+1}d_{2k+1}(P_{2n+2k+1}+P_{2k+1-2n})$$

The function e(t) is the error signal used by the system to lock the quartz crystal oscillator to the atomic resonance. The servo loop is designed to null the error signal in the servo bandwidth. It is consequently justified to admit that e(t) is small enough to be supposed equal to zero. We are thus able to express $\delta\omega_s(t)$.

$$\boxed{\delta\omega_s(t) = -\frac{2}{\alpha} \times \sum_{n=1}^{+\infty} \beta_n.q_{2n}(t)}$$

II.3. Power Spectral Density

Using the last expression of the slow component of noise $\delta\omega_s(t)$, it is possible to evaluate its one-sided power spectral density $S_{\delta\omega_s}(f)$.

$$S_{\delta\omega_s}(f) = \frac{4}{\alpha^2} \times \sum_{n=1}^{+\infty} \beta_n^2.S_{q_{2n}}(f)$$

In the considered range of frequencies the local oscillator noise is assumed to be a white phase noise. The one-sided power spectral density of frequency fluctuations is so a square function of the Fourier frequency f. Such assumptions lead to the following expression of $S_{q_{2n}}(f)$.

$$S_{q_{2n}}(f) = n^2 . S_{q_2}(f)$$

It is now possible to express the one-sided power spectral density of $\delta\omega_s(t)$ in a very simple way.

$$\boxed{S_{\delta\omega_s}(f) = \frac{4.S_{q_2}(f)}{\alpha^2} \times \sum_{n=1}^{+\infty} \beta_n^2.n^2}$$

II.4. Deteriorating factor

Let ξ be the deteriorating factor. It is introduced to numerically represent the spurious effect of the white phase noise in the system. In other words it is an evaluation of the contribution of the interrogation phase noise, drawn back into the useful signal.

$$\xi = \frac{S_{\delta\omega_s}(f)}{S_{q_2}(f)}$$

With the previous results, the expression of the deteriorating factor becomes :

$$\xi = \frac{4}{\alpha^2} \times \sum_{n=1}^{+\infty} \beta_n^2 . n^2$$

With the given assumptions, the ξ factor is defined in such a way it is a constant with no unit.

II.5. Square wave modulation

The typical modulation function used at LHA is a square waveform. This is both the easiest to synthesise and the most advantageous [5]. Using this square waveform, the expression of the coefficients P_{2k+1} are very easy.

$$P_{2k+1} = \frac{1}{2k+1} \quad \forall\, k \in \mathbb{N}$$

The resulting expression of the deteriorating factor ξ is

$$\xi = \frac{\sum_{n=1}^{+\infty} \left(\sum_{k=0}^{+\infty} \phi_{2k+1} d_{2k+1} \left(\frac{(2k+1)\, n}{(2k+1)^2 - 4n^2} \right) \right)^2}{\left(\sum_{k=0}^{+\infty} \frac{\phi_{2k+1} . d_{2k+1}}{2k+1} \right)^2}$$

III. PRESENT LIMITATIONS

The purpose of this paragraph is to determine the present limitations in passive frequency standards.

III.1. Basic frequency standard

The system to be analysed is for example a caesium beam clock with square wave modulation and demodulation function. A filter is added at the output of the atomic resonator. This filter is supposed to be very selective around the frequency f_M. All the frequency components will be cut off except the component f_M which will remain unchanged. With these assumptions, it is possible to give the value of each filtering coefficient ϕ_{2k+1}.

$$\phi_1 = 1 \text{ and } \phi_{2k+1} = 0 \quad \forall\, k \in \mathbb{N}^*$$

This last relation implies a very simple expression of the deteriorating factor ξ.

$$\xi = \frac{\sum_{n=1}^{+\infty} \left(\phi_1 d_1 \frac{2\, n}{1-4n^2} \right)^2}{4 \left(\phi_1 d_1 \right)^2}$$

Obviously, in such a system, the deteriorating factor does not depend on the chosen demodulation function as long as the first harmonic coefficient d_1 is different from zero.

$$\xi = \sum_{n=1}^{+\infty} \left(\frac{n}{1-4n^2} \right)^2$$

All the assumptions correspond to the real system that will so be considered as a reference and its deteriorating factor ξ will be written ξ°. Using the previous definitions it is obvious to determine the expression of ξ°.

$$\sum_{n=1}^{+\infty} \left(\frac{n}{1-4n^2}\right)^2 = \frac{\pi^2}{64}$$

We are able to give a numerical expression for the standard deteriorating factor ξ°. This result has already been found in a peculiar calculation of the intermodulation effect [1].

$$\boxed{\xi^\circ = \frac{\pi^2}{64} \cong 0.154}$$

III.2. Improving factor

Once a reference has been defined, it is possible to estimate the relative improvement given by a new system. Let ζ be the improving factor of the studied system relatively to the reference device whose deteriorating factor is ξ°. For a given system of deteriorating factor ξ, the resultant improving factor will be defined by

$$\zeta = \sqrt{\frac{\xi^\circ}{\xi}}$$

The definition of ζ is done in such a way it directly gives the factor by which the frequency stability limit is improved. For instance an improving factor of 2 would mean the frequency stability limit related to the limiting effect, is two times lower than the reference.

III.3. Notch filter insertion

The second case to study corresponds to the previous one except there is a notch filter inserted in the frequency synthesis. This idea has already been studied and realised at NIST [4]. We thus are able to guess that for some practical systems it is a successful way of improvement.

The purpose of this filter is to reject two narrow windows of frequency symmetrically located at $\pm 2f_M$ from the carrier in the interrogation signal. The most important term found in the expression of ξ, $S_{q2}(f)$, is then significantly reduced.

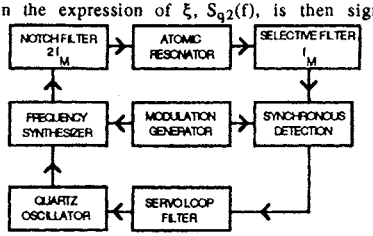


Figure 3
Bloc diagram with the notch filter

Without any filtering system at the input of the atomic resonator, the expression of the one-sided power spectral density of $\delta\omega_s(t)$ is :

$$S_{\delta\omega_s}(f) = \frac{4}{\alpha^2} \times \sum_{n=1}^{+\infty} \beta_n^2 . S_{q_{2n}}(f)$$

Both the modulation and demodulation processes are still realised with a square waveform. The very selective filter around f_M is still present at the output of the atomic resonator. The previous expression of $S_{\delta\omega_s}(f)$ is to become :

$$S_{\delta\omega_s}(f) = \frac{4}{(\phi_1 d_1)^2} \times \sum_{n=1}^{+\infty} \left(\phi_1 d_1 \frac{2}{1-4n^2}\right)^2 S_{q_{2n}}(f)$$

The one-sided power spectral density of the interrogation signal before the notch filter remains the same and it still is $S_{q_{2n}}(f)$. Let's introduce $S'_{q_{2n}}(f)$, the one-sided power spectral density of frequency fluctuations at the output of the notch filter which is also the input of the atomic resonator. The notch filter is designed to pull down the noise component at $2f_M$ which is $q_2(t)$. Let λ be the power gain for the term $q_2(t)$. The resulting expression of $S'_{q_{2n}}(f)$ is :

$$\begin{cases} S'_{q_2}(f) = \lambda . S_{q_2}(f) \\ S'_{q_{2n}}(f) = S_{q_{2n}}(f) \quad \forall\, n \in IN-\{1\} \end{cases}$$

The following property remains valid :

$$S_{q_{2n}}(f) = n^2 S_{q_2}(f)$$

It is so possible to express $S_{\delta\omega_s}(f)$ as a function of $S_{q_2}(f)$ by subtracting the original term and adding back the lowered one :

$$S_{\delta\omega_s}(f) = \sum_{n=1}^{+\infty} \left(\frac{n^2}{1-4n^2}\right)^2 S_{q_2}(f) + \frac{\lambda-1}{9} \times S_{q_2}(f)$$

From this relation, the deteriorating factor ξ can be expressed as a function of ξ^o and of the power gain λ. To remind the assumption of the notch filter at $2f_M$ this factor will be written ξ^o_n.

$$\boxed{\xi^o_n = \xi^o + \frac{\lambda-1}{9}}$$

The calculation of the improving factor $\zeta_r(\lambda)$ leads to the following expression :

$$\zeta_r(\lambda) = \sqrt{\frac{\xi^o}{\xi^o + \frac{\lambda-1}{9}}}$$

It appears that the insertion of the notch filter is a good solution to improve the stability of the system. A rejection of 30 dB of the $2f_M$ component would afford an improvement by a factor 1.889 in frequency stability. Nevertheless it is interesting to notice that the rejection of $2f_M$ gives a limited improvement. The best improvement we could expect is

$$\lim_{\lambda \to 0} \zeta_r(\lambda) = \sqrt{\frac{\xi^o}{\xi^o - \frac{1}{9}}}$$

The improving factor is thus limited to $\zeta^1_n = 1.89$. This result shows that, with our assumptions, the best improvement of frequency stability a notch filter can bring in the reference system is a factor 2. Hopefully, the limit is very quickly reached and it is not necessary to use very deep rejections. For instance with a -30 dB notch filter the limit is theoretically reached at almost 99.9%.

IV. OPTIMISATION

The expression of the one-sided power spectral density $S_{\delta\omega_s}(f)$ established in the previous paragraphs for a square waveform frequency modulation is :

$$S_{\delta\omega_s}(f) = S_{q_2}(f) \times \frac{\sum_{n=1}^{+\infty}\left(\sum_{k=0}^{+\infty} \phi_{2k+1} d_{2k+1} \frac{(2k+1)\,n}{(2k+1)^2-4n^2}\right)^2}{\left(\sum_{k=0}^{+\infty} \frac{\phi_{2k+1}\cdot d_{2k+1}}{2k+1}\right)^2}$$

The coefficients of demodulation d_{2k+1} and the filtering coefficients ϕ_{2k+1} are always implied as a product $\phi_{2k+1}d_{2k+1}$. It shows that the demodulation waveform and the filtering are working as a whole. Their effect cannot be separated in the result. Consequently as long as the demodulation function has a sine waveform it is useless to consider an other filter but the selective one around f_M. Symmetrically, it is useless to try to change the demodulation function waveform when a selective filter is used.

The optimisation principle deals with the research of an optimum demodulation waveform. Because of the previous considerations, this optimisation will obviously have to consider both the filtering device at the output of the atomic resonator and the modulation function. To simplify the expression it is then suitable to define a coefficient c_{2k+1} to replace the term $d_{2k+1}\phi_{2k+1}$. The new expression of the deteriorating factor is :

$$\xi = \frac{\sum_{n=1}^{+\infty}\left(\sum_{k=0}^{+\infty} c_{2k+1} \frac{(2k+1)\,n}{(2k+1)^2-4n^2}\right)^2}{\left(\sum_{k=0}^{+\infty} \frac{c_{2k+1}}{2k+1}\right)^2}$$

The optimisation is to find the best c_{2k+1} family which involves a minimum of the factor ξ. This will obviously give a maximum for ζ and so a minimum of the frequency stability limit.

Because it has no implication in the optimisation process the first coefficient c_1 will always be considered as a constant. It is arbitrary chosen to be 1.

IV.1. Two-harmonic optimisation

The first approach is to try to find the best demodulation function with only two harmonics. By definition, the coefficient c_1 can not be changed. The next coefficient to optimise is then c_3. Considering only two harmonics means that because of the chosen filter (or demodulation function) it is assumed that all coefficients down to c_5 are equal to zero. The expression of ξ is then a function of c_3.

$$\xi(c_3) = \xi^o \times \frac{c_1^2 + 9\,c_3^2}{\left(c_1+\frac{c_3}{3}\right)^2}$$

There is a possible optimisation for $c_3=c_{3_{opt}}$ of the deteriorating factor ξ which admit a minimum written ξ_{opt}.

$$c_{3_{opt}} = \frac{C_1}{27} \quad \text{and} \quad \xi_{opt} = x^\circ \times \frac{81}{82} \cong 0.152$$

Even if the expected optimisation exists, it is unfortunately a very small improvement. With two harmonics the reference system can only be improved of a useless factor ζ=1.0061.

IV.2. Three-harmonic optimisation

It has been shown that a raw optimisation with two harmonics is useless. Would it be so with a three-harmonic optimisation ? This time, the expression of the deteriorating factor ξ is to depend of c_3 and c_5.

$$\xi(c_3,c_5) = \xi^\circ \times \frac{C_1^2 + 9\,c_3^2 + 25\,c_5^2}{\left(c_1 + \frac{c_3}{3} + \frac{c_5}{5}\right)^2}$$

Numerical analysis will allow us to find the optimised value ξ_{opt} of the deteriorating factor. This optimum is found for a couple of c_3,c_5 with the following values :

$$c_3 = \frac{1}{27} \quad \text{and} \quad c_5 = \frac{1}{125}$$

The value of the deteriorating factor is nevertheless very disappointing so close it is from the reference value ξ°.

$$\xi_{opt} = \xi^\circ \times \frac{50625}{51331}$$

This results would give an improvement of a factor 1.00695. Once more the three-harmonic optimisation is possible but useless.

IV.3. n-harmonic optimisation

With the hypothetical model considered up to now, the optimisation with a very limited number of harmonics does not seem to be persuasive. The matrix theory used in the search of a very general optimum of the system gave once a simple result. The optimised demodulation waveform has to be the temporal inverse of the modulation waveform. This implied that the filter at the output of the atomic resonator didn't change any of the frequency components $(2k+1)f_M$, odd multiples of f_M. The matrix calculation also implied that the even components $2nf_M$ were useless for the optimisation and could eventually be cut off.

In the case studied in this paper, the frequency modulation function has a square waveform. Multiplying a square function by itself gives unity at each point. Therefore, the temporal inverse of a square waveform frequency modulation is the same synchronous square waveform function. According to the matrix theory, the square waveform modulation function should be optimised by a square waveform demodulation process, furthermore, the limitation effect should disappear.

To check this theory, we will calculate the deteriorating factor of our reference system in the case of a square waveform demodulation and a very selective filter around each $(2k+1)f_M$. The consequences of this choice will give the following values for the coefficients c_{2k+1}.

$$c_{2k+1} = \frac{1}{2k+1} \quad \forall\, k \in \mathbb{N}$$

The expression of the resulting deteriorating factor is the following one.

$$\xi = \frac{\sum_{n=1}^{+\infty}\left(\sum_{k=0}^{+\infty}\frac{n}{(2k+1)^2-4n^2}\right)^2}{\left(\sum_{k=0}^{+\infty}\frac{1}{(2k+1)^2}\right)^2}$$

The sum of the implicit variable k is convergent and its limit is 0.

$$\sum_{k=0}^{+\infty}\frac{n}{(2k+1)^2-4n^2}^2 = 0 \;\forall\, n \in \mathbb{N}$$

This result leads to the very simple conclusion that the value of the deteriorating factor ξ is zero. This optimisation would be perfect enough to theoretically suppress the limitation effect.

Unfortunately this theoretical conclusion does not correspond to reality. On one hand it has no meaning to consider a function with harmonics present up to infinity. On the second hand the behaviour of the atomic resonator is not modelled for a very wide frequency band and the assumptions might as well be incorrect.

This could make us think to cut the frequency bandwidth to a finite dimension and then try to find the best optimisation. In fact this corresponds to the n-harmonic optimisation. It has been shown that with two or three harmonics the optimisation was ridiculously small. Much more troubling it could be shown that with any number of harmonics, once they are in finite number, the optimisation would remain inefficient. With a finite number of harmonics, a useless asymptotic maximum of 3% improvement can be expected !

V. OPTIMISATION WITH NOTCH FILTER

This last chapter is to consider both the optimisation process previously explained and the adjunction of a notch filter method. The notch filter considered is still at the input of the atomic resonator. It still is the one described in III.3.

V.1. Two-harmonic optimisation & notch filter

The first optimisation to be taken into account is the two-harmonic optimisation. As it has been done previously, let λ be the power gain for the component $q_2(t)$ in the notch filter. The expression of the deteriorating factor will still be a function of the third harmonic represented by the coefficient c_3.

$$\xi(c_3,\lambda) = \xi^o \times \frac{C_1^2 + 9\,c_3^2}{4.\left(c_1+\frac{c_3}{3}\right)^2} - \frac{1-\lambda}{4} \times \frac{\frac{6}{5}c_3 - \frac{2}{3}c_1}{\left(c_1+\frac{c_3}{3}\right)^2}$$

It is interesting to notice that the function $\xi(c_3,\lambda)$ as a function of λ has no local minimum but just an asymptotic one when λ reaches zero. For a peculiar value $c_{3_{opt}}(\lambda)$ of the variable c_3, it is possible to find a minimum $\xi_{opt}(\lambda)$ of the function $\xi(c_3,\lambda)$. The expression of this optimum is a function of λ. It does mean that the optimising waveform to be chosen for the demodulation function depends on the shape of the notch filter. With the previous assumptions, the expression of this optimum is :

$$c_{3_{opt}}(\lambda) = \frac{-5\,(2048\,(1-\lambda) - 45\,\pi^2)}{9\,(2048\,(1-\lambda) + 675\,\pi^2)}$$

$$\xi^c_{opt}(\lambda) = \frac{81\,\pi^2\,(2176\,(\lambda-1) + 225\,\pi^2)}{128\,(32768\,(\lambda-1) + 9225\,\pi^2)}$$

When there is no notch filter, the power gain should be taken to $\lambda=1$. The resulting value of ξ, 0.153, is the same than the one found in the previous paragraphs and means that there is no significant improvement.

It is interesting to notice that $\xi^{c}_{opt}(\lambda)$ is very quickly converging towards a limit denoted ξ_{opt}. In fact with a 30 dB notch filter the optimisation is about 97% equal to the possible optimisation limit.

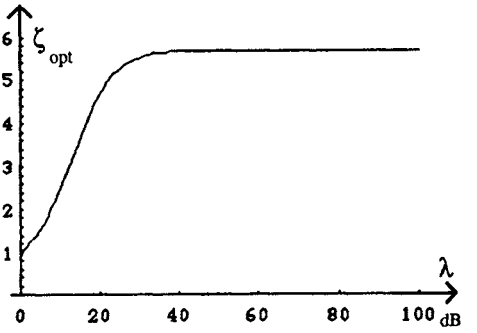


Figure 4 : Improving factor versus gain λ

The limit of the optimisation is obtained with $\lambda=0$. It does mean a total rejection of the $q_2(t)$ noise component. It is possible to calculate the value of the deteriorating factor in this peculiar situation:

$$\left|\xi_{opt}\right|_{\lambda=0} = \frac{81\ \pi^2\ (-2176 + 225\ \pi^2)}{128\ (-32768 + 9225\ \pi^2)} \cong 0.0048$$

With such a result the expected theoretical improving factor is around $\zeta_{opt}=5.7$. This result is significantly interesting, especially if one remembers the previous results. The only optimisation gave a useless $\zeta=1.0061$ and the notch filter alone a moderate $\zeta=1.89$.

Therefore these results are determinant. With an optimisation involving only one unknown parameter it is then possible to get an improvement by a factor 6 of the frequency stability limitation. The value of the parameter c_3 is explicitly given as a function of the power gain λ.

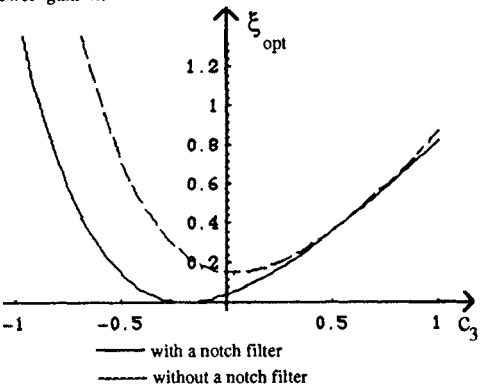


Figure 5 : Chart $\xi(c_3,\lambda)$ of versus $c_{3_{opt}}$ with $\lambda=0$.

With a total rejection of the noise component $q_2(t)$, it appears that the optimum is found for a negative value of the coefficient c_3. The variation of $c_3(\lambda)$, versus λ represented by the following chart, shows that there is a threshold $C_{3_{th}}$ which is very quickly reached.

$$C_{3_{th}} = \lim_{\lambda \to 0} C_3(\lambda) = \frac{-5\,(2048-45\,\pi^2)}{9\,(-2048+675\,\pi^2)} \cong -0.19$$

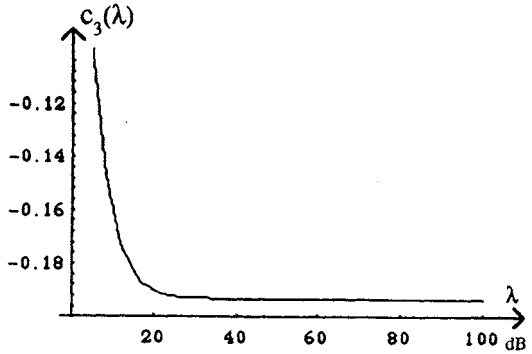


Figure 6

optimised coefficient $c_3(\lambda)$ versus gain λ

According to the above expressions, it becomes obvious that the optimisation with both the notch filter and the adapted demodulation process gives much more interesting results than each one considered apart.

V.2. Three-harmonic optimisation & notch filter

The three-harmonic optimisation consists in setting all the coefficients over c_5 to zero. The optimisation will be to find the best couple (c_3,c_5) to choose, assuming $c_1=1$. The notch filter still introduces an attenuation of the noise component $q_2(t)$ with a power gain λ. The expression of the deteriorating factor $\xi(c_3,c_5,\lambda)$ is then

$$\xi(c_3,c_5,\lambda) = \xi^o \times \frac{C_1^2 + 9\,c_3^2 + 25\,c_5^2}{4.\left(c_1 + \frac{c_3}{3} + \frac{c_5}{5}\right)^2} - \frac{1-\lambda}{4} \times \frac{\frac{10}{21}c_5 + \frac{6}{5}c_3 - \frac{2}{3}c_1}{\left(c_1 + \frac{c_3}{3} + \frac{c_5}{5}\right)^2}$$

It is still possible to find the optimum couple $(c_{3_{opt}}, c_{5_{opt}})$ by which the function $\xi(c_3,c_5,\lambda)$ has a local minimum $\xi_{opt}(\lambda)$.

$$C_{3_{opt}}(\lambda) = \frac{2504704\,(\lambda-1) + 55125\,\pi^2}{45\,(106496\,(\lambda-1) + 33075\,\pi^2)}$$

$$C_{5_{opt}}(\lambda) = \frac{49\,(22528\,(\lambda-1) + 675\,\pi^2)}{125\,(106496\,(\lambda-1) + 33075\,\pi^2)}$$

The resulting expression of $\xi_{opt}(\lambda)$ is a function of λ.

$$\xi_{opt}(\lambda) = \frac{50625\,\pi^2\,(108224\,(\lambda-1)+11025\,\pi^2)}{64\,(2140012544\,(\lambda-1)+565924275\pi^2)}$$

As it is intuitively expected, the fluctuations of the deteriorating factor ξ versus c_3 and c_5 show that c_3 is involved as a basic optimisation. The coefficient c_5 corresponds to a very fine localisation of the optimum.

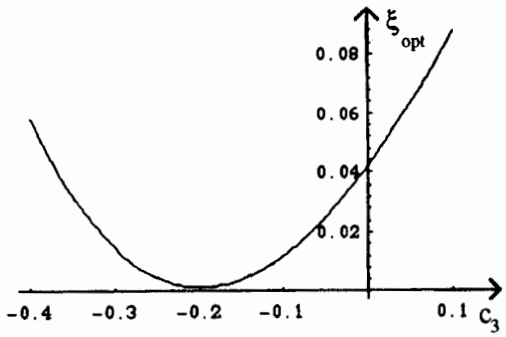


Figure 7
Fluctuation of the deteriorating factor ξ versus c_3
(c_5 is supposed to be the optimum one)

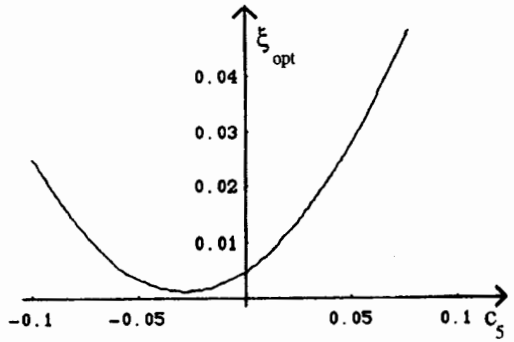


Figure 8
Fluctuation of the deteriorating factor ξ versus c_5
(c_3 is supposed to be the optimum one)

The graph which represents $\xi_{opt}(\lambda)$ shows a limit.

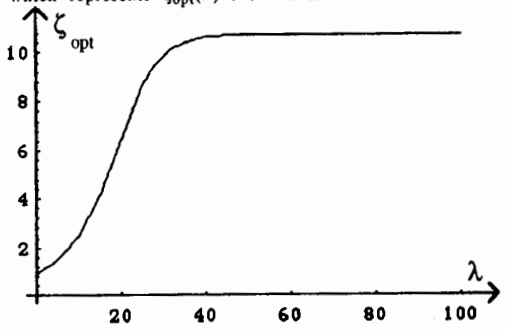


Figure 9
Improving factor ζ_{opt} versus gain λ

This limit is in fact an absolute limitation of the improvement of the three-harmonic optimisation combined with a notch filter. The absolute limit of the improving factor is ζ_{opt}.

$$\zeta_{opt} = 10.75$$

The process requiring both the optimisation of the demodulation function and the rejection of the component $q_2(t)$ is very promising. An improvement by a factor 11 is of a high interest. It could significantly reject the actual limitation effect far enough to satisfy a lot of future experiments.

V.3. Comparison

It seems necessary to sum up all the above results in a board to compare the different possible optimisations. The calculations have been done for a system using a square waveform frequency modulation. In order to allow the optimisation, no selective filter had been added at the output of the atomic resonator. Coefficients c_{2k+1} have been defined previously in such a way they gather both the filter and the harmonics of the demodulation waveform.

System (Optimised harmonics)	Improving factor ζ		
	Without rejection	Full rejection	30 dB rejection
0	1	1.89	1.89
c_3	1.00615	5.67	5.54
c_3, c_5	1.00695	10.75	9.89
c_3, c_5, c_7	1.00994	17.21	14.17
c_3, c_5,... c_{11}	1.01018	32.3599	19.73
c_3, c_5,... c_{21}	1.01037	83.653	23.81

Figure 10

Optimisation comparisons

Those results have been established still considering the quasi-static assumption and a square wave modulation.

VI. CONCLUSION

The limitation involved by the intermodulation effect is up to now a very worrying problem. In most systems, this limitation seems to be just at the edge of the actual affordable possibilities. The method exposed in this paper is a way to strongly decrease the intermodulation effect. Several cases of optimisation have been studied.

It has been shown that the combination of both the optimisation principle and the notch filtering could afford a very effective improvement. With presently available USO's at 5 or 10MHz, a frequency stability limit at the level of $\frac{1.10^{-14}}{\sqrt{\tau}}$ seems possible with a square wave frequency modulation at 40Hz. It is also proved that taken apart, these two methods are not efficient enough to significantly improve the frequency stability limit.

Certainly not so far from reality the results will anyway have to be validated by experiments. The dynamical aspects of frequency standards will have to be considered as well.

VII. ACKNOWLEDGEMENTS

We are very grateful to **DRET** (Direction des Recherches, Etudes et Techniques) who has supported this work.

We also are very grateful to **Omega Technologies** deeply involved in our studies.

We gratefully acknowledge **Professor Michel Dubois-Violette** who helped us with mathematical considerations.

VIII. REFERENCES.

[1] C.Audoin, V. Candelier, and N. Dimarcq
"A Limit to the Frequency Stability of Passive Frequency Standards due to an Intermodulation Effect".
IEEE Trans. on Instr. and Measur. 40, n°2, april 1991, pp 121-125.

[2] R.Barillet,V.Giordano, J.Viennet and C.Audoin
"Microwave interrogation frequency noise and clock frequency stability: experimental results"
Proceedings of : 6 TH EUROPEAN FREQUENCY AND TIME FORUM Noordwijk ,17-19 march 1992 (ESA SP340, june 1992)pp.453-456.

[3] G. Kramer "Noise in Passive Frequency Standards."
CPEM 1974 Digest, IEEE Conference Publication n°113,
IEEE Catalog n° CH 0770-8IM (unpublished).

[4] L.P. Lowe, F.L. Walls and R.E. Drullinger.
"Ultra-High Stability Synthesizer for Diode Laser Pumped Rubidium"
Proceedings of the 1992 IEEE Frequency Control Symposium -
Page 183

[5] C. Audoin, V. Giordano, N. Dimarq, P.Cérez, P.Petit and G. Théobald.
"Properties of an optically pumped cesium beam frequency standard with $\phi=\pi$ between the two oscillatory fields."
Proceedings of : 7 TH EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL 16-17-18 MARCH 1993.

HIGH Q MICROWAVE CAVITY USED IN A FREQUENCY STANDARD COMPARISON BETWEEN ACTIVE AND PASSIVE CONFIGURATIONS

T. Awada, V. Giordano, J. Groslambert, M. Olivier
Laboratoire de Physique et Métrologie des Oscillateurs du C.N.R.S.
associé à l'Université de Franche-Comté-Besançon
32 avenue de l'Observatoire - 25044 Besançon Cedex - France

Summary

We have studied a microwave source using as a reference a metallic cavity previously used as an hydrogen maser cavity. A high Q factor could be expected. We have choosen to operate on the TE_{121} mode of the cavity ; the frequency of resonance is in the vicinity of 1 940 MHz, and the loaded Q factor is 48 000. Two different configurations have been tested. In the first one, the transmitted signal from the cavity is used in an oscillating loop. In the second system, the high Q cavity is utilized as a frequency discriminator : an external oscillator is then frequency-locked to the cavity. We describe the design of the two systems, and we analyze the phase noise spectrum of the output signal. It is shown that the 1/F noise spectrum of the microwave loop amplifier gives the most important contribution to the phase noise spectrum of the signal obtained with the active configuration. A better result has been obtained with the passive scheme, when the limiting noise is due to the 1/F frequency noise of the cavity.

1. Introduction

A frequency reference has to be used in a frequency standard to obtain a signal with high spectral purity. This reference is usually a high Q resonator presenting a high slope of the phase-frequency characteristic. Two different configurations can be used to obtain a reference signal. If we want to obtain an active standard, the resonator is mounted in a loop where an amplifier provides enough gain to obtain sustained oscillation. We can also use the resonator as a frequency discriminator. An oscillator can then be frequency locked on the reference cavity.

The contribution of the noise sources to the phase noise spectrum of the signal depends on the configuration. The frequency noise of the cavity gives an absolute limit to the performances of the system. The noise of the loop amplifier in one case or the noise of the phase detector in the other case, that is the passive scheme, gives a contribution to the signal spectrum which may become predominant. A comparison between the performances of the two systems has been done, using as a reference two high Q metallic cavities.

2. Reference cavity

We have at our disposal for this experiment two cavities which were previously used at the L.H.A. (Laboratoire de l'Horloge Atomique - France) to build an hydrogen maser. The cavities have the same diameter $D \sim 270.10^{-3}$ m, the heights H are slightly different : $H1 = 335.10^{-3}$ m and $H2 = 328.10^{-3}$ m.

The cavities are made with quartz, the inner walls are plated with silver (50 µm) and then gold (1 µm). Different modes of oscillation $TM_{\ell mn}$ or $TE_{\ell mn}$ have been studied. We have observed that the TE_{121} mode at a frequency of 1 939 MHz gives a f_oQ_o product equal to $1.5.10^{14}$, better than the f_oQ_o product concerning the classical mode TE_{011} at 1 420 MHz.

Figure 1 shows the magnitude and phase of $S_{21}(\omega)$ of the TE_{121} mode for one cavity. The signal is coupled in and out of the cavities using small loops which dimensions and positions on the upper end cap are adjusted to obtain the optimum coupling factor.

Some parasitic modes appear in the vicinity of the resonance frequency ; they have been attenuated, using metallic plunger conveniently placed on the end cap of the cavities. A microwave filter, which bandwidth is 80 MHz, is placed at the output of each cavity. It contributes also to the attenuation of the parasitic modes.

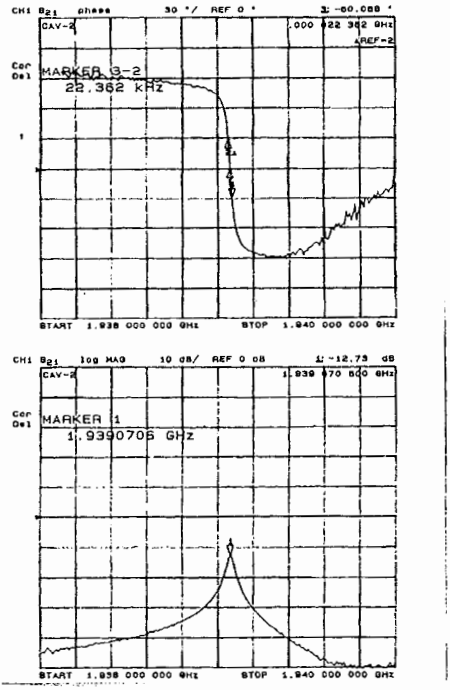


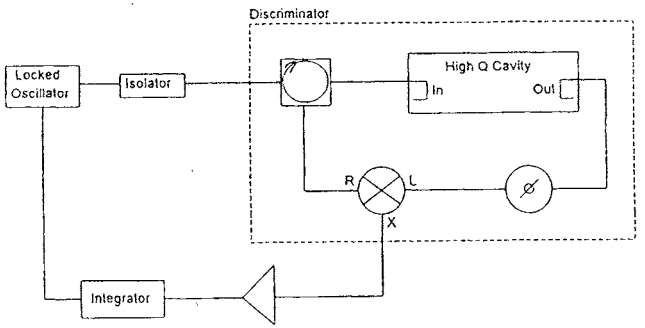
Fig. 1

3. Active system

We have realized two oscillators working at 1.94 GHz. The frequencies of oscillations are slightly different, owing to the different dimensions of the cavities. The frequency difference is of the order of 1 MHz. We use a microwave amplifier in the loop. The output power, at the output of the loop amplifier, is +13 dBm.

4. Passive system

Figure 2 represents the schematic of the passive system. We use both the reflected and the transmitted signals from the cavity in a discriminator scheme.



Frequency locked Oscillator

Fig. 2

The phase fluctuations of the reflected signal and of the transmitted signal can be expressed respectively by the following expressions ($\Delta\phi_r$ and $\Delta\phi_t \ll 1$ rad) :

$$\Delta\phi_r = 2\, Q_c \frac{\Delta f}{f_o} \frac{2\beta_1}{1 - \beta_1 + \beta_2}$$

$$\Delta\phi_t = -2\, Q_c \frac{\Delta f}{f_o} \quad ; \quad Q_c = Q_o \Big/ (1 + \beta_1 + \beta_2)$$

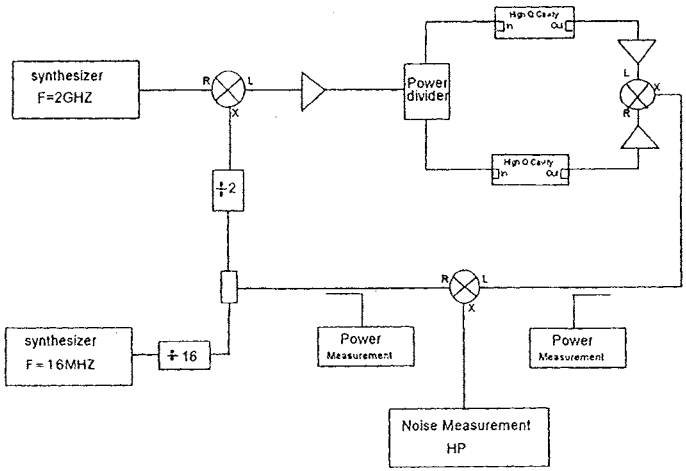
$\beta_1 = 0.2$ and $\beta_2 = 0.3$ are the input and output coupling coefficients of the cavities.

The theoretical sensibility $K_d = \Delta V/\Delta f$ of the discriminator can thus be obtained, using the measured sensibility $\mu = \Delta V / (\Delta\phi_r - \Delta\phi_t)$ of the phase detector. We have obtained $(K_d)_{th} = 1.4.10^{-5}$ V/Hz. The measured value of K_d is of the order of 9.10^{-6} V/Hz, in good agreement with the theoretical one.

5. Phase and frequency noise measurement

6.1. Cavity noise measurement

We have first measured the frequency noise of the cavity. The classical measurement system uses an oscillator for the excitation of two identical cavities placed in the two paths of a phase detector mount. In our case, the resonance frequency of the cavities being different, we have to use an auxiliary signal. We create two modulation lines separated from the microwave signal by half the frequency difference of the cavities (Fig. 3).



Cavity noise measurement

Fig. 3

The frequency noise of the microwave signal is rejected, because of the symmetry of the system. This is not the case for the noise of the auxiliary oscillator. We have used a frequency divider by 16 to enhance the spectral purity of this signal.

The phase and frequency noise of the cavity is shown in Fig. 4. We observe mainly a 1/F frequency noise. It can be attribute to the roughness of the walls of the cavity. A better polishing of the surfaces will certainly improve this

1/F noise level. The white noise which becomes preponderant for Fourier frequencies greater than 2 kHz may be due to the noise of the microwave amplifier at the output of the cavity.

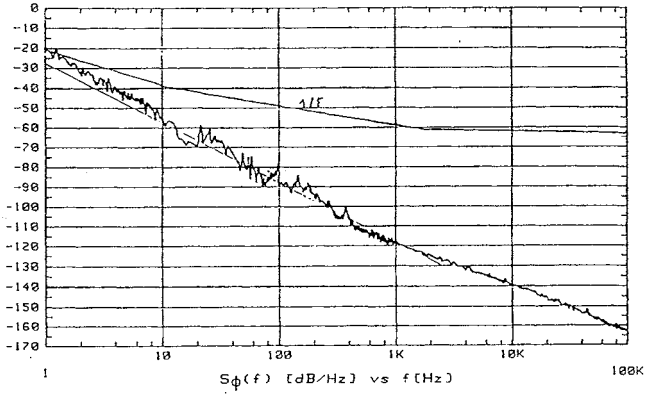


Fig. 4 : *Phase and frequency noise spectrum of the cavity*

Whe have also measured the phase noise spectrum of the measurement system of the divider and of the auxiliary signal at 1 MHz. These spectra are much better than the spectrum of the cavity (Fig. 5).

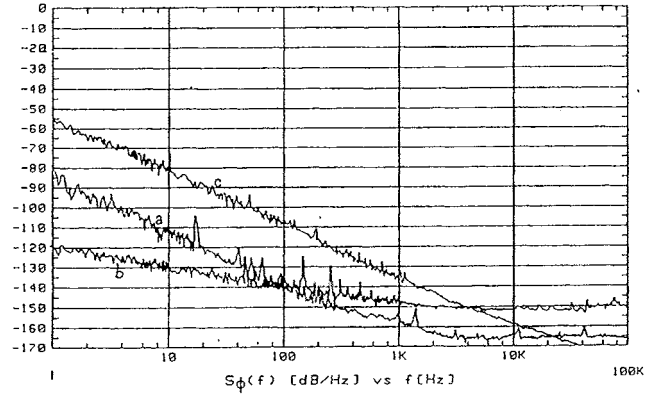


Fig. 5 : *Phase noise spectrum of the auxiliary signal (a), of the divider (b) and of the measurement system (c)*

6.2. Noise of the active system

The frequency of oscillation of the oscillators differs by 1 MHz. Here again we have to use an auxiliary signal, derived from a divided synthesizer output signal, to measure the phase noise spectrum of the oscillator (Fig. 6).

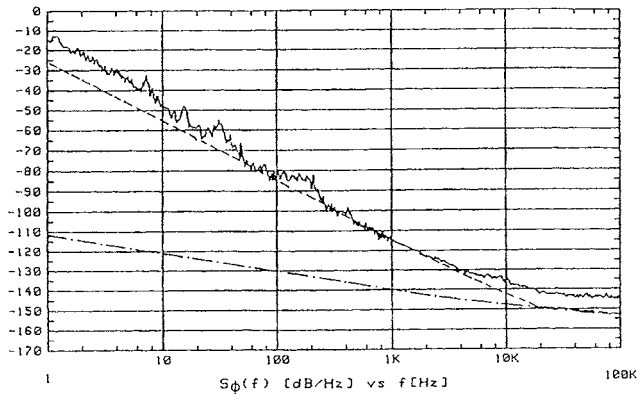


Fig. 6 : *Phase noise of an oscillator*
– – – – contribution of the noise of the amplifier – · – · – amplifier noise

We have tested the phase noise of the microwave amplifier inserted into the loop. It is essentially a 1/F noise, which is converted into a $1/F^3$ noise inside the bandwidth of the cavity. For frequencies between 10 Hz and 10 kHz, it corresponds to the noise of the oscillator.

6.3. Noise of the passive system

The phase noise of an oscillator locked on the cavity is shown on figure 7.

We have compared this noise spectrum which is a $1/F^3$ spectrum from 1 Hz to 2 kHz, with the frequency noise of the cavity, converted into a phase noise spectrum. We observe that the two spectra are quite identical, inside the locking bandwidth. That means that the noise of the reference cavity is, in this case, an absolute limit for the performances of the oscillator.

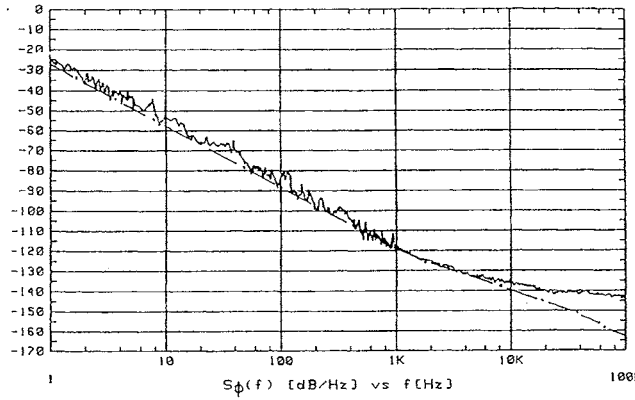


Fig. 7 : *Noise of a phase locked oscillator*
---- phase noise of the cavity

6. Conclusion

Two high Q metallic cavities, working at 2 GHz, are used to obtain a signal of high spectral purity. Two configurations, an active and a passive one, are tested.

The frequency noise of the cavity, which spectrum follows a 1/F law, limits in any case the performances of the system. The phase locked oscillator exhibits a noise spectrum corresponding to this limit. On the contrary the noise of the electronic contributes to the noise spectrum of the oscillator in the active configuration.

References

F.L. Walls and C.M. Felton (NIST) and T.D. Martin (Gravity Research Institute), "High spectral purity X-Band source", 44th Annual Symposium on Frequency Control, pp. 542-547.

G.J. Dick, J. Saunders and T. Tucker (California Institute of Technology, J.P.L. Pasadena, California), "Ultra-low noise microwave phase stabilizer using sapphire ring resonator", 44th Annual Symposium on Frequency Control, pp. 577-584.

B. Villeneuve, P. Tremblay, A. Michaud, M. Têtu (Université Laval, Québec), "RF spectrum of thermal noise and frequency stability of a microwave cavity-oscillator", IEEE Trans. on Microwave Theory and Techniques, vol. MTT-82, n° 12, dec. 1984.

Applications with Space Systems (B 6)

TIME ASPECTS OF THE EUROPEAN COMPLEMENT TO GPS: SOME EXPERIMENTAL RESULTS.

P.Uhrich, B.Juompan, G.Fréon, R.Tourde, P.Blondé.
Laboratoire Primaire du Temps et des Fréquences [LPTF]
Observatoire de Paris, Bureau National de Métrologie
61 Avenue de l'Observatoire, 75014 Paris, France

M.Brunet, J-F.Dutrey,
J.Barbier, J-L.Issler, N.Suard, T.Trémas.
Centre National d'Etudes Spatiales [CNES]
18 Avenue Edouard-Belin, 31055 Toulouse Cedex, France

Summary.
This paper presents the experimental results of the first phase of the CNES project of a European Complement to GPS /1/ involving CNES (Toulouse) and LPTF (Paris). It describes the synchronization of monitoring clocks using GPS Common-view or the C- to L-Band transponder of Inmarsat satellite. Data were collected during a few weeks of experiment through specially designed ground stations based on navigation GPS receivers. The estimated accuracy of Two-way time transfer through Inmarsat is 9ns (1σ), but the behaviour of the equipment generating the signal, which is the most degrading part in terms of uncertainty, is still under study. The restitution of GPS time through BlockI and BlockII satellites with an averaging method is also presented, with a reachable accuracy of 10ns (1σ) or better depending on the averaging time.

Introduction.
The CNES project of a European Complement to GPS (CE-GPS in the following) is dedicated to the needs of the Civil Aviation community to achieve the requirements of a non-precision approach phase with GPS being used as a sole navigation means. Many functions have to be fullfilled by such a system for which time requirements are reaching the state of the art of the techniques used by the Time Metrology community. The whole CE-GPS project started more than three years ago, and the experimental part about 18 months ago. The concept of the CE-GPS and the experimental system were presented at EFTF93 /1/.
At this point of the CE-GPS experiment, the first (calibrations of ground stations and spatial links measurements - May/June 93) and second stages (tests of the spatial links, clocks synchronization between two stations - September/November 93) have been performed, with a great amount of data to be processed. The results presented in this paper concern the calibrations of the ground stations, the synchronization of monitoring clocks using either Common-view GPS or Two-way time transfer through geostationnary satellite, and the demonstration of the feasability of GPS time restitution through multichannels GPS receivers, following the requirements of the CE-GPS project.

1. Short reminder of the CE-GPS project.

This chapter describes the main features of the European Complement to GPS. It is a summary of the paper given at EFTF93 /1/.

1.1. CE-GPS presentation.

The CNES project of a European Complement to GPS is mainly dedicated to the needs of the Civil Aviation Community. The reference mission adopted by CNES will:

- achieve the requirements of a non-precision approach phase.
- enable GPS to be used as a sole navigation means.

The functions to be fullfiled by such a system are:

- a monitoring mission: alarm of a failure on a GPS satellite within 10 seconds.
- a navigation mission: increase the GPS availability by eliminating coverage gaps.

Following proposals of satellite operators, it was agreed that the space segment would be provided by the Inmarsat III geostationnary satellites.

Because one of the system specifications is to minimize modifications on existing GPS receivers, the signal transmitted by the geostationnary payload shall be similar to a GPS signal. In addition it has to carry specific CE-GPS informations. This and other specifications have many consequences, concerning the navigation mission:

- synchronization with respect to GPS satellites. The standard deviation of the time difference between an event issued from the geostationnary payload and an equivalent event issued from a GPS satellite should be less or equal to 120ns with Selective Availability (S.A.) on. With S.A. off, this requirements drops to 20ns.
- frequency stability: better than 3.10^{-10} over 1s.
- code/carrier slippage: speed less than 0.2m/s, variations in pseudo-ranges (code/carrier) not greater than 10m over 1000s.

or concerning the monitoring mission: relative synchronization of monitoring clocks should be within 10ns (1σ) in accuracy if they are in view of the same geostationnary satellite or within 15ns (1σ) if not.

1.2. The need for an experimentation.

To prove the feasability of such a Complement to GPS, and to help estimating performances and limits of an operational system, the CNES has organized an experiment of which main objectives are:

- to confirm the ability of transmitting a signal similar to GPS from a geostationnary satellite.
- to demonstrate the feasability of synchronizing with GPS time a virtual clock on board the geostationnary satellite.
- to demonstrate the capability of GPS receivers to process the CE-GPS signal.
- to evaluate the User Equivalent Range Error (UERE) when using the geostationnary satellite.
- to synchronize the ground stations following the requirements.

The space segment of the CE-GPS experiment, beside the GPS NAVSTAR constellation, is the Inmarsat II F2 geostationnary satellite of which part of a transponder in the payload was made available free of charge to CNES for the duration of the experiment.
The ground segment is made of 3 specially designed ground stations based on 10 channels navigation GPS receivers, whose development has been entrusted to the IN-SNEC (Caen, France), and of a computer processing station for collected data located at CNES Toulouse Space Center.

1.3. Schedule of the experimentation.

The CE-GPS experimentation is divided into three stages.
A first stage in May-June 1993, with two stations located in Caen (France) at IN-SNEC in parallel with the same clock.
A second stage from September to November 1993 with one station at LPTF (Paris, France), and another one at CNES (Toulouse, France). This paper presents some of the experimental results obtained after processing of collected data.
A third stage is scheduled from May to June 1994, with three stations located at the CNES space centers of Toulouse (France), Kourou (French Guiana) and Hartebeeshoek (South Africa).

2. Time aspects of the CE-GPS.

The description of the system architecture, of the ground stations, and the discussion about time and frequency servo-control techniques, have been made elsewhere /1/. The corresponding results of the first and second stage of the experiment are shown in another paper presented at EFTF94 /2/. The time aspects of CE-GPS described in this paper are the restitution of the GPS time and the synchronization of ground stations clocks.

2.1. GPS time restitution.

The restitution of the GPS time is one of the basic feature of the CE-GPS project. Some publications have already been made on this subject /3//4/. The LPTF has proposed to use the most recently published results about the noise of GPS Block II satellites /4//5//6/. It was observed, by use of the Allan deviation on 15s sampled measurements of REF-GPS (comparison of a Cesium clock on the ground to GPS time as transmitted by one GPS satellite), that S.A. downgraded Block II satellites exhibit a white phase noise of $\sigma_x(\tau_0)$=84ns for τ_0=240s. This white phase noise behaviour last up to more than 2 hours.
By averaging [REF-GPS](τ_0) measurements made simultaneously with n satellites, over a sufficiently long time $\tau=p\tau_0$, yelding a value [REF-GPS](τ), it is then possible to reduce S.A. phase noise to the level

$$\sigma_x(\tau)=\frac{\sigma_x(\tau_o)}{\sqrt{n}\sqrt{p}}$$

Because of the low number of satellites available simultaneously, n is limited to 5 or 6 in the most general case /7/. If for example a limit of 10ns (considering S.A. noise only) is chosen for the accuracy of GPS time restitution, given 5 satellites for which REF-GPS measurements are possible simultaneously and using the numerical values given above, an averaging time of ~56mn is necessary to achieve these 10ns.

It shall be noted that the [REF-GPS](τ) averaged result is given for the mid-point of the averaging interval τ, for which it is the best estimate as the S.A. phase noise is white over τ. The stability of the clocks involved also allow the users to consider the [REF-GPS](τ) averaged result as the best prediction for the next interval of duration τ. The difference between two Cesium clocks is usually affected by white frequency noise, that is random walk phase noise. In this case the best prediction is indeed the current state of the signal.
Over one hour, the instability of the reference clock or the GPS time remain negligible. But the averaging was made considering only the S.A. noise: observation noise should also be considered for the estimation of the restitution accuracy.
The main advantage of this method is its simplicity. There are several drawbacks, most of them depending on S.A. pattern. It has been shown that some BlockII satellites may not follow the same behaviour as the others /4/. There is nothing in the STANAG /8/ which indicates the official form of S.A. noise (except its limits) for an operational constellation.
Of course, multichannels receivers are needed.

2.2. Synchronization of the ground stations through GPS.

The well known method for the synchronization of remote atomic clocks is to use the GPS Common-View technique /9/. Because GPS time receivers are included in the CE-GPS ground stations, it was decided to use the Common-View technique with the BIPM shedule as the reference for ground stations clocks synchronization, provided that a calibration of remote receivers is done.
An other way to proceed is to use the results of GPS time restitution separately calculated in remote stations /1/.

2.3. Synchronization of the ground stations through the geostationnary satellite.

Two-way time transfer through telecommunication satellite is also a well known technique /10/. It was proposed to use Inmarsat C- to L-Band transponder, and C/A GPS code for the transmitted signals /1/. Other related techniques will also be tested during the CE-GPS experiment: One way + ranging, Servo-control, or Orbitography methods /1/.

3. Experimental set-up.

The description of the ground stations will be limited to the basic equipments involved and to the items and techniques related to the results presented in this paper. The complete presentation of the stations of the CE-GPS experiment can be found elsewhere /1/.

3.1. Description of the ground stations.

Inside each ground station (figure 1) are implemented two SERCEL NR106, which are 10 channels GPS navigation receivers. They are related, thanks to the switcher 1, either to a common GPS antenna (L1 carrier), or to the receiving antenna of Inmarsat signal (L-Band) converted to L1 before the switcher.
The GPS antenna is also connected to a GPS time receiver SERCEL NRT2, which is a 4 channels receiver. The NRT2 is directly supplied with the 1pps output of the

Cesium clock of the station. This is the classical set-up for the GPS Common-view technique.
Because the navigation receivers NR106 have no input for dating external events, like the 1pps from the Cesium clock, it was necessary to built up a so-called 'GPS signals generator', which has two functions. It generates a sequence of C/A code synchronized with the 1pps output of the clock, which modulates a L1 carrier like a GPS signal in order to be dated by the internal counter of the NR106. This signal is denoted 1ppsL1-C/A. The C/A code chosen for this internal link for all CE-GPS stations is numbered 33: it is a spare for operational GPS, not used until further notice. For referencing any external signal to the Cesium clock of the station, the dating of this internal 1ppsL1-C/A signal has to be done simultaneously with the dating of the external signal. The other function of this 'generator' is to output the servo-controlled signal related to the 1pps of the clock to the transmitting antenna towards Inmarsat (C-Band). This is similar to the transmitting part of a Modem used for two-way time transfer, like MITREX for instance. The equivalent to the receiving part of a classical two-way Modem are the NR106 receivers.
There are opportunities for test links, short loop, and changes of the role of each NR106 receiver. There are other possible station configurations to be considered (see /1/), but for the Time experiments of the CE-GPS, a stable configuration was chosen for the whole period of data collection.

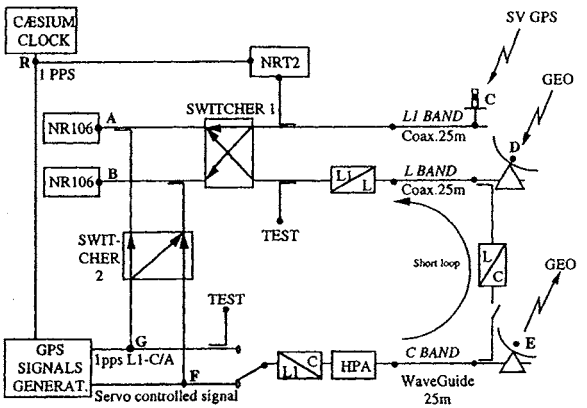


Figure 1. CE-GPS ground station.

The Cesium clocks monitoring the CE-GPS stations were manufactured by Hewlet-Packard, references 5061A option 4 at CNES and 5071A option 1 at LPTF. Both types have proven to remain stable enough to evaluate properly the performances of CE-GPS stations.

3.2. Calibrations.

One of the critical issues is the calibration of the signal generator. Examining the synchronization equations, it appears that the 1pps signal group delay through the generator must be monitored.

Frequent calibration sessions were scheduled in both stations during the whole experiment. A calibration consists roughly in measuring the time delay between the arrivals into the NRT2 of two homologous 1pps signals: a direct 1pps signal from the Cesium clock and the 1ppsL1-C/A signal supplied through the generator output. This procedure was possible because the NRT2 has proven to remain quite stable all over the experiment.
The results of the successive calibrations are presented on figure 2. The calibrations at CNES station appear more stable than the LPTF station measurements. Moreover the CNES station was switched off for a while on MJD 49266 (October 6): there are obviously two sets of data before and after this MJD, each scattered within 4ns. On the contrary the LPTF data are much more scattered within 25ns.

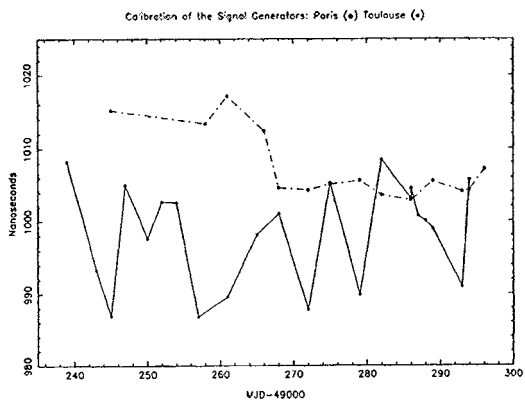


Figure 2. Calibrations of signal generators at CNES and LPTF.

It has been shown by the manufacturer of the stations that owing to the components used for this experiment, the signal generator could exhibit stepwise varying delays, each step being a multiple of 2.44ns. By processing in deferred time the dating of the 1ppsL1-C/A performed continuously by both NR106 of each station, as explained in chapter 3.1, those variations could be identified, summed and compared to the day to day calibrations of figure 2. Only a part of these records has been analized so far: some improvements could be achieved, but many problems remain under study.
For the purpose of this paper, these discontinuities in the delays are disregarded. The average values of calibration sessions were adopted along with the relevant uncertainties, giving for the generator delays:

- at CNES: Mean value=1007.5ns; Standard deviation=4.8ns.
- at LPTF: Mean value=998.5ns; Standard deviation=7.1ns.

Other internal delays of the ground stations were either measured (cables) or estimated (electronic components) following manufacturers data sheets with a good uncertainty. A short loop performed during the first stage of the experiment, along with a Two-way session with both stations connected to the same clock have shown that the measured differential delays were in good agreement with the theoretical values, given the estimated global uncertainty.

3.3 Atmospheric delays.

Either GPS or C- to L-Band radio techniques need estimations of the atmospheric delays. Models for tropospheric delays are working reasonnably well at the nanosecond level. For ionospheric delays, ionospheric calibrators are needed. AIR TEC100 (10 channels codeless receivers) were used at both CE-GPS stations for measuring ionospheric delays on GPS satellites signals. A polynomial mapping method was scheduled to be used to determine ionospheric delays in the direction of the Inmarsat geostationnary satellite.

We have experienced some problems during the running of the ionospheric calibrators. Due to the lack of time for this publication, and considering the small distance between both stations involved (~800km), it was decided to use STANAG results for ionospheric delays in the direction of GPS satellites, and a Bent model for C- and L-Band ionospheric delays in the direction of Inmarsat. Because the lines of sight of Inmarsat from both stations were very close, the difference of the C-Band ionospheric delays was negligible at the nanosecond level.

4. Clocks synchronization reference.

The GPS time receivers NRT2 of both stations were used for this clocks synchronization by the GPS Common-view technique. A calibration by transportation of two GPS receivers was organized. An other control of the results was possible thanks to the operational GPS Common-view link between the two stations involved, LPTF(Paris) and CNES (Toulouse), used for the computation of the TAF (Temps Atomique Français).

4.1. Calibration by transportation of GPS receivers.

A transportation of two GPS time receivers, one AOA TTR6 [TTR06] and one SERCEL NRT2 [NRT50], was organized during the second stage of the experiment, so that three round-trips were performed between LPTF and CNES. It has worked remarquably well, except for some local problems too long to explain here.

The analysis of the data was made in a very classical way /11//12//13/. Modified Allan deviations have justified the computation of daily average values. Because of the consistency of the averaged values, and using the estimated uncertainties of each period of measurement, it was possible to calculate a weighted average value for the calibration between the reference GPS receivers of both stations used for the computation of the TAF, LPTF[NBS51] and CNES[NRT03], giving the following results:

	Differential correction	Estimated uncertainty
UTC(OP)-UTC(CNES)	6.5ns	1.5ns(1σ)

These GPS receivers were different from those connected to the CE-GPS ground stations. But in both locations a direct comparison was possible over the whole period of the experiment, allowing to retain the above uncertainty as the differential uncertainty for GPS Common-view results between the time receivers connected to the CE-GPS stations. No temperature effect was noticeable on the set of data.

4.2. Common-view GPS between LPTF and CNES.

To recapitulate, here are the receivers located at both station (their caracteristics are well known to the Time Metrology community):

- At LPTF: NBS51 for TAF link.
  NRT28 connected to the CE-GPS station.
- At CNES: NRT03 for TAF link.
  NRT50 connected to the CE-GPS station.
  TTR06 for calibration periods.

Modified Allan deviations have justified for all the links between LPTF and CNES the computation of daily average values, as it is usual for short distances Common-views. These daily average values are presented on figure 3.

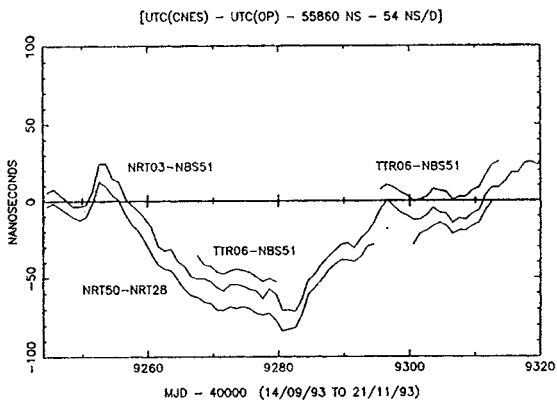


Figure 3. GPS Common-view between LPTF and CNES.

The average differences between the three curves are due to unrecorded changes of the internal delays of the receivers since last calibration, and to an additional delay of 15ns which was not taken into account between the clock and the CE-GPS station at CNES.

The state of the art for the accuracy of GPS Common-view time transfer between stations less than 1000km apart is estimated to be of the order of 3 to 5ns (1σ). It remains the operational reference for any other technique tested on the same link between two stations.

5. Two-way time transfer through Inmarsat.

Until now, Two-way satellite time transfer was performed using Eutelsat or Intelsat telecommunication system, or domestic satellites. When using MITREX Modems over short bases (~800km), an accuracy of 1.7ns has been estimated, and directly compared to equivalent GPS Common-view results /10/.

5.1. Spatial link.

Inmarsat II F2 is located -15.5°E. Both antenna coordinates are well known: within 20cm relative to ITRF90. The locations can roughly be given by: LPTF λ=2.333 °E, ϕ=48.867 °N; CNES λ=1.482 °E, ϕ=43.559 °N. The Sagnac effect for a link between LPTF and CNES is 2.2ns, with an uncertainty within 0.1ns (1σ).

Both stations are identical in terms of spatial link:

Antenna diameter C-Band: 0.6m
Uplink frequency: 6428.475 MHz (C-Band)
Uplink S/C G/T: -14dB/°K
Antenna diameter L-Band: 1.2m
Downlink frequency: 1533.475 MHz (L-Band)
Downlink on axis G/T: 1.3 dB/°K
Maximum EIRP: 39.8 dBW

At the start of the experiment, no data were available concerning Two-way time transfer performances when using Inmarsat transponder, C/A code, and GPS receivers as Modem. Because of the use of SERCEL GPS navigation receivers as Modems, the basic sampling of the data is 0.6s. It was decided to schedule four sessions per day each lasting 15mn to allow statistical analysis on a sufficient amount of data, and to detect any influence of atmospheric parameters on the performances. Sessions took place at 1:15 TU, 7:15 TU, 13:15 TU, 19:15 TU.

5.2. Analysis of the data.

On the figure 4 are presented the raw data of one 15mn session, with the related Modified Allan deviation on figure 5. These two figures are representatives of the common results obtained during nearly all the Two-way sessions. White Phase Noise is preponderant over a period long enough to allow computation of an average value: like in the case of MITREX experiments, 2mn of raw data (basic sampling 0.6s) is enough for one measurement of the difference between the clocks involved. This result will be used for the next stage of the CE-GPS experiment.

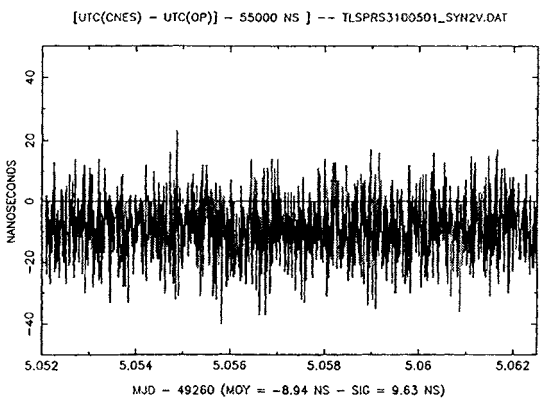


Figure 4. Two-way 15mn raw data of one session.

The precision of one 0.6s CE-GPS observation is about 8 to 9ns, as can be deduced from the Modified Allan deviation of figure 5. It could be compared to 1s MITREX observations that are 10 times better, but it remains consistent with the magnitude expected when using C/A code and standard discriminators.

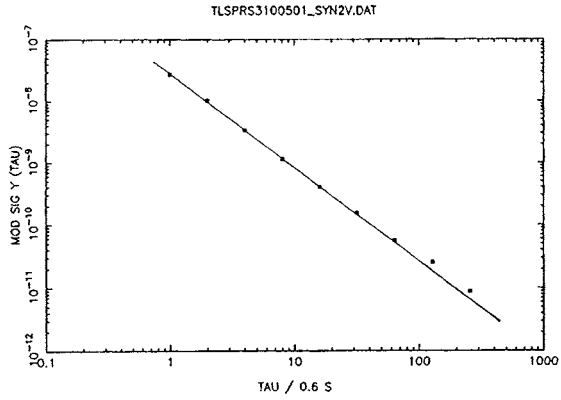


Figure5. Modified Allan deviation of the session on figure 6.

5.3. Accuracy of the method and comparison to GPS Common-view.

On figure 6 are presented the average values of the Two-way 15mn sessions, compared to the Common-view GPS daily averages between the GPS time receivers connected to the CE-GPS stations. One can see that the Two-way points are well distributed around the GPS curve, even though to much scattered.

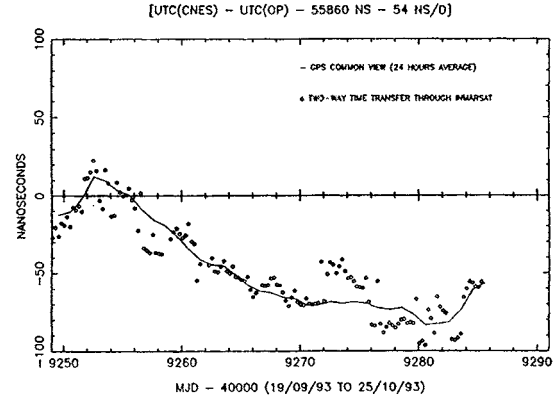


Figure 6. Two-way and GPS between LPTF and CNES.

It appears that many sessions have suffered from the above mentionned stepwise varying delays compared to the average values of the generators delays. This is a logical consequence of the choice made in chapter 3.2 in averaging the delay measurements. One can even see on figure 7 that these steps may happen during a session of measurement.

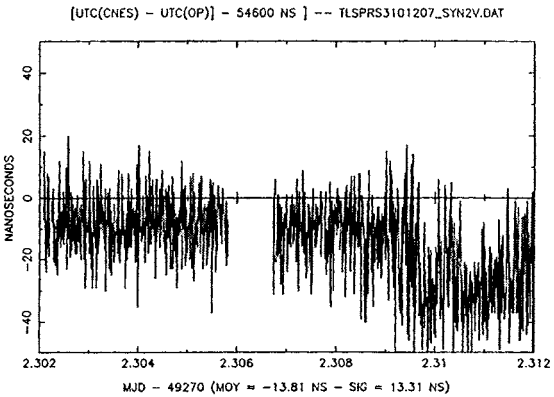


Figure 7. An example of a perturbated Two-way session.

The uncertainties of the generators delays are obviously the most degrading part for the accuracy of this experiment of Two-way time transfer through Inmarsat. We propose the following error budget in nanosecond (1σ):

Inmarsat transponder	0.0	[Global coverage]
Sagnac effect	0.1	
Ionospheric delays	2.0	[Differential/Bent model]
Tropospheric delays	0.5	[Differential/STANAG model]
Differential delays (LPTF-CNES)	8.7	[1ppsL1-C/A generator]
UTC(LPTF)-1ppsREF	0.5	[short line]
UTC(CNES)-1ppsREF	1.0	[long line]
Two-way accuracy	9.0	[Quadratic sum]

This accuracy is near the upper limit of the CE-GPS requirements (10ns). It surely will drop to a more interesting value compared to MITREX if the behaviour of the 1ppsL1-C/A generator could be withdrawn from the measurements: this is now the next step of the processing of this part of the experiment.

6. GPS time restitution.

The programming of the 10 channels NR106 receivers of both stations was done in the following way:

- One channel in each NR106 for the internal 1pps datation.
- One channel for short loop or test dating of the transmitted code.
- One channel dedicated to the remote station code.
- Two channels scheduled to track the same GPS satellites in both NR106 for redundancy measurements to control the behaviour of each NR106 compared to the other.
- One channel scheduled to track GPS satellites following the international BIPM Common-view schedule.

This leaves 11 channels for tracking other GPS satellites (including some overlapping between sessions), in order to increase as much as possible the number of GPS satellite for which a measure of REF-GPS is available simultaneously.
The basic sampling of NR106 is a mesure of pseudo-range every 0.6s. Because NR106 are navigation receivers, there is no software for processing REF-GPS data like in Time dedicated receivers. To minimize the short term effects and also the amount of data to be processed, it was proposed to perform the restitution of GPS time on [REF-GPS](15s), meaning realization of REF-GPS over 25 measurements made every 0.6s, case for which previous studies have shown analysis easiness, and corresponding to the usual short term processing of Time dedicated receivers. A software for processing the short term measured pseudo-ranges is needed, and will soon be available, but was not ready to use at the date of this publication.
To prove the capability of the method described in chapter 2.1, we have decided to use the NRT2, Time dedicated receiver, which is a 4 channels receiver.

6.1. Use of a 4 channels receiver.

Thanks to Dr C.Thomas (BIPM), a schedule of 89 tracks per day and per channel was implemented in the 4 channels of the NRT2 receiver located at LPTF. The experiment lasted from the 30th of October to the 22nd of November 1993. Each track lasts 16mn (like the usual Common-view tracks) as following: 2mn for initialization of the track, 13mn of measurements based on short-term 15s points (basic sampling 0.6s), and 1mn for computation of the result of the track.

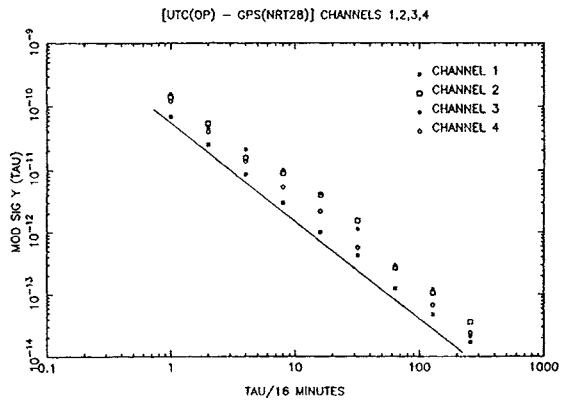


Figure 8. Modified Allan deviation of the 4 channels.

Analysis of the Modified Allan deviation [Mod$\sigma_y(\tau)$] on each channel showed for τ=16mn a value slightly above 10^{-10} (figure 8). The corresponding value of σ_x(16mn) is then slightly superior to 55.4ns. We have selected as standard deviation σ_x(16mn)$\simeq$60ns. The better Allan deviations obtained on channel 1 are explained by the fact that there are BlockI satellites on this channel only.
If we want to reduce the S.A. phase noise to the level of 10ns, given n=4 channels able to track simultaneously GPS satellites, we use the formula given at chapter 2.1

which gives the period of averaging $\tau=p\tau_0$ with p=9: for τ_0=16mn the value of the period of averaging is τ=2h24mn. This result means that if we average the measured [REF-GPS](16mn) on 9 consecutive tracks on the 4 channels, we will reduce S.A. phase noise to 10ns(1σ). It can be verified on figure 8 that the white phase noise is still preponderant for the period of averaging τ=2h24mn.
Figures available about GPS time stability show that over 2h24mn proper variations of the phase of GPS time are of the order of a few ns. It can be assumed that it remains negligible at the 10ns level in the error budget of GPS time restitution.
Again, this reduced S.A. noise is given for the mid-point of the averaging interval, but remains the best prediction for the next interval of 2h24mn. With this method, we claim to be able to perform the restitution of GPS time on the ground at 10ns(1σ), with one point every 2h24mn, considering S.A. noise only.

6.2. Comparison to Block I measurements.

One experimental way to verify this assessment is to compare the results of the averaging as described above to [REF-GPS](16mn) measurements made with BlockI satellites, which are not subject to S.A.. At the time of the experiment, there were only 3 BlockI satellites still available, PRN 3/12/13, which were included in the 89 tracks scheduled in the NRT2 on channel 1 only.
It was decided to compare each [REF-GPS](16mn) BlockI measurement to the averaging of all the [REF-GPS](16mn) placed on the 9 lines around the one containing this BlockI measurement in the results files.

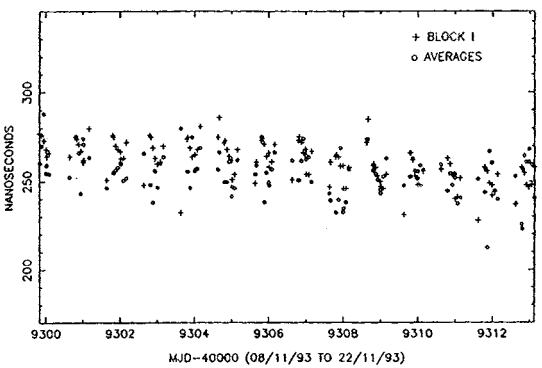


Figure 9. Plots of BlockI and averaged [REF-GPS].

On figure 9 are the plots (+) of BlockI [REF-GPS](16mn) measures and the plots (o) of averaged [REF-GPS](2h24mn). It can be seen that the scatter of the plots is roughly the same for both types of values, even if the [REF-GPS](2h24mn) values seem in average slightly above the BlockI values.
On figure 10 are plotted the difference between the BlockI [REF-GPS](16mn) and the averaged [REF-GPS](2h24mn). Obviously nearly all the points are included inside the $\pm3\sigma=\pm30$ns limits, with a few exceptions towards the end of the

experiment. Very few points were withdrawn, after checking that the bad value was caused by the BlockI satellite only, or that the number of satellites available for averaging was much too small.

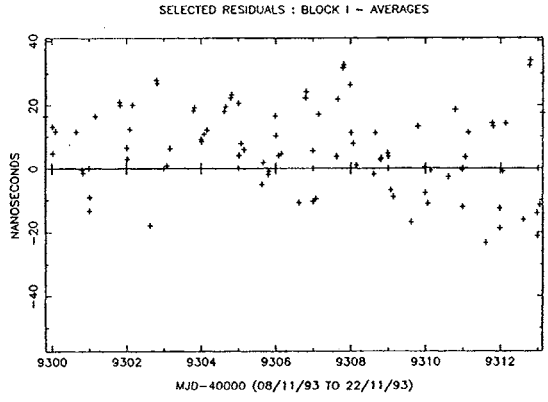


Figure 10. Residuals of averaged [REF-GPS].

6.3 Accuracy of the restitution of GPS time.

When looking carefully at figure 9, one can see that the BlockI values exhibit roughly the same behaviour from day to day, with a scatter of about 50ns. It is well known that remaining BlockI satellites are getting old now, and that the performances obtained with these satellites are not the highest possible when measuring [REF-GPS]. It has also been discovered that this type of GPS receivers were subject to daily temperature effect /14/, even if no effects of that kind were noticeable during the calibration of the GPS receivers involved in this experiment.

The presence of some residuals outside the $\pm 3\sigma$ limits (figure 10), beside the usual behaviour of the statistics, could also be explained by the fact that the processing of the data was made considering S.A. noise only. There are other sources of noise when measuring [REF-GPS] data, which could be called 'observation' noises. The scatter of BlockI data is representative of the effect of these observation noises, of which the major part cancels when processing Common-views for clocks synchronization.

Nevertheless, the results obtained are in agreement with the hypothesis made in chapter 2.1, indicating that S.A. behaviour is still the same as was observed in previous papers /4//5//6/. It has then been again demonstrated that the restitution of GPS time on the ground was possible with this averaging method, providing an accuracy of 10ns (1σ) when considering S.A. noise only. The averaging period depends directly on the number of satellites simultaneously available, following the equation given in chapter 2.1.: to reduce the averaging period, the number of satellite must increase.

With the use of NR106 data, one of the main feature of CE-GPS could be fullfilled very easily. The requirements of CE-GPS was about the GPS time as transmitted by the geostationnary satellite. To the S.A. noise mentionned above should be added the

observation noise, roughly 10ns (1σ), and the noise of the transmission of GPS time to the geostationnary payload, which could be realized within 10ns (1σ). With the S.A. noise implemented today on the GPS, it should be possible to fullfil the requirements of CE-GPS by restituting GPS time at the geostationnary payload with an accuracy of 17ns (1σ): it remains within the requirement with S.A. off (20ns(1σ)), and by far inferior to the requirement with S.A. on (120ns(1σ)).

7. Future of the CE-GPS experiment.

The processing of the collected data during the second stage of the experiment has just started, and the third stage of the experiment is about to begin.

7.1. Processing the data of the second stage.

The most urgent need is the software for computing the [REF-GPS](15s) values from the 0.6s sampled pseudo-ranges made by the NR106 receivers. It will allow computation of Common-view GPS the same way as in GPS receivers dedicated to time measurements, making the GPS time receivers NRT2 connected to the CE-GPS stations useless. Because NR106 are 10 channels receivers, it will supply data for the restitution of GPS time with the highest number of GPS satellites available simultaneously, reducing the averaging period.

Beside Two-way time transfer, other techniques are scheduled to be tested with signals transmitted by the Inmarsat transponder /1/. Among others the servo-control could be the most interesting due to its near real-time time transfer capability. In this case however the synchronization is less accurate.

7.2. Third stage of the experiment.

The third stage of the experiment is scheduled for May-June 1994. The same testing program is anticipated, with two noticeable differences:

- The distances between the stations will be roughly 10000km. In terms of Time Metrology, it is well known that performances of time transfer are going to be degraded.
- With three stations involved, a direct test of accuracy will be possible: by computing the differences of the three stations together, a measurement of the deviation from closure will be given. Even if it cannot be considered to be an evaluation of the accuracy of a two stations link, it will give precious informations about the time transfer methods involved in CE-GPS.

It should be noted that with three monitoring stations, measurements about orbitography of the geostationnary satellite could be achieved.

Conclusions.

It has been demonstrated that the time requirements of the CE-GPS could be fullfilled with ground stations comparable to those built up for the purpose of the experiment presented in this paper, at least for a distance between the stations inferior to 1000km.

Even with the calibration problems related to the signal generator as conceived for this experiment, the monitoring clocks synchronization could be done by the Two-way technique through Inmarsat with an accuracy of 9ns (1σ). It can be foreseen that an even better accuracy could be obtained by processing not yet considered data.

The restitution of GPS time should be feasible with an accuracy close to 10ns (1σ) when considering S.A. noise only (17ns on board the geostationary payload, following CE-GPS requirements) depending on the averaging time, provided the S.A. behaviour will remain similar in the future. Among other considerations, this would allow any geostationnary satellite transmitting a CE-GPS signal to be considered by the users similar to any other GPS satellite.
The next stage of the CE-GPS experiment will explore the consequences of using three monitoring stations, roughly 10000km apart from each other.

Acknowledgements.
The authors gratefully acknowledge Dr C.Thomas (BIPM) for many fruitfull discussions about the restitution of GPS time, and for having provided the schedule of 89 sessions/day over four channels for the NRT2 receiver. The authors thank J.Berthier (Observatoire de Paris) for his help in the processing of raw data.

References.

/1/ J.Barbier, T.Tremas (CNES). European Complement to GPS: presentation of the concept and experimental system. Proceedings of the 7th EFTF, Neuchâtel, March 93.

/2/ J.Barbier (CNES). CE-GPS-Experiment-Stage 1 Time & Frequency Servo Control Loops Practical Results. Same Proceedings EFTF94.

/3/ C.Thomas (BIPM). Real-time Restitution of GPS Time Through a Kalman Estimation. Metrologia 1992, 29, 397-414.

/4/ C.Thomas (BIPM). Real-time restitution of GPS time. Proceedings of the 7th EFTF, Neuchâtel 16-18 March 1993.

/5/ D.W.Allan (Allan's TIME) and A.Lepek (Tech Projects). Trends in International timing. Proceedings of the 7th EFTF, Neufchâtel 16-18 March 1993.

/6/ C.Thomas (BIPM). The use of the AOA TTR-4P GPS receiver in operation at the BIPM for real-time restitution of GPS time. Proceedings of the PTTI 1993 (to be published).

/7/ C.Thomas (BIPM). Private communication.

/8/ STANAG. Technical characteristics of the NAVSTAR GPS. June 91.

/9/ W.Lewandowski, C.Thomas (BIPM). GPS time transfer. Proceedings of the IEEE, 1991, 79, 991-1000.

/10/ D.Kirchner (TUG), H.Ressler (IWF), P.Grudler, F.Baumont, C.Veillet (OCA), W.Lewandowski (BIPM), W.Hanson (NIST), W.Klepczynski (USNO), P.Uhrich (LPTF). Comparison of GPS Common-view and Two-way Satellite Time Transfer Over a Baseline of 800 km. Metrologia 1993, 30, 183-192.

/11/ W.Lewandowski (BIPM). Rapport BIPM 91/6, November 1991.

/12/ W.Lewandowski (BIPM), M.Imae (CRL), C.Thomas (BIPM), M.Aida (CRL). Rapport BIPM 93/4, June 1993.

/13/ C.Thomas and P.Moussay (BIPM). Rapport BIPM 93/5, June 1993.

/14/ W.Lewandowski (BIPM), R.Tourde (LPTF). Sensitivity to External Temperature of some GPS time Receivers. Proceedings of the 22nd PTTI, 1990, 307-316.

8TH EUROPEAN FREQUENCY AND TIME FORUM

Technical University Munich, Germany
March 9-11 , 1994

THE COSPAS-SARSAT SEARCH AND RESCUE SYSTEM

System description and performances

M. Brunet, G. Portas, C. Augoyard, C. Gal, L. Fourcade,
CNES, Centre National d'Etudes Spatiales, Toulouse, France.

ABSTRACT

The goal of the Cospas-Sarsat satellite system is to detect and localize distress beacon transmissions from boats, planes or persons (scientific expedition, ...) in distress and to provide alert and location data to appropriate authorities responsible for Search And Rescue. Its ability to localize not only the 406 MHz distress beacons designed as part of the satellite system, but also, with some restrictions, the aviation 121.5 MHz Emergency Locator Transmitters (ELT) designed for human detection (by overflying planes or ground stations) is one of its advantages.

The system is made of a constellation of at least four low altitude near polar orbit satellites carrying Search And Rescue payloads and of a ground segment. The payloads comprise repeaters for the 121.5 MHz signals and receiver-processors for the 406 MHz. The ground segment is composed of stations to process the satellite signals for distress message detection and location and Mission Control Centres to validate locations and forward the distress data to appropriate Rescue Coordination Centres.

The Cospas-Sarsat programme was developed by four national agencies, DOC (Canada), CNES (France), MORFLOT (Soviet Union) and NASA (USA), in constant relation with the International Maritime Organization (IMO) and the International Civil Aviation Organization (ICAO). On July 1, 1988, the four founder countries signed and Intergovernmental Agreement guaranteeing system continuity for 15 years, to be extended in five-year periods (now the Russian Federation has replaced the USSR). Other countries have joined the system, either as ground segment providers or as user states.

Today, more than 3,400 persons have been saved by the Cospas-Sarsat satellite aided operations since its beginning in 1982.

This paper describes the general characteristics of the system, the performances of the receiver-processor Sarsat USOs, in orbit and on the ground, as well as the characteristics of the 406 MHz distress beacon oscillators.

1. THE COSPAS-SARSAT PROGRAMME

The rapid detection and location of an aircraft crash or marine distress is of paramount importance to the potential survivors and to Search And Rescue (SAR) teams. Studies show that while the initial survivors of an aircraft crash have less than a 10 % chance of survival if rescue is delayed beyond two days, the survival rate is over 50 % if the rescue can be accomplished within eight hours. Similar urgency applies in marine distress situations, particularly where injuries have occurred. Furthermore, rapid location can significantly reduce both SAR costs and the exposure of rescue forces to hazardous conditions.

In this context, the objective of the Cospas-Sarsat programme is to provide the SAR organizations with the detection and location of distress signals by global monitoring based on satellites in near-polar orbits. Complete coverage of the earth, including the polar regions, is thus achieved using simple emergency transmitters installed on the mobile units.

The Cospas-Sarsat system development started in 1979 after an agreement was signed between agencies in Canada, France, USA and USSR. The first Cospas and Sarsat satellites were launched respectively in 1982 and 1983. The system was declared operational in 1985 and on July 1st 1988, the four states providing the space segment signed the international "Cospas-Sarsat Programme Agreement" which ensures the continuity of the system and its availability to all States on a free of charge and non-discriminatory basis.

The International Maritime Organization (IMO) and the International Civil Organization (ICAO) now recommend that ships and aircraft carry distress beacons.

2. THE COSPAS-SARSAT SEARCH AND RESCUE SYSTEM

2.1. Cospas-Sarsat System Overview

2.1.1. System Concept

The basic Cospas-Sarsat system concept is illustrated by Figure 1 (ref. /1/). The signals, transmitted by three types of distress beacons, ELTs : aviation, EPIRBs : maritime, and LMSBs : land, are received by the Search And Rescue payloads installed on satellites. Two Cospas satellites and payloads are provided by the Russian Federation which is maintaining full responsibility over all rights and obligations of the USSR in the Cospas-Sarsat programme. Two SARSAT satellites provided by the USA, in fact meteorological spacecraft financed by NOAA, are equipped with payloads furnished by Canada and France. The distress signals are relayed by the satellites to ground receiving stations termed Local User Terminals (LUT) and allow them to determine the beacon location. Distress site location is then transmitted via a Mission Control Centre (MCC) either to another MCC or to the appropriate Rescue Coordination Centre (RCCs) to initiate the SAR activities.

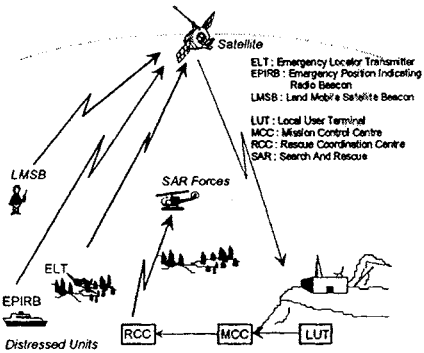


Figure 1 : Basic Concept of the Cospas-Sarsat system

A single satellite circling the earth around the poles eventually views the entire earth's surface. The orbital plane of the satellite, remains fixed while the earth rotates underneath. At most, it takes only one half a rotation of the earth (i. e. 12 hours) for any location to pass through the orbital plane. A satellite at an altitude of 850 km like Sarsat, or 1000 km like Cospas, views a "swath" of the earth over 4000 km wide as it circles the globe, giving an instantaneous "field of view" about the size of a continent. The Cospas-Sarsat constellation, which consists of four satellites, thus reduces greatly the time for a beacon to be seen by the next passing satellite (average waiting time $\approx$ 1 h 30 min).

Due to the relative motion between the satellite and the beacon, the carrier frequency received by the SAR payload varies during the mutual beacon-satellite visibility duration (generally comprised between 10 and 15 minutes). The carrier frequency transmitted by the beacon being reasonably stable during this time, the variation of the received frequency, known as the Doppler effect, allows the LUT to compute the beacon's location. The Doppler location provides two positions : the true position and its mirror image relative to the satellite ground track. This ambiguity is resolved by calculations that take into account the earth's rotation.

Two uplink frequencies are currently used :

- the 121.5 MHz aeronautic emergency frequency which is also allocated to distress beacons,
- the 406.0 - 406.1 MHz, exclusively reserved for distress beacons operating with satellite.

The first frequency corresponds to beacons already in widespread use before the implementation of the Cospas-Sarsat system. The beacons using the UHF band, designed for the Cospas-Sarsat system, have a better frequency stability and a unique digital identification code.

2.1.2. Coverage Modes

The Cospas-Sarsat system implements two coverage modes for the detection and location of beacons, namely the regional mode and the global mode. Both the 121.5 MHz and 406 MHz systems operate in the regional mode, while the later also operates in the global coverage mode providing complete earth coverage.

Regional mode : 121.5 MHz repeater
A repeater on board the spacecraft relays the 121.5 MHz signals directly to the ground. If a LUT is in view of the spacecraft, the signals can be received and processed to determine the transmitter locations. The 121.5 MHz regional mode has proven extremely effective in providing quick alert and location data for the large number of existing 121.5 MHz distress beacons in the coverage area of LUTs. However, the transmission characteristics of these beacons do not permit satellite onboard processing of the signal which would be required for global earth coverage.

Regional mode : 406 MHz processed data
Once the satellite receives the 406 MHz beacon signals, the Doppler shift is measured and the beacon digital data recovered. This information is time-tagged, formatted as digital data, and transferred to the repeater downlink for regional transmission to any LUT in view (the data are also simultaneously stored, see hereafter).

Global mode :
The 406 MHz system provides global coverage by storing data, derived from onboard processing of beacons signals, in the spacecraft memory unit. The memory is continuously dumped for reception of 406 MHz stored data by LUTs when the satellite comes into visibility. Each beacon can therefore be located by all operating LUTs, this provides ground segment processing redundancy and reduces global system processing time.

2.1.3. Distribution of Alert and Location Data

The alert and location data generated by LUTs are validated and forwarded to appropriate Rescue Coordination Centres (RCCs) through the Cospas-Sarsat MCC network. Since a single distress incident is usually processed by several LUTs, in particular in the 406 MHz global mode of operation, the alert and location data are sorted by MCCs to avoid unnecessary transmission of identical data. The Cospas-Sarsat ground network architecture allows to take advantage of the high built-in redundancy in alert data distribution.

2.2. Description of Main Sub-Systems

The figure 2 gives a block diagram of the complete Cospas-Sarsat system. It comprises three basic sub-systems :
- the beacons (ELT, EPIRB, LMSB), which transmit distress signals at 121.5 or 406 MHz,
- the satellites (COSPAS or SARSAT) which relay and process the distress signals using a downlink frequency of 1,544.5 MHz,
- the ground segment which comprises receiving processing stations (LUTs) and Mission Control Centres (MCCs) in contact with the Search And Rescue (SAR) networks.

The following description will be mainly centered on the Sarsat 406 MHz system.

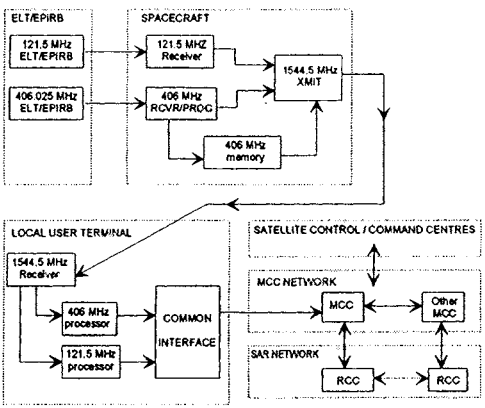


Figure 2 : Cospas-Sarsat Block Diagram

2.2.1. Beacons

The 121.5 MHz beacons :

Although no registration exists, the current number of 121.5 MHz beacons is estimated to be over 500 000 worldwide. Most of these units are used aboard light aircraft and are required to meet national specifications based on ICAO standards. These standards were not established with the aim of satellite reception. Therefore, the 121.5 MHz Cospas-Sarsat system was designed to serve the existing type of beacons, even though system performance like capacity and location accuracy would be limited. Nevertheless, the efficiency of 121.5 MHz beacons has been greatly enhanced by the use of satellite detection and Doppler location techniques.

The main characteristics of these beacons are :

- frequency : 121.5 MHz ±6 kHz
- transmitted power : 50 to 100 mW
- transmission life : 48 hours at - 20°C (less than - 20°C in special cases)
- modulation : AM by sweeping audio signal.

The 406 MHz beacons :

The development of a new type of beacon, operating at 406 MHz, began at the start of the Cospas-Sarsat project. These new units were designed specifically for satellite detection and Doppler location with the following design objectives :

- global coverage,
- improved location accuracy and ambiguity resolution,
- unique identification of each beacon allowing identification of carrier vehicles,
- increased system capacity (i. e. capability to process a greater number of beacons transmitting simultaneously in the satellite field of view).

These beacons transmit a 5 W burst of approximately 0.5 second duration every 50 seconds. The carrier frequency is very stable and the pulse is phase-modulated with a digital message. Improved frequency stability provides location accuracy, while the high peak power assures a high probability of detection. The low duty cycles provides a multiple-access capability of 90 beacons simultaneously operating in view of the satellite, and low energy consumption. An important feature of the new emergency transmitter is the addition of a digitally encoded message, which provides such information as the country where the beacon is registered and identification of the vessel or aircraft.

Because the performance of the 406 MHz system depends on actual transmission characteristics of the beacons, Cospas-Sarsat has implemented a beacon type approval procedure. It is the manufacturer's responsibility to obtain type approval of its beacons and to comply with specific national regulation.

The main characteristics of these beacons are (see § 3 for more detail) :

- frequency : 406.025 ± 0.005 MHz
- transmitted power : 5 W ± 2 dB
- modulation : phase modulation, bi-phase L data encoding
- message duration : 440 ms or 520 ms.

2.2.2. Space Segment

The space segment comprises the two Cospas and the two Sarsat satellites equipped with fully compatible SAR payloads (ref. /2/). The Search And Rescue instrumentation comprises (see figure 2 above) the following basic sub-assemblies : (i) the 121.5 MHz receivers (for regional mode), (ii) the 406.025 MHz receiver-processor and memory unit (for regional and global modes) and (iii) the 1,544.5 MHz transmitter.

The 121.5 MHz receiver :

The 121.5 MHz receiver collects signals transmitted by the beacons in a 25 kHz band. After amplification and down frequency conversion, the signal is output to the 1,544.5 MHz transmitter for downlinking.

The 406.25 MHz receiver-processor and memory unit :

The functions of the 406 MHz receiver-processor are : (i) reception and demodulation of the digital messages received from beacons, (ii) measurement of the received frequency, (iii) time tagging the frequency measurement, (iv) formatting an output messages including the message digital data and the frequency and time measurements.

All these data are included in the output signal frame for downlinking to LUTs at 2400 bps in the regional mode, and simultaneously stored in memory. The onboard memory is dumped in the same format and at the same bit rate as regional data. LUTs directly receive the stored beacons messages acquired during previous orbits implementing thus the global mode.

The receiver-processor functional diagram of the Sarsat new generation payload is shown in figure 3. It comprises :

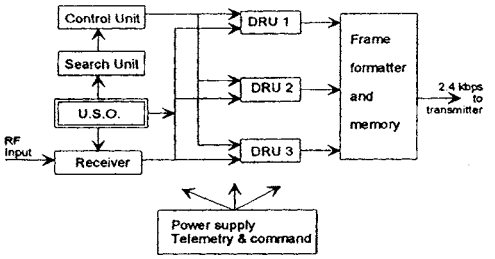


Figure 3 : Sarsat receiver-processor payload on NOAA satellites

- a UHF receiver centered on 406.050 MHz, an ultrastable oscillator (USO) which is the reference for the time counter and the Doppler frequency counter,
- a Search Unit (spectrum analyzer) for the detection of messages, three Data Recovery Units (DRU), each including a phase lock loop (PLL) to process simultaneously three messages,
- a Control Unit which monitors the processor, assigns the detected messages to the DRU available, according to an algorithm which optimises the message probability access,
- a telemetry formatter which creates the downlink message,
- a memory to store the messages.

The main characteristics and performances of the receiver-processor are summarised as follows :

- total mass of the units : 24 kg,
- power consumption : 20 W,
- receiver frequency band : 80 kHz, centered at 406.05 MHz,
- power of received messages : - 161 dBW to - 131 dBW,
- bit error rate : $\leq 2.10^{-5}$,
- Doppler frequency measurement : standard deviation $\leq$ 350 mHz,
- USO stability : $\leq 1.10^{-10}$/day and $\leq 1.10^{-10}$/100 ms (see hereafter §3).

All the Sarsat receiver-processors have been and are manufactured by the French Dassault Electronique company, from the NOAA-E through NOAA-N spacecraft, that is ten payloads.

The 1,544.5 MHz transmitter :

The 1,544.5 MHz transmitter (i) accepts input from the 406 MHz receiver-processor and receivers operating on the uplink Cospas-Sarsat band, (ii) phase modulates a low frequency carrier with the composite signal, (iii) multiplies the frequency to produce 1,544.5 MHz, (iv) amplifies the power level and transmits via the spacecraft downlink antenna.

Since 1982, ten satellites have been launched to maintain the nominal constellation of four satellites. The next launch is planned for September 1994.

2.2.3. Ground Segment

Local User Terminals (LUTs) :

The satellite downlink signal formats ensure interoperability between the various satellites and all LUTs meeting Cospas-Sarsat specifications.

In a LUT, a steered antenna and the receiving system pick up the signal which is down-converted to an intermediate frequency, and linearly demodulated to produce the composite base-band spectrum which is filtered and separated into the various bands of interest.

For the 121.5 MHz signals, each transmission is detected and the Doppler information calculated. A beacon position is then determined using these data. Processing of 2400 bps data (i. e. those generated from 406 MHz transmission) is relatively straightforward since the Doppler frequency is measured and time-tagged onboard the spacecraft. All 406 MHz data received from the satellite memory on each satellite pass can be processed within a few minutes of pass completion. In order to improve location accuracy, highly stable 406 MHz orbit determination beacons at accurately known locations are used. For ensuring that data provided by a LUT are

reliable and can be used by SAR community on an operational basis, Cospas-Sarsat has developed a LUT performance specification and a procedure for LUT commissioning.

The Mission Control Centres (MCCs) :

The MCC main functions are : (i) to collect, store and sort the data from LUTs and other MCCs, (ii) to exchange data within the Cospas-Sarsat system, and (iii) to distribute alert and location data to associated Rescue Coordination Centres (RCCs). Alert data comprise the beacon location and, for 406 MHz beacons, other coded information. Other operational information is also exchanged between MCCs and RCCs as required to assist search operations. All MCCs in the system are interconnected through appropriate public communication networks to ensure the reliable distribution of system information and alert data. The French MCC in Toulouse has a master beacon slaved to a cesium clock.

2.3. Cospas-Sarsat Performances

After a first phase of engineering tests, the overall performances of the Cospas-Sarsat system was determined during a Demonstration and Evaluation phase which was carried out by the agencies responsible for Search And Rescue operations in the various participating countries. Additional information was obtained through operations of the system in actual distress events and world-wide exercises conducted to evaluate data exchange procedure and MCC communication links.

The following parameters are particularly significant for the users :

- *the beacon detection probability* = the probability of detection by a LUT of at least one beacon message with a correct code-protected section for the first tracked satellite (this parameter applies to 406 MHz only),
- *the beacon location probability* = is defined, at 406 MHz, as the probability of detecting and decoding at least four individual messages during a single satellite pass so that a Doppler curve estimate can be generated by the LUT; at 121.5 MHz, beacon location probability is defined as the probability of location during a satellite pass above 10° elevation with respect to the beacon,
- *the beacon location accuracy* = is the difference between the location calculated by the system and the actual location as reported by the field personnel,
- *the capacity of the system* = the number of beacons in common view of the satellite that the system can process simultaneously.

The Cospas-Sarsat main system performances can be summarized as :

	121.5 MHz	**406 MHz**
Detection probability	(not applicable)	≥ 0.98
Location probability	≥ 0.90	≥ 0.98
Location accuracy, 90% within	17 km (typically = 10 km)	5 km (typically = 1 km)
Capacity in the satellite field of view	10 beacons	90 beacons.

3. SARSAT USOs SPECIFICATIONS AND PERFORMANCES

3.1. Onboard and orbitography USOs

3.1.1. SARSAT USOs specifications

The main characteristics required for the onboard (OB) USOs and orbitography determination beacons (ODB) USOs are given in table 1.

OB and ODB USO's frequency are about 5 MHz (5.203205 MHz and 5.0753125 MHz respectively). Frequency stabilities versus DC supply, load, pressure, vibrations are very classical.

Magnetic field sensibility of OB USO should be less than 5.10^{-11} per gauss, due to two components :

- the magnetic field generated by the Earth, at the altitude of the NOAA spacecraft,
- the magnetic field generated by magneto-torquers of the spacecraft.

In order to have a precise orbit determination of the satellite (± 100 m), we need :

- the short term stability, expressed in ALLAN variance : $\sigma_y(\tau) \leq 1.10^{-10}$ over 100 ms, which is about the duration of the board Doppler count,
- the SARSAT medium term stability, defined by the slopes (min^{-1}) and residuals at 1 sigma, of 40 samples of frequency measurements made over 100 ms every 30 seconds.

Slopes (S) and Residuals (R) are calculated by the least squares method. They must be compliant over all environmental conditions : temperature, magnetic field, radiation in orbit, and temperature, pressure for ODB USOs. Of course, OB USO's must satisfy all space environmental conditions (vibrations, acceleration, vacuum, radiation, EMC, ...) and space quality assurance (Hi-Rel parts, materials, quality control, reliability, analyses, ...).

Parameters	OB USO	ORBITO. USO
. Frequency	~ 5 MHz	~ 5 MHz
. DC supply variation	$\pm 5.10^{-11}$ (1 %)	$\pm 5.10^{-11}$ (1 %)
. Load variation	$\pm 5.10^{-11}$ (20 %)	$\pm 5.10^{-11}$ (20 %)
. Pressure variation	$\pm 1.5.10^{-8}$/bar	-
. Magnetic variation	$\pm 5.10^{-11}$/gauss	-
. Vibrations (sine, random)	$\pm 5.10^{-9}$	-
. Temperature variation	$\pm 2.10^{-10}$ (air, vacuum, 5 to 35 °C)	$\pm 2.10^{-10}$ (air, 5 to 35 °C)
. ALLAN Stability τ = 0.1 s	1.10^{-10}	1.10^{-10}
. SARSAT mid term stability with 5 °C/hour		
max. slope	$\pm 5.10^{-12}$/min	$\pm 2.5.10^{-12}$/min
residual (1 σ)	5.10^{-10}	2.10^{-10}
. Aging		
per day, after 1 month	$\pm 1.10^{-10}$/day	$\pm 1.10^{-10}$/day
10 years	$\pm 2.10^{-7}$	$\pm 2.10^{-7}$
. Phase noise in dBc/Hz		
10 Hz	-120	-120
1 kHz	-150	-150
. DC power		
air, +20 °C	5 W	2.5 W
vacuum, +20 °C	2 W	-
. Power supply voltage	12 V	12 V
. Random vibration	16,6 g rms	-
. Sine vibration	20 to 500 Hz : 10 g 500 to 1000 Hz : 10 to 20 g 1000 to 2000 Hz : 20 g	-
. Radiations (3 years)	7.5 krads	-

Table 1

3.1.2. SARSAT USOs procurement

The SARSAT OB Flight Model (FM) USOs have been procured from FEI (USA) since 1979 : 12 single FM including redundancy for 9 receiver-processors. From NOAA-N, FM USOs are procured from THOMSON-CEPE in France.

3.1.3. OB and ODB USO performances

OB USOs are tested in detail, first by the suppliers, then by CNES. Main characteristics (short and mid term stabilities, consumption, thermal vacuum, vibrations) are measured with the receiver/processor in France, and with the satellite in USA before launch.

The table 2 gives typical performances of the OB and ODB USOs.

The French MCC situated at Toulouse process onboard Doppler data using specific software. Several parameters are adjusted :

- orbit elements,
- dynamic parameters,
- frequency deviations between OB and ODB USO, for each pass when the satellite is within a range of beacon. The absolute frequency deviation of OB USO is calculated when the satellite is over the CNES at Toulouse (France), because we have installed a Master Beacon slaved to a cesium clock. The accuracy of frequency determination by this method is estimated to 3.10^{-9}. Several passes per day of SARSAT satellites over Master Beacon permit to know the drift of each OB USO in orbit.

Parameters	FEI OB USO	CEPE OB USO	CEPE ODB USO
. ALLAN stability τ = 0.1 s	3.10^{-12}	3.10^{-12}	7.10^{-12}
. SARSAT mid term stability with 0.1 °C/min			
Slopes	$\pm 1.10^{-12}$/min	$\pm 1.10^{-12}$/min	$\pm 3.10^{-12}$/min
Residuals	1.10^{-11}	$1.6.10^{-11}$	4.10^{-11}
. Aging after 1 month	1 to 5.10^{-11}/day	3.10^{-11}/day	$< 1.10^{-10}$/day
. Temperature sensitivity 5 to 35 °C			
in air	-	-	$1.5.10^{-10}$
in vacuum	1.10^{-10}	$6.5.10^{-11}$	-
. Pressure sensitivity	1 to 5.10^{-9}/bar	5 to 10.10^{-9}/bar	-
. Magnetic sensitivity	$< 1.10^{-11}$/gauss	$\sim 1.10^{-11}$/gauss	-
. Phase noise in dBc/Hz			
10 Hz	-125	-130	-125
1 kHz	-150	-150	-150
. DC power at 20 °C			
in air	1.6 W	5.3 W	1.6 W
in vacuum	1.3 W	2.0 W	-

Table 2 - Typical SARSAT OB and ODB USO performances

Figure 4 give an example of drift in orbit (SARSAT 4 USO). We can see a negative slope versus time due to irradiation on quartz crystal. The mean slope is $-1.5.10^{-9}$/month.

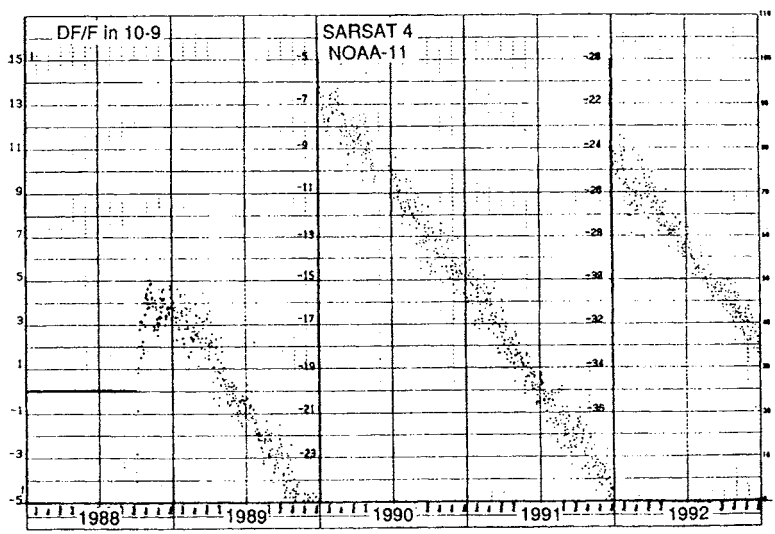


Figure 4 - Drift in orbit of SARSAT 4 OB USO

3.2. COSPAS-SARSAT 406 MHz distress beacon USOs

3.2.1. Specifications

Specifications of distress beacons (DB) USOs are given in the following table 3.

Parameters	DB USO spec.
. Frequency	~ 5 MHz or 10 MHz
. Initial setting frequency	$\pm 5.10^{-6}$
. ALLAN stability $\tau = 0.1$ s	2.10^{-9}
. Medium term stability	
max. slope	$\pm 1.10^{-9}$/min
max. residual (1 σ)	3.10^{-9}
. Aging (5 years)	$\pm 1.10^{-5}$
. Phase noise in dBc/Hz	
10 Hz	-95
1 kHz	-130
. DC power	
+20 °C	≤ 200 mW
-40 °C	≤ 400 mW
. Power supply voltage	5 to 18 volts
. Operating conditions	
Max. temperature range	-40 to +55 °C
Thermal gradient	± 5 °C/hour
Thermal shock	± 30 °C
Pressure	50 to 1000 mbar
Humidity	95% H at +40 °C
Vibrations	RCTA or Eurocae ED62
Shocks	500 g -4 ms or 100 g -23ms
	25 g -6 ms 1000 shocks

Table 3 - DB USO specifications

The nominal frequency of DB USOs depends on the frequency plan of the transmitter of the beacon.
The short term stability and the aging specified require good quartz crystal, and the frequency of DB USOs was chosen at 5.0753125 MHz or at 10.150625 MHz. Frequency stability specifications and operating conditions are derived from international recommendations /3/, /4/, /5/, which are negotiated by COSPAS-SARSAT council, in relation with the International Maritime Organisation (IMO) and the International Civil Aviation Organisation (ICAO). In order to have a good location precision (± 1 km), short term and medium term stabilities are required. Mid term stability is defined by the slopes (min^{-1}) and residuals at 1 sigma of 18 samples of frequency measurements made over 100 ms every 50 seconds. Slopes (S) and Residuals (R) are calculated by the least square method. S and R must be compliant over all operating conditions of distress beacons (ships, aircraft, land vehicles) of table 3. S and R must be compliant 30 min after power on and when a thermal shock of ± 30 °C is applied at the same time. Power of DB are very limited and the consumption of DB USOs must be less than 400 mW at -40 °C.

3.2.2. SARSAT 406 MHz DB USO procurement

In 1980's, it appeared to CNES that TCXO (analog or digital) cannot satisfy S and R specifications. It would be necessary to use oven controlled crystal oscillators, but with a very low consumption.
In 1984, the Laboratoire de Physique et Métrologie des Oscillateurs (LPMO) in Besançon-France under a CNES contract with the help of ANVAR, studied a miniature ovenized quartz crystal oscillator named "MINOS", based on watch quartz and ASIC electronic /6/, /7/. Quartz resonators used are COMODOR of VIBROMETER and CX-AT of MICROCRYSTAL company. ASIC was designed to have several electronic functions : oscillator loop and output amplifier, temperature probe, oven control and heater, voltage regulator. Several prototypes were tested and results were compliant to main specifications of table 3.
In 1986, EUROWAVE-SOREP company in France was chosen to industrialize SARSAT DB USOs at very low cost (about 100 US dollars for 1000 units per year).

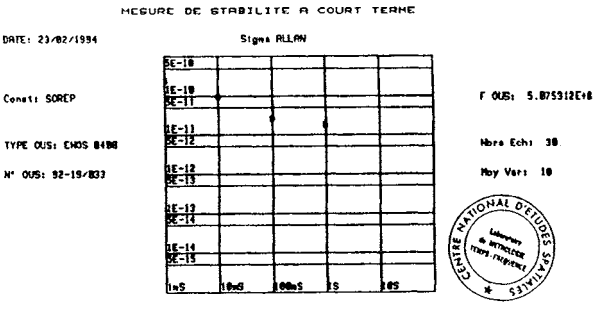
3.2.3. SARSAT 406 MHz DB USO performances

Table 4 give typical performances of DB EUROWAVE USO measured by CNES.

Parameters	EUROWAVE DB USO
. Short term stability ALLAN τ = 0.1 s	2 to 7.10^{-11}
. Medium term stability Slopes S Residuals R with ± 5 °C/hour within -40 to +55 °C	0.5 to 1.10^{-9}/min 0.2 to 1.10^{-9}
. Max. time to respect S and R from turn on, with ± 30 °C	20 min
. Aging	$\leq 5.10^{-7}$/year
. Temperature sensitivity -40 °C to +55 °C	1 to 4.10^{-7}
. Phase noise in dBc/Hz 10 Hz 1 kHz	-95 -135
. DC power +20 °C -40 °C	125 mW 250 mW
. Output level	TTL or HCMOS
. Size	30.2 x 24.2 x 12.5 mm

Table 4 - SARSAT EUROWAVE DB USO performances

Figures 5, 6, 7 give performances of EWOS 400 unit S/N 033 oscillator.



VARIANCE D'ALLAN SUR 30 ECHANTILLONS

TAU	Minimum	Moyenne	Maximum
1mS	1.8E-10	2.3E-10	3.8E-10
10mS	8.E-11	9.3E-11	9.8E-11
100mS	2.1E-11	2.5E-11	3.E-11
1S	1.4E-11	1.7E-11	2.2E-11
10S	0	0	0

Figure 5 - Short term stability of EWOS 400 USO

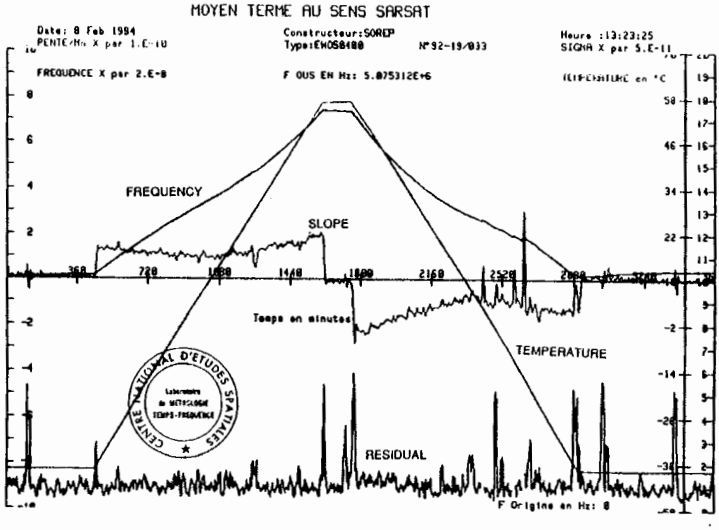


Figure 6 - Medium term stability with ± 5 °C/hour, -40 °C to +50 °C

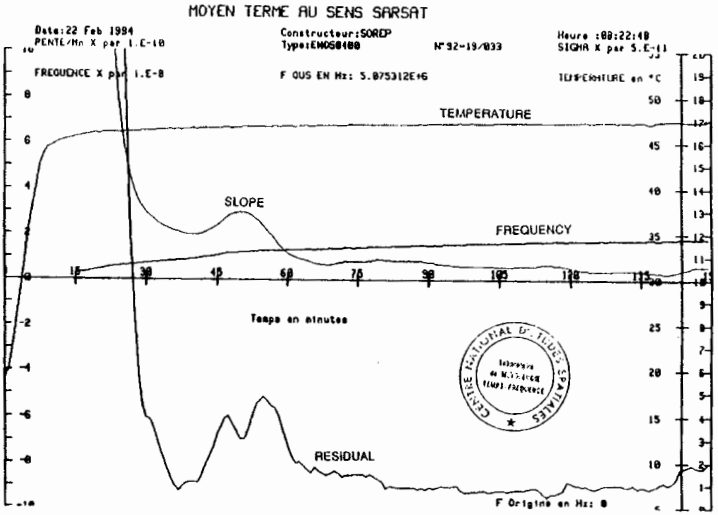


Figure 7 - Medium term stability from turn-on and 30 °C thermal shock

Since 1988, about 40,000 units have been produced and sold on the world market for airplanes, boats and ground distress beacons. From 1994, a new SARSAT DB USO will be produced in DIL case with the same performances, and at lower cost.

4. CONCLUSIONS

The Cospas-Sarsat system began to operate on an experimental basis after the launch of the first satellite in June 1982. It was declared operational in July 1985, when the complete satellite constellation and ground segment were established. On July 1, 1988, the four founder countries (Canada, France, USA, Soviet Union) signed an Intergovernmental Agreement guaranteeing system continuity for 15 years, to be extended in five-year periods. The system is open to all states on a free-of-charge and non-discriminatory basis.

At the end of 1993, twenty-three states participate in the system either as space or ground segment providers, or as user states. Sixteen MCCs and twenty-eight LUTs are in operation or almost reaching the operational state all over the world. To the about 500 000 aeronautic 121.5 MHz beacons already existing worldwide, the Cospas-Sarsat system has today added almost 100 000 new type 406 MHz beacons specifically designed to provide global coverage, increased system capacity and improved location accuracy.

In order to have a good location precision, 406 MHz beacon USOs must be stable during all distress operating conditions which are very severe. A miniature oven controlled crystal oscillator which satisfies all COSPAS-SARSAT specifications has been developed in France. This oscillator features very low power consumption and low cost.

Since the first rescue of three airmen in the Canadian forest in 1982, more than 3,400 persons have been saved by the COSPAS-SARSAT satellite aided operations.

REFERENCES

/1/ Introduction to the Cospas-Sarsat System
C/S G.003, Issue 3, June 1989.

/2/ Description of the Cospas-Sarsat Space Segment
C/S T.003, Issue 2, December 1992.

/3/ Specifications for COSPAS-SARSAT 406 MHz DISTRESS BEACONS
C/S T.001, Issue 2, rev. 5, September 1993.

/4/ COSPAS-SARSAT 406 MHz DISTRESS BEACON TYPE APPROVAL STANDARD
C/S T.007, Issue 3, rev. 1, December 1993.

/5/ ITU CE8 Geneva
Recommendation 633-1.

/6/ G. Marianneau, D. Hauden, J.J. Gagnepain
"Integrated oven controlled quartz crystal oscillator",
40th Ann. Freq. Con. Symp. (1986).

/7/ G. Marianneau, D. Hauden
"Integrated quartz oscillator for SARSAT buoys"
Int. Chronometry Symposium, Stuttgart (1989).

Advanced Transponder Concept for Deep Space Missions up to EHF

Ottfried Bätz, Willi Lindemer, Klaus-Peter Wagner
Deutsche Aerospace AG (DASA)
P.O. Box 80 11 69, D-81663 Munich

Abstract

Mission planning is seriously considering extreme high frequencies up to 32 GHz. Increasing frequencies and tracking performance requirements give rise to new considerations on versatile frequency synthesizer concepts in deep space transponders.

DASA has performed a comprehensive study of satellite transponders tailored to the specific requirements of coherent receive/transmit frequency turn-around. General rules have been figured out as to pave the way for novel transponder frequency plans.

The importance of this analysis has been stressed by the fact that standard frequency plans suffer from their significant phase noise created at SHF and, in particular, at EHF. The source of excessive phase noise is usually a frequency agile voltage-controlled or numerically-controlled oscillator along with its subsequent frequency multiplier chains.

As standard TCXO´s exhibit much lower phase noise characteristics, a transponder frequency plan has been derived which needs only to multiply the low-noise TCXO outputs while the frequency controlled oscillator needs almost no multiplier. In this way, the harmful enhancement of VCO/NCO phase noise can be avoided.

The implementation of this concept considers the use of numerically controlled oscillators, NCO´s, in place of VCO´s. A frequency plan has been developed that displays an attractive commonality of transponder modules for a 7/8 GHz transponder and a 7/32 GHz transponder, thus offering a cost-effective design approach.

The use of a NCO in the receiver of the transponder constitutes a means to precisely measure the receive frequency aboard the spacecraft for the purpose of Doppler measurement at scientific missions. The NCO in the transmit path of the transponder is employed for both, ultra-linear phase modulation and binary phase shift keying. This feature cannot be easily presented by standard transponder concepts.

All digital implementation of the carrier/subcarrier demodulation and bit detection attracts future incorporation of forward error correction on the uplink.

The study has been performed for ESA / ESTEC. Breadboarding of the receiver concept is currently in progress at DASA (Deutsche Aerospace) in Ottobrunn with subcontract to Satellites International Ltd..

1 Introduction

1.1 Outlook to Future Mission Requirements

Future exploration missions into deep space are increasingly driven by significant growth in telemetry link capacity beyond 2 Mbit/s along with low bit error rates. Precise navigation of deep space vehicles require precise tracking performance that could be supported by wideband Delta-VLBI (ΔDOR) techniques. Plasma effects can be considerably reduced at frequencies above 8 GHz. Probe tracking from an orbiting vehicle call for precise Doppler tracking capability.

While keeping antennas and transmitters to a reasonable size, mission planning is seriously considering extreme high frequencies up to 32 GHz. Increasing frequencies tend to magnify phase noise and spurious signal generated inherently in oscillators and synthesizers, thus impacting on the carrier acquisition and tracking performance of the transponder.

On-board frequency measurement is going to attract investigators being interested in precise Doppler detection, e.g. CASSINI/HUYGENS mission.

The current TT&C standard using subcarrier modulation for data transmission suffers from wasting valuable signal power due to significant modulation losses. Similar loss is imposed on tone ranging as defined in the ranging standard. Advanced modulation techniques come seriously under consideration emphasizing the benefits of direct carrier phase-modulation, e.g. B-PSK Q-PSK or derivatives. This consideration includes the substitution of tone-ranging by pseudo-random noise codes which, in addition, may serve as "subcarrier" for low data rates (see TDRSS). This trend is calling for precise 90° and 180° phase shifters needing techniques other than for linear phase modulators currently used in TT&C transponders.

The discussions raised so far leaves open the applicable radio frequency, modulation standards and inherent techniques. This open situation attracts versatile concepts with modular implementation offering a cost-effctive approach with short response times to specific requirements.

1.2 Constraints of Standard Transponder Architectures

The fundamental feature of transponder architectures is reflected through their frequency plan and the position of the downlink modulator. Figure 1 shows a typical frequency plan for a 7 GHz uplink and a 32 GHz downlink. Automatic-gain-control (AGC) is not shown as it is not relevant in this context.

The receiver includes two down-converters and a coherent demodulator. Full compensation of Doppler and oscillator instabilities is achieved by a long loop through the first down-converter. The first and second IF (intermediate frequency) are fixed-frequency. Two basic frequency sources are implemented, i.e. a stable oscillator (TCXO or external ultra-stable oscillator) and a frequency-

agile oscillator (VCO) controlled by the carrier aquisition and tracking unit. The exciter (forming, with a power amplifier, the entire transmitter) includes at least two up-converters. The specific turn-around ratio, r_{TA}, is met if the multipliers M satisfy the relation:

$$r_{TA} = \frac{M4 + M5}{M1} = \frac{M3}{1 + M2} \qquad \text{(equ. 1)}$$

Three optional locations are shown for the modulator. Option 1 is preferred if the modulator is operated with low modulation index magnified by the multiplier M3 as to increase the modulation linearity. This way is useless for B-PSK and QPSK modulation. Option 2 and 3 require the modulators to be operated with the full modulation index associated with all problems of linearity; digital modulation can be easily accomplished.

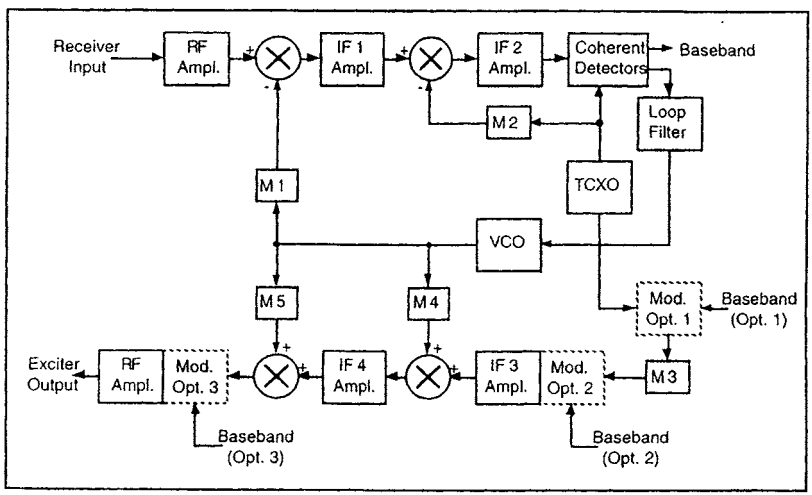


Figure 1: Generalized Concept of Conventional Transponders (AGC not shown)

Standard transponders reflect a fundamental handicap imposed by the frequency turn-around ratios established by NASA and ESA, see Table 1. The ratio is tailored to the technology of the sixties and seventies, and is aiming at a VCO frequency around 9 MHz common to the receive and transmit path. Each RF frequency calls for a tailored frequency plan implemented in dedicated synthesizers. Re-use of transponder elements at different RF frequencies is almost not possible.

As the VCO output is to be multiplied by a factor up to 3344 (at EHF: 32 GHz), the inherent phase noise power is magnified by considerable 70 dB.

		RX band	S	X	Ka
		RX frequ.	2.1 GHz	7.1 GHz	34 GHz(*)
		Multiplier, M	221	749	3583(*)
TX band	TX frequ.	Multiplier, N			
S	2.3 GHz	240	O		
X	8.4 GHz	880	O	O	
Ka	32 GHz	3344	O	O	(open)

Table 1: Multipliers for Coherent Turn-around; Receive (RX), Transmit (TX); (*) still under discussion

1.3 Objectives of Advanced Transponder Concepts

The features outlined above become a burden if we want to exploid modern digital oscillators for the sake of an universal concept with the following properties:

- suitability for linear and digital modulation,
- re-use of most of the transponder modules if different RF frequencies are to be implemented,
- digital signal processing for performance enhancement.

Modern technology of digital processing and numerically controlled oscillators (NCO) pave the way for advanced transponder techniques. Instead of one common TX/RX VCO, two NCO's can be assigned to RX and TX, which are tied together through a common clock derived from a TCXO. This approach opens the view to more flexibility in designing an almost universal frequency plan.

2 Generalized Frequency Plan for Coherent Transponders

2.1 General Considerations

Whichever transponder concept is being derived from the options of short-loop through the long-loop principles, it always can be reduced to the schematic shown in Figure 2. For practical reasons, the sequence of fixed frequency-conversion (TCXO) and tracking frequency-conversion (VCO, NCO) can be exchanged. The sequence is not relevant for the general consideration hereafter. The blocks "a" to "d" render the effective composites of frequency multipliers and conversion stages (mixers). The factors "a" through "d" can be positive or negative, depending on the frequency-conversion scheme in receiver and exciter, because the output frequency of any converter is the sum or difference of the RF/LO input frequencies. This virtually is the same as taking only the sum of the input frequencies and assuming the multiplier factors to be positive or negative depending on whether the lower or the upper sideband is being used after the conversion.

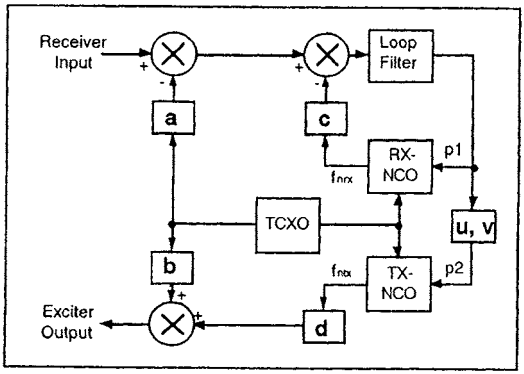


Figure 2: Equivalent Schematic of Transponder Concepts

If a VCO or VCXO is used like in standard transponder concepts, the RX-VCO (for RX-NCO) has also to feed the exciter when the transponder is in the coherent mode, $f_{ntx} = f_{nrx}$. In the non-coherent mode, f_{ntx} must be derived from a fixed-frequency source, say a TCXO. Note that this fact restricts the possible locations of the modulator in the exciter.

If two NCO's are used as shown, the RX-NCO is controlled by the receiver carrier tracking processor (loop filter) while the TX-NCO is controlled by the same control data with the transponder in coherent mode, or by a fixed control word memorized in the carrier tracking processor, with the DST in non-coherent mode. This concept offers an attractive design possibility: If the transponder turns from the coherent mode into the non-coherent mode, the last control word valid in coherent mode can be held and hence applied in the non-coherent mode. In this way, the switch-over does not cause any phase or frequency jump, and the earthstation will not lose carrier lock.

2.2 Design Parameters for Coherent Transponders

The NCOs are controlled by digital words p1 and p2, integers of wordlength n, each. The term $p/2^n$ denotes the fractional output frequency relative to the NCO clock, f_o. The respective oscillator output frequencies are

$$f_{nrx} = p1 \cdot f_o / 2^n \text{ and}$$
$$f_{ntx} = p2 \cdot f_o / 2^n .$$

The functional box labeled "u, v" before the TX-NCO relates the control word p2 to p1

$$p2 = u \cdot p1 + v, \qquad \text{(equ. 2)}$$

With equ.2,

$$f_{ntx} = (u' \cdot p1 + v') \cdot f_o \qquad \text{(equ. 3)}$$

where $u' = u/2^n$ and $v' = v/2^n$.

The turn-around ratio, $r_{TA} = f_{out} / f_{in}$, leads to: $r_{TA} = (b + d \cdot u' \cdot p1 + d.v') \cdot f_o / f_{in}$ and with $p1 = (f_{in} / f_o - a) \cdot 2^n/c$,

$$r_{TA} = \frac{d\,u}{c} + \left(b + d \cdot v' - \frac{a\,d\,u}{c}\right) \cdot \frac{f_o}{f_{in}}$$

The ratio r_{TA} must be independent of the input and output frequency, what requires the term in (···) = 0. Then one obtains:

$$u = c\, r_{TA} / d, \qquad \text{(equ. 4)}$$

$$\text{and } v' = (a \cdot r_{TA} - b) / d \qquad \text{(equ. 5)}$$

While the factors "a" through "d" are realized in the analogue RF/IF domain, the factors "u" and "v" can be set in the digital domain.

Conventional transponder concepts as defined in section 1.2, feature pure feedback schemes where the factors

- "a","b" and in turn "v' " virtually are "zero",
- c,d are properly chosen according to the required turn-around (see equ. 1), and
- u corresponds to 1 if a common RX/TX VCO is used.

For all general cases equation 4 offers a means to optimize "c" and / or "d" with respect to technical simplifications while disregarding constraints due to r_{TA} . Finally, the digital "u" can be chosen to match with r_{TA}. The same design option is offered through equation 5 for "a" and / or "b"; here again, the digital value "v" can be finally matched with r_{TA}. This advanced baseline thus attracts by virtue of its great flexibility and design freedom in favour of optimal performance and properties stipulated in section 1.3.

2.3 Design Approach

Performant digital demodulation and processing, the option of linear and binary modulation capability and the synthesizer commonalities to various RF frequencies have been the major drivers for chosing NCO's. Phase noise and, above all, spurious signal outputs of the NCO's are 50 to 70 dB down w.r.t. the carrier power. Figure 3 shows a sample of the NCO, STEL 1172, with 32 bits for frequency control, 10-bit phase quantisation and 8-bit amplitude. The spurious level is about -50 dBc down and depends on the ratio of clock-to-output frequency. A ratio around 4 and multiple thereof have turned out an optimal spurious suppression. An improved version is offered by STEL 1175.

The spurious level requires to keep "c" and "d" as close to 1 as possible. Thus, the turn-around ratio is mainly determined by the factor "u".

The value "v" is compensating for minor TX-RF deviations incurred by simplification of the technical implementation of the synthesizer with a high degree of commonality for different RF frequencies. Table 2 shows the issue of a detailed study.

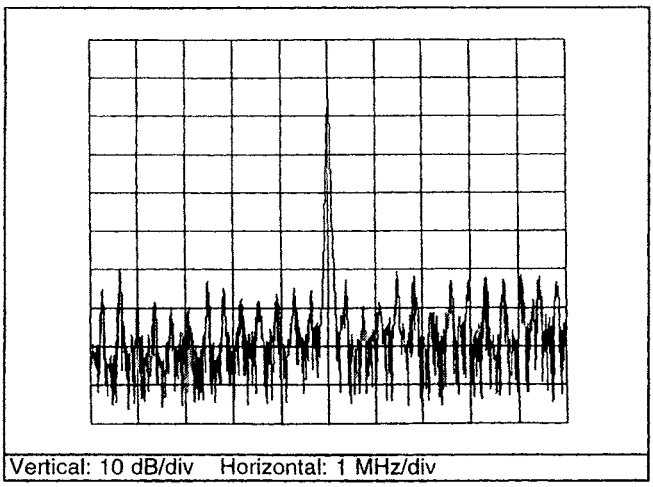


Figure 3: Typical Spurious Output Spectrum of an NCO (STEL-1172 from Stanford Telecom)

Transponder Type	X/X-Band	X/Ka-Band
"a"	111.375	111.375
"b"	131	498.5
"c"	1	1
"d"	1	1
"u"	≈ 1.1749	≈ 4.4646
"v´ "	≈ -0.145517	≈ -1.253

Table 2: Condensed Design Parameters for an X/X-Band and X/Ka-Band Transponder

2.4 The Effect of Digital Quantisation

The numbers of "u" and "v´ " in equation 4 and 5 usually are infinite fractional numbers. Their representation in the digital processors will result in digital truncation errors depending on the word length of the processors. This may lead to a small bias error of the turn-around ratio and of the output frequency.

If, at the earthstation, no allowance is made for the truncation of the parameters "u" and "v´ ", the relative bias error of the transponder output frequency will not exceed $3 \cdot 10^{-11}$ with a 32-bit implementation of equation 2, and at optimal TCXO frequency.

In fact, the truncation is deterministic and thus can be taken into account at the receiving earthstation, so that the received frequency need not deviate from the expected value. Therefore, Doppler integration can still be performed without sacrificing the precision.

3 Architecture of the Transponder

3.1 Main Characteristics

Following a conceptual design study for ESA/ESTEC performed by DASA and its subcontractor Satellites International (GB), the team has recently completed a detailed design of a Deep Space Transponder, DST, which now is going to be substantiated by receiver breadboarding. Figure 4 depicts the block diagram of the X/X-band and X/Ka-band transponders. The DST accommodates two NCO´s clocked at about 72 and 64 MHz respectively. All demodulation, carrier aquisition and tracking is performed in the Digital Unit of the transponder. A/D conversion is performed at a receiver IF of 24 MHz. In order to cope with the receiver dynamic range of some 70 dB, a non-coherent AGC is implemented in the IF section, and a coherent AGC is in the Digital Unit.

The overall frequency plan exhibit a great deal of commonality of the X-band and the Ka-band transponder. Two separate exciter branches are shown after the fourth IF amplifier. Note that only one branch can be operated in a time; the TX-NCO parameters memorized in the preceding control converter, have to be set accordingly.

Design provisions have been made for accommodating a wideband ΔDOR modulation. Normal telemetry and ranging modulation is performed through the TX-NCO. Future design upgrades will take into consideration that the DAC-ADC combination in the ranging path is fully implemented in digital techniques. The following section describes the specific properties and benefits of the design.

3.2 RF Technology

The RF circuitries are realized in MIC co-planar technolgy on ceramic substrate. The low-noise amplifier front-end uses HEMT transistors followed by MMIC devices. The receiver noise-figure is 1.9 dB at room temperature.

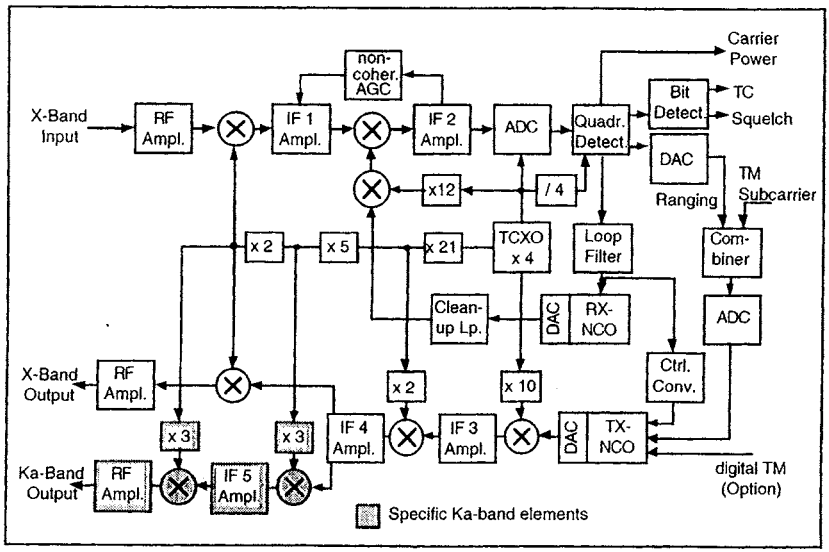


Figure 4: X/X-Band Transponder Architecture. Modifications and Add-ons for X/Ka Band are indicated through Shadowed Blocks

3.3 The Digital Receiver

The A/D converter applies moderate 4/7-undersampling of the second IF. The Nyquist double-sided bandwidth is 8 MHz. This is sufficient to demodulate the maximal range-tone of 1.5 MHz; practically the range-tone will not exceed 500 kHz. The modulation bandwidth due to the telecommand subcarrier and its sidebands will not exceed 30 kHz.

The Digital Unit also performs the subsequent PSK-demodulation of the subcarrier (8 or 16 kHz) and the final bit timing and bit detection. A squelch function informs the user about the data quality.

The detected range-tone is D/A converted for further analogue filtering and leveling and final feeding of the downlink modulator. It is planned in future design upgrades to perform the ranging signal processing fully in digital techniques. Thus the double conversion (digital-to-analogue-to-digital) of the ranging baseband will later be discarded in favour a complete ASIC technology. Stipulation of all digital ranging techniques, e.g. PRN codes, instead of tone-ranging would result in a specific functional design.

Carrier tracking is performed through a second order loop. Provisions are made to operate a third order loop for enhanced tracking capability. The acquisition threshold will be -138 dBm referred to X-band receiver input port. The

theoretical tracking threshold (i.e. 0 dB in the loop bandwidth) is predicted at about -158 dBm.

The functions are implemented in a μ-processor (3 1750), an ASIC (application specific integrated circuit), an ADC and a DAC. On breadboard level, the functions of the ASIC are preliminarily implemented in FPGA technology (field programmable gate array).

3.4 The Digital Modulator

The downlink carrier modulation is performed at an IF of about 16 MHz. The modulator is fed by a composite of a ranging signal and a telemetry subcarrier, thus complying with the current ESA standard. Future application may also consider direct digital modulation, say B-PSK, Q-PSK and the like.

The phase-modulation is accomplished through phase-control of the TX-NCO. The device has been selected from Stanford Telecom: STEL-1175. The spurious level is expected to be below -65 dBc.

For linear modulation (better than 0.5%) the NCO is controlled via an A/D converter. For digital modulation the respective control words are applied E.g. 0°, 180° (B-PSK) could be precisely shifted through the sign-bit.

3.5 The Reference Oscillator

All NCO-clocks and fixed-frequency local oscillators are derived from a common reference mandatory for a coherent transponder. The reference is a TCXO at around 8 MHz. The 64 MHz clock is generated through a VCXO-based PLL locked onto the TCXO. Provisions are made to switch the PLL to an external USO (ultra stable oscillator).

3.6 Phase Noise and In-Band Spurious Signals

The specific frequency plan prevents phase noise and spurious signals of the NCO's from being magnified in both the receiver and exciter. In this way is minimized the risk of false lock onto spectral lines close to the carrier. Figure 5 and 7 show the predicted off-carrier level bound of phase-noise only at the DST output port. For comparison, the phase-noise characteristics of the TCXO is also drawn. Below 1 kHz off carrier, one can clearly identify the effect of the receiver PLL noise bandwidth. Above 100 kHz, the output phase-noise is the phase-noise of the TCXO enhanced by the multipliers in the TX branch of the DST.

Figure 6 and 8 show the bound of phase-noise and spurious signal stemming from the NCO's. The spurs level above 100 kHz off carrier is determined by the assumed -65 dBc level from the TX-NCO source; below 1 kHz the contribution from the "copied" RX-NCO spurs is added. Any multiplier in the output branches of the NCO's would raise the spurious signals to unacceptable levels.

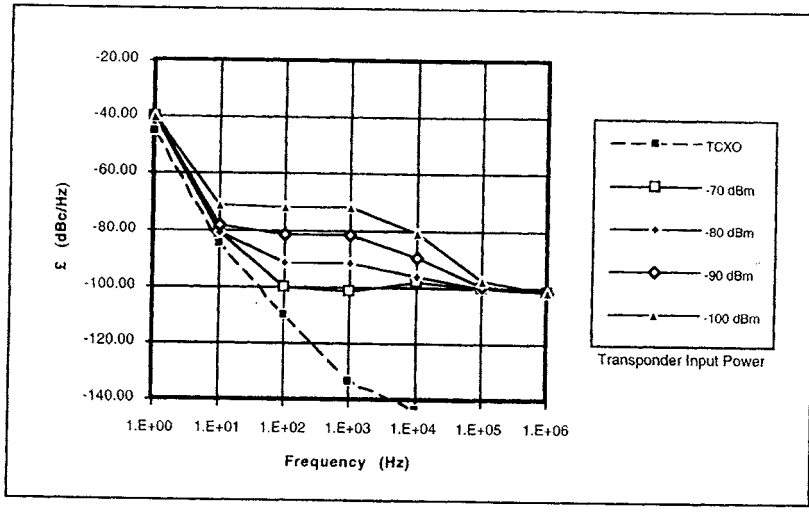


Figure 5: Phase Noise Prediction at **X**-Band Output (Noise sources are TCXO phase noise and receiver thermal noise)

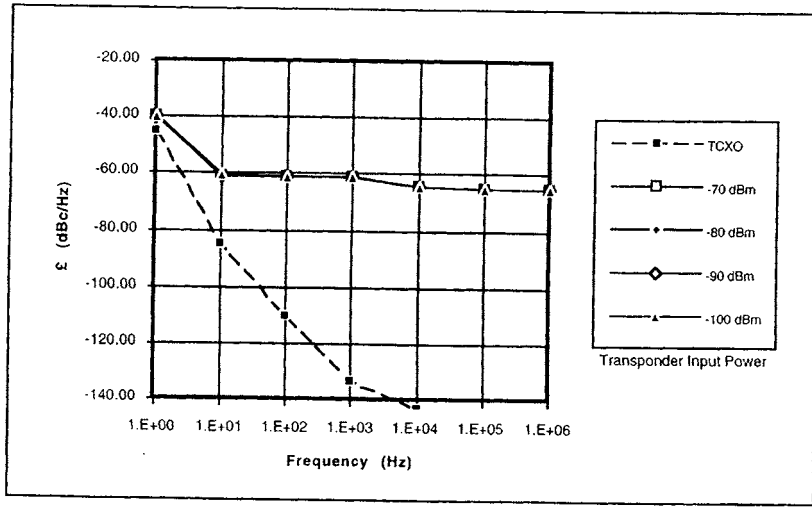


Figure 6: Phase Noise and In-Band Spurious Level Prediction at **X**-Band Output (Noise sources are TCXO phase noise, receiver thermal noise and NCO spurious signals)

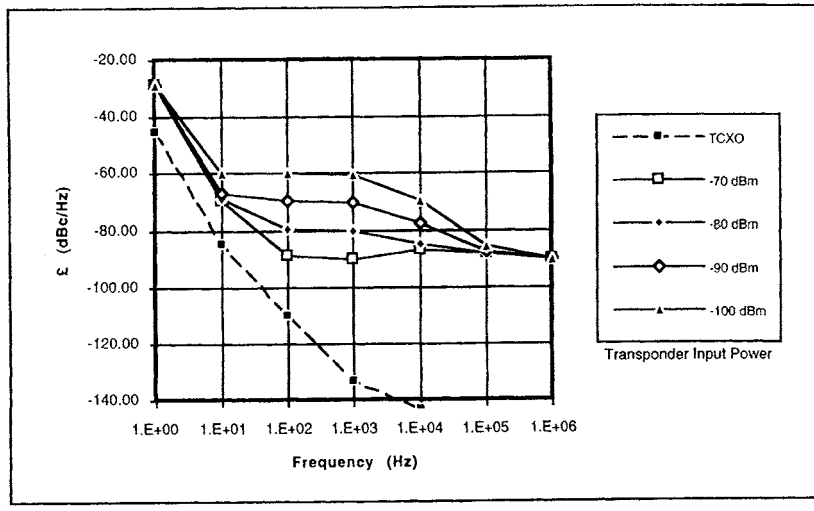


Figure 7: Phase Noise Prediction at **Ka**-Band Output (Noise sources are TCXO phase noise and receiver thermal noise)

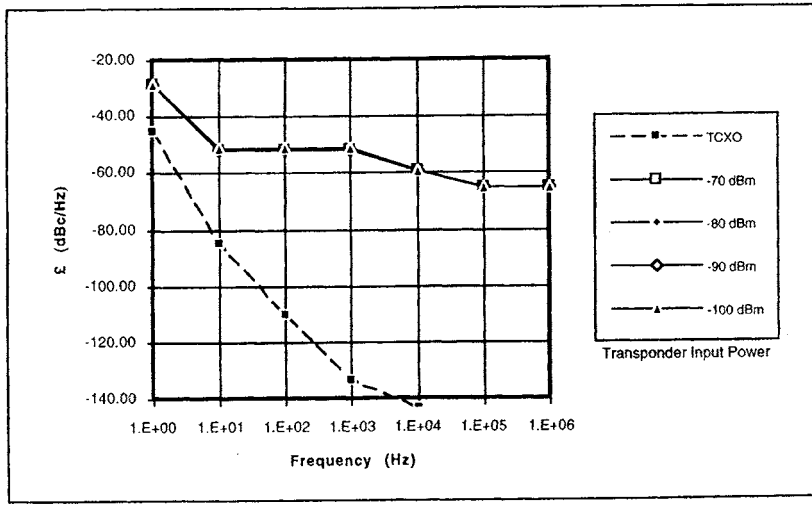


Figure 8: Phase Noise and In-Band Spurious Level Prediction at **Ka**-Band Output (Noise sources are TCXO phase noise, receiver thermal noise and NCO spurious signals)

3.7 Hardware Modularity

The DST is partitioned into a number of modules housed in frames. The frames are accommodating receiver, synthesizer, digital unit, signal generator and exciter. A specific frame will constitute the control and monitoring interface to the "outer" world in the spacecraft. Thanks to the particular frequency plan, the exciter and synthesizer module need only minor modifications and add-ons if an X-band exciter is to be upgraded to Ka-band, see Figure 4.

4 Conclusion

An alternative to standard transponder concepts has been derived which aimes at low spurious signal generation and a versatile frequency generation. The presented architecture attracts future EHF applications in that the modulator can be operated in both a linear and a strict digital mode, the receive capabilities can be matched through software adaptations, and the implementation requires minimal modifications and add-ons.

5 Acknowledgements

The authors gratefully acknowledge the valuable support by ESTEC in Noordwijk and their staff member Luitjens Popken for valuable discussions and reviews.

Performance of GPS Disciplined Rubidium Oscillators under Selective Availability (SA)

Gerhard Hübner, Ball Efratom Elektronik GmbH

1. Abstract

The Global Positioning System (GPS) can be used to provide a precise time and frequency reference to control local oscillators. The paper gives an overview of the hardware design and firmware concept of a rubidium oscillator which is disciplined both in time and frequency by reference signals derived from a C/A code GPS receiver.

Error sources and accuracy limitations of GPS based one-way time transfer under signal degradation by selective availability (SA) will be discussed particularly with respect to absolute timing accuracy. The paper addresses also the influence of local oscillator stability to overall performance.

Performance data and verification methods will be presented.

2. GPS Related Error Sources

The GPS system, although providing unprecedented accuracy, does have some limitations. The following paragraphs address some of the error sources which limit the time transfer accuracy particularly if common view techniques cannot be applied.

2.1 Broadcast Ephemeris

Each satellite broadcasts ephemeris parameters describing the orbit during a curve fit interval of normally four hours. The parameters are obtained via least square curve fits of the predicted ephemeris /1/. We know little about the accuracy of the orbits calculated from these data. Orbits calculated from subsequently transmitted data sets typically fitted to two or three meters at the time the new data set was available to the receiver. Thus the change in apparent receiver clock delay is less than 10 nanoseconds and so buried in the noise floor. We suppose there are no intentional biases in the broadcast ephemeris. If averaged over time and satellites time transfer errors due to ephemeris errors should be minimized. Theoretically ephemeris errors vary slowly with time as the satellite travels in the real orbit rather than on a least square curve fit.

2.2 Ionospheric Delay

The satellites also broadcast ionospheric parameters which allow a single frequency (L1 or L2) receivers to utilize a model to estimate propagation delay of the GPS signals in the ionosphere. It is estimated that the use of this model will provide at least a 50 percent (RMS) reduction of this error /1/. Typical estimates are around 30 nanoseconds resulting in a 15 nanoseconds RMS error.

2.3 Satellite Clock

The clocks in the GPS satellites are basically free running. Each satellite broadcasts parameters of a clock correction polynomial which allows the receiver to estimate the effective PRN code phase offset with respect to GPS system time /1/. It is difficult for

us to give an estimate for the accuracy of this correction but again we assume the error will vary randomly from satellite to satellite and increase slowly with the age of the data.

2.4 Selective Availability

To deny access by Standard Precision Service (SPS) users to the full system accuracy the signals are transmitted from the satellites with errors added. SPS users will be unable to remove these errors and will receive a reduced standard of service. Planned accuracies are 174 m for 3D position and 363 ns for timing, respectively. SPS accuracies may be varied should a security threat exist to US/NATO interests /2/.

Naturally no details have been published as to the nature of the added errors. It appears however that the errors are added primarily to the C/A code phase SPS receivers are tracking. The code phase jitter appears to be of (pseudo) random nature. Our analysis shows that for a block II satellite reporting a User Range Accuracy (URA) index of 7 (24 m ... 48 m) the C/A code phase is modulated with white noise at modulation frequencies extending from less than 50 μHz to approximately 2 mHz. The spectral density is around 2800 ns*Hz$^{-1/2}$. This is illustrated in figure 1. In a 2 mHz bandwidth this evaluates to 125 ns RMS. Hence, if Selective Availability (SA) shall be removed at least partially one must use phase locked loops with a cutoff frequency significantly lower than 2 mHz. This in turn determines the requirements on the inherent stability of the oscillator to be disciplined.

2.5 Receiver Position

Determining the precise receiver position in the coordinate system used by GPS (WGS84) is of utmost importance. While horizontal errors tend to average out if satellites from all directions are being received, vertical errors translate into timing errors proportional to the sine of elevation. Since elevation is always positive the errors are not being averaged out.

In stationary applications one can develop a position average over time. There seems to be no bias in any one of the C/A-code GPS receivers we have tested so far. Figure 2 shows raw position fix data and Figure 3 shows how the average becomes more and more stable over time. When averaged over 24 hours or more, position has always been found to be within 3 meters (3-D) relative to our adopted position. This has been proven also at carefully surveyed locations.

2.6 GPS vs UTC(USNO) time

GPS system time is maintained using all the clocks in both the control and space segment. It is not a independent time scale but is steered towards UTC(USNO). The control segment shall maintain GPS time relative to Coordinated Universal Time (UTC) to within 176 ns (95%) plus accumulated leap-second jumps. The difference between GPS time and UTC shall be included in the navigation message /2/. Along with the number of current and future leap seconds the parameters of a first order polynomial are given to estimate the the difference between GPS time and UTC(USNO). In one way time transfer these parameters must be used if timing relative to UTC(USNO) is desired.

When we monitored the reception of these data sets we discovered that apparently not all satellites are transmitting the same data set at the same time. This causes the receiver to frequently use outdated data sets. Figure 4 plots the difference between GPS time and UTC as estimated using the parameters as received over a period of 100 days. The marks along the X-axis indicate the time a new data set was received.

3. Receiver Related Error Sources

GPS receivers are available as board level OEM products from several manufacturers. Most of them are designed as position sensors. The time transfer capability of GPS is not supported in each model. In many cases the position fix calculation is transparent to the user i.e. the user does not know which of the satellites being tracked are used for the fix.

Normally receivers derive all timing from an onboard crystal oscillator, typically a TCXO. The 1pps output, also being derived from the TCXO, is disciplined towards GPS time by introducing steps of typically one microsecond or less. As the TXCO frequency error varies over temperature the frequency of the 1pps correction varies. Unless additional information is available from the receiver this limits its accuracy to the step size.

Signal propagation delay in the receiver does not degrade position fix accuracy as long as it is independent of the satellite being tracked. However, it goes directly into the time transfer accuracy of the receiver. Normally, board level receivers are globally specified ("1 μs to UTC") and calibration data are not available. The errors introduced by such receivers can easily dominate those discussed above.

Modern multi-channel GPS position sensors are able to simultaneously track up to 12 satellites. Currently (March 1994), with 24 active satellites between 5 and 9 are continously visible in Munich. The typical sensor selects the four satellites providing the best PDOP and calculates position fix and clock bias. Any error in the pseudo-range measurement translates into position fix and clock bias error. TDOP (Time Dilution Of Precision), depending on satellite geometry, relates range to time error. The timing accuracy then becomes:

$$TA = TDOP * (URA_1^2 + URA_2^2 + URA_3^2 + URA_4^2)^{1/2} / c$$

With TDOP = 1.5 and the URA set to 32 m for all block II satellites TA evaluates to:

$$TA = 1.5 * (4*32^2)^{1/2} \text{ m} / 3*10^8 \text{ m s}^{-1} = 320 \text{ ns}$$

Figure 5 shows the reported clock bias over a period of $2^1/_2$ days. Figure 6 shows how satellites are used during this period. Note that 4 satellites were available thoughout the oservation time.

The receiver we use is a modification of a standard position sensor to make it more suitable for time transfer applications. It allows an external standard to substitute the onboard crystal oscillator and disciplines the 1pps to the nearest millisecond only thus minimizing the number of steps necessary. Position fix and clock bias data are available via a RS-232 interface along with a variety of GPS satellite related data such as azimuth, elevation, URA, signal level, ephemeris status etc. In stationary applications we can not only average the position fixes over time to obtain a more accurate receiver position. At the same time it is possible using the satellites apparent places to relate the clock bias of the position fix to the individual apparent satellite clock bias. We can then average over the set of satellites to improve accuracy. Assuming the individual satellites errors are not correlated the timing accuracy then becomes:

$$TA = TDOP * ((URA_1^2 + URA_2^2 + URA_3^2 + URA_4^2)/4)^{1/2} / c$$

With TDOP = 1.0 for a single satellite and the URA set to 32 m for all block II satellites TA evaluates to:

$$TA = 1.0 * (32^2)^{1/2} \text{ m} / 3*10^8 \text{ m s}^{-1} = 107 \text{ ns}$$

Figure 7 shows the clock bias data after averaging over the 4 satellites and correction for the fixed position.

4. The Disciplined Rubidium Oscillator

The performance of a typical rubidium oscillator and of the GPS system as a frequency reference is shown for various conditions in figure 8 in terms of Allan Variance. Ambient temperature variations affect the rubidium oscillator stability basically at averaging times of 1000 seconds or longer. The stability of GPS depends on SA and wether corrections can be applied in stationary applications. The disciplined oscillator shows the typical rubidium performance up to 10000 seconds. At longer times the control loop adds some noise from the GPS reference. Ultimately, as averaging time approaches infinity, the disciplined oscillator assumes the stability of the reference.

Figure 8 shows a block diagram of the receiver/controller module which was developed to discipline the rubidium oscillator. The 10 MHz reference frequency from the oscillator passes a digital phase shifter with a resolution of approx. 4.6 ns. It is then converted into a 16368 kHz clock and a 1pps for the GPS receiver. The 1pps is also available at the output. The GPS receiver operates in normal position fix mode using four satellites to determine 3-D position and clock bias as described in para. 3. It is allowed to fall back to 2-D using only three satellites or even a position hold mode if less satellites are available. Position and clock bias data as well as other GPS related parameters are reported via an RS-232 interface to the microprocessor. The processor averages the position data, corrects the clock bias data for the running average and calculates least square linear fits to the corrected data over a user selectable time interval. The duration of the time interval can be set individually for time (1pps) and frequency (control voltage) corrections. At the end of the time correction interval the processor relates the calculated clock bias, which is in GPS time, to UTC(USNO) and steers the 1pps using the phase shifter. At the end of the frequency correction interval the processor calculates the change in control voltage necessary to steer the rubidium oscillator frequency offset towards zero.

5. Performance

5.1 10 MHz frequency output

Unit Under Test (UUT) was a GPS disciplined rubidium oscillator model GPS-RR. The UUT was placed in a normal laboratory environment with a temperature of (22±2)°C. The averaging time was set to 4 hours for time correction and 24 hours for frequency correction, respectively.

A passive hydrogen maser model CH-76 was used as a reference standard. It provides a one to two orders of magnitude higher stability than a rubidium oscillator for averaging times of hours to days. The maser accuracy was checked against the United States Naval Observatory (USNO) Master Clock using GPS common view technique. The USNO tracks GPS satellites at times when they are also visible in Europe. Block I satellites were selected from the USNO tracking schedule and included in our own schedule. USNO results were retrieved from the Automated Data Service (ADS) via telephone. Figure 10 shows a block diagram of the test setup.

The phase difference between the UUT and the maser was recorded in 2 ways. A phase meter with strip chart recorder was connected during the entire test period. The chart was then manually sampled at 12 hour intervals to generate a file on the PC for further processing. Figure 11 presents these results. A time interval counter with a 1 ns resolution was available for a limited period of 5 days only. Time interval measurements were taken at 10 second intervals between the maser 1 PPS and the UUT 10 MHz. Figure 12 presents these results.

Although the phase difference was measured at 10 s intervals the 1 ns resolution prevents Allan Variance calculations for averaging times of less than 1000 seconds. At

300 s, for example, a 1 ns digitizing noise causes sigma = $1.6*10^{-12}$, which is worse than what is expected from the UUT. Frequency difference was estimated by applying a linear least squares fit to the phase data from the time interval counter. Figure 13 shows the result for an averaging time of 3600 seconds.

At longer averaging times frequency difference was estimated by a linear interpolation of the phase data from the strip chart recorder. Figure 14 shows the result for an averaging time of 1 day.

Similar calculations have been done for other averaging times. A summary of results is given below:

Accuracy: $0.87*10^{-12}$ with respect to UTC(USNO)

Stability (Allan Variance)	
1000 seconds	$0.60*10^{-12}$
3600 seconds	$0.40*10^{-12}$
10000 seconds	$0.45*10^{-12}$
12 hour	$0.82*10^{-12}$
1 day	$1.05*10^{-12}$
2 days	$1.30*10^{-12}$
3 days	$1.16*10^{-12}$

5.2 1 pps timing output

While it is no problem to verify accuracy and stability of a oscillator if a precise reference oscillator such as a passive maser is available it is much more difficult to evaluate timing accuracy. We carried out tests at the Deutsche Forschungsanstalt für Luft- und Raumfahrt (DLR) who maintain a precise timescale traceable to UTC(USNO). Figure 15 shows the accuracy and stability of the GPS-RR.

We were also able to verify the timing accuracy of the GPS-RR at the Physikalisch-Technische Bundesanstalt (PTB) in Braunschweig. The antenna was installed at a precisely known position. After a few days of averaging the instrument had found its own position estimate. The result is summarized below.

	PTB Survey	GPS-RR Estimate	Difference
Latitude	47°17.7680'	47°17.768'	0.0 m
Longitude	10°27.5816'	10°27.581'	0.7 m
Altitude	130.2m	131m	0.8 m

Figure 16 shows the difference between the GPS-RR 1pps output and UTC(PTB). During the first test period the UUT was operated in the air-conditioned clock building. During the second period it was located in one of the laboratories which only have central heating. The difference in stability is marginal demonstrating the excellent temperature performance of the unit. The step between the two periods is due to different lengths for both the antenna and the 1pps cable. Therefore the two periods were analysed separately by first order least square fits.

Period	Epoch	Rate
MJD 49168 - 49188 MJD 49190 - 49201	-2217ns -2602ns	$-1.1*10^{-14}$ $-1.0*10^{-15}$

The cable delays were known only for the second period. Referring the test results to UTC(USNO) we can write:

UTC(PTB) - [T(GPS-RR) - 128ns - 169ns] = -2602ns

UTC(PTB) - T(GPS-RR) = -2899ns

UTC(PTB) - UTC(USNO MC) = -2929ns (per BIPM Circular T67)

UTC(USNO MC) - T(GPS-RR) = 30ns

6. Acknowledgements

The author wishes to thank H. Nau and P. Hetzel for their help to obtain the accuracy and stability data.

/1/ Interface Control Document GPS (ICD-GPS-200), Arinc Research Corporation

/2/ Technical Characteristics Of The Navstar GPS, NATO Navstar Technical Support Group

GEOSYNCHRONOUS SATELLITE LONGITUDINAL DRIFT

DETERMINATION USING TIME MEASUREMENTS

I. Kardos (*) , P. Pâquet (**)

(*) Satellite Geodetic Observatory, Penc, Hungary
(**) Observatoire Royal de Belgique, Av. Circulaire 3,
1180 Bruxelles, Belgium

SUMMARY

The longitudinal drift of geostationary satellites could provides useful informations about time variations of the Earth gravity field. With this objective an International Campaign for Optical Observations of Geosynchronous Satellites (COGEOS) is carried out. One of the techniques used to locate a geostationary satellite is the one way range measurement obtained by recording at ground sites the epoch of arrivals of signals broadcasted by the satellite; a network of 10 stations is presently operational.

The data are processed according two different methods. The present results show a precision of the order of 100 meters for Y and Z absolute satellite coordinates, while the daily drift is probably one order of magnitude better due to averaging.
This precision has still to be improved by a factor 2 at least; this objective seems now obtainable in particular by improving some steps of the data processing related to data smoothing which is not yet used in the present analysis.

1. INTRODUCTION

Earth satellites are widely used for various scientific investigations. The position and motion, i.e. the orbit, of these satellites must be determined accurately enough to reach the requirements defined by the scientific goals. The high precision orbit determination can help to improve the geodynamic model and gives a considerable contribution to our geophysical information. High precision results of satellite geodesy led to discover effects such as the temporal variation of the J_2 coefficient of the Earth gravity field; it has been determined from the analysis of LAGEOS laser tracking data. Next largest coefficients of the geopotential are those related to geosynchronous orbits. To investigate the temporal variation of these coefficients, geostationary satellites

are appropriate; the measurement of their motion, basically the longitudinal drift, is the objective of the so called COGEOS project. The Royal Observatory of Belgium and the Satellite Geodetic Observatory of Penc are taking care of that part of the project related to positioning of geostationary satellites using Time transfer measurements.

From the first campaign organized in December 1991 we learned that a connection of the time scale of participating laboratories to the GPS time scale was mandatory; two campaigns organized in 1993 provided data which became progressively operational. For most of the stations, observations are of good quality but for some of them technical improvements are still needed.

In the next sections we report on two methods used for data processing; some results are given about the data quality and longitudinal drift of the satellite.

2. COGEOS RESEARCH PROJECT

The COGEOS project (International Campaign for Optical Observations of Geosynchronous Satellites) has been established by IAG and COSPAR as a CSTG subcommission [1]. It's main task - partly supported by the European Economic Community - is to detect the time variation of the larger coefficients of the geopotential using optical observations of geosynchronous satellites.

From experiences gained from optical tracking of geostationary satellites, the first method of observation has been extended to other ones; one of them, founded on the difference of range (DOR) from TV time measurements, has been accepted for satellite tracking [2].

The tasks to be carried out in the frame of COGEOS are the followings :

- optical observations (photographic, CCD)
- laser tracking
- radioelectric tracking of active TV and Meteosat satellites
- carrying out the data analysis

The organization of the satellite tracking is the main task for the Royal Observatory of Belgium in collaboration with the Satellite Geodetic Observatory, Penc. The first step was to test the feasibility of the proposed technique, which has the advantage to be of limited cost and free from weather conditions.

3. TIME SYNCHRONIZATION USING SATELLITE MEASUREMENTS

There are several methods to compare time scales between time service; in the recent years - thanks to space-based systems - it can be carried out globally with high accuracy. In our case let us

consider the "one way" method : terrestrial stations have receivers only and receive radio signals transmitted by satellites. If the signal structure is appropriate and can be decoded by the receiver, then the following measurements can be taken :

- code phase (or pseudo range)
- Doppler shift
- carrier phase

The most simple is to measure the code phase which in our case is the arrival time of TV pulses. The synchronization pulse - transmitted by the satellite at epoch t_0 - is received at two stations at epoch t_A and t_B respectively. Let us write $T_A - T_B$ the time difference between time scales of stations A and B, τ_{AB} the differential propagation delay; the satellite is located on a hyperboloid of focal points A and B and the following relationship can be written :

$$T_A - T_B = t_A - \tau_{AB} - t_B$$

The measurements for these quantities can be analyzed from two different points of view :

1. if τ_{AB} is known, the time difference between the two time scales can be determined, and so the time synchronization is accomplished;

2. if $T_A - T_B$ is known, the differential propagation delay is deduced and the difference of range (DOR) can be computed :

$$D_{AB} = R_A - R_B = c\tau_{AB}$$

where R_A and R_B are distances from the satellite to stations A and B.

Recent developments in time transfer technique may allow the high accuracy clock synchronization. The highest accuracy - sub nanosecond - is achievable by VLBI and satellite laser technique [6]. Potentially, the time synchronization based on GPS time transfer, can also reach the nanosecond level of accuracy [7]. Thus if the two stations are equiped with GPS time receivers, we can consider that $T_A - T_B$ is known and the difference of range can be determined with a good accuracy. If more stations are involved in such measurements, then the satellite orbit of synchronous observations can be determined from DOR measurements by purely geometric method [3]. Let us pointed out that from measurements, the range rate can be also computed :

$$c\ (t_{Ai+1} - t_{Ai}) = PSR_{i+1} - PSR_i = \Delta PSR;$$

its accuracy can be higher than the DOR one, since it does not depend on the time scale differences; however it has not yet been investigated in the present COGEOS experiment.

4. MEASUREMENTS IN THE COGEOS CAMPAIGNS

The method proposed for COGEOS is based on the interferometric principle, by measuring at 3 or more stations the arrival time of the synchronous pulses of the TV signal broadcasted by the satellite. Stations are equiped with commercial satellite TV receivers, and high stability time services. To make test measurements tracking campaigns have been organized by ORB in 1991 and in 1993. In the first campaign several methods were used to synchronize the station clock with respect to UTC (local TV, satellite TV, travelling clock, GPS receiver).

At the end of the first campaign, organized in December 1991, valuable results were not obtained because the time scales between all participating laboratories were not known with an acceptable precision which must be at least of the order of 100 nanoseconds. As conclusions of this first campaign it appeared that:

(1) all laboratories participating to such an experiment must be connected to a GPS time receiver;
(2) the delay of the clock used for the TV link should be known with respect to GPS time scale.

These two conditions were fulfilled for the two other campaigns organized in 1993. Having the GPS measurements, the time of arrival of TV time signal can be computed for all stations.

The stations involved in the COGEOS campaigns organized till now are :

- Royal Observatory of Belgium (ORB)
- Observatory Borowiec, Poland (AOS)
- Observatoire de Besancon, France (BE)
- Satellite Geodetic Observatory, Hungary (SGO)
- Osservatorio Astronomico di Cagliari, Italie (CAO)
- Finnish Geodetic Institut, Finland (MET)
- Real Instituto y Observatorio de la Armada, Spain (SF)
- Inst. of Radio Engineering and Electronics, Rep. Tcheque (CSA)
- Instituto Elettrotecnico Nazionale, Italie (TO)
- National Physical Laboratory, United Kingdom (TE).

The geographic distribution is given in Fig. 1.

5. SATELLITE LONGITUDINAL DRIFT ESTIMATES

They are two ways to determine the longitudinal drift of the satellite :

(1) to use the absolute coordinates of the satellite or
(2) to deduce it from the time measurements performed at two stations.

Both solutions have been tested.

References

1. Nobili, A. : An International Campaign for Optical Observations of Geosynchronous Satellites - COGEOS. CSTG Bulletin, No 9. 1987.

2. Kardos, I. : Geosynchronous satellite range variation measurements using its transmitted radio signal. COGEOS Workshop, Brussels, 1990.

3. Kardos, I., Frey, S. : Satellite location determination using time difference measurements. Geodézia és Kartográfia, 1993. to be published.

4. Davis, J.A., Pearce, P.R., Memary, P., Hutekins, M., Sagin, M. : Time transfer by passive use of direct broadcasting satellite signals. 5th European Frequency and Time Forum. Besancon, 1991.

5. Meyer, F., Brunet, M., Granveaud, M., Vernotte, F., Vincent, M.: Time transfer using a geostationary direct TV satellite. 5th European Frequency and Time Forum. Besancon, 1991.

6. Veillet Ch. : LASSO, two-way and GPS time comparisons COGEOS. Cogeos Workshop, Brussels, 1990.

7. Imae et al. : Precise time transfer using GPS. IRIS 93. Tokyo.

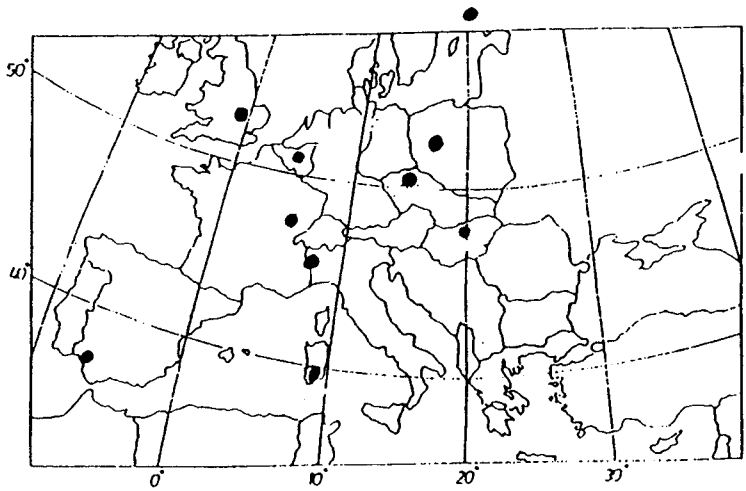


Fig. 1 : Geographic distribution of the stations participating to the COGEOS experiment.

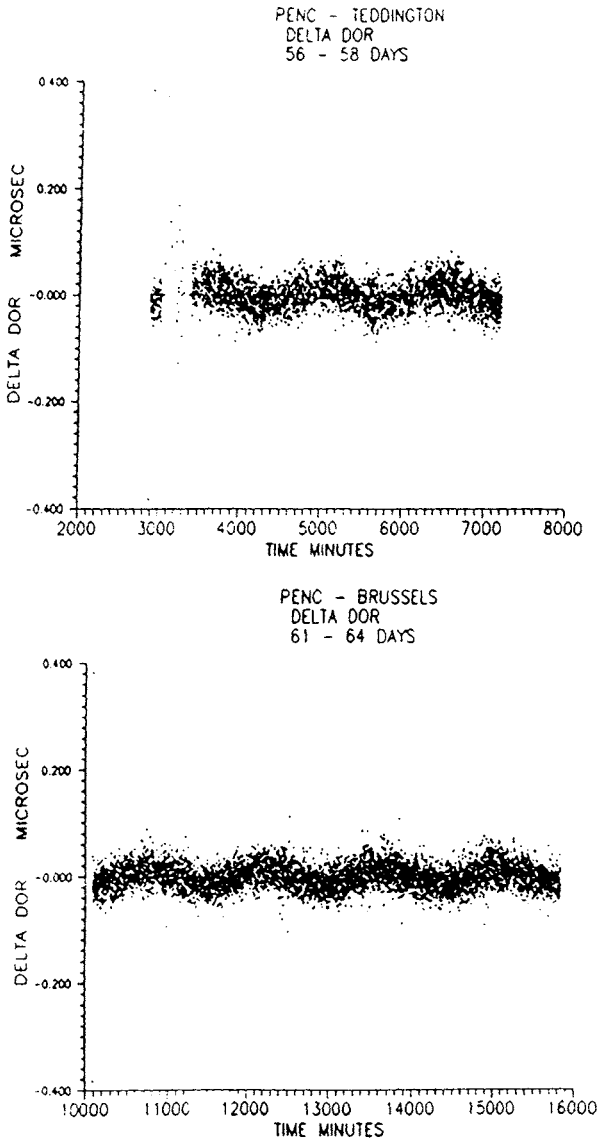


Fig. 2 : Noise of the difference of measurements.
Horizontal axis are minutes elapsed since the beginning of the campaign (23/02/1993-05/03/1993).

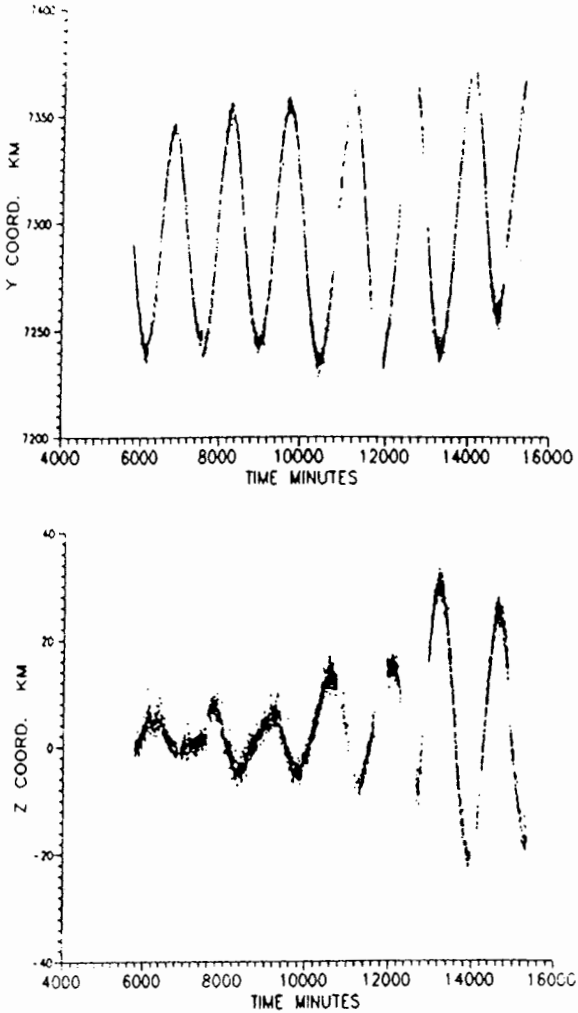


Fig. 3 : Y(top) and Z(bottom) absolute satellite coordinates as determined during the third campaigns (02/08/1993-13/08/1993).
Horizontal axis are minutes elapsed since the beginning of the campaign.

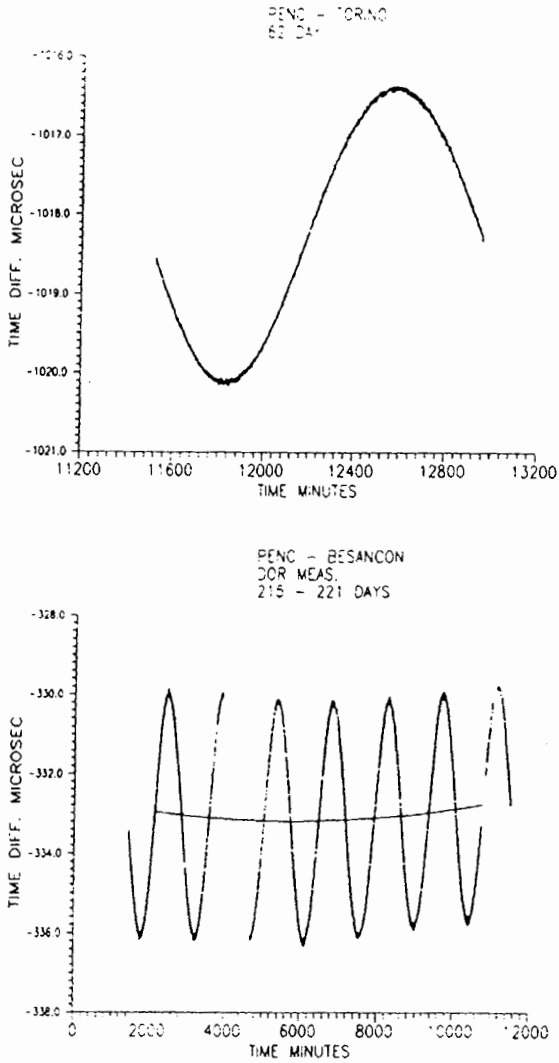
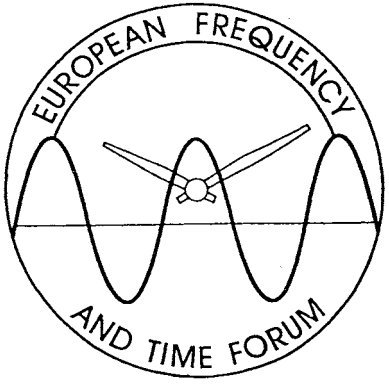


Fig. 4 : (1) the top figure shows time differences between two stations (Penc-Torino).
(2) the bottom figures shows the measured longitudinal drift (continuous line close to an horizontal one) as deduced from measurements of 2 stations.

8TH EUROPEAN FREQUENCY AND TIME FORUM WITH EXHIBITION

PROCEEDINGS
Volume II
Poster Contributions



EFTF 94

MARCH 9-11, 1994

TECHNICAL UNIVERSITY MUNICH
WEIHENSTEPHAN, GERMANY

8th European Frequency and Time Forum

Technical University Munich

Weihenstephan, Germany

March 9-11, 1994

Conference organizer and proceeding publisher:

VDI-PROJEKT UND SERVICE GMBH
Postfach 10 11 39
D-40002 Düsseldorf

Phone: int. +49/211/6214-583
Fax: int. +49/211/6214-164

CONTENTS

Volume I

page

Introduction

Local Organizer, Acknowledgement III

Executive Committee III

Scientific Committee IV

Verbal Contributions

Opening Session 3

A1: New Developments in Atomic Frequency Standards 19

B1: Resonators 73

A2: Telecommunications and Teletiming 149

B2: Materials 203

A3: Time Transfer 275

B3: Atomic Frequency Standards I 359

A4: Time Scales 407

B4: Oscillators 439

A5: Atomic Frequency Standards II 497

B5: SAW Devices 533

A6: Interrogation Signals in Atomic Frequency Standards 627

B6: Applications with Space Systems 683

Volume II

Poster Contributions 749

Local Organizer

VDI-PROJEKT UND SERVICE GMBH, Düsseldorf, Germany
Lutz Zühlke und Angelika Bohnen

1994 Chairman of the Scientific Committee

Dr. Klaus Dorenwendt, Physikalisch-Technische Bundesanstalt, Braunscheig, Germany

Acknowledgement

We wish to thank The Robert Bosch Foundation GmbH for the kind support which made it possible for scientists from Central and East Europe to participate.

EFTF Executive Committee

Audoin C.	C.N.R.S.-L.H.A., Orsay	France
Besson R.	E.N.S.M.M.-L.C.E.P., Besançon	France
Ecabert M.	FSRM, Neuchâtel	Switzerland
Gagnepain J.-J.	C.N.R.S., Paris	France
Kartaschoff P.	Cormondrèche	Switzerland
Schlueter B.	Keres Engineering Ltd., Neuchâtel	Switzerland

Honorary member of the Executive Committee

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

EFTF Scientific Programme Committee

Aubry J.P.	Thomson CEPE, Argenteuil	France
Audoin C.	C.N.R.S.-L.H.A., Orsay	France
Ballato A.	US Army Research Laboratory Fort Nonmouht	U.S.A.
Bauch A.	PTB, Braunschweig	Germany
Besson R.	E.N.S.M.N. - L.C.E.P., Besançon	France
Boloix R.	Real Observarorio Armada San Fernando	Spain
Brunet M.	C.N.E.S., Toulouse	France
Busca G.	Observatoire Cantonal, Neuchâtel	Switzerland
Cutler L.	Hewlett-Packard Co., Palo Alto	U.S.A.
De Jong G.	Van Swinden Laboratory, Delft	The Netherlands
De Marchi A.	Politecnico di Torino, Torino	Italy
Defranould Ph.	Thomson Sintra ASM, Valbonne	France
Deyzac F.	ONERA, Châtillon-sous-Bagneux	France
Dorenwendt, K.	PTB, Braunschweig	Germany
Dowsett J.	C-MAC Quartz Crystals Ltd. Harlow Essex	U.K.
Duchaussoy H.	D.R.E.T., Paris Armées	France
Feltham St.	ESA/ESTEC, Noordwijk	The Netherlands
Graf E.	Oscilloquartz S A., Neuchâtel	Switzerland
Granveaud M.	LPTF-Observatoire de Paris, Paris	France
Handrek H.	Quarzkeramik GmbH, Stockdorf	Germany
Hartl Ph.	Universität Stuttgart	Germany

Hauden D.	L.P.M.O. - C.N.R.S., Besançon	France
Heighway J.	Advanced SAW Products S.A., Bevaix	France
Hilty K.	Swiss Telecom PTT, Bern	Switzerland
Hruska C.K.	York University, Toronto	Canada
Kalliomäki K.	Technical Research Centre of Finland, Espoo	Finland
Kartaschoff, P.	Cormondrèche	Switzerland
Kihara M.	NTT Transmission Systems Laboratories, Yokosuka	Japan
Kirchner D.	Technische Universität Graz, Graz	Austria
Knight D.J.E.	National Physical Laboratory, Teedington	U.K.
Krempl P.	AVL List GmbH, Graz	Austria
Leschiutta S.	Polytecnico di Torino, Torino	Italy
Nakagiri K.	Radio Research Laboratory, Tokyo	Japan
Neubig B.W.	Tele Quarz GmbH, Neckarbischofsheim	Germany
Nieuwejaar P.	Norwegian Naval Materiel Command Haakonsvern	Norway
Nottarp K.	Institut für Angewandte Geodäsie Rodgau	Germany
Prost L.	Offive Fédéral de Métrologie Wabern	Switzerland
Quinn T.J.	Bureau International des Poids et Mesures, Sèvres	France
Reid R.	British Telecom, London	U.K.
Rutmann J.	FIEE Féd. Ind. Electrique et Electroniques, Paris	France
Starker S.	DLR, Weßling	Germany

Steele J.	National Physical Laboratory Teddington	U.K.
Thomann P.	Observatoire Cantonal, Neuchâtel	Switzerland
Uljanov A.	Kvarz Institute of Electronic Measurements, Nizhni Novgorod	Russia
Unverdross R.	Unverdross Technik, Wörthsee	Germany
Vanier J.	National Research Council Canada, Ottawa	Canada
Walls L.	National Institute of Standards & Technology, Boulder	U.S.A.
Winkler G.	US Naval Observatory, Washington	U.S.A.
Yannoni N.	Hanscom AFB Bedford	U.S.A.
Zampetti G.	Telecom Solutions, San José	U.S.A.
Zingg W.	Microcrystal div. of ETA, Grenchen	Switzerland

Poster

LEIBNIZ - D'ALEMBERT ENERGY FORMULATION, APPLICATION TO PIEZOELECTRIC MEDIA WITH INTERFACES.

N. DAHER
Laboratoire de Physique et Métrologie des Oscillateurs du C. N. R. S.
associé à l'Université de Franche-Comté-Besançon
32 avenue de l'Observatoire - 25044 Besançon Cedex

Summary

This paper deals with a mechanical and electromagnetic continuum theory with interfaces where different spaces are in interaction (volumes, surfaces, lines and points). Relations associated with the geometry of interfaces (curvatures, Torsion) and with Stoke's and transport theorems are established. A non Lagrangian energy formulation compatible with non integrable systems in view of accomodating with irreversible processes is proposed. This theory heavily lies on extrinsic and intrinsic invariance principles. The first type (gauge and rotational invariances) imposes restrictions on the physical fields and makes a distinction between the electromagnetic and the mechanical systems. The second type governs the interactions between the different spaces. By avoiding the degenerescence of the vectorial method and the difficulty associated with the multiplication of distributions encountered in energy like formulations, this approach applies in a variety of circumstances where the previous formulations reach the limit of their validity. A special attention is devoted here to deformable piezoelectric media.

Introduction

The present work develops some ideas related to continuum field theory (mechanics and electromagnetism) including discontinuities and interfacial properties. The interest of introducing interfacial fields in addition to the well-known jump or boundary conditions is now a well established fact. Numerous works have been performed in different contexts (pure mechanics, electromechanics, pure electromagnetism, thermodynamics etc.). Two types of difficulties appear in this field of investigations. The first is technical since

one must account simultaneously for jump relations linking the two adjacent volumes and for surface spatio-temporal variations including both physical entities and geometrical features such as the curvature of the interface. The second is more fundamental and concerns the mathematical construction such as the impossibility associated with the multiplication of distributions or the degenerescence relative to the extension of the basic postulates in the vectorial approach. [see eq. (5.1)]

In order to show the need for a general and new physico-mathematical approach of interfaces, different procedures are recalled and various inconsistencies and ambiguities are placed in evidence. In brief, it is shown that actually, the introduction of interfacial properties much resembles to Ptolemy's epicycles where new ones are invoked and added when the old set fails to account for new facts. The author's discovery of the numerous discrepencies, errors and particular cases convinced him that one should not treat interfacial properties locally and as something decreed from the outside. The consideration of a general procedure including all possible surface properties (charges, currents, mechanical and electromagnetic fields, polarization, magnetization etc.) does not lead just to richer structures as compared to previous works, but it also avoids errors by keeping symmetry with volume fields. Such a correspondance between volume and surface fields is of a great help for a better understanding of the system as a whole. In brief, one needs a deeper theory equally valid for mechanics and electromagnetism based on conceptually simple and economical principles of universal applicability. It will be shown that one may construct an energy formulation that satisfies the above mentioned requirements. The latter procedure goes back to Leibniz, Bernouilli and d'Alembert. It is not to be confused with the widespread Lagrangian formulation which constitutes a particular case valid only for integrable systems. To distinguish it from the latter procedure as well as from d'Alembert's method which is of a purely mechanical nature we call it "Leibniz-d'Alembert principle". Indeed, although Leibniz have much contributed to the development of mechanics, by emphasizing the importance of conservation principles and particularly the energy concept [1] his name was overshadowed by Newton's system of the world. He paid a high price for his opposition to Newton and his followers. His ideas had to wait until the advent of relativity theory to be justified and recognized on a rational ground [2, 3]. His attitude towards theory and experiment is very actual. Indeed, he

favored the multiplicity of points of views claiming that there is no one to one correspondance between the informations or facts delivered by nature and the theory constructed in view of explaining these facts. According to Leibniz, the economy in Nature is associated with universal principles and not with their manifestations that may be of a great variety. This program was revived particularly, by Einstein who was not afraid of introducing myriads of components (ten component entity instead of a one-component gravitational charge) because he knew the deeper principle behind his construct.

In the present context, numerous ideas satisfy Leibniz's methodology (Principle of sufficient reason, passage from the possible to the actual, multiplicity of points of views, invariance requirements etc.). However, the development of these ideas in a detailed deductive manner is quite impossible in a brief space. At best we can only outline some of the important features and results. In particular, the attention will be focussed on different situations that cannot be treated by the methods presently available. Then, new procedures will be given. The latter open new perspectives and pave the way for further physical investigations.

1. Inconsistencies and limitations of classical procedures.

In Ref. [4] dealing with deformable piezoelectric semiconductors, although a number of interesting results are obtained, different inconsistencies appear in the interfacial equations. In particular, on account of equations (3.39), (6.12), (3.35), (6.19), (3.11) and (8.14), one may notice different problems. The authors introduce a surface electric charge, in Gauss' law keeping Ampere's law unchanged. Since the interfacial equations are not independant one is led to a contradictory result. In addition, it is usual to postulate Gauss' law, Ampere's law and the law of conservation of charge independently and apply the pill box method to obtain interfacial conditions. However, since the three statements are not independent one must verify the global mathematical coherence. To do so, relations are needed about the spatio-temporal variations of the normal to the interface. But these informations are absent from these treatments, and one is unable to make a verification of the correctness of the postulates.

In more theoretically oriented works the Pill-box method is replaced by generalized transport and stoke's theorems so that one gets more acurate

results, but does not solve the problem completely. In particular, the risk of contradiction remains as long as one postulates dependent relations. This situation may be seen in Ref [5] where a general theoretical procedure is proposed but a discrepancy occurs as to the introduction of surface polarization. (See eq. (3.9.3) and (3.9.7) of Ref. [5]). The solution to this problem requires the establishment of new mathematical theorems actually absent from the present approaches. In particular, here not just the spatio-temporal variations of the normal to the interface has to be determined, but also the variations of the binormal to that interface (See the second term of eq. (5.1) which is absent in Ref. [5]). Another school of thought have treated the problem of interfacial electromagnetism from a local point of view through a thin interphase layer approximated by a singular surface [6, 7]. This method disregards the possible influence of the curvature of the interface and postulates strong requirements so that one is led to particular results. In addition, it applies only when the volume fields are known and cannot be used to generate new field equations needed for our purpose. In brief one may say that although many procedures are used none of these is satisfactory. In the next Sections, we shall divide the presentation into two mains parts : the first one is purely mathematical and the second concerns physical principles and applications. The mathematical part will include two subsections related to the establishment of (i) relations associated to curved surfaces and lines and (ii) new theorems of a general interest and particularly useful here to the solution of the inconsistency related to surface polarization. As to the physical part, it contains three subsections : the first concerns the establishment of Ampere's law through an integration procedure minimizing the number of fundamental postulates and avoiding thus a possible contradiction. The second shows the interest and the need for a new energy like formulation. The third is practical and concerns the modelization of deformable ferroelectric semiconductors.

2. General Dynamical interfaces.

In order to study general interfaces which may be singular surfaces, lines or points, we shall adopt a Leibnizian methodology : "Explanation of local states by global arguments". Thus a line is neither regarded as an infinite number of points nor as the motion of one point but rather as the intersection of two dynamical surfaces represented by $f(x,t)$ and $g(x, t)$. This procedure may be

extended to account for the motion of a singular point along a line element. In this case, a point is obtained through the intersection of three surfaces. A linear combination of the three dynamical surfaces allows to determine a three dimensional coordinate system ($\boldsymbol{n}$, $\boldsymbol{b}$, $\boldsymbol{\tau}$) as well as three motions $\hat{v}.\boldsymbol{n} \equiv \hat{v}_n$, $\bar{v}.\boldsymbol{b} \equiv \bar{v}_b$ and $\bar{v}.\boldsymbol{\tau} \equiv \bar{v}_\tau$ associated with the singular surface, line and point. We shall restrict ourselves here to $\boldsymbol{n}$ and $\boldsymbol{b}$. Thus after some calculations one gets :

$$\boldsymbol{n} = \nabla f \Big/ \frac{\partial f}{\partial n} \qquad \hat{v}_n = -\frac{\partial f}{\partial t} \Big/ \frac{\partial f}{\partial n} \qquad \frac{\partial}{\partial n} \equiv \boldsymbol{n}.\nabla \tag{2.1}$$

$$\boldsymbol{b} = \hat{\nabla} g \Big/ \frac{\partial g}{\partial b} \qquad \bar{v}_b = -\frac{\hat{\partial} g}{\partial t} \Big/ \frac{\partial g}{\partial b} \qquad \frac{\partial}{\partial b} \equiv \boldsymbol{b}.\nabla \tag{2.2}$$

From which one may deduce the following relations :

$$\hat{\nabla} \times \boldsymbol{n} = 0 \qquad \hat{\nabla}.\boldsymbol{n} = -2\overset{m}{\Omega} \qquad \frac{\hat{\partial}\boldsymbol{n}}{\partial t} + \hat{\nabla}\, \hat{v}_n = 0 \tag{2.3}$$

$$\bar{\nabla} \times \boldsymbol{b} = \frac{\boldsymbol{b}}{Tg} \qquad \bar{\nabla}.\boldsymbol{b} = -2\overset{n}{\Omega} \qquad \frac{\bar{\partial}\boldsymbol{b}}{\partial t} + \bar{\nabla}.\bar{v}_b = \boldsymbol{K} \tag{2.4}$$

with

$$K = \left(\frac{\partial \hat{v}_n}{\partial b} + 2\left(\overset{m}{\Omega} - \overset{n}{\Omega} \right) \bar{v}_b \right) \boldsymbol{n} - \hat{v}_n \frac{\boldsymbol{\tau}}{Tg} \;,\; \frac{1}{Tg} = \boldsymbol{n}.\frac{\partial \boldsymbol{b}}{\partial \tau} \;,\; \bar{\nabla}.\boldsymbol{n} = -2\overset{n}{\Omega} \tag{2.5}$$

The surface and line spatial and temporal operators are defined by

$$\hat{\nabla} = \nabla - \boldsymbol{n}\frac{\partial}{\partial n} \qquad \frac{\hat{\partial}}{\partial t} = \frac{\partial}{\partial t} + \hat{v}_n \frac{\partial}{\partial n} \tag{2.6}$$

$$\bar{\nabla} = \hat{\nabla} - \boldsymbol{b}\frac{\partial}{\partial b} \qquad \frac{\bar{\partial}}{\partial t} = \frac{\hat{\partial}}{\partial t} + \bar{v}_b \frac{\partial}{\partial b} \tag{2.7}$$

The parameters $\overset{m}{\Omega}$, $\overset{n}{\Omega}$, $\overset{s}{\Omega}$ and Tg denote respectively the mean, normal, geodesic curvatures and the geodesic torsion.

Equations (2.3)-(2.4) generalize those given in Ref. [7] where no restrictions are imposed on normal derivatives and where we have additional relations governing the curvatures, torsion and dynamic of singular lines.

3. Other interfacial operators and a new line transport theorem.

The surface divergence $\hat{\nabla}.\hat{J}$ is the counterpart of the volume one $\nabla.J$, only when the normal component of $\hat{J}$ vanishes. Otherwise, one may show in different ways that the surface counterpart of ∇ is the following operator.

$$\tilde{\nabla} = \hat{\nabla} + 2\,\overset{m}{\Omega}\, n \quad \Rightarrow \quad \tilde{\nabla}.n = 0 \tag{3.1}$$

As to the line operator one obtains

$$\bar{\nabla} = \tilde{\nabla} + 2\,\overset{n}{\Omega}\, n + 2\,\overset{s}{\Omega}\, b \quad \Rightarrow \quad \bar{\nabla}.b = 0 \tag{3.2}$$

These new operators have their temporal counterparts as follows

$$\frac{\tilde{\partial}}{\partial t} = \frac{\hat{\partial}}{\partial t} - 2\,\overset{m}{\Omega}\, v_n \qquad , \frac{\bar{\partial}}{\partial t} = \frac{\tilde{\partial}}{\partial t} - 2\,\overset{n}{\Omega}\, v_n - 2\,\overset{s}{\Omega}\, v_b \tag{3.3}$$

The latter spatio-temporal operators appear in a natural manner when dealing with Green-Gauss and transport theorems.

Notice that if one identifies $\tilde{\partial}/\partial t$ with $\tilde{\nabla}_4$ and $-\hat{v}_n$, with n_4, then space and time variations may be set in a compact 4-dimensional notation. This greatly simplifies the formulas. In addition, this suggests a relativistic treatment with Lorentzian signature which is given elsewhere and from which the actual presentation may be derived in the limit of small velocities.

Before ending this paragraph, let us give an essential result, which is necessary to the establishment of Maxwell equations with interfaces and which is missing in previous works leading to a partial result. (see Ref. [5])

$$\frac{\hat{d}}{dt}\int \hat{D}.b\,d\ell = \int \left[\frac{\tilde{\partial}\hat{D}}{dt} + \tilde{\nabla}\times(\hat{D}\times\hat{v}) + \hat{v}(\tilde{\nabla}.\hat{D}) + (n.\hat{D})\left[(\tilde{\nabla}\otimes\hat{v})\,n \right] \right].b\,d\ell \tag{3.4}$$

This result has been obtained here (i) by establishing a transport theorem on singular lines for any scalar field ϕ and any motion $\bar{v}$, (ii) by letting the motion of the singular line coïncide with that of a boundary (i.e $\bar{v} = \hat{v}$) and (iii) by setting $\phi = \hat{D}.b$. Obviously, one needs to know how the binormal varies. Consequently, this clearly shows the necessity for the previous section absent from previous works.

4. Generalized Ampere's law through a Leibnizian procedure. (Minimum of hypotheses - Maximum of generality).

By use of surface Green - Gauss and transport theorems, one may establish the surface counterpart of Gauss' law and the law of conservation of electric charge as follows :

$$\tilde{\nabla}.\hat{D} + n\,[\![D]\!] = \hat{q} \qquad , \frac{\tilde{\partial}\hat{q}}{\partial t} + \tilde{\nabla}.\hat{J} + n.[\![J - q\hat{v}]\!] = 0 \tag{4.1}$$

As to Ampere's law, it may be derived by integrating equation $(4.1)_2$. One starts by proposing a general solution including physical as well as geometrical spatio-temporal variations in addition to the influence of the adjacent bulk fields. Then, an identification procedure is performed leading to a particular form compatible with eq. (4.1). One starts with 66 degrees of freedom *(H_{ij}, K_i, C_{ijk}, C_{ij}, F_{ij} and S_{ij})*. After identification, only three degrees of freedom survive *($H_{ij} = -H_{ji} = \varepsilon_{ijk}\,\hat{H}_k$)*. The others combine to already defined parameters as follows :

$$\begin{aligned} \hat{J} = \tilde{\nabla}\times\hat{H} - \frac{\tilde{\partial}\hat{D}}{\partial t} + (\,\tilde{\nabla}\otimes n\,)\Big(\, n\times\hat{H} + \hat{v}_n\,\hat{D}\,\Big) \\ + \frac{\tilde{\partial} n}{\partial t}\,(\,n.\hat{D}\,) + n\times[\![H]\!] + \hat{v}_n\,[\![D]\!] \end{aligned} \tag{4.2}$$

This procedure of integration which remains valid for singular lines is typically Leibnizian since on the one hand it satisfies the minimality of the basic principles and on the other hand a passage is performed from the possible to the actual by restriction principles.

5. Various reasons in favor of a generalized energy like formulation.

Actually, the vectorial approach remains the most general one in continuum physics in reason of its application simultaneously to irreversible processes and to both mechanics and electromagnetism. However, in the present context, where one is concerned with complex media such as microstructures or deformable ferroelectric semiconductors with interfaces, this approach presents two drawbacks. Firstly, it cannot be used safely in the derivation of the various couplings between the different continua because of the absence

of a general guiding principle such as the principle of objectivity (invariance under rigid body motion) introduced by Germain, Maugin and co-workers by use of d'Alembert's principle [8]. Secondly, when interfaces are taken into account, then the vectorial approach reaches rapidly the limit of its validity. See Ref. [5]

$$\frac{d}{dt}\int \boldsymbol{D}.\boldsymbol{n}da + \frac{\hat{d}}{dt}\int \hat{\boldsymbol{D}}.\boldsymbol{b}\, d\ell + \oint \mathcal{H}\boldsymbol{x}d\ell + ? = \int \mathcal{J}.\boldsymbol{n}da + \int \hat{\mathcal{J}}.\boldsymbol{b}\, d\ell \quad (5.1)$$

This is not the case in an energy like formulation where a unified mode of expression is possible through the following form: (only the pertinent contribution is reproduced here).

$$\int_D K_{ij}\, \delta R_{i,j}\, dv = 0 \qquad \forall \delta R_i \quad (5.2)$$

where $\delta R_{i,j}$ is the gradient of an infinitesimal vector δR_i.

The latter may be identified with an infinitesimal displacement in the mechanical case and with an infinitesimal vector potentiel in the electromagnetic case. Invariance properties place restrictions on the dual field K_{ij} (rotational invariance $\Rightarrow K_{ij} = K_{ji}$, gauge invariance $\Rightarrow K_{ij} = -K_{ji}$). These questions have been discussed elsewhere by the author [9].

More important to us here, is the problem of singular surfaces and lines. Unlike equation (5.1), eq. (5.2) is directly generalizable as follows :

$$\int_{D-\Sigma} K_{ij}\, \delta R_{i,j}\, dv + \int_{\Sigma-\Gamma} \hat{K}_{ij}\, \hat{\delta}\hat{R}_{i,j}\, da + \int_{\Gamma} \bar{K}_{ij}\, \bar{\delta}\bar{R}_{i,j}\, d\ell \quad (5.3)$$

Such a procedure remains compatible with non integrable systems since Kij does not necessarily derive from a potential. As to the interaction between the different spaces, one may start with a general form, then by use of an invariant principle obtains, the following relation

$$\delta W_{int} = \int_{\Sigma-\Gamma} \left[\!\!\left[K_i \left(\delta R_i - \hat{\delta}\hat{R}_i \right) \right]\!\!\right] da \quad (5.4)$$

Such a contribution has not its counterpart in the vectorial approach. Indeed while a vector (derivable from a potential or not) can be manipulated on an

equal footing as a vector function or as a distribution, a scalar product (energy) cannot. Consequently, this extra term may be regarded as an argument that compensates and avoids the difficulty associated with distribution products. It is somehow a measure of our ignorance. "We know that something happens at the limit between two media but we do not know precisely what". The vector Ki is actually unknown, it is simply introduced as the dual of (δRi $-$ $\hat{\delta}\hat{\text{Ri}}$). It will be uniquely determined at a later stage by the application of the virtual energy principle.

6. General scheme for the study of deformable ferroelectric semiconductors :

In this last section we single out an important example where the conventional methods do not apply safely. This is the case of a deformable medium presenting ferroelectric properties, semiconduction phenomena and interfacial effects. Schematically, such a problem may be modelized as follows :

Electromagnetism

First gradient theory with respect to volume and surface vector potentials : in this description, Maxwell equations are governed by the principle of gauge invariance.

Electro-mechanics

First gradient theory with respect to volume and surface material velocities, polarizations and motions of charge carriers : the interactions of the different continua are governed by the principle of objectivity (invariance under rigid body motion).

Thermodynamics of irreversible processes :

Combination of the electromechanical equations with the total energy conservation principle and total entropy production, in view of obtaining reversible and irreversible constitutive relations.

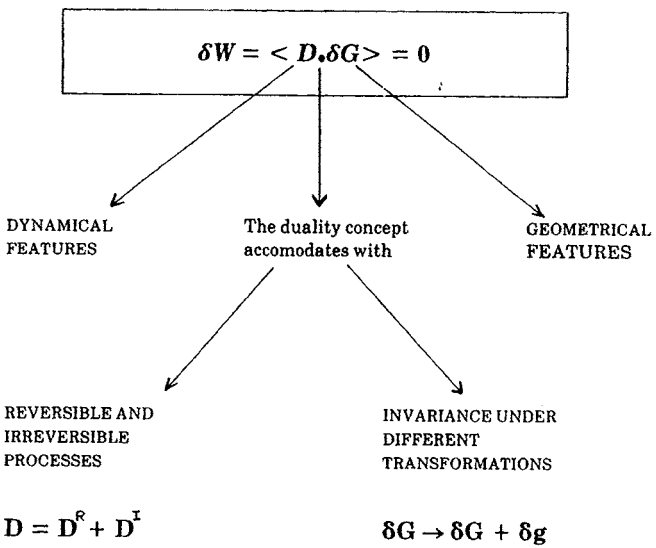
Aknowledgment :

In addition to the qualitative ideas borrowed from Leibniz and d'Alembert philosophies of nature, the author is much indebted to professor Maugin whose work on d'Alembert principle in a Galilean and Einsteinian frameworks and in complex media constituted a precious tool and a rational basis for the author investigations.

LEIBNIZ - D'ALEMBERT GEOMETRO - DYNAMICS

Orthogonality relation including

VOLUME ENERGY	VOLUME-SURFACE INTERACTION ENERGY	SURFACE ENERGY



CONSTITUTIVE RELATIONS :

$$D \rightarrow \begin{cases} D^R = \partial \mathcal{L} / \partial G & \text{(Integrable)} \\ D^I \ / \ D^I . \delta G \geq 0 & \text{(Non integrable)} \end{cases}$$

$D^I \equiv 0 \Rightarrow$ LEIBNIZ - D'ALEMBERT PRINCIPLE $\Leftrightarrow$ MAUPERTUIS - LAGRANGE PRINCIPLE

INVARIANCE PRINCIPLES :

$\delta g \rightarrow$

Intrinsic Invariances :

Multi-space interactions. (Volume - Surface, Surface-Line, Line-point)

Extrinsic Invariances :

- **Gauge transformations** (ELECTROMAGNETISM)
- **Objectivity Requirement** (ELECTRO-MAGNETO-MECHANICAL INTERACTIONS)

REFERENCES

[1] N. Daher, "Relativité Leibnizienne : philosophie relationniste intrinsèque et caractéristique universelle multiple. Lien avec les fondements des physiques newtonienne et einsteinienne", EPIPHYMATHS, Université de Franche-Comté, UFR des Sciences et Techniques, Besançon (1992).

[2] H. Reichenbach, "The philosophy of Space and Time" Dover Publications, INC. Newyork (1958).

[3] M. Jammer, "Concepts of Space, the history of theories of Space in Physics". Harvard University Press (1954).

[4] H. de Lorenzi and H. F. Tiersten, "On the interaction of the electromagnetic fields with heat conducting deformable semiconductors", J. Math. Phys. 16-958-953 (1975).

[5] A. C. Eringen, G. A. Maugin, "Electrodynamics of Continua I, Foundations and Solid Media" Spring Verlag (1989).

[6] P. A. Wolf, A. M. Albano, "Non-equilibrium thermodynamics of interfaces including electromagnetic effects" Physica, 98A, 791-508, (1979).

[7] A. M. Albano, D. Bedeaux and J. Vlieger, "On the description of interfacial electromagnetic properties using singular fields, charge density and currents at a dividing surface", Physica, 102, A 105-119 (1980).

[8] G. A. Maugin, "The method of virtual power in continuum mechanics. Application to coupled fields", Act. Mech. 35, 1-70 (1980).

[9] N. Daher, "On a general energy method in continuum physics with electromagnetic interactions and interfaces". IUTAM Symposium - Victoria - Canada, August (1993).

ERRORS IN THE NONLINEAR CONSTANTS OF α-QUARTZ

Carl K. Hruska, Peggy Ng and Xu Fan
Piezoelectricity Research Laboratory
Department of Mathematics and Statistics, York University
North York, Ontario M3J 1P3, Canada

While the values of the traditionally calculated standard errors for the nonlinear constants of α-quartz are in many cases very large, they are much smaller than the differences among the values of these constants obtained by different methods. This conclusion now holds even when the standard errors are increased by incorporating other (usually disregarded) sources of random errors: the linear material constants and misorientation of the quartz specimens used in determining the nonlinear constants. Consequently, the present conflicting values of the nonlinear constants are not only difficult to accept as reliable but to explain as well. Further attempts to remove these difficulties will require a systematic reconsideration of the past experiments, their interpretation, and a search for new independent sources of data.

1. Introduction

Some of the values of the third-order nonlinear electromechanical constants of α-quartz obtained independently and by different methods are in sharp disagreement [1,2]. In some cases the disagreement amounts to an order of magnitude and includes the sign as well. While such conflicting values are difficult to recommend for acceptance and possible use, the disagreement would be easier to explain and its origin simpler to trace if the standard errors of these values were sufficiently large.

It has been pointed out that traditionally calculated standard errors of the nonlinear constants reflect only the random errors in the experimental data used for their calculation. More realistic and larger standard errors are obtained if random errors in other quantities entering into the calculation are also taken into account. A method to do so has been described in [3].

This paper reports the results of an attempt to explain the large differences in the values of the nonlinear constants by recomputing their standard errors. It was undertaken in order to discover whether the recomputed standard errors will increase sufficiently to span the differences between the conflicting values of the nonlinear constants. The calculations made apply to the values of the third-order nonlinear electromechanical constants of α-quartz independently determined by the resonator and the transit-time method [1].

All nonlinear material constants appearing in this paper are related to the state of zero strain and zero electric field. Their numerical values, as well as those of other quantities, are stated for right-hand quartz at room temperature and the frame of reference according to the IEEE Standard 176 of 1978 [4].

2. Disagreement among nonlinear constants

Table I shows the values and standard errors of seven nonlinear constants of α-quartz independently obtained using the resonator method and the transit-time method. They have been taken from [1] and computed

TABLE I. Disagreement in the nonlinear constants of α-quartz obtained independently using the resonator and the transit-time method.

parameter	resonator method	transit-time method
k_{315}	-0.79 ± 0.03	-0.89 ± 0.02
k_{11}	-3.19 ± 0.91	4.92 ± 6.04
k_{13}	-8.62 ± 2.59	13.13 ± 3.75
k_{14}	-2.43 ± 0.54	-5.93 ± 2.54
k_{31}	-12.40 ± 2.99	0.85 ± 3.18
k_{41}	-4.42 ± 0.67	-1.44 ± 2.21
$q_{111}.10^{20}$	-0.29 ± 0.24	3.72 ± 2.41

k_{ijk} are in N/(V.m), k_{ij} are dimensionless, q_{111} is in F/V. The errors are standard errors reflecting the experimental errors pertaining to each method.

using the least-squares fit while carefully avoiding the pitfalls of calculations described in [5].

TABLE II. Definitions of the nonlinear constants of α-quartz in Table I.

$$k_{315} = f_{315}$$

$$k_{11} = l_{11} + 1.205\ .10^{-12}\ c_{111} - 0.603\ .10^{-12}\ c_{112} - 0.379\ .10^{-12}\ c_{114} + 0.379\ .10^{-12}\ c_{124} + 0.060\ .10^{-12}\ c_{144} - 0.603\ .10^{-12}\ c_{222}$$

$$k_{13} = l_{13} + 2.00\ l_{44} + 1.205\ .10^{-12}\ c_{113} - 1.205\ .10^{-12}\ c_{123} - 0.759\ .10^{-12}\ c_{134} + 0.060\ .10^{-12}\ c_{344}$$

$$k_{14} = l_{14} - 2.411\ .10^{-12}\ c_{124} - 0.379\ .10^{-12}\ c_{144} + 0.379\ .10^{-12}\ c_{155} + 0.060\ .10^{-12}\ c_{444}$$

$$k_{31} = l_{31} + 2.00\ l_{44}$$

$$k_{41} = l_{41}$$

$$q_{111} = \kappa_{111} - 2.045\ .10^{-23}\ l_{12} - 2.465\ .10^{-35}\ c_{111} + 1.233\ .10^{-35}\ c_{112} + 0.776\ .10^{-35}\ c_{114} - 2.329\ .10^{-35}\ c_{124} - 0.367\ .10^{-35}\ c_{144} + 0.244\ .10^{-35}\ c_{155} + 1.234\ .10^{-35}\ c_{222} + 0.038\ .10^{-35}\ c_{444}$$

The constants are combinations of the fundamental third-order nonlinear elastic constants c_{ijk} in N/m^2, electroelastic constant f_{315} in N/(V.m), electrostrictive constants l_{ij} (dimensionless) and third-order dielectric constant κ_{111} in F/V.

Most of the constants in Table I are linear combinations of the fundamental third-order nonlinear electromechanical constants of quartz. Some of the combinations are quite complex. Their detailed definitions, in terms of the fundamental constants, are given in Table II, where f_{315} is one of the electroelastic constants, l_{ij} are the electrostrictive constants, c_{ijk} are the third-order elastic constants and κ_{111} is the only independent third-order dielectric constant of

quartz. These fundamental constants are recorded using the matrix index notation and the traditional choice indices. The electrostrictive constants l_{ij} are the elements of the total electrostrictive tensor which includes the vacuum electric-stress tensor [6].

There are three reasons, at least one of them present in each of the seven cases shown in Table I, why none of the values of the nonlinear constants in Table I are easy to accept: (1) the standard errors are too large suggesting that the values of the nonlinear constants are unreliable; (2) the values produced in parallel by the resonator method and the transit-time method are substantially different (up to one order of magnitude and also in sign); and (3) the values produced in parallel do not comply with the traditional rule of thumb that the disagreement among values of the same quantity determined by different means should not exceed the size of their standard errors. In the absence of a more exact criterion this is regarded as an indication of significant disagreement between results obtained by different methods.

The first two reasons may be a consequence of poor accuracy of one or both of the two methods or of the unsuitable choice of quartz samples for the experiments conducted under the methods. The third reason is different. It suggests that there may be a fundamental and irreconcilable contradiction between the results produced by the two methods. It casts doubt on their capacity to produce consistent results (however inaccurate) and, therefore, on their soundness (for whatever reason). Concerned primarily with the third reason and attempting to remove it, this article concentrates on the size of the standard errors.

3. More realistic standard errors

The standard errors given in Table I have been determined using the least-squares algorithm and, as such, reflect only the errors from a single source. The source is the random errors in the experimental data used to determine the nonlinear constants. More realistic and necessarily larger standard errors are obtained if other sources of errors are also taken into account. Since sufficiently large standard errors would remove the principal objection to the acceptability of the

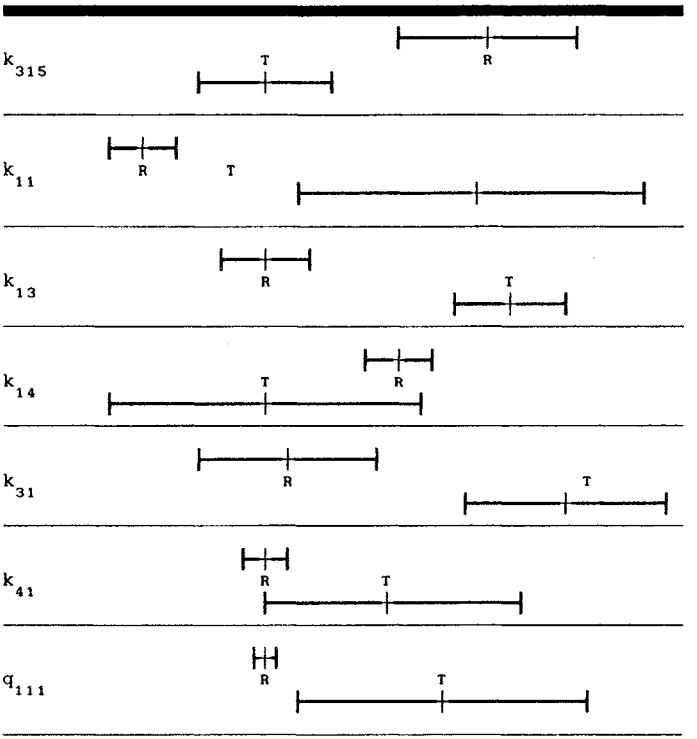


Fig.1. The values of the nonlinear constants and their standard errors including all sources of random errors. The points marked R and T represent the values obtained by the resonator and the transit-time method, respectively. Segments plotted on both sides of R and T are their standard errors. The standard errors are not large enough to contain both R and T. In order to see this, each pair or related plots must projected on the same horizontal axis. To accomodate the widely diferring values of R and T of the seven nonlinear constants all in one diagram, the scale used for each constant is different. The choice of the scale units is irrelevant and not shown.

nonlinear constant values, i.e., their inconsistency, the search for these standard errors is the obvious course to take.

The principal sources of additional errors in the nonlinear constants are twofold: the misorientation of the crystal specimens used to produce the experimental data from which the nonlinear constants are calculated and the uncertainty in the linear constants used during the calcuation. A method to obtain the size of the standard errors in the nonlinear constants due to these two sources has been described in [3].

As there exists no exact expression to combine the additional errors with the experimental errors given in Table I to produce the total standard error of the nonlinear constants, a simplified formula has been suggested in [3]. It was tested in preparation for this paper for a number of cases using a suitable randomization procedure. It was found to produce results which are systematically smaller by about 5% than they should be.

Taking this into account, the total standard error of all fourteen nonlinear constant values in Table I have been estimated and plotted in Fig.1. The plot displays the size of the standard errors in relation to the difference between the values of the nonlinear constants for each pair of related values. It demonstrates that the standard errors are generally not large enough to simultaneously contain both values of the nonlinear constants obtained by the resonator (R) and transit-time (T) method. Consequently, enlarging standard errors by the incorporation of the effects of uncertainty in the linear constants and of the misorientation of experimental samples does not remove the inconsistency problem.

4. Conclusion

The disagrement among the third-order nonlinear electromechanical constants of α-quartz is too large to be conveniently explained by means of random errors. Its cause must then be more fundamental and systematic in nature. Consequently, in order to obtain a single, complete and reliable set of third-order nonlinear electromechanical constants of α-quartz it is now necessary

(1) to reconsider the theory of interpretation of the current experiments; and

(2) to look for possible sources of systematic errors in these experiments.

While the standard errors of the nonlinear constants are too small to make the existing discrepancies readily explainable, they are, at the same time, too large for any value of the related material constant to be considered reliable. Therefore, concurrently with the above, an additional effort must be directed towards reducing the size of the standard errors and the general uncertainty regarding their values. This should be attempted by

(1) using a better design of the quartz samples in the current experiments; and

(2) seeking new sources of experimental data to provide new opportunities of an independent verification of their values.

Positive steps in this direction are believed to be new observations currently conducted by Xu Fan using the lateral dc field interactions with the thickness mode of the precision custom-made quartz plates supplied by Micro Crystal, Division of ETA S.A., Switzerland, as well as experimenting with new crystal units provided by Tesla a.s., Czech Republic, and designed to reduce the standard errors of the electrostrictive constants.

The solution of the above problem should not be viewed just as a search for the third-order nonlinear electromechanical constants of α-quartz. It is a search for suitable and reliable methods to determine the nonlinear constants in crystals in general.

Acknowledgements

The authors wish to acknowledge the support of this work by a grant from the Natural Sciences and Engineering Research Council of Canada and the assistance of Pauline Hruska who helped with the final stages of the preparation of this paper.

References

[1] Hruska, Carl K.: J.Appl.Phys., vol.73 (1993), pp. 343/7.

[2] Hruska, Carl K.: J.Appl.Phys., vol.74 (1993), pp. 7005/7.

[3] Hruska, Carl K.: J.Appl.Phys., vol.75 (1994), pp. 77/83.

[4] IEEE Standard on Piezoelectricity, IEEE Std 176-1978. Available

from The Institute of Electrical and Electronics Engineers, Inc., 345 East 47th Street, New York, NY 10017.

[5] Hruska, Carl K.: J.Appl.Phys., vol.72 (1992), pp. 2432/9.

[6] Nelson, Donald F.: J.Acoust.Soc.Am., vol.63 (1978), 1738/48.

OPTIMIZATION OF LFE SC-CUT RESONATORS OSCILLATION CIRCUITS

B.Gniewińska, A.Masiukiewicz
Piezoelectric Department
Tele and Radio Research Institute, Ratuszowa 11, 03-450 Warsaw

ABSTRACT

Well known advantages of LFE resonators (e.g. lack of the electrodes in the active part of the blanks) combined with double rotated SC-cut resonator features, lead to new possibilities in the area of high stability oscillators. The results of resonator parameters measurements and some problems concerning oscillation circuits and oven optimization are discussed.

1.INTRODUCTION

Lateral field excitation resonators have a few advantages compared with typical high stability resonators /ref. 1,2,3,4,5,6 /. The most important features are as follow:
- better long term stability due to removing of the electrodes from the active part of the blanks,
- higher quality factor of unloaded resonator as a result of surface free from the metal layer,
- more efficient suppresion of unwanted modes (e.g. B mode of SC-cut resonators).

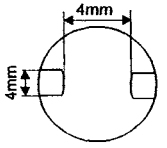
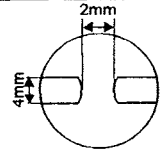
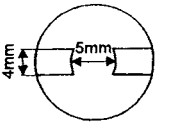
But on the other side lateral excitation change basic parameters of the resonators. Very high as for this frequency range motional resistance of the resonator, complicates the construction of the oscillation circuit which should assure high short-term frequency stability and small motional capacitance considerably reduce the frequency adjustment range.

Design of the oscillation circuit for the 5 MHz SC-cut lateral field excitation resonator, which guarantees the electronic circuit influence on short-term frequency stability lower then E-13 for averaging time $\tau \geq$ 1sec is presented in this paper. The construction and parameters of three chamber oven of very small temperature fluctuations at the level of E-7 are also presented.

2.PARAMETERS OF LFE SC-CUT RESONATORS

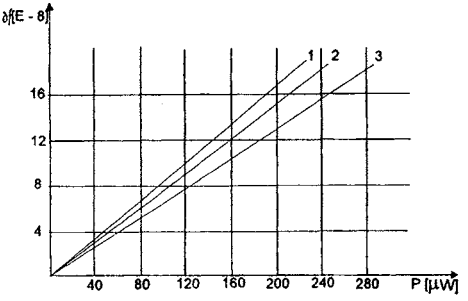
A few types of 5 MHz SC-cut resonators operating at 3rd overtone mode have been realized. Parameters achieved for three types of electrodes configuration are presented in Table 1./ref. 7/

Table 1 LFE resonator parameters

Elektrodes configuration	Motional resistance [Ω]	Motional capacistance C1 [E - 18]	Uniloaded resonator Q factor [E6]	B mode suppresion [dB]
	1100 - 1300	average 9	2.4 -2.6	> 30
	400 - 650	average 18	2.8 - 3	> 40
	1500 - 1800	average 8	2.2 - 2.6	-

Besides basic parameters, dynamic temperature behaviour and nonlinearity of resonators were measured. Typical measured, dynamic temperature coefficient of the LFE resonator is within the range of ±2E-7 (the measurement was carried out according to the method presented in reference 8, basing on estimation of resonator response to step temperature change).

The characteristic of nonlinear behaviour of the resonator is presented in figure 1. According to expectation this characteristic is very similar to the characteristic obtained for typical SC-cut resonators.



1 SC-cut resonator R1=70Ω (nr 258)
2 LFE SC-cut resonator R1=500Ω (nr 15)
3 LFE SC-cut resonator R!=600Ω (nr 13)

Fig.1. Frequency versus drive power characteristic

3.OSCILLATION CIRCUIT

Considering resonator parameters, application of following oscillation circuits were taken into account:

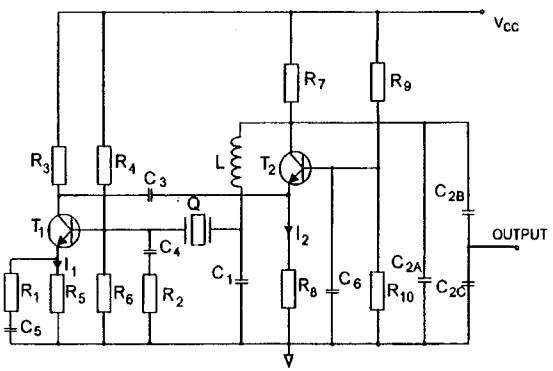
- bridge circuit with differential amplifier,
- Butler with the resonator connected through the transformer to the emitter circuit,
- Butler with the resonator in the base circuit.

Essential features of studied circuits are presented in Table 2.

Table 2. Features of oscillation circuits

Type of Circuit	Features
Bridge	No reactance elements in the feedback circuit. Improved efficiency of phase noise reduction, because of bridge circuit influence. Very effective mechanism of drive level limitation.
Butler with the resonator in the emitter	Neccesity of matched impedance circuit application. Phase noise reduction due to negative feedback. For achieving proper drive level limitation, the additional circuit is required.
Butler with the resonator in the base	The resonator and the input impedance of the amplifier can be matched without any transformation circuits. Negative feedback for parametric noise reduction can be applied. For achieving proper drive level limitation, the additional circuit is required.

Considering the phase noise measurement results in the open loop (which are similar for these three types of circuits) and simplicity, the Butler configuration with the resonator in the base was chosen for further experiments. The diagram of this circuit is shown in figure 2.



I1≈ 1mA R1 negative feedback
I2≈ 0.2mA R2= 400÷1200Ω

Fig. 2. Diagram of the oscillation circuit

Two stage amplifier with second stage built around T2 transistor (operating as limiting circuit) and negative feedback in the emitter circuit of T1 transistor, guarantee very good phase noise characteristic in the open loop for a broad range of equivalent resonator resistance from 400 to 1200 Ω.

Basing on methods describe in ref. 9 the open loop measurements were made and coefficients A and B were calculated. The A coefficient which characterize $\frac{1}{f}$ noise is about 3.7E-14 and the B coefficient defining f^0 noise is 3.5E-16. These values were measured under conditions, which guarantee drive level lower than 5 μW and efficiency of the amplitude limiting system as shown in figure 3.

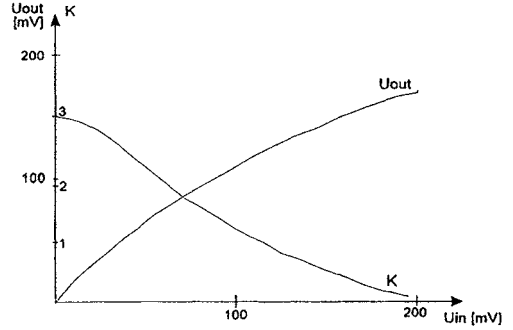


Fig.3. Amplitude limiting efficiency (for R1=30Ω)

Basing on following equation /ref.9/:

$$S_\phi(f) = (\frac{A}{f} + B)[1 + (\frac{f_0}{2Qf})^2]$$

the spectral density of phase noise fluctuations, for resonator of Q factor close to 3E6 was calculated. The results are presented in figure 4.

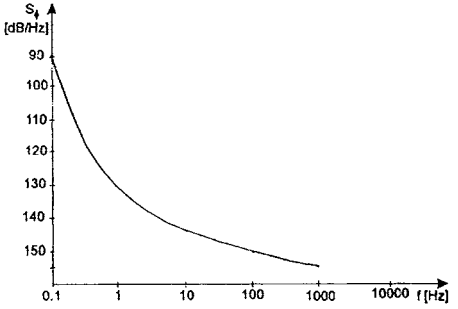


Fig.4. Spectral density of phase fluctuations

4. OVEN CONSTRUCTION

Basic principles of the oven construction are presented in figure 5. The oven consists of three chambers of stable temperature, where the resonator and electronic circuits of higher influence on frequency are placed.

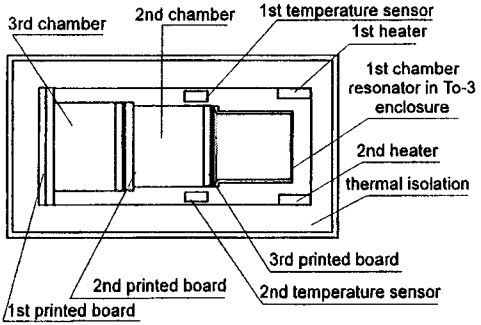


Fig.5. Oven construction

The results of the oven temperature stability are shown in Table 3.

Table 3. Oven temperature stability

Temperature range	-10÷50°C
Quality factor : I chamber	>2000
II chamber	>500
III chamber	>100
Short-term temperature stability of I chamber at 25°C	
τ=1s	5E-7
τ=10s	4E-7
τ=100s	1E-6

Temperature stability n the first chamber was measured by the means of B mode of the Sc-cut resonator. It was thus assured that the temperature sensor was placed exactly at the same place as the LFE resonator.

To obtain the high resolution of measurements, the high stability oscillator of the same frequency as B mode was built. Due to such solution, typical equipment for short term frequency stability measurement could be used. Measurements in second and third chambers were made using precise thermistors.

Analysis of oven temperature stability depending on power loss in the resonator itself was made according to reference 10. The change of loss power from 5 to 40 μW decrease two times the temperature fluctuations. The results for loss power equal 5 μW are sufficient for obtaining short term frequency stability (caused by temperature fluctuations) better than E-13 for τ=1s and E-14 for longer averaging time.

5.CONCLUSION

Applying the proposed oven construction and carrying out further optimization of the LFE resonators operating conditions it is possible to obtain short-term frequency stability better than E-13 for averaging time longer than 1s.

6. REFERENCES

/1./ A.Ballato, E.R.Hatch: Lateral Field Excitation of Quartz Plates. 34th AFCS, 1983

/2./ R.C.Smythe: An Approximate Expression for The Motional Capacitance of A Lateral Field Resonator. 41st AFCS, 1987

/3./ A.W.Warner: Lateral Field Resonators. 39th AFCS, 1985

/4./ A.W.Warner: Measurement of Plano Convex SC Quartz Blanks Using Lateral Field Excitation. 42nd AFCS, 1985

/5./ M.M.Driscoll: Fabrication and Methods for Evaluation And Circuit Utilization of Prototype Lateral Field Resonator. 41st AFCS, 1987

/6./ B.Capelle: Synchrotron Radiation X Ray Topography Study of Lateral Field Resonators. 6th EFTF, 1992

/7./ W.Szulc, K.Weiss: High Stability Oscillator Using LFE Resonator. Part I. Model of LFE Resonator. Contract B3-2, 1993

/8./ B.Gniewińska:Effets Thermigues Dans les Oscillateurs a Quartz. Seminaire "Etalons de Frequence". Besancon, 1981

/9./ D.B.Leeson: A Simple Model of Feedback Oscillator Noise Spectrum. Proceedings of the IEEE, 1966

/10./ J.P.Valentin: Analyse de Quelqes Aspects du Comportement Thermique des Resonateures a Quartz a Ondes de Volume. These a L'Universite de Franch Comte, 1983

A DIRECT REFLECTION METHOD AND SIX-PORT MEASURING SYSTEM FOR ACCURATE DETERMINATION OF PIEZOELECTRIC RESONATOR BASIC PARAMETERS.

M.Wójcicki, A. Lisowiec
Tele-and Radio Research Institute
Poland
03-450, Warsaw, Ratuszowa 11

Summary

A direct reflection method and special measuring system has been elaborated at Tele - and Radio Research Institute. Basing on measurement results of only a few points of reflection coefficient frequency response all parameters of the investigated resonator may be calculated directly. The heart of the measuring system constitutes the special six-port reflectometer circuit in which only amplitude measurements are necessary to characterize complex reflection coefficient value. The advantages of this circuit are lack of phase measurement, wide frequency range and simplicity.

1. **Introduction.**

According to actual IEC recommendation the reference methods for determination of piezoelectric resonator basic parameters are network analyzer methods basing on measurements of
- scattering parameters of two-port containing tested resonator (transmission methods), or
- reflection coefficient in transmission line loaded with tested resonator (reflection methods) [1].

The reflection method is characterized by potentially higher measurement accuracy than transmission methods because the measured resonator, treated as a one-port device, is better defined electrically and may be directly substituted by reference immittance standards [2].

Reflection methods measurement procedures used until now are considerably

sophisticated. They typically involve multipoint reflection coefficient measurements, conversion of measurement data to admittance values and fitting of admittance circle. On the base of admittance circle all parameters of resonator equivalent circuit are determined.

At Tele - and Radio Research Institute a direct reflection method of much more simpler but effective measurement procedure has been elaborated [3]. The method bases on measurements of certain new characteristic frequency of the resonator, closely related to the series resonance frequency. For last ten years further work on this method has been carried out [4], [5], [6]. It has been applied in many systems intended for measurements of quartz resonator parameters and their changes caused by environmental influences. Measurement procedure of the method has been a subject to progressive improvements in each new application. The paper presents the most actual optimized version of the measurement procedure.

In network analyzer methods two-channel conventional vector voltmeter is typically used as a transmission or reflection measuring circuit. Such circuits feature considerably high and difficult for identification phase measurement error arising when levels in both channels are changing relatively. To avoid above mentioned inconvenience one proposes to introduce in piezoelectric metrology area a six-port measurement technique used for several years in microwave measurements.

This technique allows for achievement of potentially highest absolute measurement accuracy in simple and easy for optimization measuring circuits.

In the paper, an example of six-port technique application in system for determination of piezoelectric resonator parameters is presented.

2. **The direct reflection method.**

2.1. The essence and characteristic features of the method.

In this approach piezoelectric resonator is treated as a one-port device of well known four-element equivalent circuit. Resonator terminates a uniform lossless transmission line and the reflection coefficient existing at the connection plane is measured at several selected frequencies. Then, using simple but accurate analytical relations, all parameters are calculated. The graphical interpretation of the method is given in fig. 1. A circular frequency response of the resonator reflection coefficient is shown on the complex plane. At characteristic frequency $f\Gamma$ the reflection coefficient exhibits a specific value; the magnitude has a minimum value Γm and phase angle $\varphi\Gamma$ has the same (or shifted by 180°) value as outside the resonance region (φo). The frequency responses of the magnitude and phase of the reflection coefficient are shown in fig. 2 [6]. As results from fig.1 and fig.2, at the frequency $f\Gamma$ there exist simultaneously two useful criteria. Because of very high slope ($S\varphi = d\varphi/df \mid max$) of the phase response in the vicinity of the $f\Gamma$, very high measurement resolution may be achieved. The frequency $f\Gamma$ is placed very close and is uniquely related to the series resonance frequency fs (see Appendix, rel..A.6.). Analysing relation A.6 one can observe that the difference between $f\Gamma$ and fs does not depend on motional resistane R1 of the resonator, which value is difficult to predict, but only on motional (C1) and parallel (C0) capacitances much easier to a priori evaluation.

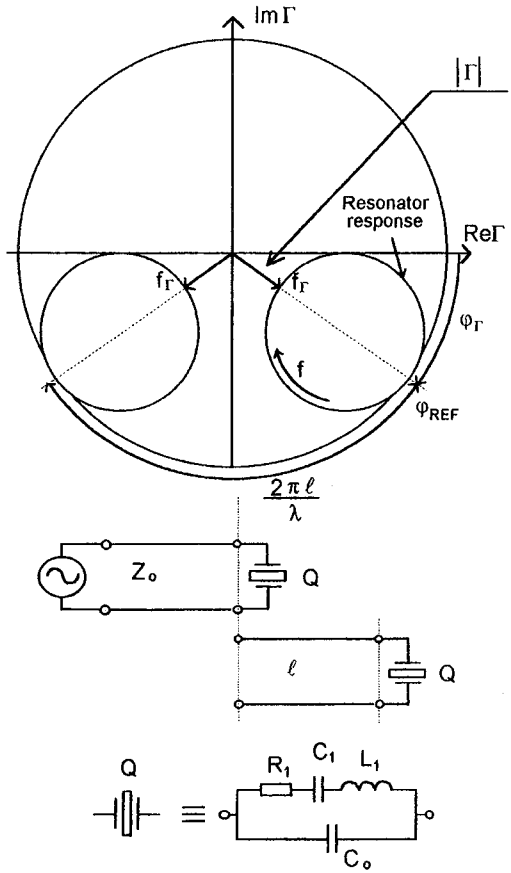


Fig. 1 Graphical interpretation of the direct reflection method.

The fΓ frequency may be determined on the base of measurements of only three points of the magnitude response (see App. rel. A.1..). Basing on four measured values - φo , Γm, Sφ, fΓ, all parameters of resonator equivalent circuit - fs, C1, Q, C0, R1 may be calculated with the use of simple analytical formulas (see App. rel. A.2, A.3, A.4, A.5, A6).

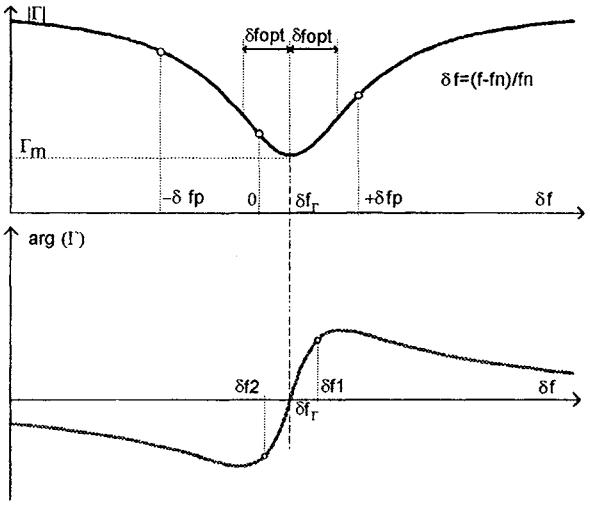


Fig.2 Frequency characteristic of reflection coefficient in transmission line loaded with piezoelectric resonator.

An interesting property of the presented method is that, in general, frequency fΓ is insensitive to the lenght of the transmission line section inserted between measuring port and the resonator under test (fig. 1). It allows for remote measurements in which quartz resonator is placed at some distance from measuring circuit, for example inside test chamber.

2.2. Measurement procedure.

Basic measurement procedure uses both magnitude and phase frequency responses of reflection coefficient. A simplified algorithm of the procedure is given in fig.3 and fig.4. An input information constitute resonator nominal frequency fN and its tolerance.

The tested resonator is connected to the measuring circuit, previously calibrated. Then the reference value of the reflection coefficient (Γo, φo) is measured at frequency placed a little lower than resonance region of the resonator.

Changing the frequency with the step δfp equal to 1/2 of the tolerance of fN three consecutive measurement values of the magnitude of Γ are obtained with the middle one having the lowest value. From A.1 the first approximation - fΓ ' of fΓ is then computed as well as the optimum deviation δfopt of measurement frequency (from fΓ) for which the magnitude characteristic exhibits an inflection point (fig.2). Next, three consecutive measurement values of the magnitude of Γ are obtained for frequencies fΓ '(1±δfopt) and fΓ ' from which the second approximation - fΓ " of fΓ is computed.

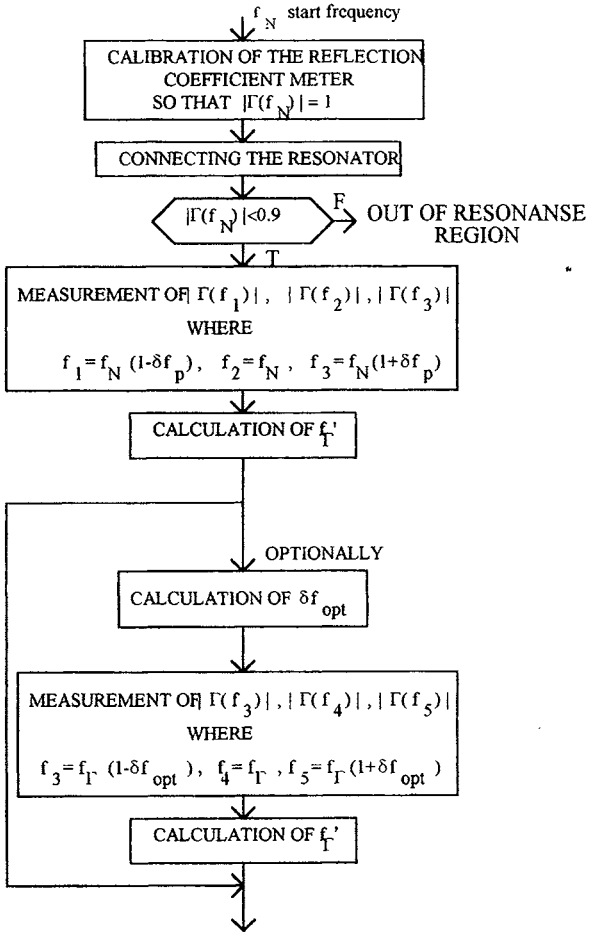


Fig. 3 Simplified algorithm of the first part of the measurement procedure

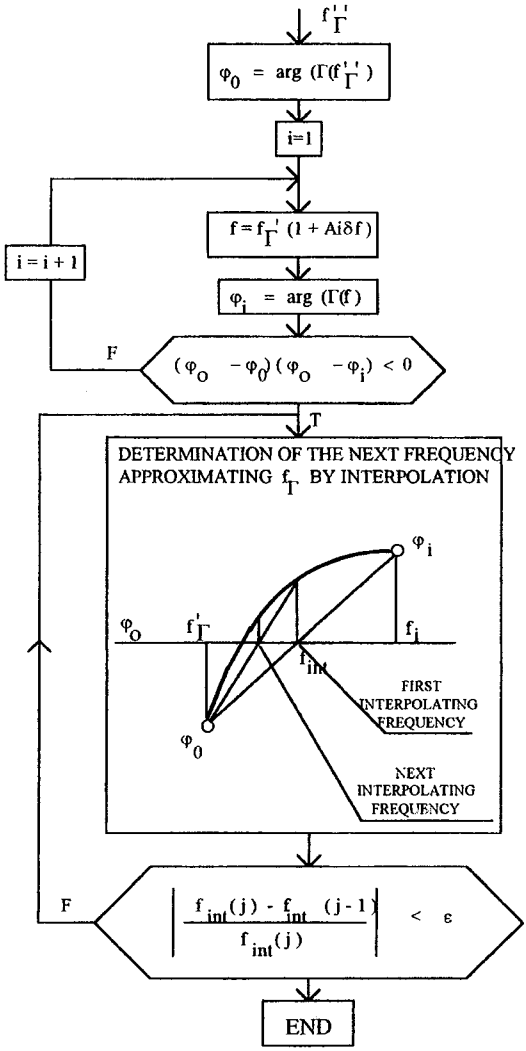


Fig. 4 Simplified algorithm of the second part of the measurement procedure.

The next step of the measurement procedure is a determination, basing on the phase frequency dependence, the exact value of fΓ. For that purpose the phase at fΓ " is measured. Then the frequency is changed by δf to obtain two phase measurement results for which one lies below and the other above the reference phase φo (or φo - 180°) which is the measurement criterion. From these two results the next approximation fint(1) of fΓ by linear interpolation is computed.

The part of the measurement procedure described above which makes use of the phase characteristic is repeated iteratively until sufficiently low difference between two consecutive approximations of fΓ is obtained. The final value of fΓ is computed by a method of linear regression from the set of measurement points which lie on the linear part of phase frequency characteristic.

From the values of characteristic frequency fΓ, the slope of the regression line (equal to the Sφ), reference phase φo, and the minimum value of the magnitude of Γ - Γm, the parameters fs, R1, C1, C0, Q of the resonator are computed with the use of the formulas A.2, A.3, A.4, A.5, A.6.

Sometimes, particularly for resonators with highest resonant frequencies, it is necessary to take into account additional parameter of the resonator, namely the parallel conductance G0. In that case the value of G0 is determined directly from the measured values of magnitude and reference phase. The procedure for obtaining the rest of the resonator parameters doesn't essentially change. The relation A.6 between fΓ and fs must be substituted by a modified relation A.7.

3. **Measurement circuit.**

3.1. Application of six-port technique in reflection measurement method.

The measurement procedure that has been presented in 2.2 has been successfully put to use in a number of measurement systems for investigating the parameters of quartz resonators. In those systems the conventional measurement circuit with two-channel heterodyne network analyzer was used. For obtaining the maximum measurement accuracy the measurement circuit is initially calibrated with the use of precision immitance standards. Assuming linearity, the circuit can be calibrated with three standards. In that case the absolute accuracy depends on the quality of the standards and the resolution and stability of the measurement circuit. However, in practice in the above mentioned vector analyzer there exists a nonlinear phase measurement error connected with the difference of signal levels in individual channels. This error is hard to identify because of the lack of sufficiently accurate broad-band variable attenuators with precisely defined phase characteristics. As the dynamics of the relative levels in quartz resonator measurements is very high, the influence of the phase error on measurement accuracy can be substantial.

Taking that into account it was proposed to make use of six-port measurement technique which had been successfully applied for several years in microwave metrology.

Six-port is a linear, passive electrical circuit with six ports, one of which is an input

for measurement signal, the other is for connecting the measured one-port device and the other four ports are connected to power meters. With the help of such circuit it is possible to determine with high accuracy the magnitude and phase of the reflection coefficient only from power measurements.

The calibration procedure of six-port is slightly more complicated than the calibration procedure of conventional voltmeter but it is of little consequence when the system is computer controlled.

The graphical interpretation of six-port technique in the measurement of resonator reflection coefficient is presented in fig. 5. The six-port is usually characterized by three complex parameters, the so called q-points (q1, q2, q3). One of the power meters (which is being used as the reference meter), measures only the power incident on the resonator under test. The normalized signals from the rest three meters are proportional to the distances between q-points and the end of the reflection coefficient vector. The magnitude and phase of reflection coefficient is computed from simple formula (1).

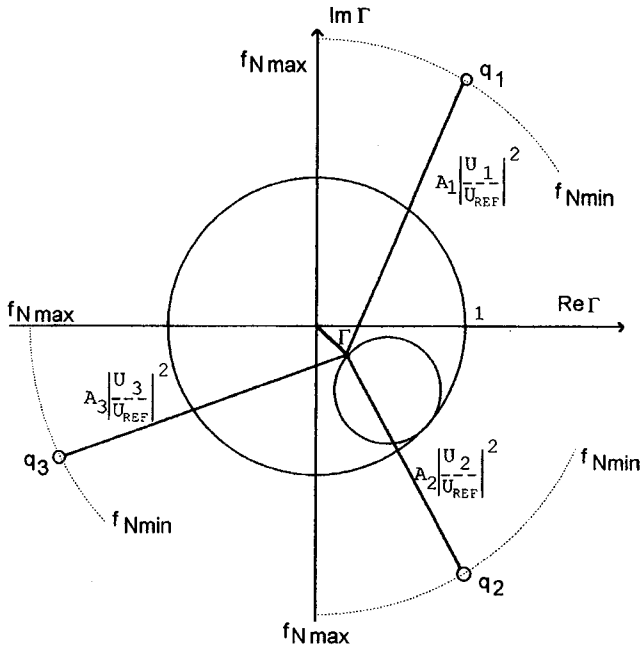


Fig. 5 Interpretation of six-port technique

$$\Gamma = \frac{k_1 * \left[\frac{U_1}{U_{REF}}\right]^2 + k_2 * \left[\frac{U_2}{U_{REF}}\right]^2 + k_3 * \left[\frac{U_3}{U_{REF}}\right]^2 + k_4}{m_1 * \left[\frac{U_1}{U_{REF}}\right]^2 + m_2 * \left[\frac{U_2}{U_{REF}}\right]^2 + m_3 * \left[\frac{U_3}{U_{REF}}\right]^2 + 1} \qquad (1)$$

$$k_1, k_2, k_3, k_4 - complex$$

$$m_1, m_2, m_3 - real$$

It turns out [8] that it is relatively easy to construct the six-port circuit from lumped elements working in wide frequency range (1÷1000 MHz) which makes possible measurements of most actually fabricated piezoelectric resonators.

Besides potentially higher measurement accuracy (elimination of nonlinear phase error), the essential advantage of circuit with six-port is simple construction and so lower equipment costs.

3.2. An example of measurement circuit realization with six-port.

For practical realization of proposed technique there has been developed a laboratory measurement circuit, the block diagram of which is presented in fig. 6. The use has been made of a modified (in comparison with [8]) six-port circuit with lumped elements. The circuit has an input for a signal from frequency synthesizer, a measurement input for connecting by means of a special adapter the investigated resonator and two measurement outputs, one of which is a reference. Measurement outputs are connected to two-channel, narrow-band detection circuit made of two balanced mixers, switch, IF signal filter, linear detector and fast AD converter of high resolution. The frequency synthesizer is equipped with additional signal output shifted by fp in frequency relative to measurement frequency, which feeds mixer inputs. The six-port has a three-position switch that modifies the inner structure of the circuit so as to obtain for each position the proper value of parameter q and associated with it the signal at the measurement output (see fig. 3). The whole measurement circuit is computer controlled. All the parameters that characterize the measurement circuit are obtained during calibration procedure with the use of a set of immittance standards.

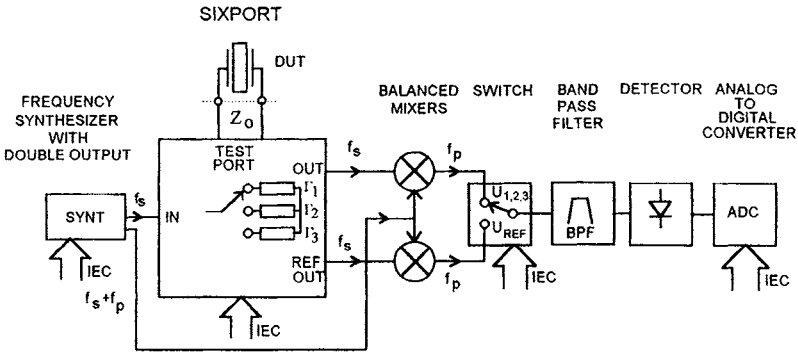


Fig. 6 Simplified block diagram of six-port measuring circuit.

Determination of reflection coefficient with investigated resonator consists of successive measurements of three normalized voltages and then making use of formula (1).

An analysis of the influence of voltage measurement resolution on the measurement resolution of $f\Gamma$ of resonators in frequency range 4÷400 MHz has been carried out. With relative voltage resolution of 1E-4 and typical resonator parameters the frequency measurement resolution is better than 2E-8. That was confirmed by the experimental investigation of the resonators carried out in the measurement setup just described.

At present the research effort is directed towards optimizing the calibration procedure of proposed circuit and analyzing the absolute measurement accuracy .

APPENDIX.

Relations used in direct reflection method.

$f\Gamma$ is computed from the equation

$$f\Gamma = fN(1+\delta fp/2)\frac{(1-\Gamma_{m1}^2)(\Gamma_{m2}^2-\Gamma_{m3}^2)}{2(\Gamma_{m2}^2+\Gamma_{m1}^2\Gamma_{m3}^2)(1+\Gamma_{m2}^2)(\Gamma_{m1}^2+\Gamma_{m3}^2)} \qquad \text{A.1}$$

where $\delta fp=(f2-f1)/fN=(f3-f2)/fN$ and Γ_{m1}, Γ_{m2}, Γ_{m3} are values of the magnitude of the reflection coefficient for frequencies f1, f2, f3.

From the minimum of the magnitude of reflection coefficient Γm, the slope of phase characteristic at $f\Gamma$ - $S\varphi$ and reference phase φo the parameters of the resonator can be computed from the following relations.

$$C0 = \frac{tg(\varphi o/2)}{2\pi fk Z_o}, \qquad \text{A.2}$$

where fk is a frequency below the resonance region;

$$C1 = 2\pi f\Gamma R1Q\ , \qquad \text{A.3}$$

$$R1 = Z_0 \frac{1 + A\Gamma m}{(1 - A\Gamma m)(1 + tg^2(\varphi o/2))} \qquad \text{A.4,}$$

A= -1 if R1 < Z_0 and A = 1 if R1 > Z_0,

$$Q = |S\varphi| \frac{\Gamma m}{1 - \Gamma m^2}\ . \qquad \text{A.5}$$

The relation between fΓ and fs is given by

$$(f\Gamma - fs)/fs = \frac{C1/C0}{2([1 + \frac{1}{(2\pi fsC0Z_0)^2}])} \qquad \text{A.6}$$

If the parallel conductance G0 is assumed to have a non zero value, the relation between fΓ and fs is given by the equation

$$(f\Gamma - fs)/fs = \frac{(C1/C0)(1 + 2G0R1)}{2\{[1 + \frac{1}{(2\pi fsC0Z_0)^2}][1 - \frac{G0^2}{1/Z_0^2 + (2\pi fsC0)^2}]\}} \qquad \text{A.7}$$

References.

[1] IEC Document 49 (Central Office) 248, 1992
[2] IEC Document 49 (Central Office) 268, 1993
[3] M. Wójcicki "Method for measurement of basic parameters of HF quartz resonators", Works of Tele-and Radio Research Institute, No 84/1980
[4] W. Horton, R. Smythe "A New frequency for piezoelectric resonator measurement", Proc. of the37th Annual Symp. on Frequency Control, 1983
[5] M. Wójcicki "A new single port reflection method for measurement of crystal unit equivalent parameters in wide frequency range", Proc. of International Conference on Frequency Control and Synthesis, The University of Surrey, Guildford, 1987
[6] M. Wójcicki "A method for measurement of quartz resonator parameters in wide frequency range", Doctor thesis, The Warsaw Technical University 1989
[7] G. Engen "The Six-Port Reflectometer: An Alternative Network Analyzer", IEEE Transactions on Microwave Theory and Techniques, Vol. MTT-25, No 12, Dec. 1977
[8] V. Bilik, V. Raffaj, J. Bezek "A new extremely wide-band lumped six-port reflectometer", Proc. of 20th European Microwave Conference 1990, Budapest

THE EFFECTIVE COUPLING OF THE ELASTICALLY COUPLED Y-CUT $LiTaO_3$ TWO AND THREE RESONATOR STRUCTURE

J. Zelenka*, I. Mateescu**, J.Kosinski***

* Technical University of Liberec, Liberec, Czech Republic ,
** Institute of Physics and Technology of Materials, Bucaresti, Romania,
*** US Army Research Laboratory, Fort Monmouth, USA .

SUMMARY

This paper describes the method the computing of the effective interresonator coupling coefficient of the elastically coupled multi-resonator structure realized on the Y-cut plate of Lithium Tantalate crystal. The method is based on the solution of the wave propagation along x_3 in an infinite piezoelectric plate and takes into account the influence of the electroded and unelectroded part of the plate on the wave propagation. The three dimensional equations of linear piezoelectricity are used in the derivation of the relations.

1. INTRODUCTION

The method of calculation of the effective coupling of the elastically coupled n-resonator structure realized on the AT cut quartz plate was described by Beaver [1]. Beaver's method makes it possible to calculate the coupling realized by means of the trapped fundamental thickness-shear or thickness-twist mode of vibration. Mateescu and Kosinski [2] experimentally found an interesting value of the effective coupling of the two resonator structure realized on the Y-cut $LiTaO_3$ plate near the cut-off frequency of the third harmonic of the thicness shear vibration. The aim of this paper is to describe the method which allows one to calculate the effective elastic coupling of two or three resonator structure realized on Y-cut $LiTaO_3$ plates driven near the third harmonic of the thickness shear vibrations.

Waves and vibrations of Y-cut $LiTaO_3$ plates were studied by Lee and Syngellakis [3]. They found (by consideration of the symmetry of the plate) that the variables describing the vibrations and wave propagation are separated in two uncoupled groups. In the first group the main variable is the displacement u_1 and there is no coupling with the electrical potential. This group governs the face-shear and thickness-twist vibrations which may not be driven by lateral excitation. The second group of equations includes the remaining variables (displacement u_2 and u_3 and

potential ϕ) and governs the flexure, extension, thickness-shear and thickness-stretch vibrations.

The second group of equations is important for the describtion of the coupled n-resonator structure and only this group of equations will be considered in this paper.

2. WAVES PROPAGATING ALONG THE x_3 AXIS AND DESCRIBED BY DISPLACEMENTS u_2, u_3 AND POTENTIAL ϕ

Consider an infinite Y-cut $LiTaO_3$ plate of thickness $2b$ with its middle plane taken as the x_1, x_3 plane and let the faces at $x_2 = \pm b$ be free of tractions. In a rotated Y-cut $LiTaO_3$ plate the arrays of the material constants (elastic stiffnesses c_{pq}, piezoelectric stress constants e_{ip} and permitivities ϵ_{ij}) referred to the plate axes have the folloving form

$$c_{pq} = \begin{bmatrix} c_{11} & c_{12} & c_{13} & c_{14} & 0 & 0 \\ c_{12} & c_{22} & c_{23} & c_{24} & 0 & 0 \\ c_{13} & c_{23} & c_{33} & c_{34} & 0 & 0 \\ c_{14} & c_{24} & c_{34} & c_{44} & 0 & 0 \\ 0 & 0 & 0 & 0 & c_{55} & c_{56} \\ 0 & 0 & 0 & 0 & c_{56} & c_{66} \end{bmatrix}, \quad e_{ip} = \begin{bmatrix} 0 & 0 & 0 & 0 & e_{15} & e_{16} \\ e_{21} & e_{22} & e_{23} & e_{24} & 0 & 0 \\ e_{31} & e_{32} & e_{33} & e_{34} & 0 & 0 \end{bmatrix}$$

$$\epsilon_{ij} = \begin{bmatrix} \epsilon_{11} & 0 & 0 \\ 0 & \epsilon_{22} & \epsilon_{23} \\ 0 & \epsilon_{23} & \epsilon_{33} \end{bmatrix} \tag{1}$$

Consider the governing equations of the linear theory of piezoelectricity and a harmonic plane wave solution of the form

$$\begin{aligned} u_i &= A_i \cos(\eta x_2 + \xi x_3 - \omega t), \\ \phi &= A_4 \cos(\eta x_2 + \xi x_3 - \omega t), \qquad i = 2, 3, \end{aligned} \tag{2}$$

where η and ξ are wave vectors in the direction of wave propagation and ω is the angular frequency.

Substitution of equations (2) into the constitutive equations and strain-mechanical displacement relations of the linear theory of piezoelectricity yields

$$\begin{bmatrix} \Gamma_{22} - \varrho\omega^2 & \Gamma_{23} & \Gamma_{24} \\ \Gamma_{23} & \Gamma_{33} - \varrho\omega^2 & \Gamma_{34} \\ \Gamma_{24} & \Gamma_{34} & \Gamma_{44} \end{bmatrix} \cdot \begin{bmatrix} A_2 \\ A_3 \\ A_4 \end{bmatrix} = 0, \tag{3}$$

If λ_i and λ_{i+1} are complex conjugate the amplitudes A_{2i} and $A_{2(i+1)}$, A_{3i} and $A_{3(i+1)}$, A_{4i} and $A_{4(i+1)}$ are also complex conjugate and can be obtained from equations (6) when A_{2r}, A_{3r}, A_{4r} and λ_r are considered as a complex quantities.

The general solution of the wave propagation therefore becomes

$$u_i = \cos(\xi x_3 - \omega t)\sum_{r=1}^{6} L_r A_{ir}\cos(\xi\lambda_r x_2) - \sin(\xi x_3 - \omega t)\sum_{r=1}^{6} L_r A_{ir}\sin(\xi\lambda_r x_2), \qquad i = 1,2$$

$$\phi = \cos(\xi x_3 - \omega t)\sum_{r=1}^{6} L_r A_{4r}\cos(\xi\lambda_r x_2) - \sin(\xi x_3 - \omega t)\sum_{r=1}^{6} L_r A_{4r}\sin(\xi\lambda_r x_2). \quad (7)$$

3. BOUNDARY CONDITIONS

Let the plate be partly unplated and partly plated with thin electrically conducting electrodes.

When the plate faces at $x_2 = \pm b$ of the unplated part of the plate are traction free then at $x_2 = \pm b$

$$\begin{aligned} T_{22} &= c_{22}u_{2,2} + c_{23}u_{3,3} + c_{24}(u_{2,3} + u_{3,2}) + e_{22}\phi_{,2} + e_{32}\phi_{,3} = 0, \\ T_{23} &= c_{24}u_{2,2} + c_{34}u_{3,3} + c_{44}(u_{2,3} + u_{3,2}) + e_{24}\phi_{,2} + e_{34}\phi_{,3} = 0, \end{aligned} \quad (8)$$

and the corresponding electrical conditions are

$$D_2 = e_{22}u_{2,2} + e_{23}u_{3,3} + e_{24}(u_{2,3} + u_{3,2}) - \epsilon_{22}\phi_{,2} + \epsilon_{23}\phi_{,3} = 0. \quad (9)$$

When the plate faces at $x_2 = \pm b$ are coated with thin electrically conducting films with the thickness $2b'$ and density ϱ' and a uniform, time dependent voltage is applied across the thickness of the plate then at $x_2 = \pm b$

$$\begin{aligned} T_{22} &= c_{22}u_{2,2} + c_{23}u_{3,3} + c_{24}(u_{2,3} + u_{3,2}) + e_{22}\phi_{,2} + e_{32}\phi_{,3} \pm 2b'\varrho'\ddot{u}_2 = 0, \\ T_{23} &= c_{24}u_{2,2} + c_{34}u_{3,3} + c_{44}(u_{2,3} + u_{3,2}) + e_{24}\phi_{,2} + e_{34}\phi_{,3} \pm 2b'\varrho'\ddot{u}_3 = 0, \end{aligned} \quad (10)$$

and the electrical face conditions are

$$\phi = \phi^{\pm}(t) = \pm\phi_o e^{i\omega t}. \quad (11)$$

For vibrations modes with long wavelength we may ignorate the spatial variations of field variables u_2, u_3 and ϕ by setting their derivatives with respect to x_3 equal to zero and from

$$D_{2,2} = e_{22}u_{2,22} + e_{24}u_{3,22} - \epsilon_{22}\phi_{,22} = 0 \quad (12)$$

we have

$$\phi = \frac{1}{\epsilon_{22}}[e_{22}u_2 + e_{24}u_3 + A(t)x_2 + B(t)] \quad (13)$$

and

$$D_2 \mid_{\pm b} = \epsilon_{22}A(t) \quad (14)$$

where

$$A(t) = \frac{1}{2b}(\phi^+ - \phi^-) - \frac{1}{2b\epsilon_{22}}[e_{22}(u_2^+ - u_2^-) + e_{24}(u_3^+ - u_3^-)],$$
$$\frac{1}{2b}(\phi^+ - \phi^-) = \frac{1}{b}\phi_o e^{i\omega t}$$

and $u_i^{\pm}(t) \equiv u_i(\pm b)$ are the surface displacements.

For the shorted electrodes $\phi^+ = \phi^- = 0$ and the electrical face conditions at $x_2 = \pm b$ are

$$D_2 = e_{22}u_{2,2} + e_{23}u_{3,3} + e_{24}(u_{2,3} + u_{3,2}) - \epsilon_{22}\phi_{,2} + \epsilon_{23}\phi_{,3} + \\ + \frac{1}{2b}[e_{22}(u_2^+ - u_2^-) + e_{24}(u_3^+ - u_3^-)]. \quad (15)$$

Case (a): Unplated part of the plate

By substituting (7) into boundary conditions (8) and (9), one obtains for $i = 1, 2, 3$

$$\sin(\xi x_3 - \omega t)\sum_{r=1}^{6} L_r M_{ir}\cos(\xi\lambda_r b) \pm \cos(\xi x_3 - \omega t)\sum_{r=1}^{6} L_r M_{ir}\sin(\xi\lambda_r b) = 0, \quad (16)$$

where

$$M_{1r} = (c_{22}A_{2r} + c_{24}A_{3r} + e_{22}A_{4r})\lambda_r + c_{24}A_{2r} + c_{23}A_{3r} + e_{32}A_{4r},$$
$$M_{2r} = (c_{24}A_{2r} + c_{44}A_{3r} + e_{24}A_{4r})\lambda_r + c_{44}A_{2r} + c_{34}A_{3r} + e_{24}A_{4r},$$
$$M_{3r} = (e_{22}A_{2r} + e_{24}A_{3r} - \epsilon_{22}A_{4r})\lambda_r + e_{24}A_{2r} + e_{23}A_{3r} - \epsilon_{23}A_{4r}. \quad (17)$$

The equations (16) lead to the transcendental frequency equation

$$\begin{bmatrix} M_{ir}\cos(\xi\lambda_r b) \\ M_{i)r}\sin(\xi\lambda_r b) \end{bmatrix} = 0, \qquad i = 1, 2, 3; \quad r = 1, 2, ..., 6. \quad (18)$$

Case (b): Plated part of the plate

In this case the variables will be indicated by a prime, that is $\bar{\xi}$, $\bar{\lambda}$, $\bar{M}_{ir}$, etc. After substituting (7) into boundary conditions (10) and (15), one obtaines for $i = 1, 2, 3$

$$\sin(\bar{\xi}x_3 - \omega t)\sum_{r=1}^{6} \bar{L}_r \bar{M}_{ir}\cos(\bar{\xi}\bar{\lambda}_r b) \pm \cos(\bar{\xi}x_3 - \omega t)\sum_{r=1}^{6} \bar{L}_r \bar{M}_{ir}\sin(\bar{\xi}\bar{\lambda}_r b) = 0, \quad (19)$$

where

$$\bar{M}_{1r} = (c_{22}\bar{A}_{2r} + c_{24}\bar{A}_{3r} + e_{22}\bar{A}_{4r})\bar{\lambda}_r + c_{24}\bar{A}_{2r} + c_{23}\bar{A}_{3r} + e_{32}\bar{A}_{4r}$$

$$
\begin{aligned}
&\pm\ R\varrho b\omega^2\bar{\xi}^{-1}\bar{A}_{2r}\tan(\bar{\xi}\bar{\lambda}_r b),\\
\bar{M}_{2r} &= (c_{24}\bar{A}_{2r}+c_{44}\bar{A}_{3r}+e_{24}\bar{A}_{4r})\bar{\lambda}_r+c_{44}\bar{A}_{2r}+c_{34}\bar{A}_{3r}+e_{24}\bar{A}_{4r}\\
&\pm\ R\varrho b\omega^2\bar{\xi}^{-1}\bar{A}_{3r}\tan(\bar{\xi}\bar{\lambda}_r b),\\
\bar{M}_{3r} &= (e_{22}\bar{A}_{2r}+e_{24}\bar{A}_{3r}-\epsilon_{22}\bar{A}_{4r})\bar{\lambda}_r+e_{24}\bar{A}_{2r}+e_{23}\bar{A}_{3r}-\epsilon_{23}\bar{A}_{4r}\\
&\pm\ (e_{22}\bar{A}_{2r}+e_{24}\bar{A}_{3r})\frac{b}{\bar{\xi}}\tan(\bar{\xi}\bar{\lambda}_r b),\\
\bar{M}_{4r} &= (c_{22}\bar{A}_{2r}+c_{24}\bar{A}_{3r}+e_{22}\bar{A}_{4r})\bar{\lambda}_r+c_{24}\bar{A}_{2r}+c_{23}\bar{A}_{3r}+e_{32}\bar{A}_{4r}\\
&\pm\ R\varrho b\omega^2\bar{\xi}^{-1}\bar{A}_{2r}\cot(\bar{\xi}\bar{\lambda}_r b),\\
\bar{M}_{5r} &= (c_{24}\bar{A}_{2r}+c_{44}\bar{A}_{3r}+e_{24}\bar{A}_{4r})\bar{\lambda}_r+c_{44}\bar{A}_{2r}+c_{34}\bar{A}_{3r}+e_{24}\bar{A}_{4r}\\
&\pm\ R\varrho b\omega^2\bar{\xi}^{-1}\bar{A}_{3r}\cot(\bar{\xi}\bar{\lambda}_r b),\\
\bar{M}_{3r} &= (e_{22}\bar{A}_{2r}+e_{24}\bar{A}_{3r}-\epsilon_{22}\bar{A}_{4r})\bar{\lambda}_r+e_{24}\bar{A}_{2r}+e_{23}\bar{A}_{3r}-\epsilon_{23}\bar{A}_{4r}. \qquad (20)
\end{aligned}
$$

The equatios (19) lead to the transcendental frequency equation

$$
\begin{bmatrix} \bar{M}_{ir}\cos(\bar{\xi}\bar{\lambda}_r b) \\ \bar{M}_{(i+3)r}\sin(\bar{\xi}\bar{\lambda}_r b) \end{bmatrix} = 0, \qquad i=1,2,3; \quad r=1,2,...,6. \qquad (21)
$$

and

$$
R=\frac{2\varrho' b'}{\varrho b}.
$$

The previous equations allow one to compute the wave number ξ for the unplated part of the plate and $\bar{\xi}$ for the plated part of the plate for a given orientation angle of the plate, frequency ω and mass-loading R.

4. TWO RESONATOR STRUCTURE EXAMPLE

Consider an infinite plate of thickness $2b$ on which two pairs of strip electrodes are deposited. The widths of the electrodes in the x_3 direction are a_1 and a_2 and the distance between the electrodes is d. We will consider the narrow frequency range near the cut-off third harmonic thickness-shear frequency where the wave number ξ of the unplated part of the plate is pure imaginary and the wave number $\bar{\xi}$ of the plated part of the plate is pure real. Imaginary value of ξ correspond to decaying (or growing) standing waves in unelectroded part of the plate. The real values of $\bar{\xi}$ correspond to trigonometric standing waves in electroded regions of the plate. The values of wave numbers ξ and $\bar{\xi}$ which yields the frequency ω we know from the relations given in paragraphs 2 and 3.

The description of the elastic waves in the plated and unplated part of the plate let be the same as was shown by Beaver [1]. The boundary condition to be imposed are that the displacements u_3 and $\bar{u}_3$ and also the derivation of the displacements $u_{3,3}$ and $\bar{u}_{3,3}$ be continuous at the boundaries between the electroded and unelectroded regions of the plate. After substituting into this conditions, one can receive the

following relation

$$\begin{bmatrix} B_{11} & B_{12} & 0 & 0 & 0 & 0 & 0 & 0 \\ B_{21} & 0 & B_{23} & 0 & 0 & 0 & 0 & 0 \\ 0 & B_{32} & B_{33} & B_{34} & B_{35} & 0 & 0 & 0 \\ 0 & B_{42} & B_{43} & B_{44} & B_{45} & 0 & 0 & 0 \\ 0 & 0 & 0 & B_{54} & B_{55} & B_{56} & B_{57} & 0 \\ 0 & 0 & 0 & B_{64} & B_{65} & B_{66} & B_{67} & 0 \\ 0 & 0 & 0 & 0 & 0 & B_{76} & B_{77} & B_{78} \\ 0 & 0 & 0 & 0 & 0 & B_{86} & B_{87} & B_{88} \end{bmatrix} = 0, \tag{22}$$

where

$$\begin{aligned} B_{11} &= 1, & B_{55} &= e^{-\xi l_2}, \\ B_{12} &= -1, & B_{56} &= -\cos(\bar{\xi} l_2), \\ B_{21} &= \xi, & B_{57} &= -\sin(\bar{\xi} l_2), \\ B_{23} &= -\xi, & B_{64} &= \xi e^{\xi l_2}, \\ B_{32} &= \cos(\bar{\xi} l_1), & B_{65} &= -\xi e^{\xi l_2}, \\ B_{33} &= \sin(\bar{\xi} l_1), & B_{66} &= \bar{\xi} \sin(\bar{\xi} l_2), \\ B_{34} &= -e^{\xi l_1}, & B_{67} &= -\xi \cos(\xi l_2), \\ B_{35} &= -e^{-\xi l_1}, & B_{76} &= \cos(\bar{\xi} l_3), \\ B_{42} &= -\bar{\xi} \sin(\bar{\xi} l_1), & B_{77} &= \sin(\bar{\xi} l_3), \\ B_{44} &= \bar{\xi} \cos(\bar{\xi} l_1), & B_{78} &= -e^{-\xi l_3}, \\ B_{44} &= -\xi e^{\xi l_1}, & B_{86} &= -\bar{\xi} \sin(\bar{\xi} l_3), \\ B_{45} &= \xi e^{-\xi l_1}, & B_{87} &= \cos(\bar{\xi} l_3), \\ B_{54} &= e^{\xi l_2}, & B_{88} &= \xi e^{-\xi l_3} \end{aligned} \tag{23}$$

and $l_1 = a_1$, $l_2 = a_1 + d$, $l_3 = a_1 + d + a_2$.

The inter-resonator coupling coefficient k can be calculated as a relative distance between two lowest frequencies f_1 and f_2 which yields the relation (22)

$$k = 2 \frac{f_2 - f_1}{f_1 + f_2}. \tag{24}$$

5. REFERENCES

[1] Beaver W.D.: Analysis of elastically coupled piezoelectric resonators. J. Acoust. Soc. Amer 43 (1968), pp 972-981.

[2] Mateescu I., Kosinski J.: Eight pole monolithic filters using lithium tantalate. IEEE International Freq. Control. Symp., Salt Lake City, 1993.

[3] Lee P.C.Y., Syngellakis S.: Waves and vibrations in an infinite piezoelectric plate. Proc. 29th Annual Symp. Freq. Control (1976), pp. 184-190.

A SINGLE BRIDGE RESONATOR

Serge GALLIOU, Marc MOUREY and Bernard DULMET

ENSMM / LCEP
16 , Route de Gray - La Bouloie
25030 BESANÇON CEDEX FRANCE

ABSTRACT

At the moment, TCXO performances are limited by the thermal hysteresis phenomena as well as the added noise introduced by the compensation system. A temperature controlled oscillator using the Single Bridge Resonator (S.B.R.) which is described in this paper offers features close to the TCXO's, i.e, low power consumption and instantaneous nominal frequency. The S.B.R. design is the result of the association of two constraints : minimal stresses induced by the quartz supports which minimize hysteresis, high thermal insulation which authorizes a temperature control with low power consumption.

Design and basic concepts are described from some realistic structures. One of them is especially presented and analysed in detail.

A dedicated thermal model is established in order to estimate the conductive exchanges as well as the radiative ones. In addition, it allows to verify that power consumption (Max. 200 mW) and warm-up time (Max. 3 mn) do not exceed the maximum fixed values within the external operating temperature range : -20° C, +70° C. Results of simulations using a finite elements method illustrate thermal and mechanical properties of the analysed structure.

1. INTRODUCTION

In this work we present a miniature crystal oscillator so-called Tactical Miniature Crystal Oscilator (TMXO) /1/ . Such a device is aimed to gather advantages of Temperature Compensated Crystal Oscillator (TCXO) and Oven Controlled Crystal Oscillator (OCXO). The former provides small size, very low power consumption and its output frequency is immediately usable at turn-on. The latter gives better performances but requires to be ovenized : This is the origin of recent works /2/ /3/ /4/ /5/. Among them, one approach consists in a direct resonator heating /6/ /7/ which is one characteristic of the device described below. Design goals are stated to less than 3minutes for the warm-up time (when the output frequency achieves its final value within an accuracy of $\pm$ 5 10^{-9}), a maximum power consumption of 200 mW in an external temperature range of [-20°C,+70°C] with a maximum frequency deviation of 5 10^{-9} over this range.

2. BASIC CONCEPTS

The basic idea consists in heating the minimum mass surrounded with the best thermal insulation.This results in low power consumption and fast warm-up provided the most temperature sensitive element of the crystal oscillator -i.e the quartz resonator- is also the closest one to the heaters.
Conventional quartz crystal resonators are usually set into a sealed enclosure with a high-quality vacuum which eliminates a large amount of heat transfer by gas conduction and convection. This

leaves only conduction through the quartz resonator supports and radiation as thermal coupling between the quartz bulk and the outside environment.
Therefore direct crystal heating is obviously the most efficient basic solution if we put aside the temperature sensitivity of the electronics : heaters close to the quartz bulk are controlled in order to lock the resonator at its turn-over point temperature whatever the enclosure temperature is. Then assuming that radiative exchanges are negligible (this will be discussed later), power consumption and warm-up time only depend on the thermal quality of the crystal supports.
More generally, frequency stability is tightly dependent of the thermal gain of the temperature control system which is also tightly dependent of the mechanical structure. Indeed, it is important to keep in mind that an ideal feedback control system is able to lock the probe temperature and only it, i.e to say a very local temperature. In steady state, it is required that the vibrating part be isothermal to ensure equal probe and resonator temperatures ; thus, there should be no internal thermal flux.
In addition, the thermal flux must obviously vanish all over the surface of this isothermal volume except where it may be controlled i.e in a limited zone where the temperature probe is set on. This is the first argument for a unique support of the quartz resonator as a way to channel heat flow.
A second argument for the unique support is related to thermoelastic considerations. The structure must provide a thermal and mechanical link between the quartz resonator and its enclosure without inducing internal stresses : The isothermal quartz bulk must also be free to expand. Furthemore, this aspect is better achieved with a monolithic crystal structure.
This idea of a unique support made from quartz -a "bridge"- may be extended considering that the quartz is a poor thermal conductor but a quite stiff material. This is the opportunity to ally a good insulation with a compact resonator support when one of its end is hot (at the temperature controlled quartz bridge) and the other is cold (at the room temperature) : the quartz bridge may be extended by a quartz ring clamped at its cold end to the enclosure.

3. EXAMPLES OF DESIGNS

Both main designs of Single Bridge Resonators (S.B.R.) are shown in figure 1 and figure 2. In these designs we stuck to a constant section of the quartz ring.

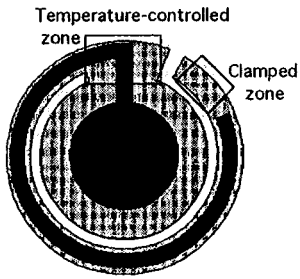


Figure 1.: The open ring SBR

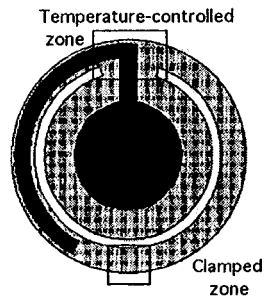


Figure 2.: The closed ring SBR

Heaters and thermal probe must be set on the quartz bridge where the concentrated conductive flux can be controlled. This hot region named "temperature controlled zone" is strongly isolated from the cold end of the quartz ring (the "clamped zone") in contact with the enclosure. The design of fig.1 emphasizes thermal insulation whereas the design of fig.2 improves mechanical stability.

Electronics including the oscillator circuit with its output amplifier in one hand, and the heating control system in the other hand is also temperature sensitive : without any care its effect on frequency shift can reach up to a few 10^{-8} within the ambient temperature range.
The electronics itself must be at least temperature compensated, but it is better to temperature-control it. We think of future improvements aimed to put all the required electronics inside the enclosure.

4. A SIMPLE THERMAL ANALYSIS

It is useful to set up an ideal thermal model as a design goal. Using the electrical analogy, we propose the model shown in figures 3a-3b.

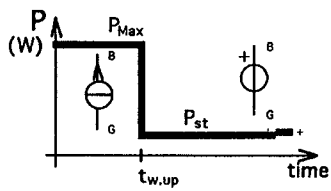


Figure 3-a : controlled power from turn-on

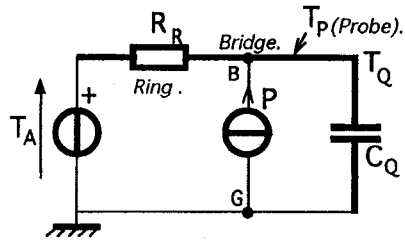


Figure 3-b : ideal thermal model

Figure 3a describes the power consumption behavior of an ideal temperature controlled system with respect to time from turn-on. During the transient time the two-terminal component BG is working as a constant current generator because the error signal (function of the difference between the probe temperature T_P and the reference temperature T_{Ref}) is such that the system saturates at P_{Max} (the maximum power supply). At $t = t_{w.up}$ the system reaches its steady state when $T_P = T_{Ref}$ (in the ideal case) and needs to supply a constant power P_{st} if the ambient temperature T_A is constant. An ideal control system is able to lock T_P at T_{Ref} that is to say that the BG generator then turns into a constant voltage generator.

In figure 3b, the thermal behavior of the S.B.R. is described with the simplest model which just takes into account the quartz ring as a thermal resistor R_R and the quartz resonator as a thermal capacity. Heat flow in the quartz bridge is controlled by the probe temperature T_P which is also equal to the quartz temperature T_Q in that case. Then, from $t = 0$ at turn-on to $t = t_{w.up}$, we have :

$$T_Q = R_R.P_{Max}.(1 - e^{-\frac{t}{\tau}}) + T_A \qquad \text{where} \quad \tau = R_R.C_Q \qquad (1)$$

and from $t = t_{w.up}$ in steady state :

$$T_Q = T_{ref} \qquad \text{with} \qquad P_{st} = (T_{Ref} - T_A) / R_R \qquad (2)$$

Moreover, here the warm-up time will be approximatively equal to $t_{w.up}$ that is :

$$t_{w.up} = \tau.\ln[P_{Max} / (P_{Max} - P_{st})] \qquad (3)$$

For example, let us take a 10 MHz - 5th overtone AT-Cut crystal 0.8 mm thick with a diameter of 10.6 mm and a quartz opened ring 1.5 mm width. This gives a thermal resistance of about 3400 K/W and a time constant of 510 s (about 85 minutes). Taking a maximum power supply PMax = 200 mW and assuming the reference temperature TRef is exactly set at the crystal turnover point of 80° C, yields

at	T_A = - 20° C	P_{st} = 30 mW	and	$t_{w.up}$ = 82 s
at	T_A = + 20° C	P_{st} = 18 mW	and	$t_{w.up}$ = 48 s

All these values must be understood as ultimate performances.

The ideal model of figure 3-b can be slithly improved to match a more realistic behavior as shown in figure 4. Then the thermal capacity of the quartz ring C_R does not remain negligible because it is of the same order than the resonator's ! Moreover, the quartz bulk has its own inner thermal resistance which operates when temperature is not homogeneous i.e. when thermal flux exists. The conductive flux into the resonator is not zero during transient periods and also in steady state, if we consider the radiative exchanges between the quartz surface and its enclosure. Without writing the 3D-analytical

expression of temperature in the quartz bulk we can admit that the main heat flow component comes from the bridge and keeps the same direction when flowing through the resonator.
In this one-dimensional model the quartz resonator is then cut out in N slices (see figure 4 where N = 4) of thickness Δx and cross section S(x) with

$$\Delta x = D/N \text{ and } S(x) = 2e\sqrt{x\,(D-x)} \qquad (4)$$

where D and e are the quartz diameter and thickness respectively. In the above calculations the quartz anisotropy is obviously neglected. It may be noticed that thermal conduction into metallic thin films like the electrodes is negligible.

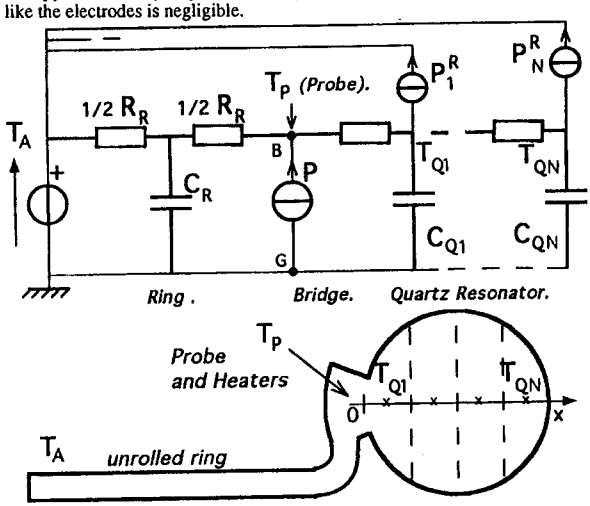


Figure 4 : a more realistic thermal model

In figure 4, radiative exchanges are simply a constant fraction of the global radiative flux in each slice. The global flux is approximated by

$$P^R = S_Q\, f_{QE}\, \sigma\left(T_Q^4 - T_E^4\right) \quad \text{where} \quad f_{QE} = \frac{\varepsilon_Q\, \varepsilon_E}{\varepsilon_E + \varepsilon_Q \frac{S_Q}{S_E}\left(1 - \varepsilon_E\right)} \qquad (5)$$

S_Q denotes the total quartz resonator surface which exchanges with the enclosure surface S_E, assuming that each of them is at the mean temperature T_Q and $T_E = T_A$ respectively. The emissivity ε_E of the enclosure which is a conventional nickel coated copper cap is estimated to $\varepsilon_E = 0.3$ (this is a quite good reflector)

Quartz material has a complex radiative behavior which can be simplified in two zones : a first one for wavelength down to 4 µm where quartz is transparent and a second one for wavelength greater than 4 µm where it can be considered a black body (we assume that Kirchhoff's law is valid).
98 % of the gray body emissive energy is within the range $[\lambda_m / 2,\ 8\lambda_m]$, where λ_m is the wavelength at the peak of emission : according to the Wien's rule for a quartz bulk at 80° C (353 K) λ_m = 8.2 µm. Hence, we can use the straightforward formulation stating the quartz be equivalent to a black body(ε_Q = 1)instead of the band approximation /8/.

Since gold deposited on a polished surface is a very good reflector the S_Q surface in equation (5) can be restricted to non-metallized quartz surfaces.

Clearly, the radiative fluxes induce a shift of the average temperature as well as thermal gradients in the vibrating crystal area and consequently create frequency shifts. Such spurious fluxes could be canceled by a gold coating of the crystal and/or by making the enclosure inner surface perfectly reflective which would give $\varepsilon_E \approx 0$ in equation (5) for example.

Figure 5 shows some results of simulation which illustrate the behavior of the network described in figure 4 with and without radiative exchanges in order to point out their influence. Four nodes represent the quartz resonator (cuts out in four slices) : just two temperatures T_{Q1} and T_{Q4} are shown, i.e. respectively temperatures at the closest node to the bridge and the farthest from it. In this simulation, the power control loop is defined by $P = K.(T_{Ref} - T_P)$ with K = 200 W/K and T_{ref} = 80° C ; the ambient temperature T_A is set at 20° C. Figure 5 clearly points out that the non-controlled radiative flux increases the steady state power consumption and the warm-up time while creating a temperature gradient in the resonant part of the quartz.

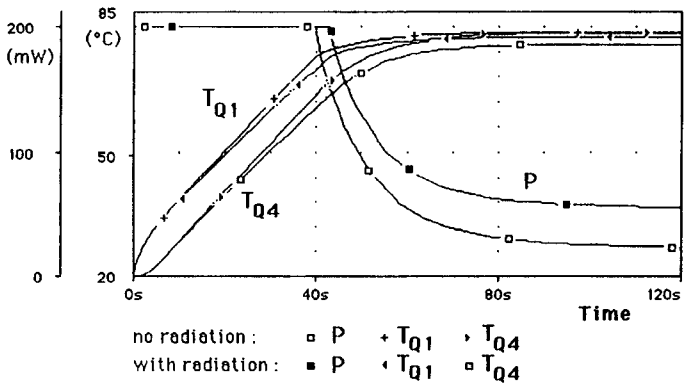


Figure 5 : simulation results : Warm-up

5. MECHANICAL ASPECT

The quartz resonator with its ring must be mechanically coupled up to the enclosure. There are many solutions for this. For the moment, we suppose a stiff link which means the ring is clamped at its end for the open ring and in the middle for the closed one.
The quartz structure has been meshed for a detailed modal analysis with a finite elements method : with the above-mentionned dimensions the first predicted eigen frequency occurs at about 400 Hz for the open ring S.B.R.,which could be judged too low for some applications, whereas we predicted a first eigen frequency close to 1100 Hz for a closed ring S.B.R. with a ring which is twice as thin as in the opened design. Let us notice that in the latter design the thermal resistance of the ring is also divided by two.
Those first eigen modes are shown in figures 6 and 7.

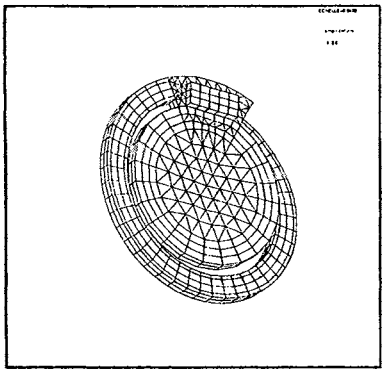


Figure 6 : First eigen mode (at 400 Hz) of the open ring SBR

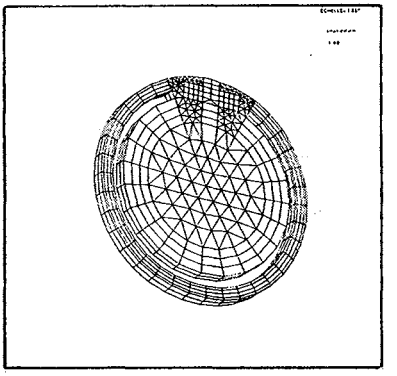


Figure 7 : First eigen mode (at 1100 Hz) of the closed ring SBR

6. ANALYSIS OF A FIRST PROTOTYPE

At this time, we just made a basic prototype : the quartz resonator is a 5th overtone-10MHz AT-cut heated by a single conventional surface mounted resistor simply pasted on the quartz bridge.We chose an open ring design clamped at its end in a metallic holder. The temperature probe -a thermistor- is pasted on the heating resistor. Figure 8 shows this assembly.

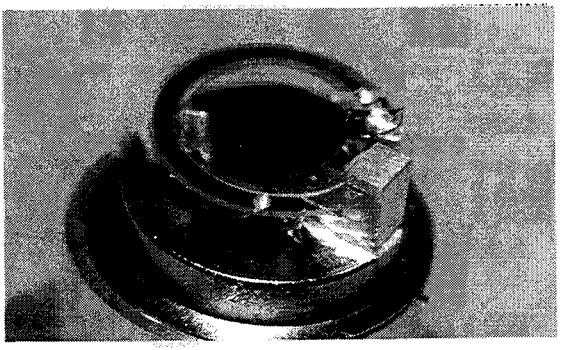


Figure 8 : Open ring SBR prototype

It is obvious that the C mode of an SC-cut would be better because it is less temperature sensitive than the AT cut's. The latter is a willful choice for tests (frequency is ten times more temperature sensitive for the AT-cut).

Results are in accordance with simulation's for warm-up time and power consumption but very disappointing concerning the frequency deviation within the ambient temperature range. No attention was paid to avoid radiative exchanges in the structure and the electronic was not temperature controlled in the first experiments : a frequency change of a few 10^{-8} to 1.10^{-7} was then expected within the temperature range (- 20° C, + 70° C). In fact, in these conditions,

experimental results gave 7.10^{-7} peak to peak with the 3rd order behavior of the resonator f-T characteristic !

The explanation comes from the fact that the measured temperature in the controlled loop is not really the quartz bridge temperature : it was easier to paste the thermistor on to the heating resistor instead of the quartz. To take into account this reality we must disjoint the probe net from the bridge's (point B in figure 4) and insert one thermal resistor R_{al} in the schematic of figure 4 : this resistor R_{al} can be estimated between 20 K/W to 100 K/W considering the thermal path into alumina -the substrate of the thick film resistor- and mainly the contact surface.

Moreover, other experimental parisitic thermal fluxes such as those of the electrical wires can be added in the schematic.

Simulations using this electrical analogy as well as the finite elements method provide explanations on the observed defects. The former has the advantage to be easy to build and perfectly works in closed loop. The latter provides more extended results for the temperature distribution but its application is more cumbersome especially to model the dynamic behavior of the temperature feedback in the probe area.

Figure 9 illustrates the use of the electrical analogy when the reference temperature $T_{réf}$ is adjusted like in practice : the room temperature is about 25° C and the adjustment consists in finding the quartz turn-over point in steady state. With the AT-cut that we use, this adjustment leads to simulated frequency deviations shown in figure 9, because the quartz temperature is never equal to the probe temperature ($T_P = T_{réf}$). Only one static temperature frequency effect is taken into account in that figure as if the quartz temperature was homogeneous (here equal to T_{Q2}). In fact, thermal gradients ($T_{Q1} \neq T_{Q2} \neq ...$) add an extra frequency shift.

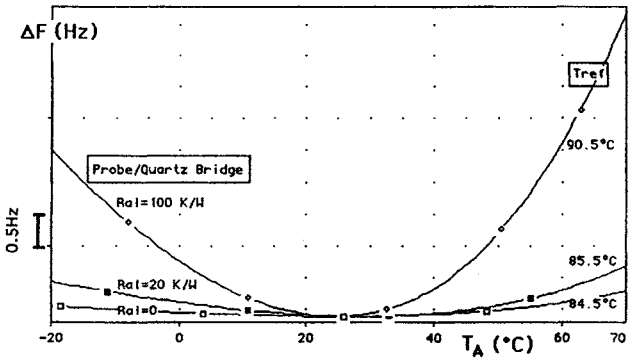


Figure 9 : Frequency deviation versus ambient temperature.T_A with respect to the thermal resistor R_{al} between the probe and the quartz bridge. Usually, the reference temperature T_{ref} of the controlled loop is adjusted at room temperature Ta = 25°C.

The effect of thermal gradients is actually double. Firstly, the temperature itself locally modifies the effective elastic coefficients governing the sound speed along the plate thickness. Secondly, thermal gradients are turned into thermoelastic stresses through thermal expansion properties of the crystal. Whatever is their origin, induced static stresses in turn locally modify the speed of sound, according to the well-known non linear mechanism widely studied in lots of papers dealing with force-sensitivity of high frequency vibrations /9/. The frequency shift arising from a purely mechanical biasing state can be evaluated performing a perturbation integral /10/ over the whole vibrating area. Indeed, a correct modelling of the global frequency shift arising from thermal gradients should take into account both locally cumulated effects /11/. Assuming a relatively strong trapping of the

vibration at the plate center, we can get a reasonable estimation just averaging the cumulated double effects at a few regularly spaced points of this restricted central region.

Using SYSTUS finite elements software, we were able to model our first prototype in an almost integrated way, according to the following steps :

- The first one consisted in computing the temperature distribution in the quartz crystal, and especially the vibrating region. This was made under various sets of assumption, for comparison: 1) purely conductive behavior, 2) conduction + simplified radiation obtained linearizing Stefan's law, 3) pure conduction + full radiation, additionaly requiring to mesh the enclosure.

-The second step consisted in computing induced stresses and strains from the previous sets of data, with help of thermoelastic capabilities of SYSTUS.

- In the third step, we evaluated the corresponding stress-induced frequency shift, following the perturbation integral procedure /12/ upon reduction to a line along the thickness, we added it to the direct local shift due to the temperature difference between the considered point and the plate center, and we averaged the obtained cumulated results.

The whole procedure was greatly facilited by the fact that we used the same mesh for the purely thermal analysis (first step) and the thermoelastic analysis (second step). Linearizing Stefan's law around the probe temperature provided with larger gradients than the full radiative model, requiring non linear solution algorithms, as can be checked on the following table and figures 11 and 11.

Model \ TA	- 20 °C	+ 20 °C	+ 70 °C
Pure conduction	- 0.35	- 0.21	- 0.04
+ linearized radiation	- 10.15	- 6.52	- 1.98
+ full radiation	- 3.50	n.a	n.a
Δ f (Hz) vs temp. of clamping zone			

7 - CONCLUSION

The main weakness of the first prototype is certainely the ill-positioning of the temperature probe, which must be put in direct contact with the crystal bridge itself, instead of the heating element. Furthermore, the detailed analysis expectedly shows the rather important influence of radiation on the global frequency shift. This can be drastically improved using a radiative shield, consisting either in goldening the crystal surface or the inner surface of enclosure.
Otherwise, it seems promising to take advantage of the strong anisotropy of the static stress-frequency effect to minimize the gradient-induced frequency shift through a proper orientation of the bridge. Finite element modelling can be an efficient tool to extensively considered the case of arbitrarily oriented designs.

Eventually, we still plan to reach significant improvements in further designs using the here-presented principles..
This device is now patented /13/.

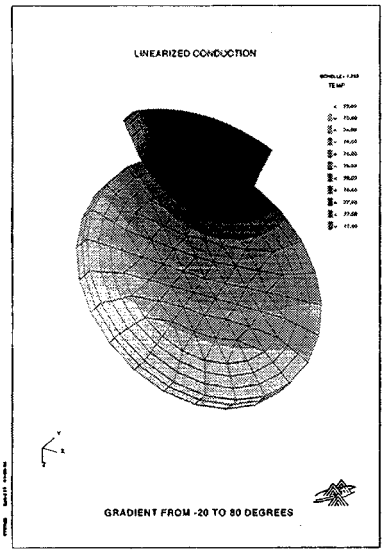


Figure 10

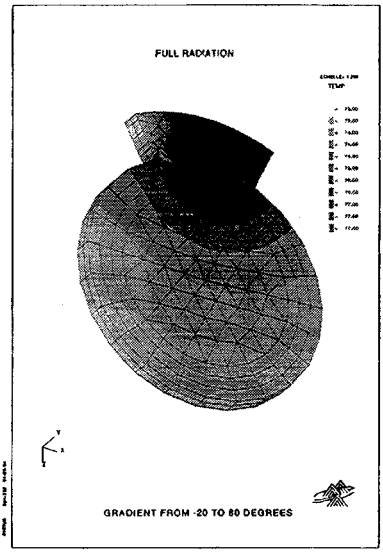


Figure 11

8. REFERENCES

/1/ D.A. SYMONDS and D. ALLEN, "An update on the TMXO" - Proceedings of the 44th Annual Symposium on Frequency Control, pp 615-621, 1990.

/2/ H.W. JACKSON, "Update on the tactical miniature crystal oscillator program" Proceedings of the 36th Annual Frequency Control Symposium, pp 492-498, 1982.

/3/ J.P. AUBRY, G. WAGNER and V. CANDELIER, "Miniaturization and stability : a new generation of ovenised oscillators" - Proceedings of the 6th European Frequency and Time Forum, pp 173-177, 1992.

/4/ Y.L. VOROKHOVSKY and B.G.DRAKHLIS, "High-stability quartz oscillators on internally-heated quartz resonators with AT and SC cuts" - Proceedings of the 45th Annual Symposium on Frequency Control, pp 447-450, 1991.

/5/ I.V. ABRAMSON and A.N. DIKIDSHI, "Improvement of characteristics of quartz resonator thermostate with direct heating piezoelement" - Proceedings of the IEEE Frequency Control Symposium, pp 499-504, 1992.

/6/ B. LONG and G. WEAVER, "Quartz crystal oscillators with direct resonator heating Proceedings of the 45th Annual Symposium on Frequency Control, pp 384-392, 1991.

/7/ S. GALLIOU, "Chauffage rapide avec résonateurs BVA tactiques", contrat DRET N° 85/180, pp 31-43, 1987.

/8/ M.N. ÖZISIK, "Radiative transfer and interactions with conduction and convection" John Wiley and Sons, New-York, pp 158-166, 1973.

/9/ J.M. RATAJSKI, "Force frequency coefficient of singly rotated vibrating quartz crystal", IBM J. Res. Dev. 12 n° 1, 92, 1968.

/10/ H.F. TIERSTEN, "Perturbation Theory for Linear Electroelastic equation for small Fields superposed on a bias" J. Ac. Soc. Am, Vol 64, n° 3, pp 832-837, 1978.

/11/ B.K. SINHA, H.F. TIERSTEN, "First temperature derivatives of the fundamental elastic constants of quartz" J. Appl. Phys., Vol 50, n° 4, pp 2732-2739, 1979.

/12/ B. DULMET, R. BOURQUIN, "Force sensitivity of trapped energy vibration in a contoured resonator" Proc. of 41th A.F.C.S., pp 295-302, 1987.

/13/ FRENCH PATENT N° 92 03175, BESSON R., GALLIOU S., MOUREY M., "Résonateur pour oscillateur thermostaté à faible consommation et chauffage rapide" PATENT PENDING EUROPE: N° 93 400638.8 ; U.S.A. : N° 08/033,092 ; JAPAN : N°5-57467

INFLUENCE OF THE SAMPLE GEOMETRY ON TWINING GENERATION IN QUARTZ : EXPERIMENTAL MEASUREMENTS AND FINITE ELEMENTS ANALYSIS

P.L. GUZZO, J.J. BOY, B. DULMET, P. MAITRE

ENSMM/LCEP - La Bouloie - Route de Gray - 25030 BESANÇON CEDEX - FRANCE

ABSTRACT

As an example of ferrobielasticity, it is well known that Dauphiné twins can be generated under stress in quartz crystals. The present work investigates the propagation of ferrobielastic twinning in X-cut samples of different geometries by means of compressive experimental conditions and finite elements approach. Synthetic quartz samples were cut in parallelepipedic and pyramidal geometric shapes. The faces to be stressed were rotated of 45° about the X-axis. The switching from one twin state to the other at room temperature was obtained using an uniaxial stressing equipment where the measuring system is isolated from the hydraulic press. From experimental results it was observed that the coercive stress at switch-over seems to be dependent on twinning-cycle. The Gibbs free energy difference between existing state and twinned state was determined and was used as twinning threshold condition in a bielastic stress-strain relationship introduced into the plasticity model for nonlinear transient analysis in SYSTUS© FEM program. As it has been experimentally observed, the finite elements approach indicates that the switching transient behavior is less abrupt in samples in which the stress distribution is less homogeneous.

I. INTRODUCTION

Dauphiné twin, often called electrical twin, is a type of domain state which is usually observed in α-quartz. It is well-known that the twinned parts are dissimilar geometrically just by a rotation of 180° around the Z-axis and are of the same hand [1]. The crystal axis are parallel but the electrical polarity of piezoelectric X-axis is reversed in the twinned parts. Therefore, the existence of the two twin domains in quartz crystals deteriorates the performances of quartz resonators and must be avoided.

Usually, the existence of electrical twins is generated by growth conditions or by the loss of symmetry that takes place at the β-α phase transition. However, electrical twinning in quartz is also induced by mechanical stresses [2]. This fact is an example of ferrobielasticity and is named ferrobielastic switching [3-5]. As previously noticed [6, 7], the growth of electrical twin under mechanical or thermal stresses can be observed during the manufacturing process of quartz resonators. Thus, the study of twinning under mechanical stresses presents a considerable interest for technological applications.

Ferrobielasticity [3, 4] is a second order effect in which the strain difference between the two orientation states is induced by applied stress. It can be evaluated through the determination of the Gibbs free energy difference between the two domains. From previous works [3, 4], twinning in quartz is governed by the second order term involving elastic compliance tensor and mechanical stresses. Then, the Gibbs free energy for the ferrobielastic switching can be written as follows :

$$\Delta G = -\frac{1}{2}\, s_{ijkl}\, T_{ij}\, T_{kl} \qquad (1)$$

Due to the symmetry of quartz crystal, only the sign of the elastic compliance coefficient s_{1123} changes between domain states [4]. Thus, the Gibbs free energy difference between the existing state (domain I) and the twinned state (domain II) can be noted as :

$$\Delta G = G_{II} - G_I = 4\, s_{1123}\, (T_{11}\, T_{23} - T_{22} T_{23} + 4\, T_{12}\, T_{13}) \qquad (2)$$

where s_{1123} is assumed positive for the existing state ($s_{1123} = 4.5 \times 10^{-12}$ m^2/N). G_I and G_{II} are the energy of domain I and domain II, respectively. T_{ij} is the applied stress expressed in the crystallographic axes XYZ. If we consider a right-handed coordinate system X"Y"Z" rotated by the angles φ, θ and Ψ around Z, X' and Y", respectively, the equation (2) should be expressed in the following manner for an uniaxial compressive stress (-P) applied in the Z" direction [7] :

$$\Delta G = 4\, s_{1123}\, P^2 \cos\Psi \cos\theta\, (\cos\Psi \sin\theta \cos\varphi - \sin\Psi \sin\varphi)$$
$$[3 \sin 2\Psi \sin\theta \sin 2\varphi + \sin^2 \Psi (6 \cos^2 \varphi - 1) + \cos^2 \Psi \sin^2 \theta\, (6 \sin^2 \varphi - 1)] \qquad (3)$$

According to the principal of minimal energy, the ferrobielastic switching will occur when a threshold condition is achieved. Domain II will replace domain I only when ΔG becomes negative. Besides the crystal orientation and the temperature [5] , the coercive stress required to achieve the threshold condition is supposed to depend on several parameters such as quartz quality, strain rate and maybe stress distribution itself [7].

In the present work, the ferrobielastic twinning is investigated by means of compressive experiments and finite elements analysis in quartz samples of different geometries. Firstly, the threshold condition for twinning was determined experimentally. In order to observe the behavior of the coercive stress, several cycles of direct switching and reverse switching have been made at very slow strain rate. Secondly, the experimentally obtained threshold value was used in nonlinear transient analysis performed in a plasticity model of SYSTUS© software. The switching transient behavior is discussed in relation to the stress distribution due to the geometry of the sample.

II. EXPERIMENTAL PROCEDURE

In the present work compressive stress is used to switch the ferrobielastic domains in synthetic quartz. The experimental arrangement is illustrated in Fig. 1. The axial force is obtained by means of a hydraulic piston with short displacement and small friction. The hydraulic pressure is given by a hydropneumatic convertor and is checked by a manometer and monitored by a mechanical valve which regulates the air flow rate. The applied force is measured by a strain gauge load cell type SEDEME SBM50 with an output sensitivity of 3.12 mV/V in the full range of measurement (0-50 kN). The signal is amplified by a PIV amplifier. The strain gauge load cell is set between the piston and a steel ram which is guided by a ball bearings. In the end of the ram is a steel ball that maintains the force in the axial direction. The sample is positioned between two platens which were previously hardened.

The sample displacement is determined by using two TESA inductive gauge heads positioned near the sample and in contact with the two platens. The output signal is the differential displacement between the signals of each gauge head, amplified by a TESATRONIC TTA20 instrument. The error of the analog output is less than 0.3 % refering to the full range of measurement (200 µm). The device used to measure the sample displacement is placed in the internal cage which is free of applied and reaction forces. Data acquisition is made by a 16-bit resolution (152.6 µV) analog input board. Amplified output signals of force and displacement are stored with the assistance of a quick-basic program driven by a personal computer. Force and displacement signals have been read every 0.4 second.

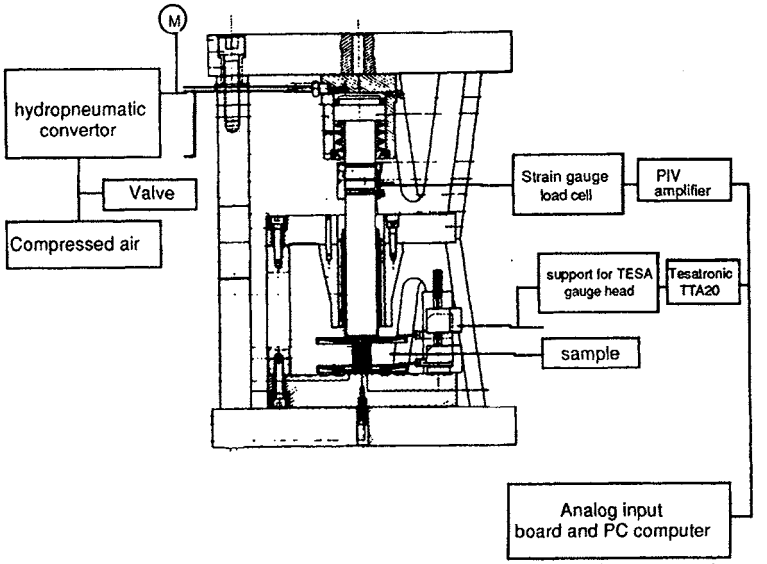


Fig. 1 - The experimental arrangement used to apply a compressive stress in quartz samples.

Synthetic right-handed quartz grown at SICN Company (France) was used in the present experiments. The samples were cut in the Z-growth region free of inclusions, twins and macroscopic defects. The crystallographic orientations were used following the IRE convention [8]. Fig. 2 shows the procedure used to define the sample orientation according to equation (3). In order to keep the Y' axis normal to the plane, the X-cut was obtained by rotating a Y-plate of 90° around Z-axis ($\varphi = 90°$). The plate was turned about Y' of 45° in the non-trigonometric sense ($\psi = -45°$), and so $\theta = 0°$.

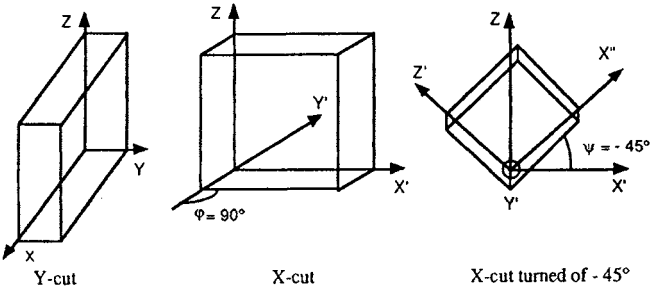
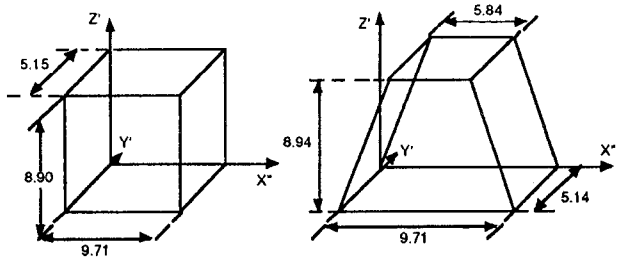


Fig. 2 - The relationship between the crystallographic axes XYZ and the coordinate system X"Y'Z' in which the force was applied parallel to the Z'-axis and X"-axis.

Two samples have been prepared in two different geometries. Their shapes and dimensions are showed in Fig. 3. All surfaces were lapped parallel and optically polished.



(a) parallelepipedic shape (b) pyramidal shape

Fig. 3 - The two different geometric shapes of X-cut samples.

The sample is centered along the force axis of the press. Firstly, the force is slowly applied in Z' direction and after turning the sample, it is applied in X" direction. In order to optically observe the development of the stripe-shape domains under stress, a polarizer and an analyzer were positioned in Y' direction. The polarity of X-faces, normal to Y' axis, were checked before and after each experiment by means of a charge amplifier.

III. EXPERIMENTAL RESULTS

The parallepipedic sample shown in Fig. 3a has been submitted to three successive twinning cycles. Each cycle consists of two switching experiments corresponding to two perpendicular directions of the applied force. The direct switching replaces the existing state (domain I) by the alternate state (domain II) and occurs when load is applied along Z' axis. The reverse switching takes place when load isapplied along X" axis. When the switch-over is achieved, the polarity of electrical axis becomes identical to this of the previous experiment. So, after each experiment, the sample has switched from one twin state to the other. In this work, the coercive stress is deduced from the measured force just before switch-over. The threshold condition is then evaluated using this coercive stress in Eq(3).

Force-displacement curves for three successive twinning cycles of the same sample are presented in Fig. 4. We observe that the force level and the way in which ferrobielastic switching occurs are different from one twin state to another. In the first and second cycles, the switch-over is abrupt when load is applied along Z' axis but exhibits several steps when load is applied along X"axis. On the other hand, for the third cycle, the switch-over becomes stepped when load is applied in Z' axis and is abrupt when load is applied along X" direction. So, from these preliminary results, it is not easy to decide wether the switching transient behavior is related or not to the direction of experiment. It should be pointed out that the slope of the force-displacement curves was not always reproductible. Thus some difference from the theoretically predicted slope occured. In addition, contact problems created non linearities at the beginning of the curves.

The values of coercive stress (σ_C) and threshold energy (ΔG) are presented in Table 1. As was previously reported [5], it is observed that coercive stress and threshold energy decrease when the sample is repeatedly twinned. Up to now, this behavior cannot clearly be explained. We can just think of two reasons to explain it:

1. the residual stress that remains in quartz lattice once the load is released,
2. the interactions between twin propagation and dislocations or impurity-related point defects.

The coercive stresses shown in Table 1 are much lower than previously reported ones [5,9] for synthetic quartz samples with the same orientation. In [9], the order of magnitude of the coercive stresses at switch-over was estimated to 450 MPa for a strain rate of 20×10^{-5} mm/s. In the present work, the strain rate is always less than 3×10^{-5} mm/s. As previously quoted in [7], the coercive stress increases with the strain rate, but till now, we have no quantitative estimation of this behavior.

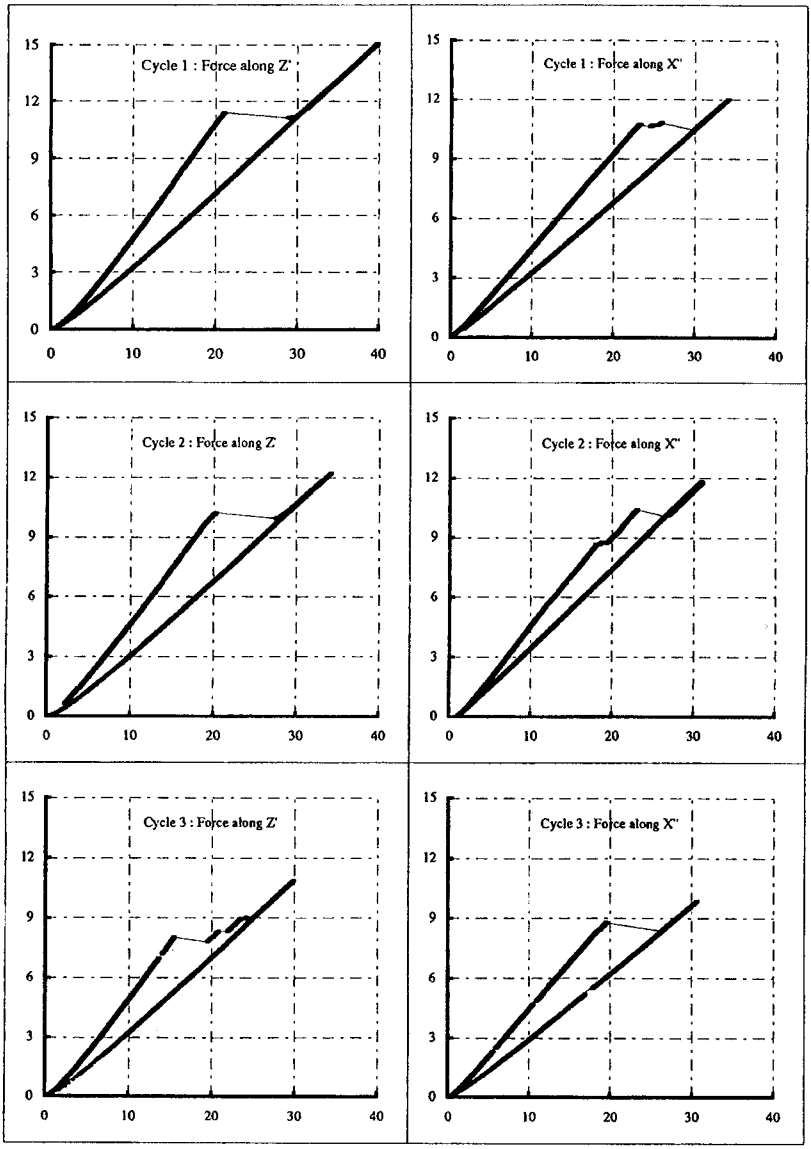


Fig. 4 : Force-displacement curves of twinning cycles for the parallelepipedic sample
(Force in kN and displacement in microns)

Table 1 : Strain rate, coercive stress and threshold energy for the parallelepipedic sample in the three sucessive twinning cycles.

Switch parameters versus cycle	Cycle 1 F along Z'	Cycle 1 F along X''	Cycle 2 F along Z'	Cycle 2 F along X''	Cycle 3 F along Z'	Cycle 3 F along X''
Strain rate (mm/s)	$3*10^{-5}$	$2*10^{-5}$	$1*10^{-5}$	$1*10^{-5}$	$2*10^{-5}$	$1*10^{-5}$
Coercive stress (MPa)	230	240	210	230	180	190
Threshold Energy (N/mm^2)	-0.23	-0.25	-0.19	-0.23	-0.15	-0.17

Fig. 5 shows comparative force-displacement curves for parallelepipedic and pyramidal samples when load is applied along the same direction (here Z' axis). It is clearly observed that the switching transient behavior is abrupt for the parallelepipedic sample and stepped for the pyramidal one. This is probably due to the stress distribution with respect to the geometrical shape of the sample. In case of pyramidal shape, the stress distribution is less homogeneous, and thus a continuous switching could be expected. In fact, we obtained an initially progressive switching, thereafter turning into a neat step.

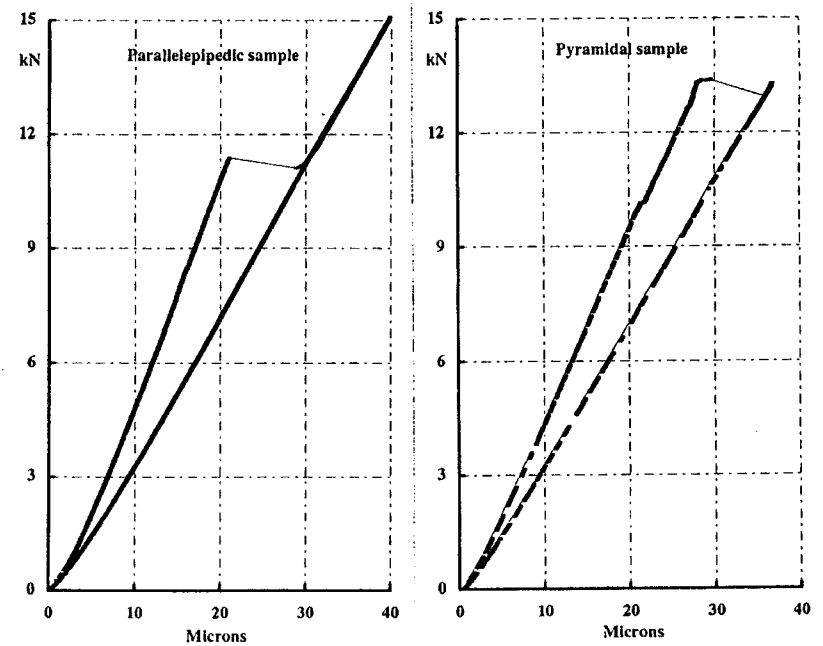


Fig. 5 : Force vs displacement curves for two sample geometries.
The force is applied along Z' axis

IV. FINITE ELEMENTS ANALYSIS

The basic assumption of the treatment remains unchanged from Ref [7] : we stated that the local stress-strain relationship is linear. Each possible twin state is associated to a stiffness coefficients matrix. The local bielastic switch from one matrix to another is ruled by a threshold condition on the energy difference between twins. Indeed, the global behavior of the whole structure is strongly non linear : although the twinning condition is assumed purely local, the switching of a given elementary domain actually modifies the stress it exerts upon its neighbors. Thus a truly non linear algorithm is required to get a convergent solution in the whole structure. In the framework of plasticity models available in SYSTUS program, the stiffness matrix for bilinear analysis is computed altogether with the tangent matrixes required to incrementally build the solution for a given loading. Because commonly encountered implementation of FEM are definitely displacement methods, the stress relaxation arising at switching time would not allow to grant convergency if the computation would be performated at imposed stress. Thus, we had to use imposed displacement loadings in all of our simulations, although the real experiments have been made while imposing the net force. This fact should be definitely kept in mind to compare simulated and experimental results.

Another consequence of the displacement approach in FEM is the way that the "locality" of the threshold condition is taken into account, and this requires a few complementary explanations. Since the unknowns are the displacement components at each node of the mesh, no other primary quantity is directly computed at these points. As a consequence, stresses, strains and energy are not computable at the nodes. In fact, the stresses are commonly computed at the center of the most elementary integration domain for a given discretization. In SYSTUS program, all volume integrals required to compute elementary stiffness matrixes are evaluated following Gauss' method, i.e. the integrands are computed at well defined points located inside the volume elements. The energy threshold condition is evaluated at those so-called Gauss points. The computed energy is not actually Gibbs' free energy, but is the stored internal energy $U=1/2\mathbf{T}:\mathbf{S}$ instead, which can be readily obtained once you get $\mathbf{T}$ and can access the stress-strain relationship. The required programming was made by Dr P. Conraux at Framasoft's. No significant change had been made from Ref. [7]. It should be remarked that the inner programming is made symmetrically, so that it can be used for both direct and reverse switching.

Since Gauss points are located inside the volume elements they are more numerous than the nodes, which are shared by adjacent elements, whereas Gauss points are not. In addition, the number of Gauss points per element depends on the polynomial interpolation degree of unknown fields inside the elements. This may result into huge data files in case of high degree interpolation, because SYSTUS records all Gauss points results as permanent data on the mass storage device. Thus, we preferred to stick to simple linear interpolation elements in most of our analyses.

Stress distributions are rather homogeneous in compressed parallelepipedic samples. Because the twinning threshold condition tends to be simultaneously satisfied in the whole structure, numerical simulations tend to exhibit a strong bistable behavior. It is then very difficult to obtain continuous convergency towards intermediate partially twinned states. For instance, on Fig 6a, corresponding to the sample geometry of Fig. 3a, the switching transient behavior starts with a neat step, followed by a small range in which the stress-strain relationship is characterized by a stable intermediate value of the global stiffness. The final stiffness value is then jumped to through a second step, smaller than the former one.

Since stress distributions in pyramidal samples tend to linearly vary along the height, we found interesting to also simulate the behavior of such samples. The corresponding predicted results are produced in Fig 6b, which effectively exhibits a continuous progressive behavior. Such continuous results indicate a stable convergency, from the numerical point of view, whereas the stepped results predicted for the parallelepipedic sample correspond to a strong instability, which is confirmed by the increased number of needed iterations before convergency.

Further improvements still remain to be done in our simulation model to take into account the decrease of energy threshold experimentally observed along successive twinning cycles.

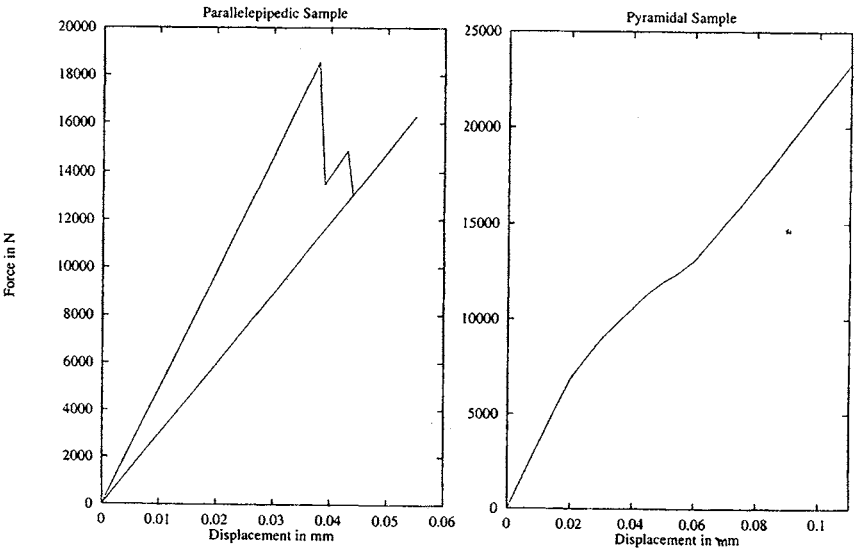


Fig. 6 Predicted Force vs displacement curves for parallelepipedic and pyramidal geometries submitted to imposed displacement loading.

V. CONCLUSION

The effect of the sample geometry on the ferrobielastic twinning has been studied by means of compressive experiments and finite elements analyses. We can summarize our preliminary investigations by the following items :

1. It is observed from experimental results that the switching transient behavior is less abrupt for the pyramidal sample than for the parallelepipedic sample. Finite elements simulations have shown a continuous progressive switching for the pyramidal sample, in which stress distribution is less homogeneous. In case of the parallelepipedic sample, the switching transient behavior is characterized by two neat steps. This is a suggestive result for further investigations aimed to understand the influence of sample geometry in ferrobielastic switching.

2. It is confirmed that coercive stress decreases when the number of twinning cycles increases.

3. Force-displacement curves do not exhibit evident relation between switching transient behavior and twinning cycles.

4. The coercive stress shown in Table 1 is about twice lower than previously reported values. Until now, it is not clear whether strain rate can actually affect coercive stress or not. Further experiments still remain to be done to precise this influence.

ACKNOWLEDGMENTS

The authors whish to thank Dr. P. Conraux, from Framasoft-Lyon, for the introduction of bielasticity in SYSTUS plasticity model. One of the authors (P.L. G.) is grateful for scholarship provided by CNPq (Brazil).

REFERENCES

[1]. C. FRONDEL ,"The sytem of mineralogy", Silica Minerals, Vol 3 Wiley&Sons, N.Y. ,1962.

[2]. L.A.THOMAS , W.A. WOOSTER "Piezocrescence-the growth of Dauphiné twinning in quartz under stress", Proc. Roy. Soc. London, **A208**, pp 43-62, 1951.

[3]. K.AIZU , "Second order ferroic state shifts", J. Phys. Soc. Japan, **34** (1) pp 121-128, 1973.

[4]. R.E.NEWNHAM, "Domains in minerals", Am. Miner. , **59**, pp 906-918, 1974.

[5].T.L. ANDERSON, R.E. NEWNHAM, L.E.CROSS, "Coercive stress for ferrobielastic twinning in quartz", 31th Annual Symposium on Frequency Control, pp 171-177, 1977.

[6].E.P. EERNISSE, R.W. WARD, "Quartz resonator sensors in extreme environments", 45th Annual Symposium on Frequency Control, pp 254-260, 1991.

[7]. J.J. BOY, B. DULMET, P. CONRAUX, "Quartz crystal twinning under mechanical stresss: Finite elements analyis", 6th European Frequency and Time forum, pp 323-327, 1992.

[8]. "Standards on Piezoelectric Crystals", Proc. I.R.E., **37**, pp.1378-1395, 1949.

[9]. J. MARKGRAAFF, "Elastic behavior of quartz during stress induced Dauphiné twinning", Phys. Chem. Minerals, **13**, pp 102-112, 1986.

Transient characteristics of a composite piezoelectric resonator

H. Nowotny*, M. Gröschl**, R. Spolenak**, W. Burger**

*) Institut für Theoretische Physik, Technical University Vienna, Austria
**) Institut für Allgemeine Physik, Technical University Vienna, Austria

We consider the behaviour of a composite piezoelectric resonator by a sudden change of the frequency of the driving generator, especially for a frequency jump from one overtone resonance frequency to an other. The building-up time is not only calculated using the electric circuit admittance $Y(\omega)$ of the composite piezoelectric resonator but also determined experimentally. Both results show that the transient response of the whole system is determined by the resonator cell itself rather than by the electronics.

1 Introduction

A composite piezoelectric resonator driven at an overtone resonance frequency can be used to obtain a standing ultrasonic wave in a liquid. It is well known that the acoustic radiation pressure drives small particles dispersed in the liquid towards the antinodes or nodes of the standing wave field. Changing the space location of these nodes or antinodes in a suitable way by changing the driving frequency (frequency jump from one overtone resonance frequency to an other) gives a systematic flow of the particles and can therefore be used as a separation technique (Drifting Resonance Field cell) /1/. For the optimization of the separation process the exact knowledge of the transient characteristics of the composite resonator in case of such a step in the driving frequency is essential.

In the present work the building-up time is calculated using the electric circuit admittance $Y(\omega)$ of the composite piezoelectric resonator and assuming a step function in the driving frequency. First the admittance $Y(\omega)$ is calculated in a one-dimensional approximation using a transfer-matrix method and adjusted to measured values. In a second step an analytical approximation for the admittance function is done to perform a Fourier-transformation and to describe the transient process as a function of time. This analytical approximation to the admittance function corresponds to a generalized Butterworth–Van-Dyke equivalent-circuit of the piezoelectric resonator.

2 Composite piezoelectric resonator

2.1 Electrical admittance

The piezoelectric resonator under investigation is shown schematically in fig. 1b. There is a piezoelectric ceramic plate (thickness 0.987 mm, Area 625 mm^2) between two electrodes and above the glass carrier (thickness 2.71 mm) of one electrode there is a water layer of thickness 29.05 mm. On the other end of the water layer there is a reflecting aluminum plate (thickness 1.99 mm).

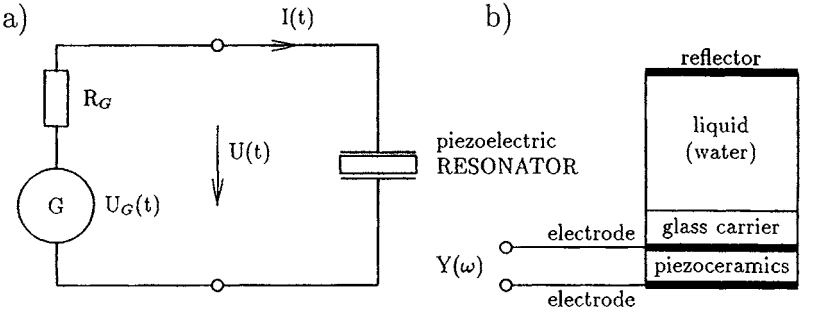


Fig. 1 Composite piezoelectric resonator
a) as part of an electric circuit with a generator G
b) schematic construction of the resonator

Using this resonator as part of a harmonic electric circuit (see fig. 1a) one can measure the admittance $Y(\omega) = I(\omega)/U(\omega)$ /2/. The experimental result of this measurement is shown in fig. 2a where the absolute value of $Y(\omega)$ is given on a logarithmic scale as a function of the frequency f. In the shown frequency range there are many overtone resonances caused by the water layer. Looking at the maxima of these resonances one detects two so called eigenresonances of the transducer caused by the piezoceramic plate in connection with the glass carrier.

Using a one-dimensional description /3,4/ we have also calculated the electric admittance of this composite resonator. The result is shown in fig. 2b.

2.2 Equivalent electric circuit

The electrical behaviour of this composite piezoelectric resonator can be described by means of a generalized Butterworth-Van-Dyke equivalent circuit /5, 6/ as shown in fig. 3. From the calculated admittance shown in fig. 2b one can approximate the values R_i, C_i and L_i describing the i-th resonance at frequency f_i of the whole

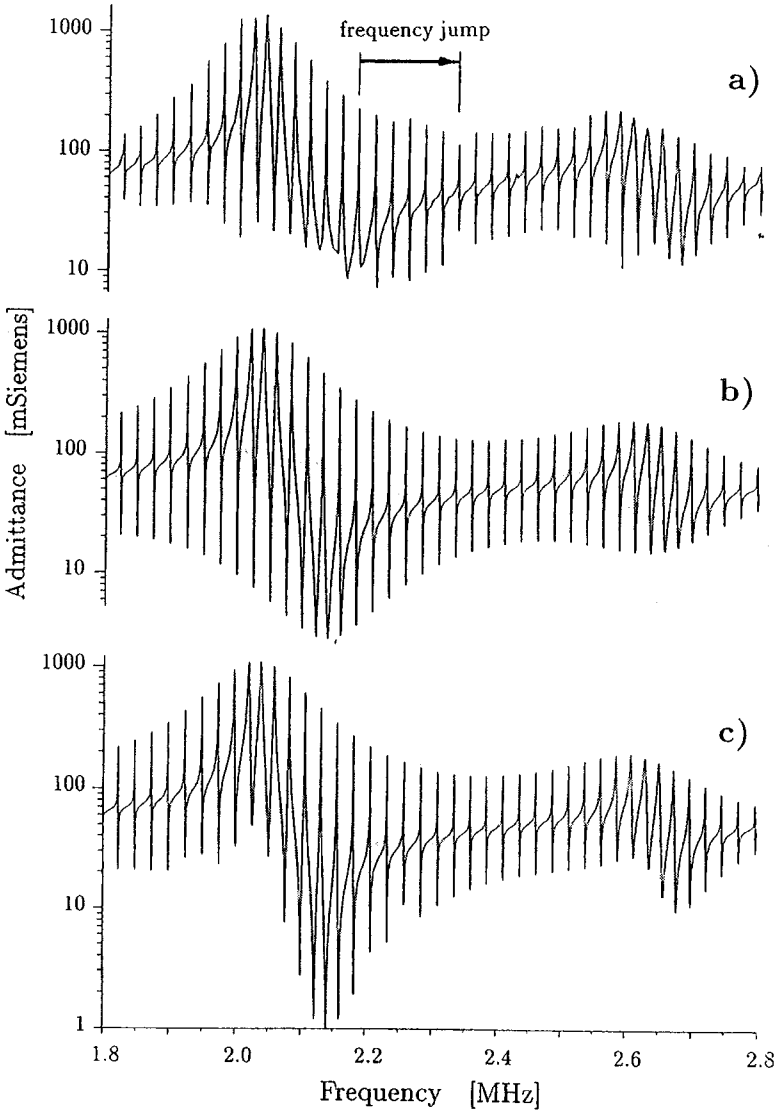


Fig. 2 Electric admittance $Y(\omega)$ of the composite piezoelectric resonator
a) measured, b) one-dimensional model, c) equivalent electric circuit

admittance. Values for these parameters R_i, C_i and L_i are given in table 1 for some resonances. The main capacity C_o was 3.9 nF.

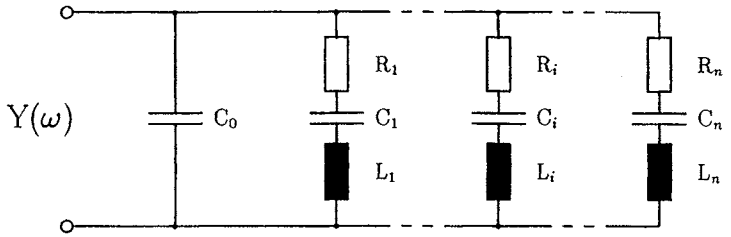


Fig. 3 Equivalent electric circuit of the composite piezoelectric resonator describing n overtone resonance frequencies

f [MHz]	...	2.156	2.181	2.206	2.232	2.257	2.283	2.308	2.336	...
R [Ω]	...	2.797	3.564	4.393	5.243	6.070	6.831	7.488	8.006	...
C_i [pF]	...	4.06	2.91	2.22	1.79	1.51	1.32	1.20	1.12	...
L [mH]	...	1.343	1.832	2.340	2.838	3.294	3.681	3.974	4.153	...

Table 1 Some values used for the description of the piezoelectric resonator by an equivalent electric circuit

3 Transient response (theoretical)

3.1 Mathematical description

For a non-harmonic time dependent generator voltage $U_G(t)$ we obtain from fig. 1a the following relation for the Fourier-component $I(\omega)$ of the current:

$$I(\omega) = \frac{U_G(\omega)}{R_G + Z(\omega)} \quad , \tag{1}$$

whereby $Z(\omega) = 1/Y(\omega)$ describes the piezoelectric resonator and R_G is the output resistance of the generator. The time-dependent current follows from equ. 1 by a Fouriertransformation

$$I(t) = \int_{-\infty}^{+\infty} \mathrm{d}\omega \, \mathrm{e}^{j\omega t} \frac{U_G(\omega)}{R_G + Z(\omega)} \quad . \tag{2}$$

If we assume for the voltage $U_G(t)$ a frequency jump at $t = 0$ from frequency ω_- to frequency ω_+

$$U(t) = U_- \, \mathrm{e}^{j\omega_- t} \, \Theta(-t) + U_+ \, \mathrm{e}^{j\omega_+ t} \, \Theta(t) + c.c. \quad , \tag{3}$$

we can calculate the Fourier transformed voltage

$$U_G(\omega) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} d\omega\, e^{-j\omega t}\, U_G(t)$$
$$= \frac{j}{2\pi}\left(\frac{U_-}{\omega-\omega_- + j\delta} + \frac{U_-^*}{\omega+\omega_- + j\delta} - \frac{U_+}{\omega-\omega_+ - j\delta} - \frac{U_+^*}{\omega+\omega_+ - j\delta}\right) \quad (4)$$

(the limit $\delta \to +0$ has to be taken).

The next step is the calculation of $I(t)$ by means of equ. 2. We can use the residue theorem if we close (for $t > 0$) the integration path in the upper half-plane:

$$I(t) = 2\pi j \left(\text{sums of enclosed residues}\right) \quad . \quad (5)$$

The result of this calculation is the sum

$$I(t) = I_+(t) + \sum_p I_p(t) \qquad (\text{ for } \quad t > 0\ ) \quad , \quad (6)$$

wherein $I_+(t)$ is a harmonic current

$$I_+(t) = \frac{U_+}{R_G + Z(\omega_+)}\, e^{j\omega_+ t} + c.c. \quad (6a)$$

and $I_p(t)$ are particulary currents disappearing for $t \to \infty$ (because all ω_p have positive finite imaginary parts):

$$I_p(t) = \left(\text{Res}\frac{1}{R_G + Z(\omega_p)}\right)\left(\frac{U_-}{\omega_p-\omega_-} + \frac{U_-^*}{\omega_p+\omega_-} - \frac{U_+}{\omega_p-\omega_+} - \frac{U_+^*}{\omega_p+\omega_+}\right) e^{j\omega_p t} \quad (6b)$$

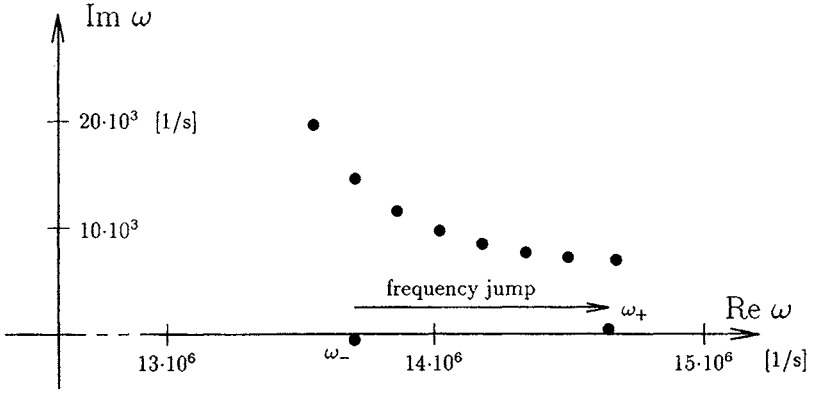


Fig. 4 Poles of $I(\omega)$ for the calculation of $I(t)$
(take notice of the different scales for Re ω and Im ω)

The complex frequencies ω_p are determined by the zeros of the function

$$R_G + Z(\omega_p) = 0 \tag{7}$$

The location of some of these frequencies for $R_G = 50\ \Omega$ are shown in fig. 4. The residues are given by

$$\text{Res}\frac{1}{R_G + Z(\omega_p)} = \lim_{\omega \to \omega_p} \frac{\omega - \omega_p}{R_G + Z(\omega)} \quad . \tag{7a}$$

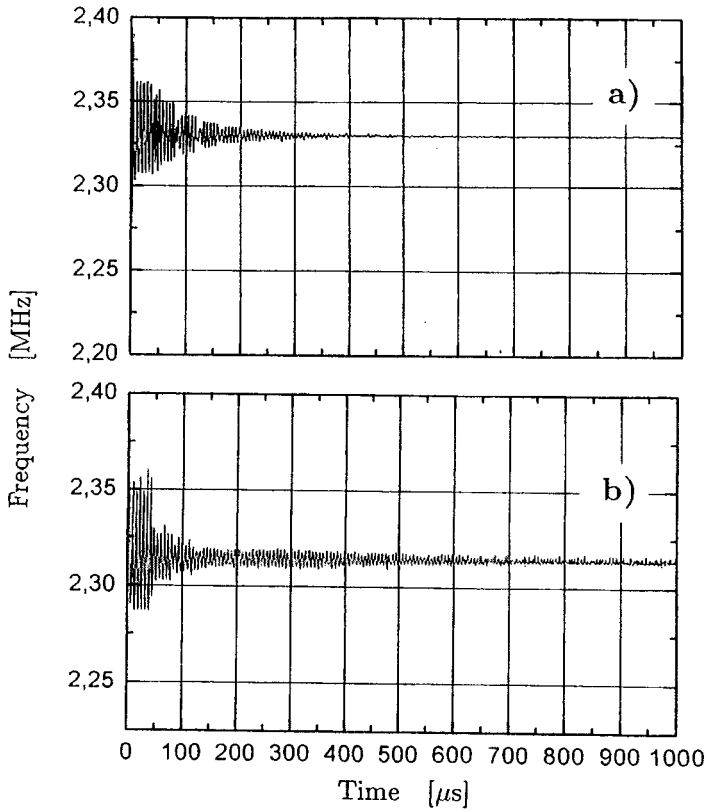


Fig. 5 Transient response of the piezoelectric resonator for a frequency jump from 2.18 to 2.33 Mhz: a) calculated b) measured

3.2 Numerical results

Using the value $R_G = 50\ \Omega$ and the condition $U_- = U_+$ we have calculated $I(t)$ for some chosen frequencies ω_- and ω_+. The separation technique with a Drifting Resonance Field cell (DRF cell) needs a cyclic frequency jump over 5 or more resonances which are located between the eigen frequencies of the transducer /1/. Therefore we have considered a frequency jump from 2.18 to 2.33 MHz in more detail (see fig. 2a).

The result is given as a time-dependent instantaneous frequency $f(t)$, which is defined by successive zero points of the current $I(t)$ with the same sign for the slope of $I(t)$:

$$f\left(\frac{t_{n+1}+t_n}{2}\right) = \frac{1}{t_{n+1}-t_n} \quad \text{with} \quad I(t_n) = I(t_{n+1}) = 0\,, \quad \dot{I}(t_n)\cdot\dot{I}(t_{n+1}) > 0 \quad . \tag{8}$$

In fig. 5a the result of a calculation is shown were we have used 28 resonances for the equivalent electric circuit. One can see that the instantaneous frequency reaches the driving frequency ω_+ after a time of some hundreds of microseconds. An other result is given in fig. 6. In this case the driving frequency ω_- is so close to a (complex) resonance frequency ω_p that the corresponding damped current $I_p(t)$ is dominant for a long time and the change of the instantaneous frequency from ω_- to ω_+ starts with a delay of 1800 μs.

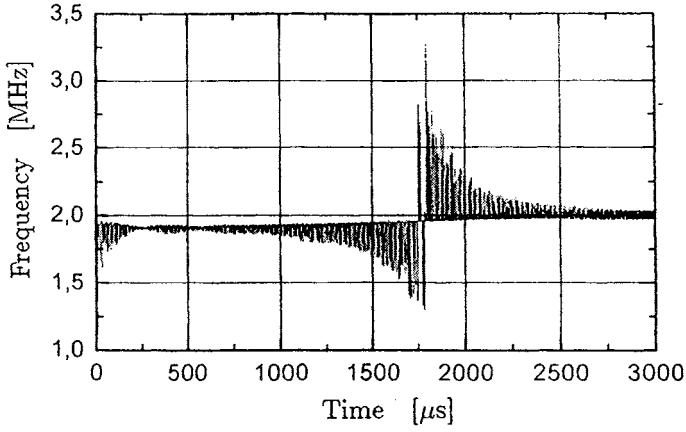


Fig. 6 Calculated transient response of the piezoelectric resonator for the case of driving frequency ω_- nearby a (complex) resonance frequency

4 Transient response (experimental)

4.1 Experimental Setup

Experimentally we have measured this transient process by means of a high resolution frequency and time interval analyzer (HP 5372A), whose input-signal, taken across a low ohm serial resistor in the electric circuit, was proportional to the current (see fig. 7). This analyzer provides the time progress of the instantaneous frequency of its input signal according to the definition given in equ. 8.

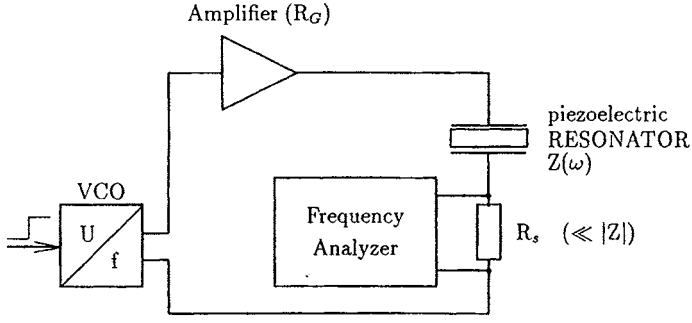


Fig. 7 Experimental setup for the measurements of the transient response

A voltage controlled oscillator (VCO, type EXAR XR 205) was used as frequency generator and connected via a power amplifier (ENI 2100L) to the piezoelectric resonator. The frequency jumping was generated applying an appropriate voltage transient with a slew rate greater than 10 V/μs at the VCO input.

4.2 Experimental results

The transient behaviour of the electronic driving circuit (without resonator) was investigated by means of replacing the resonator by an ohmic load resistor (dummy load). The result of this measurement is shown in Fig. 8 and gives a settling time less than 5 μs for the frequency jump of the generator (pay attention to the different time-scale in this figure compared to the other figures).

The measured transient behaviour of the investigated resonator cell for a frequency jump of the VCO from 2.18 to 2.33 Mhz is shown in fig. 5b. There is a good agreement with the corresponding calculation given in fig. 5a.

In the series of our measurements we sometimes observed a behaviour similar to the theoretical result shown in fig. 6. One of these experimental results is shown in

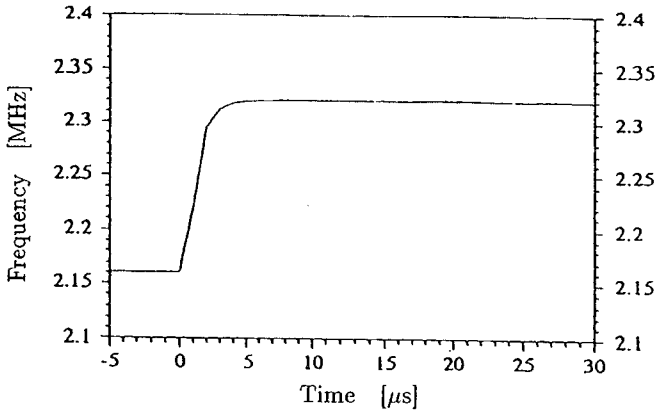


Fig. 8 Measured transient response of the driving electronics without the piezoelectric resonator

fig. 9. We think that in these cases the condition ω_- close to a resonance frequency ω_p is experimentally fulfilled. A remarkable fact is, that in all experiments done no delay time larger than 100 μs was found. The reason may be the form of the driving voltage $U_G(t)$ given in equ. 3. The sudden jump from the driving frequency ω_- to ω_+ corresponds not exactly to the experimental condidtions where a time of 2–4 μs was necessary for this frequency jump (see fig. 8).

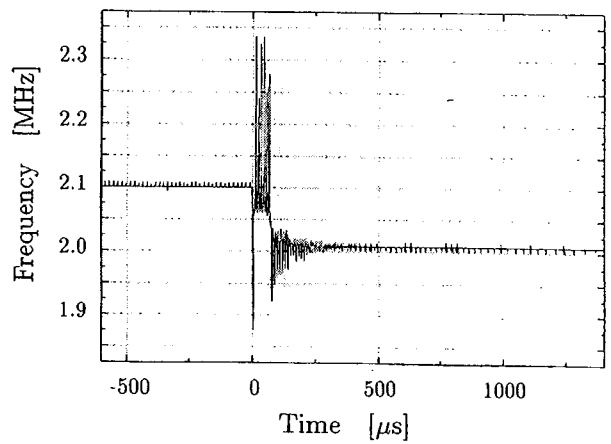


Fig. 9 Measured transient response of the piezoelectric resonator with a delay in changing the instantaneous frequency to the new driving frequency

5 Summary and Conclusion

The results (theoretical calculations as well as measurements) showed that the transient response of the whole system is determined by the resonator cell itself rather than by the electronics, because the building-up time of standing waves is between 100 and 1000 μs and is therefore much greater than the transient time of the used VCO for a frequency jump. The agreement between calculated and measured build-up times is satisfying in most cases. Only for some calculations it seems to be necessary to describe the driving frequency jump with a finite step time to get a good agreement with the measurements. But also in these cases a step-function can be used for the frequency jump if one changes the boundary conditions appropriately (this is an approximative method for taking into account the finite step time of the voltage generator).

References

/1/ E. Benes, F. Hager, W. Bolek, M. Gröschl, in: Separation of dispersed particles by drifting ultrasonic resonance fields, Ultrasonics International 91 – Conference Proceedings, Butterworth-Heinemann, Oxford, UK (1992), pp. 167/70

/2/ IEEE Standard on Piezoelectricity, No. 176 (1978)

/3/ H. Nowotny and E. Benes, General one-dimensional treatment of the layered piezoelectric resonator with two electrodes. J. Acoust. Soc. Am. vol. 82 (1987), pp. 513/21

/4/ H. Nowotny, E. Benes and M. Schmid, Layered piezoelectric resonators with an arbitrary number of electrodes (general one-dimensional treatment). J. Acoust. Soc. Am. vol. 90 (1991), pp. 1238/45

/5/ K. S. van Dyke, The electric network equivalent of piezo-electric resonators. Phys. Rev. vol. 25 (1925) pp. 825

/6/ IEEE Standard, Definitions and methods of measurements for piezoelectric vibrators, No. 177 (1966)

8th European Frequency and Time Forum, Munich, march 9-11, 1994

SYNCHRONIZATION OF OSCILLATORS THROUGH NONLINEAR RESONANCE : APPLICATION TO WAVE STABILITY IN ACOUSTIC RESONATORS, JOSEPHSON JUNCTIONS AND BIOLOGICAL NEURONS

M. Planat*, P. Koch**

*Laboratoire de Physique et Métrologie des Oscillateurs du C.N.R.S.
associé à l'Université de Franche-Comté-Besançon
32 avenue de l'Observatoire - 25044 Besançon Cedex - France

** Centre de Calcul - Université de Franche-Comté-Besançon - France

Summary :

Synchronization through nonlinear resonance is the process by which nonlinear multiparameters system organize itself.

The understanding of these coherence has grown recently thanks to the interplay of two concepts : that of periodicity (or quasiperiodicity, or torus...) and that of fractality (or self-similarity, or scaling...).

The revelance of these concepts has been demonstrated recently by studying ordinary differential equations and in particular the injection locking of oscillators and the stability of resonances in the context of ion cooling and Josephson junctions /1,3/.

In this paper we extend the analysis to the synchronization of globally coupled oscillators. It is applied to wave excitation, propagation and stability in Josephson junctions, biological neurons and acoustic resonators.

Introduction :

A general method of classifying resonances of dynamical systems has been introduced. It is based on the idea of studying them simultaneaously from the classical point of view (i.e. that of periodicity and Fourier transforms) and from the point of view of complexity theory (i.e. that of self similarity and fractals). In many cases (Josephson junctions /2/, modulated phase locked loops, injection locked oscillators /1/, the frequency locked states satisfy a simple map of the circle, the Arnold map.Variants of that model have been used to study bichromatic excitation in a radio-frequency trap /1/ and neuromimetic features in a simple model of neuronal jonction /3/.

The generic exemple of a resonant coherent structure is the phase locking of oscillators : resonance occurs whenever the frequency $p\omega_1$ $(p \in N)$ of one oscillator approaches some harmonic $q\omega_2$ $(q \in N)$ of another, so that we expect mode locking in frequency ranges around rational frequency ratios $\Omega = \omega_2/\omega_1 = p/q$. Synchronization of resonances of two coupled pendula has been observed by Huyghens in the 17th century ; mode locking of an oscillator to an external periodic driving signal has been described by Van der Pol in 1927.

The modelocking range, the stability of resonances and the transition to chaotic states at high amplitude are modeled by the Arnold nonlinear phase mapping, which is essentially the cross section of the continuous solution at each period Ω. One primitive task to perform is to derive such a map from the starting continuous dynamical model. The second task consists in measuring the stability of windings on the circle, i.e. to compare the effective frequency ratio v to its best rational approximations p_i/q_i. Modelocked zones are thus defined from our ability to fill in the parameter space with good convergent rational approximations. These stable regions are embedded into an unlocked sea of non rational zones where upon roundoff errors dominate /2/.

At low amplitude, resonance results from a single rational frequency ratio. At higher amplitude nonlinear couplings are responsable for the overlap of resonances and this leads either to bifurcation towards multistable hysteretic modelocked states or to chaotic behavior. In our recent publications we have introduced two methods of discriminating the

bifurcation sets. One is based on a scaling law relating the frequency ratio v at two closed points in parameter space. The other one is based on a scaling law relating v at two adjacent time samples. At a bifurcation point of the modelocked zone the winding number v and the scaling exponents undergo a discontinuity. In other words the scaling law is some kind of microscope (an equivalence relation) acting discontinously on the modelocked zone and it is used to tile this zone into connected subsets (with similar distribution of bifurcation levels).

Until now this strategy has been used to model driven ordinary differential equations. In this paper we will extend the method to an assembly of N oscillators slaved to a common external source.

The synchronism of those oscillators may be analysed either from the winding number v^p at one site or from the wavenumber $k_{pq} = (\Phi_n{}^p - \Phi_n{}^q)/(2\pi n)$, $n \to \infty$ connecting two distant sites in the assembly. In both cases the scaling law microscope helps to classify the parameter space into classes of equivalent degree of stability.

1. Capacitive coupling of Josephson transmission lines :

Single long Josephson junctions emitting radiation in the millimeter wave region are of great practical interest /4/. The quantum phase in a long Josephson junction driven by a microwave source of frequency ω obeys the following phase equation :

$$\Phi_{\tau\tau} - \Phi_{xx} + (2/Q)\,\Phi_{\tau} + \sin\Phi = i_1 \cos(\Omega_o^{-1}\,\tau) \qquad (1)$$

where $\omega_o = (2e\,I_c/\hbar C)^{1/2}$, $\Omega_o = \omega_o/\omega$, $i_1 = I_1/I_c$ and $Q = 2\,(2e\,I_c\,R^2\,C/\hbar)^{1/2}$, e is the electron charge, I_c the critical current and R and C are the shunt resistance and capacitance. In the limit of overdamped oscillations ($C = O$) of a short Josephson junction (Φxx term lacking in (1)), the dynamics is described by Arnold generic model :

$$\Phi_{n+1} = \Phi_n + 2\pi\Omega - c\,\sin\Phi_n \,(mod\,1) \qquad (2)$$

with $\Omega = \Omega_o\, i_1$ *and* $c = 2\pi\Omega_o$. The Poincaré section (2) has been derived by assuming synchronized solutions of the form $\Phi = \Delta\Phi/\Delta\tau = (\Phi_{n+1} - \Phi_n)/\Delta\tau = C^{te}$, taking $\Delta\tau = 2\pi\Omega_o$, the period of the external source and studying the properties of the map at large time $\tau = 2\pi\Omega_o n$, $n \to \infty$.

A well known result in the study of Arnold map is the existence of steps in the perturbed frequency ratio $\nu = (\Phi_n - \Phi_o)/2\pi n$ versus the bare frequency ratio Ω /1-3/.

For the Josephson junction, we expect steps at the driving voltages $V_r = R\, I_1 = r\,(\hbar\,\omega/2e)$, $r \in Q$. Fundamental resonances $r = n \in \mathbb{N}$ have in fact been used to achieve a voltage standard with a relative uncertainty of 10^{-7}. Moreover subharmonic steps and transition to chaos have been observed as well /5/.

The same discretization technique transforms equ. (1) into :

$$\begin{aligned}&(\Phi^p_{n+1} - 2\Phi^p_n + \Phi^p_{n-1})/(\Delta\tau)^2 + (2/Q)(\Phi^p_n - \Phi^p_{n-1})/\Delta\tau + \sin(\Phi^p_n) \\ &= (\Phi^{p+1}_n - 2\Phi^p_n + \Phi^{p-1}_n)/(\Delta x)^2 + i_1 \cos(\Omega_o^{-1}\tau)\end{aligned} \tag{3}$$

The study of the two dimensional mapping (3) can be performed by using the wavenumber parameter $K = \Delta\tau/\Delta x$ in addition to the frequency ratio $\Omega = 2\pi\,\Omega_o{}^2\, i_1$. Equ. (3) transforms into :

$$\begin{aligned}\psi^p_{n+1} &= \psi^p_n(1-2\varepsilon) + 2\pi\Omega - c\sin\Phi^p_n + K^2\left(\Phi^{p-1}_n - 2\Phi^p_n + \Phi^{p-1}_n\right) \\ \Phi^p_{n+1} &= \Phi^p_n + \psi^p_{n+1}\end{aligned} \tag{4}$$

with $\varepsilon = 2\pi\,\Omega_o/Q$, $\Omega = 2\pi\,\Omega_o{}^2\, i_1$, $C = 4\pi^2\Omega_o{}^2$ and where $\psi_n = \Phi_n - \Phi_{n-1}$ is the phase jump.

In the non ***propagative limit*** ($K = 0$) (this corresponds to the capacitive coupling of a short Josephson junction), mapping (4) is the dissipative form ($\varepsilon \neq 0$) of the so-called "standard" mapping. Standard mapping ($K = \varepsilon = 0$) was initially introduced as a model of a charged partide in a cyclotron. It accounts well for the hierarchical structure of resonances in energy preserving (hamiltonien) systems and as such it has been well studied /6/.

The structure of orbits versus initial conditions is shown in Fig. 1. The strengh parameter c = 0.972 corresponds to the critical regime where the last invariant torus at $v = (\sqrt{5} - 1)/2$ is lost. The plot has been obtained from the scaling law relating the winding number v at two closed points in parameter space.

The line $\psi = 0$ corresponds to the devil's staircase of mapping (2) and all steps except the fondamental ones at 0/1, 1/1... have been destabilized at the critical case. As an important result of this analysis we mention that the winding number is mostly discontinuous all over the modelocking zone (black regions). The preserved connected regions manifest nice circle packing structure. A second noticable feature is the unstability of second order islands (or subharmonics p/q, $q > 1$) with respect to dissipative perturbations ($\varepsilon \neq 0$).

In the undriven case ($i_1 = 0$), the propagative equ. (1) is the dissipative form of the so-called Sine Gordon equation. Simple solutions of Sine Gordon equation are in form of localized excitations (solitons or fluxons) propagating undistorted and unattenuated.

The single fluxon solution has the well known form :

$$\Phi_{(x,\tau)} = 4 \tan^{-1} \left\{ exp\left[-(x - u\tau - x_0)/(1 - u^2)^{1/2} \right] \right\} \qquad (5)$$

in which x_0 is the initial position of the fluxon and u its constant velocity of propagation.

Phase locking of fluxons in the long Josephson tunnel junction of length L can be analyzed with help of the discretization technique just presented. As an exemple we considered an array of 19 oscillators with open end conditions $\Phi(0,t) = \Phi(L,t) = 0$ and the following values of parameters : $c = 0.85$; $\varepsilon = 0.1$ and $K^2 = 0.5$. All oscillators were initially set in the same initial condition $\phi_{0\,p} = \phi_0$, $\psi_{0\,p} = \psi_0$, $\forall p = 2,17$ except end oscillators. Fluxons were found modelocked mainly in the central white channel shown in fig. 2. The bifurcation set of the solitons is shown in black. At the bifurcation set the localized wave is modelocked to the external signal but its frequency is

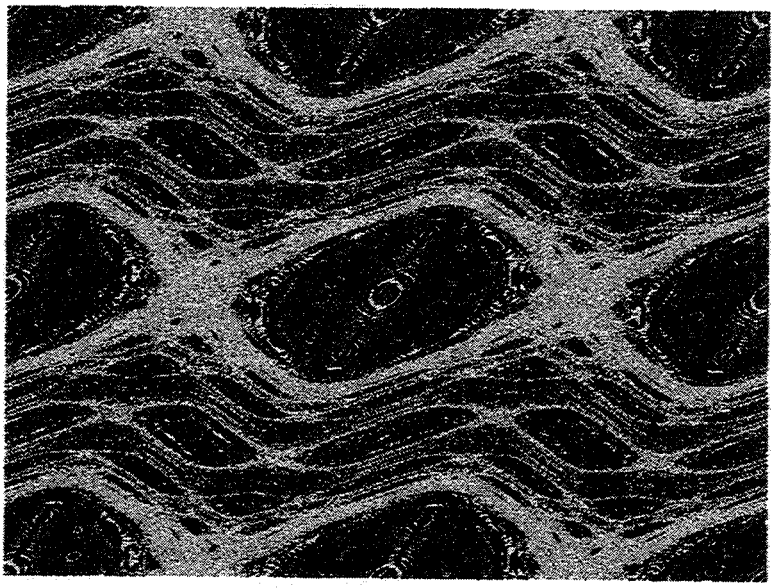


Fig.1. *Phase space stability for the standard mapping (4)*
($\varepsilon = 0$, $c = 0.972$, one oscillator)
X-axis : $-1 < \phi_o < 1$; *Y-axis* : $-1 < \Psi_o < 1$

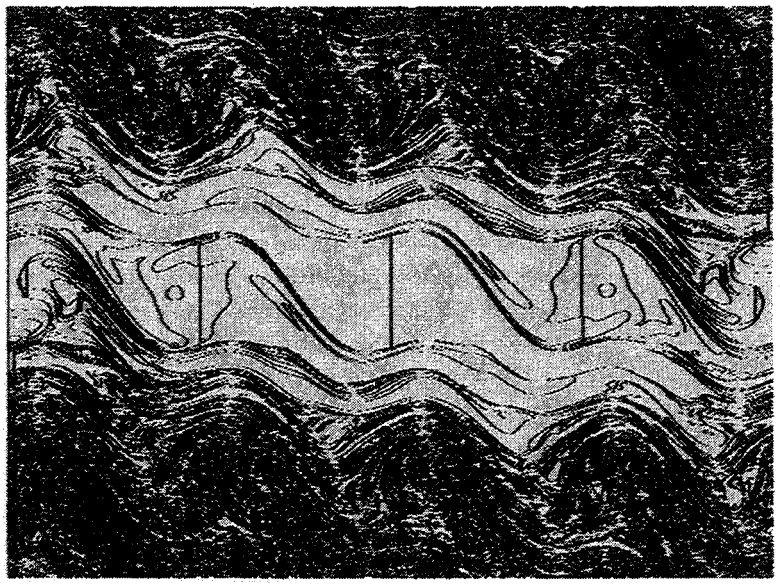


Fig.2. *Phase stability for the standard mapping (4)*
($\varepsilon = 0.1$; $C = 0.85$; $K^2 = 0.5$; 19 oscillators and open end conditions)
X-axis : $-2 < \phi_0 < 2$ *; Y-axis :* $-2 < \Psi_0 < 2$

discontinuous all over the set and is structural stability has been lost. The bifurcation set outside the central region was found to be connected so that diffusion is long range in the region. Infinitally many islands of stability for the fluxons were found embedded into the discoutinuous sea.

2. Dynamics of biological neurons :

Brain information is conveyed by transient electrical signals all produced by one basic mechanism : brief alterations in the electrical properties of the resting potential membrane. The membrane potential arises from the unequal distribution of ions on either side of the living cell membrane, together with the relatively high ionic selectivity of the membrane. The neuronal membrane is a mosaic of lipids and proteins about 10 nm thick ; its surface is formed by a double layer of lipids ; embedded within this continuous layer are various proteins. The lipids within the membrane are very hydrophobic ; in contrast, the ions of extracellular space (Na^+, $C\ell^-$) and intracellular space (K^+) are hypodrophilic. The interior of the lipid bilayer is a medium of low dielectric constant and therefore represents a high barrier for the passage of small ions such as Na^+ and K^+. An energy several tens the thermal energy kT would be necessary.to cross the barrier.

Many basic mechanisms were studied by Läuger /7/ as candidates for the ions transport and then for the neuron activity. Since small ions are hydropholic, transport may be ensured through alkali-ion complexes, with the exterior of the complexe being hydrophobic. As a result the total interaction energy of the ion with the membrane contains two terms, one arising from the electrical barrier just mentionned, and the other one originating from the hydrophobic interaction. The superposition of both leads to a double potential well with deep energy minima at the membrane interface. Hydrophobic ions spend most of the time in these energy wells and occasionnaly may channel across the central barrier to the other side. These ideas may be given a vivid illustration with help of a driven Düffing equation ; it is obtained by replacing pendulum equation (1) by a double well potential. Typical chaotic solutions are in form of randomly up and down states. Experimentally the bistable current flow through a single voltage gatad channel can be recorded using the patch champ technique /8/ and mimic the chaotic solution of driven Düffing equation.

In some ion transport systems, like the sodium channel, the transport is by far higher than that involved by mobile carriers. Läuger developed a model of protein-mediated transport in which the channel is represented by a sequence of binding sites which are separated by activation energy barriers. In the typical case, the multistable potential is approximately periodic as that of the driven pendulum equation (1).

They are at least two distinctive features in signal propagation and transmission by biological neurons. One is the propagation of action potentials which are very similar to solitons and propagates undistorted and unattenuated along the axon. The other is that transmission of information through the synapse occurs through minipotentials or quanta. Although quantal transmission has been demonstrated at all chemical synapses so far examined, it is very difficult to reconciliate it with the standard (Hodgkin and Huxley) model of propagation. Soliton propagation and minipotentials are two distinctive features which arise naturally in our model. A typical analysis of soliton stability in a synaptic type potential is shown on Fig.3.

3. Stability of nonlinear standing waves in a quartz crystal resonator:

Shear vibrations in an AT-cut quartz crystal resonator with two parallel stress-free boundaries located at $x = 0$ and h are described by the following driven Düffing equation /9/.

$$a_{tt} + \mu a_t + \omega_o^2 a + \delta a^3 = a_1 \cos \omega t \qquad (6)$$

with $\delta = 3\Delta\ k_o^4/4\rho$, $\mu = k_o^2\ \eta$, $k_o = \omega_o/V_T$, $V_T = (C_{66}/\rho)^{1/2}$, $\eta = h\ V_T/(NQ\pi)$ and $\Delta = (1/2)\ (C_{22} + C_{266} + C_{6666})$, the c's are elastic constants, ρ is the mass density, Q is the quality factor and N is the number of half periods in the cavity.

Nonlinear standing waves can be described by adding the propagative term $- V_T^2\ a_{xx}$ to the l.h.s of equ. (1). A typical stability diagramm in phase space a_n, $b_n = a_n - a_{n-1}$ is shown in Fig. 4.

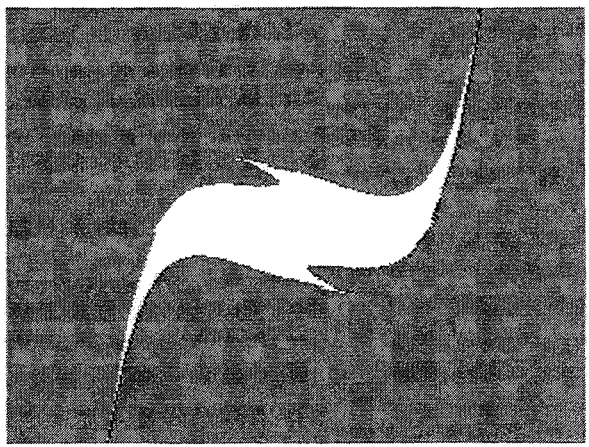


Fig.3. *Modelocked zones of a model synapse. Outside region has its basin of attraction at infinity*

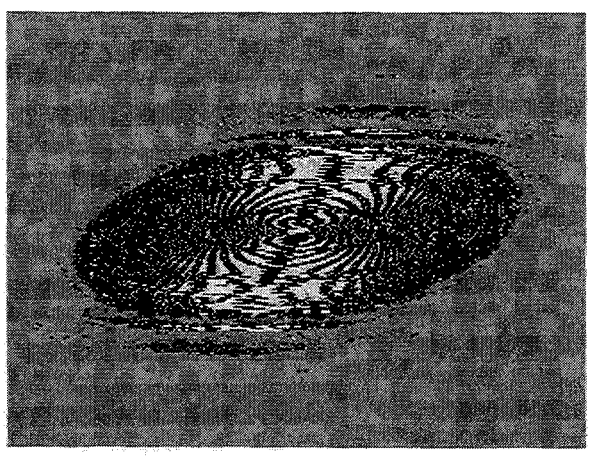


Fig.4. *Modelocked zones of a typical high Q acoustic resonator. Outside region has its basin of attraction at infinity*

/1/. M. Planat. Proc. of the European Frequency and Time Forum held in Besançon (1991, p. 372), Noordwijk (1992, p. 333) and Neuchatel (1993, p. 349). IEEE Transactions on Circuits and Systems (submitted).

/2/. M. Planat and P. Koch. Fractals, in Fractals in Natural Sciences, edited by M. Matsushita, M. Schlesinger and T. Vicsek, Vol 1, n° 3 (1993). In Nonlinear Coherent Structures in Physics and Biology, edited by F.G. Mertens, Plenum Press (1993).

/3/. M. Planat. Conference on Neural Modelling. Lyon. June 20-24 (1994).

/4/. M. Cirillo and coauthors. J. Appl. Phys. Vol 66, p. 1772 (1989).

/5/. K.K. Likharev. "Dynamics of Josephson junctions and circuits" Gordon and Breach Science. Publishers, 1986.

/6/. A.J. Lichtenberg and M.A. Liberman. Regular and Chaotic Dynamics. Second Edition. Springer Verlag. 1992.

/7/. P. Läuger, in "Membranes and intercellular communication", Les Houches, North Holland (1981), p.433.

/8/. C. Hammond and D. Tritsch, "Neurobiologie Cellulaire", Doin, (1990).

/9/. H.F. Tiersten. J. Acoust, Soc. Am. 59, 866 (1976).

INFLUENCE OF PART STABILITY ON AGING OF QUARTZ CRYSTALS

JP. MICHEL*, M. ZAFRANY*, V. CANDELIER*, J. LAMBOLEY*, L. XAVIER*

F.A. KUHNAST**, M. NASSIK**, J.M. FIORANI**

* THOMSON/CEPE
44, Avenue de la Glacière - BP 165
95105 ARGENTEUIL - FRANCE

** Laboratoire Thermodynamique Chimique Métallurgique (LTCM)
Université de Nancy 1 - BP 239
54506 Vandoeuvre Nancy

1. SUMMARY

The aging of quartz crystals is characterized by a frequency drift due to several physical phenomenons acting either in serie or in parallel. The main models involved such as stress relaxation, oxydation and physisorption are overviewed.

The subject of this paper has been to evaluate the individual instability of each constitutive part of a quartz crystal resonator.

Adhesive polymer used in the low aging crystal manufacturing are characterized by their mechanical and electrical properties under different annealing conditions.
The influence of this conditions are reviewed.

An important part of the paper is dedicated to an approach of the aging of the resonator induced by relaxation under stress of the mounts of the resonators. An extensive experiment has been performed on various materials for mounts.

The resulting relaxation behaviour under real temperature profile has been modelized on a thermodynamical basis in order to simulate the aging of the resonator. Preferential materials have been evidenced and an experimental validation has been performed.

2. INTRODUCTION

Aging of crystal for high precision oscillators has been for long the subject of continuous improvements. On the manufacturer point of view these improvements must have two aims :

firstly, the oscillators must show a low aging rate when they are delivered, in accordance to the customer requirements;

secondly, this required aging rate must be achieved in the shortest possible time for productivity purpose as well as for the target of short lead time delivery.

The to-day need for the telecommunication market is in the range $2\ 10^{-9}$ to $5\ 10^{-10}$ per month for 5 to 10 MHz. The main problem for that range of performance is not exactly feasibility but to reach it in a short time (between two and four weeks). For that reason the need appeared of a separate study of the possible causes of aging and of a better understanding of their own kinetics in order to improve them. The various causes of aging have been listed previously by /1/. They are :

1/ mass loading of the plate leading to a frequency decrease; the possible mechanisms are a/ chimisorption b/ oxydation c/ condensation; they act by the presence of residual impurities in the normally vacuum atmosphere of the package resulting from uncomplete outgassing before sealing especially for the bonding conductive glue.

2/ diffusion of the metal electrodes through the quartz, the mechanical constants of the quartz being superficially modified leading to a frequency change.

3/ stress relaxation in the resonator assembly; the unavoidable stresses occuring during assembly will vary with time leading to an evolutive stress field applied to the plate, then to a frequency shift. The possible loci for stress relaxation are : a/ package b/ mounts c/ bonding glue d/ electrodes c/ quartz.

In order to improve the manufacturing process of the resonators, these possible causes of aging must be classified regarding their importance compared to the whole aging target. The classification is not the same whether it is focused on the beginning of aging or on the ultimate achievable aging rate. This paper is more concerned by the short lead time for oscillator delivery and so by the beginning of the aging curve. One attempt to make this classification has been performed by /2/ in a backward manner : departing from an

actual aging curve they try to find the best fit with the characteristical time law for each of the possible mechanisms. So they can determine what was the major physical mechanism of aging at each time of the aging period and determine its amplitude. This method is efficient. However it only gives an indication of the involved mechanisms but does not give information about its locus: for example all the possible causes of stress relaxation will lead to the same shape of aging curve. In addition this method is no longer applicable when there are simultaneously two or more mechanisms of the same order of magnitude.

The approach which is exposed in this paper has been to study separately different possible causes of aging in order to determine their own amplitude. Focus has been given up to now on stress relaxation.

3. STRESS RELAXATION IN THE CONSTITUTIVE PARTS OF A RESONATOR

As stated above the possible loci of stress relaxation are : a/ package b/ mounts c/ bonding glue d/ electrodes e/ quartz.

Quartz is not concerned with this paper: stress in quartz is the result of the various inclusions and dislocations in it. There is no reason why the relaxation of this stress - if there is - should be more important at the beginning of the aging curve of the resonator. Impurities may play a significant role in aging but it is not a stress relaxation mechanism. Machining of the quartz may however be a non negligible parameter, but this is not developed in this paper.

The stress mechanism in the electrodes has been developed previously /3/. This paper will focus on package, mounts, bonding glue. The experiments have been performed with thermodynamical methods : differential scanning calorimetry (DSC) and thermomechanical analysis (TMA) /4/.

4. PACKAGE

The package of the resonator is a metallic can for cold welding. The holder is of metal- glass type with kovar feedthroughs. The cover is made of copper. During welding a strong stress is applied to the package. It may cause a remaining stress of the basis. This stress is transmitted to the crystal plate through the mounts in such a manner that any resulting deformation of the package may lead to a frequency shift. Glass and cover have been examined by DSC /5/. The curve 1a shows the variation of the specific heat ΔCp between a first heating and a second heating of the glass of the holder. Should it be no

changes resulting from the first heating the second ΔCp curve would have been the same and the variation shown would have been flat. The curve 1b shows the same variation between the second and the third heating: then the curve becomes flat. This demonstrates the interest of a proper stabilization of the glass holder in order to evacuate irreversible phenomenon. The same experiment has been performed on the copper cover. The curve 2a shows the first heating : an exothermic peak is observed; it is attributed to a phenomenon of recristallisation of the cold wrought alloys resulting from forming and welding. The curve 2b shows the second heating : the unstability has been evacuated. The temperature of the transition (360°C) is not so far from the using and processing temperature range where it may be cause of a slow relaxation along time.

5. BONDING GLUE

Regarding aging coming from stress relaxation, the bonding glue is a major centre of interest because it is the vector of stress transmission to the plate. If it shows a plastical behaviour, the permanent strain resulting on the glue spot will be cause of a variation of stress applied to the plate. In order to evidence this plastical behaviour a TMA analysis has been performed /5/ : the length variation of a glue sample at a constant temperature of 100°C has been observed over one day. The glue had been stabilized with a suitable preannealing. The curve 3a shows the result.

The behaviour of the glue without preannealing has not been figured: its whole extend is approximately 8%. It can be seen that a high temperature preannealing around 300°C improves the stability of the glue; however it remains evolutive. A higher temperature of 350°C leads to a degraded performance due to the glue damaging. This inconvenience may be overcome by the choice of a proper glue; the curve 3b shows the kind of result which may be observed: a preannealing temperature between 300 and 400°C eliminates any length variation.

6. MOUNTS

It can be established that mounts are always in stress conditions. It results from

1/ initial bending for shape forming.
2/ clamping of the plate.
3/ thermal stress due to the differential extension between glue annealing temperature and working temperature.

4/ strains in the package which are transmitted to the plate through the mounts.

So a special emphasis must be given to the study of the stress relaxation in the mounts. Two possible ways of improvement could be thought of :

1/ define the conditions (choice of material and manufacturing conditions) to minimize the stress relaxation rate.

2/ define the conditions (choice of material, manufacturing conditions and annealing or stabilization steps) to maximize the stress relaxation in order to have it completed in a short time, typically a few weeks.

In order to determine these conditions, the approach has been to study the stress relaxation of metallic ribbons away from the resonator by a thermomechanical analysis and then to deduce the corresponding aging behaviour of the resonator in its working environmental conditions.

6.1. Thermomechanical analysis

In order to improve the crystal stability, two types of materials with favorable characteristics have been tested :

- Non magnetic copper based alloys hardened either by structural hardening or by spinodal decomposition (Table 1).
- Iron based alloys hardened either by cold work (Inox 18/8, Dilver P and Invar) or structural hardening (Phynox)(Table 1).

6.1.1. Alloy characterization by creep measurements

The different alloys have been submitted to three point flexure bending tests at temperatures ranging from 200°C to 400°C for 8 hours. The applied loads were chosen between 0.2 and 1.2N. Creep curves were fitted by logarithm creep modelisation and three parameters A, B and n were deduced :

A : characterizes the creep rate at constant temperature and load, according to the equation $d\varepsilon/dt = A/t$ (ε is the deformation and t the time); the smaller A is, the better relaxation behavior is.

B and n : give the load effect according to the equation $A = BF^n$ (F is the applied load).

For instance, figure 4 shows curves of A versus load for both types of materials. B and n values are reported in table 1 /6/.

Table 1 : Values of the exponent n and the constant B at 200°C

	Chemical composition (Wt %)	n	B ($\mu m/N^n$)
Nickel	Ni pur	1,00	0,72
Copper alloy 1		0,50	0,74
Copper alloy 2		1,12	2,42
Copper alloy 3		0,56	1,80
Inox 18/8	Fe Cr 18 Ni8	0,72	0,88
Invar	Fe Ni36	0,90	0,72
Phynox	Fe Co40 Cr20 Ni16Mo7	0,45	0,55
Dilver P.	Fe Ni29 Co17	0,75	0,13

These results underline that iron based alloys have the best creep behavior at 200°C.

6.1.2. Stress relaxation analysis

Three point flexure experiments were carried out on the selected materials Dilver P, Inox 18/8 and Phynox in order to observe stress relaxation : a given flexure deformation Y is applied to the sample during an annealing performed at the temperature T for a time t, then the final remaining deformation Y (T, t) is measured at room temperature after stress removal /7/. The degree of transformation Pr = Y (T,t)/Y was plotted versus log t.

The theoretical analysis of stress relaxation is based ont the DNRL model (Distributions of non linear relaxation) /7, 8/. Which allows us to determine the kinetics parameters by fitting the experimental and theoretical curves (fig. 5). Then one can predict the aging behaviour of the alloys under operating conditions.

At high temperature (550°C), the faster relaxation is obtained for the Dilver P and Phynox alloys. For lower temperature (< 250°C), a complete relaxation cannot be observed /6/.

So there was no hope to define a thermal treatment in order to relax the internal stresses of the resonator before its use on oscillator.

The only possible way to improve aging remains to reduce the relaxation rate at working temperature.

6.2. Aging simulation

Let us suppose that the mounts apply a force F on the diameter of the plate. F is the resulting value of all possible causes of stress listed above. The mounts are clamped to the package output pins at one end. At the other end, F can also be considered as the result of a displacement Y_0 of the mount from its free equilibrium position.

$$F = \beta Y_0$$

where β is the stiffness of the mount depending on its geometry and Young's modulus. Regarding the model for stress relaxation, the time dependancy of the displacement Y will be :

$$dY/dt = - BF^n/t$$

where B, n have been defined previously. The force F will vary in order to compensate for the variation of Y :

$$dF/dt = - \beta BF^n/t \qquad (1)$$

When F is applied on the diameter of the plate the frequency f of the resonator varies from its f_0 free value according to the following formula /9/ :

$$\frac{\Delta f}{f_o} = \frac{KF}{pD/f_o} \qquad (2)$$

where D is the diameter of the plate, p the anharmonic range, K a constant depending on the cut angle of the plate and the orientation of its bonding points. The aging rate V of the resonator is defined as :

$$V = \frac{d}{dt}\left(\frac{\Delta f}{f_o}\right) \qquad (3)$$

integrating (1) and using (2) and (3), it comes :

$$V = \frac{-\beta B}{t}\,\frac{K}{pD/f_o}\,\left(F_o^{1-n} + (n-1)\,\beta B\,\mathrm{Log}\,\frac{t}{t_o}\right)^{\frac{n}{1-n}}$$

This formula has been applied to a 6 MHz resonator working on its fifth harmonic with a 14 mm diameter plate for the various mount materials studied above by the thermomechanical analysis at a working temperature of 75°C. The results of this simulation are given on the curve 6. It can be seen that all material lead in one month to an aging rate better than $2\ 10^{-9}$ per month. The optimization of the choice of the material for mount is only efficient in the range of 10^{-10} per month. These conclusions rely on an

initial choice of the value of F, which highly depends on the thermomechanical history of the resonator. It has not been possible to measure it in the real condition of a resonator beginning its aging time. However we have selected a value of 1.5 N in this simulation. A lower initial value will lead to lower aging rates.

7. CONCLUSION

Stress relaxation is an important cause of aging in the crystal resonators. In order to improve aging the various relaxation mechanisms have been studied separately. Such approaches as DSC and TMA have been used to evaluate the effects of this mechanisms. It has been shown that they could come either from the package (cover, glass from the holder) or from the bonding conductive glue or from the mounts. A particular emphasis has been given to that later subject: various possible metallic ribbon (nickel, iron alloys, copper alloys) have extensively been studied by thermodynamical means in order to first determine a creep model appliable over a wide range of temperature and of applied load and then to give a theoretical analysis of stress relaxation by using the DNRL model. Using these results a simulation of the aging of the crystal resonator has been determined in the usual operating temperature and stress conditions. Variable aging rates between 0.5 and 2 10^{-9} after one month have been shown according to the choice of the material for the mounts.

When applying on high stability resonators manufacturing process providing very low initial mounting stresses and material with low thermal relaxation, (including stabilization of initial parts) one can achieve ageing better than 5.10^{-10}/month after a few weeks (typically 6 to 10) of operation, either at 5 MHz or at 10 MHz.

Bibliography

/1/ R.L. Filler, J.A.Kosinski, V.J.Rosati, J.R.Vig, 38th Annual Frequency Symposium (1984) 225 - 231

/2/ M.R.Miljkovic, G.L.Trifunovic, V.J.Brajovic, 42nd Annual Frequency Control Symposium (1988) 404 - 411

/3/ M.Zafrany, F.Meyer, 7th European Frequency and Time (1993) 305 - 309

/4/ J.M. Fiorani, thèse de l'Université de Nancy I, 1992

/5/ J.M. Fiorani, F.A. Kuhnast, Ch. Cunat; Rapport interne

/6/ F.A. Kuhnast, M. Nassik, J.P. Hilger, J. Hertz, J.M. Fiorani; Ch. Cunat, Rapport interne.

/7/ J.M. Fiorani, A. Aharoune, F.A. Kuhnast, Ch. Cunat, J. Non - Cryst. Solids 156 - 158 (1993) 188

/8/ Ch. Cunat, J. Non - Cryst. Solids. 131 - 133 (1991) 196

/9/ J.M. Ratajski, 20th Annual Frequency Control Symposium (1966) 33 - 49

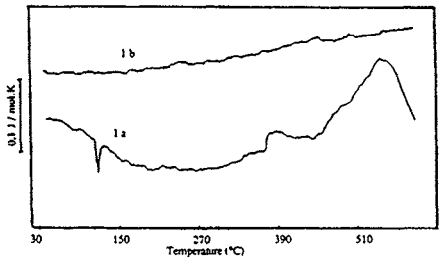


Figure 1 : Glass basis specific heat Δ Cp for successive subtracted heating runs versus temperature (dT/dt = 20°C/min)
1a : 1-2 1b : 2-3

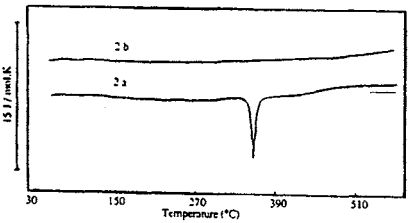


Figure 2 : Copper cover specific heat Δ Cp versus temperature (dT/dt = 20°C/min)
2a : first heating 2b : second heating

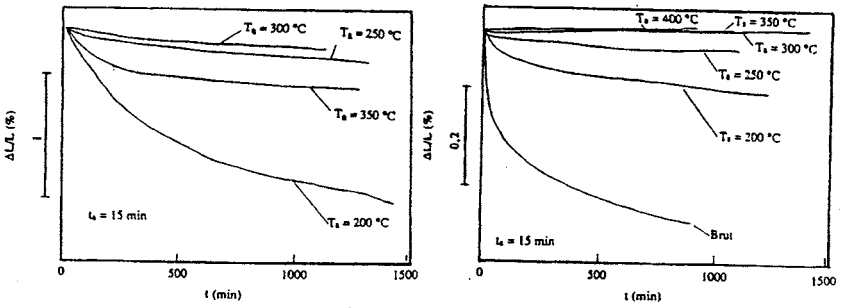


Figure 3 : Length changes ΔL/L versus time for two different bonding glues standed by at 100°C after preannealings at different temperatures Ta for ta = 15 min.

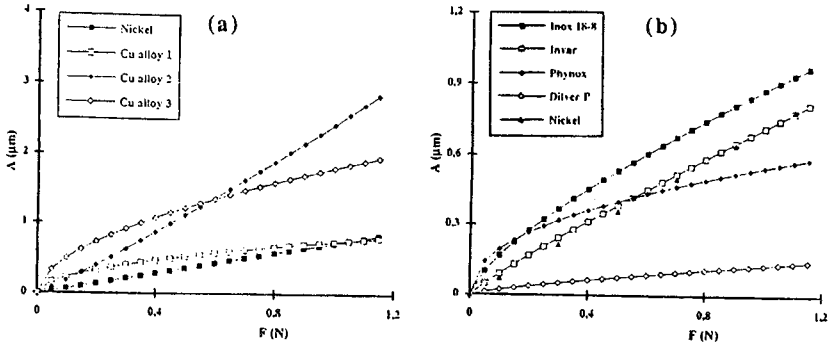


Figure 4 : The parameter A versus the applied load F at 200°C :
(a) Ni and Cu based alloys (b) Ni and Fe based alloys

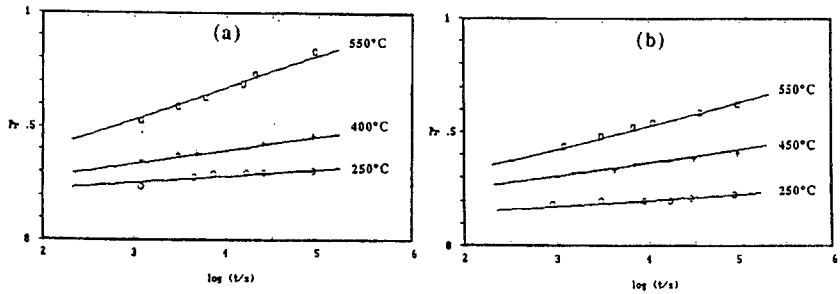


Figure 5 : Examples of fitting between the experimental and the computed Pr curves :
(a) Dilver P (b) Phynox

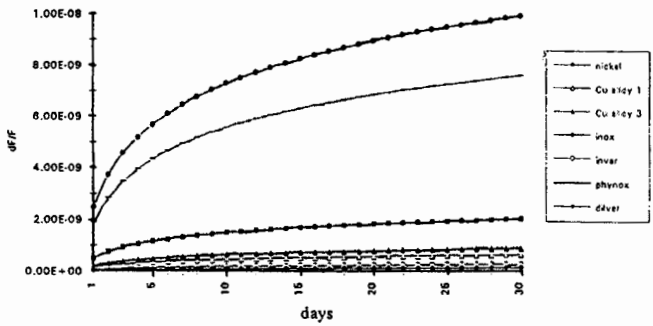


Figure 6 : Computed dF /F curves versus time (Ta = 75°C)

CRYSTAL OSCILLATORS SIMULATION SOFTWARE

A.BAILLY, P.COLLET

C.Q.E. - TEKELEC
11, rue Keller, 10150 Troyes FRANCE

ABSTRACT. A good knowledge of logic circuit used as analog amplifier allows accurate simulations of crustal oscillators. This paper deals with the behavior of CMOS and ECL circuits alone and used with crystals in the feedback loop. Various schematics are presented which allow operations at up to 70 MHz for CMOS circuits and 200 MHz for ECL ones. Selection of overtones are also discussed.
A simulation software has been specially developped. It performs dedicated computations in order to define with the needed accuracy the oscillating frequency, together with the quality factor. Additional fuction allows the optimisation of components value in order to get the best characteristics for quality factor, overtone selection, pulling linearity and so on.

1 - Scope

A simulation software has been developped in oder to suit the crystal oscillators designer needs. The aim of this software is to be directly adapted to the dedicated field of crystals, where classical requirements are often treated : High quality factors, different modes of oscillation, pulling capacity and linearity, effect of the temperature on the oscillator.

A good simulation needs accurate models for the active elements used in the oscillators. This formed a first part of the work, and was done on two typical components used in the crystal oscillators : One CMOS gate and one ECL gate. These elements are used in classical pierce oscillator or other structures as amplifiers, and have been characterized in this mode of running.

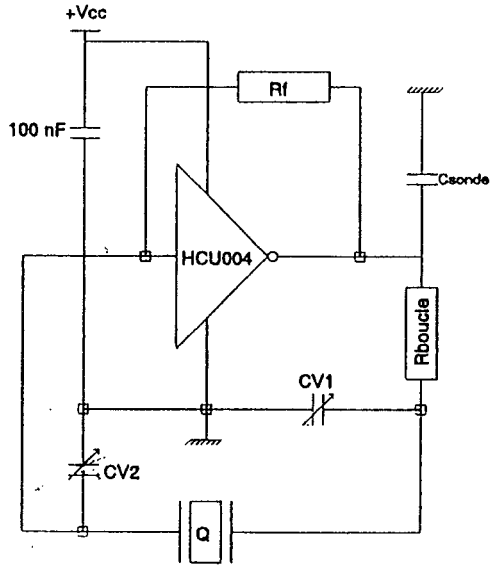
The other task deals with the software itself. It is based on a principal routine which is in charge of the analysis of the oscillator on a given range of frequency, together with the definition of the oscillation frequency. The crystal current and the quality factor of the design is then defined at this locus.

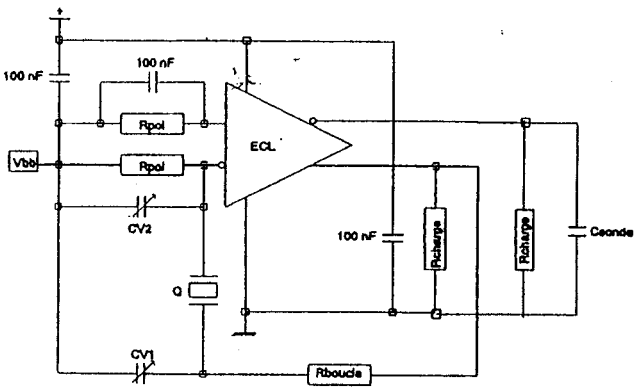
This software is now used for the internal CQE need and is constantly under improvement. Some of the foreseen functions are not yet implemented. But its heart is

developped and is satisfactory. The following paragraphs describe the results obtained during this work for the two parts defined here above.

2 - Active elements characterization

Two types of gate have been characterized : A CMOS type (74HCU04) and an ECL type (MC10H116). Theses elements are used in the following manner :



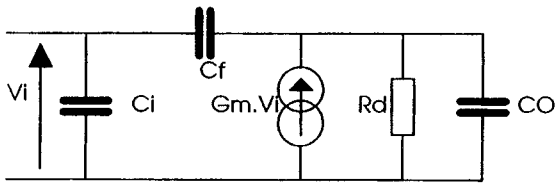


It can be seen that these types of oscillators are easily simulated in open loop condition, with a special care for impedances where the loop is considered to be open. The mode of simulation will be described later, but the needed elements are the gain of the active component, and its input and output impedances.

This has been defined for various conditions of power supply voltages, temperatures and manufacturer, except for ECL gate which is specified for a 5.2V voltage, and has only one manufacturer. These data allow worst case analysis simulation, by changing these conditions of running.

2.1 CMOS characterization

The following model has been taken for the CMOS gate :

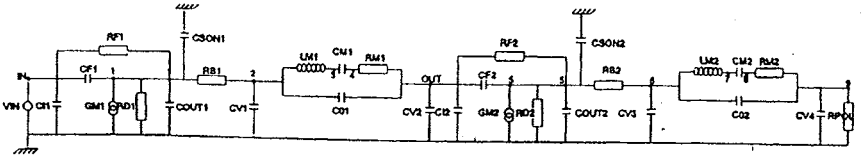


Measurements are made with a network analyser type HP4195A, and the following table has been deduced, for two types of manufacturer, two supply voltage values and at ambiant temperature :

Manufacturer	National Semiconductor		Texas Instrument	
Power supply voltage	2,5 V	5 V	2,5 V	5 V
Gm in ms	0.011	0.0245	0.009	0.018
Rd in Ω	2550	607	4125	984
Ci in pF	6	6	3,9	3,9
Cf in pF	1,7	1,7	2,65	2,65
CO in pF	9	14	14	14

A spice simulation has been made with the above values, and with the following schematics, to take into account the problem of input and output impedance where the loop is open. For this purpose, too open loops are put in serie, and the transfer

function is taken between points 1 and 6, and with crystal parameters corresponding to a frequency oscillation of 7 MHz :



The obtained phase and gain curves allow the determination of the oscillation frequency. Different simulations are made with different components values, and the corresponding frequency variations are computed. Same experiments are made on a breadboard, with the observation of these frequency variations. The comparison between the two results is given in the following table :

Supply voltage	ΔF measured	ΔF simulated	Error
2.5 V	189.6	159.5	16 %
	186.4	174.5	7 %
	56.5	38.5	32 %
	183	195	7 %
5 V	66.6	5.5	92 %
	153	128.5	16 %
	192.1	189.5	2 %

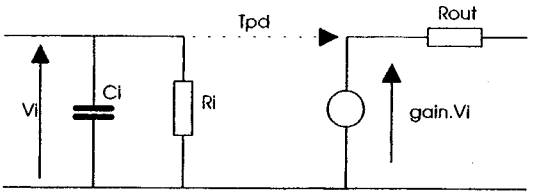
The results are not identical and the errors are rather high. The reason is that the output impedance of CMOS devices vary a lot according to the output level (Small or large signal conditions). Since the model has been built on small signal hypothesis, it is clear that the results will differ a lot if the output impedance changes between two experiments, due to input and output levels.

An another series of experiments dealt with oscillating conditions, and not oscillation frequency. The criterion was taken on the limit of oscillation for a given component : The frontier value was observed and put in the simulation software, in order to compute the overall gain of the loop in this condition. In this case, the comparison is very good (gain comprised between -1 and 0;5 dB), which shows that the small signal conditions is well modelized.

The current version of the software will not be able to foresee correctly the oscillating frequency of a CMOS oscillator, but can be used to determine if a design will oscillate or not. A later version will take into account the variation of the output impedance of such devices, versus the input level of the gate.

2.2 ECL characterization

The ECL model of a gate is simpler than the CMOS one and lead to the following schematics :



At ambiant temperature, the following values have been measured :

Ci : 3 pF
Ri : 10 kΩ
Tpd : 1,5 ns
Gain : 5
Rout : 7 Ω

No frequency cut off has been implemented, since this device exhibits open loop bandwidth in excess of 200 MHz. A propagation time of 1,5 ns has been chosen instead of a first order frequency cut off, because it corresponds better to the behavior of the gate which is working in the saturated mode.

The simulation schematics is also simpler than for the CMOS gate, since the loop is naturally open between the input impedance and the output generator in the model here above.

The same measurements than previously have been made with this circuit with the following comparison table :

Frequency	ΔF measured	ΔF simulated	Error
12.9 MHz	1795	1794	< 1%
	966	956	1%
	417	426	2.1%
38.8 MHz	197	208	5.6%
64.7 MHz	207	236	14%
	83	88	6%

These results are better than those obtained for CMOS oscillators. This proves that it is possible to have an accurate simulation, provided that the active element is correctly

modelized. This is more achievable in the case of the ECL gate. The CMOS one demands a more complex model.

3 Simulation software

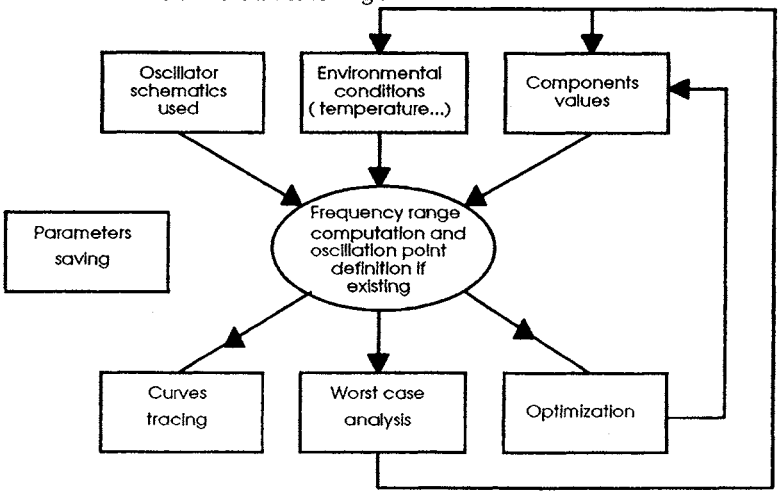
3.1 Structure

This software has been developped around a simple routine which is in charge of the simulation of a well defined circuit and components values, in a given frequency range. If a zero phase crossing occurs in this range, the oscillation frequency is also determined, together with the associated gain. If this gain is too low, the routine exits a "No enough gain" condition. In the other case, a "No phase" condition is given.

This routine uses internal complex matrix computation. The standard T and Π matrices are available for easy programming of new design schematics. All the user has to do is to transform its schematics in an assembly of such T and Π strucutres, and to put it in a succesion of calls of matrix building routines. These matrices are then multiplied in order to get the overall transfer function with input and output impedance in open loop. The system is then solved for a nul current at the output and an ipunt stimulus voltage of 1V.

In the case where an oscillation frequency is found, the crystal current and the quality factor of the design is definned. This last parameter is computed at the oscillation locus, by making varying the frequency around this point. The resulting phase variation is then divided by this small frequncy range, and is compared to the quality factor of the crystal itself. The ratio of this two quantities is given, which is classicaly called Ql/Q.

This simple routine is associated with the main functions which used it. The overall structure of this software is the following :



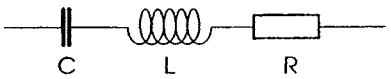
Optimization options allow the determination of components values for the better quality factor of the oscillator design, with constraints, as :
- Limited ranges for components values.
- Minimum gain at oscillation frequency.
- Rejection of an another mode of the crystal.

A worst case analysis module (Not yet developped) will give the assurance of a good behavior of the oscillator versus temperature, components tolerances and so on.

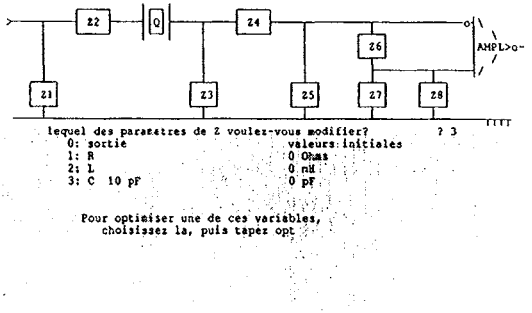
Parameters saving allow the saving and the retrieving of data (either components value, than active elements parameters, or frequency range simulation) on disks.

3.2 Functions description

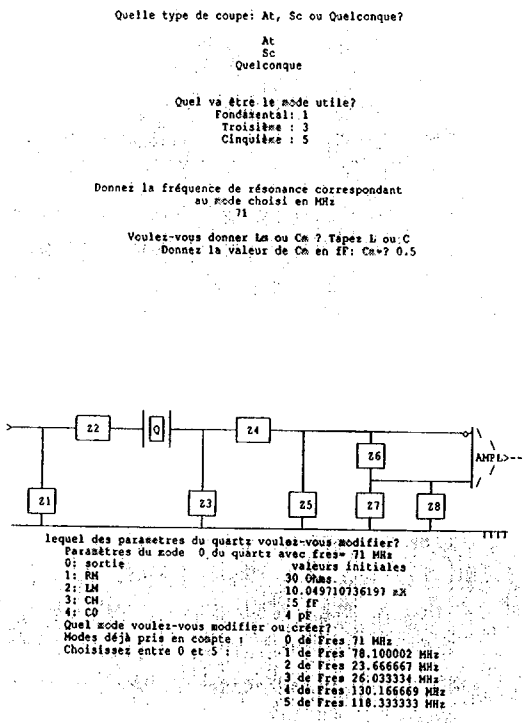
The following menu allows components values setting. Each branch of the circuit assigned with a "Zi" on the menu schematics is constituted of the impedance circuit given here below :



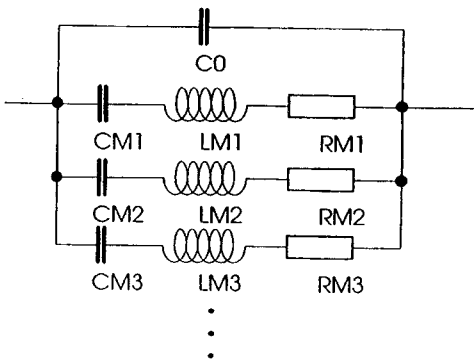
The values R, L and C are independant and are freely settable. When nul, it corresponds to a short for the corresponding capacitive, inductive or resistive element.



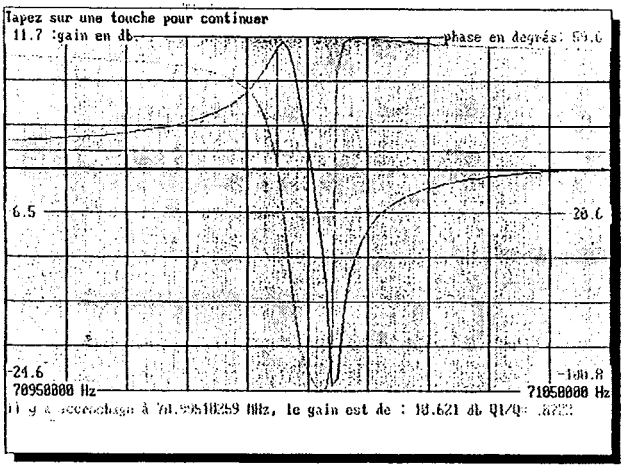
The quartz itself is defined with a special menu which allow automatic generation of associated modes, for AT and SC cuts :



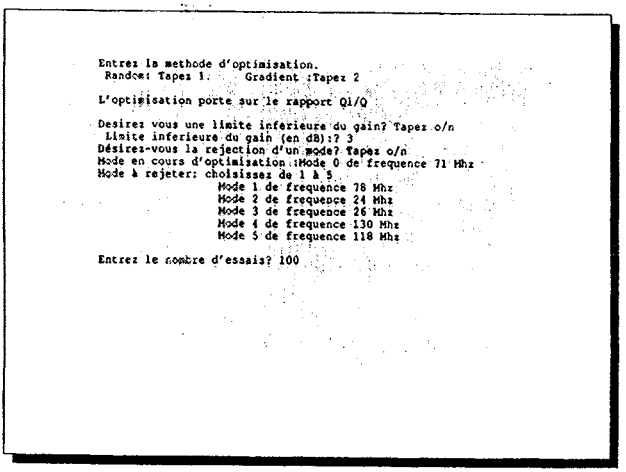
The resulting model for the quartz is the following :



Here is a result in curve mode. This is directly issued from the central routine, without further treatment :



The menu here below optimize the Ql/Q of the desing for the parameters which were designed as "Opt" in the components values setting menu. Various options are available to define a minimum oscillation gain, and the undesirable crystal mode.



THERMAL HYSTERESIS STUDY OF A MCXO

C. Boudy, P. Collet and F. Vignaud

Components Quartz Electronic TEKELEC,
11, rue Keller, BP 16
10151 Pont Sainte Marie

ABSTRACT. The hysteresis effect in quartz resonators induces a limitation of frequency stability in the MCXO (Microcomputer Compensated Crystal oscillator), between -55 and 125°C.
To surround the physical causes of this effect, we studied several materials quality and origin, to compare the physics-chemicals characteristics and the material behaviour from the point of view of thermal hysteresis.
Furthermore, we tried to verify the effect of the enclosure and annealing.

INTRODUCTION

While crystal oscillators are known for their good short and long term stability. The frequency of oscillation is also very temperature sensitive.
To stabilize the frequency temperature change, there are two methods:

-Maintain the oscillator at the same température close to the turning point of the crystal (OCXO).

- Compensate the frequency deviation (TCXO or MCXO).

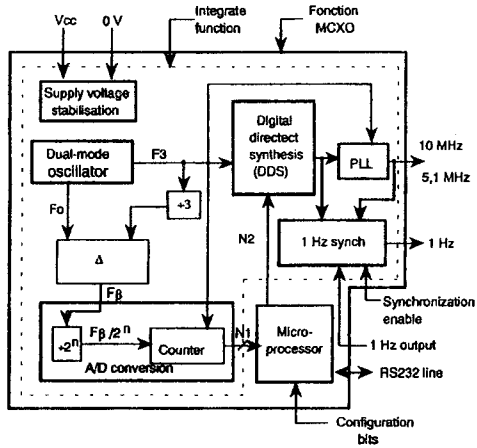
In a MCXO, the crystal oscillator operates on two modes simultaneously, the fundamental (f_{1c}) and their overtone (f_{3c}) divided per three and substracted from f_{1c} to produce a beat frequency f_β. This beat frequency is used as a temperature indicator to correct the frequency of the third overtone. This method avoids the thermal lags and temperature gradients between the sensor and the crystal resonator.

With the advent of this oscillator, thermal hysteresis has become the dominant factor limiting the stability achievable with temperature compensated oscillators.

The purpose of this paper is to study the effect of the material quality (impurities and dislocations), annealing and motional parameters on the thermal hysteresis in MCXO.

DESCRIPTION OF THE MCXO

The figure 1 gives the MCXO block diagramm developed by Tekelec/CQE (1).



- With the phase lock loop (PLL), we can transform a low frequency signal in a high frequency signal with a good stability.
- The function "Direct Digital Synthesizer" frequency (DDS) consists in obtaining an output frequency from an input frequency and a numerical value. This value depends on the micro computer correction.
- The microcomputer drives the DDS from the values generated by the digital function. This data is proportionnal to the quartz temperature.
- The dual mode oscillator gives simultaneously the fundamental and overtone modes. The beat frequency between the fundamental mode and the overtone mode divided by three is a first order linear function of the temperature.
- A mixed ASIC incorporating all the electronic functions is now realised.

An initial calibration is necessary to determine the relationship between the temperature and the beat frequency. This information is then used to programm the DDS.

ORIGIN AND DESCRIPTION OF THE THERMAL HYSTERESIS

The military specification for crystal oscillators (MIL-O-55310B) defines "hysteresis" as the maximum value of the nonrepeatability in the f versus T characteristics during a temperature cycle, i.e, as the difference between the up-cycle and the down-cycle f versus T characteristics at the temperature where that difference is maximum. Hysteresis is determined during a complete quasistatic temperature cycle between the specified temperature limits.

The frequency stability of the MCXO fluctuates between 10^{-7} and 10^{-8} depending on the resonator, for the temperature range: -55°C; +125°C.

The calibration has done from the high temperature to the low temperature and we meausured the f-T characteristics in the two directions. To know the hysteresis value, we substract the curves in keeping the same nominal frequency.

EXPERIMENTAL SCOPE

We studied four manufacturing lots, two differents enclosures and resonators with and without annealing.

The resonator manufacturing process was the same for all the units, each experiment was conducted using 10 resonators.

The calibration has done with a rate of 0.2°C/min between 126°C to -56°C.and the up and down cycles with a rate of 1°C/min between 125°C and -55°C.

COMPARISON BETWEEN MATERIAL PHYSICS CHEMICAL CHARACTERISATION AND HYSTERESIS

We measured impurity concentration, etch channel density and dislocation quantities, absorption coefficient at 3500 cm^{-1} (proportional to the OH quantity) for four different manufacturing sources.
Three of these are considered high quality material blocks (B, C, D). The fourth is standard quality block (A).
Blocks A and B have the same origin.

Chemical analysis

Block referenc	Na	Li	K	Fe	Al	Ca
A	0.79	0.16	0.41	0.74	0.96	4.12
B	0.77	0.19	0.4	1.15	2.7	3.92
C	0.51	0.03	0.12	1.3	2.72	9.79
D	0.69	0.38	0.3	0.51	2.91	3.06

Etch channel density and dislocation quantities

	A	B	C	D
Dislocations /cm^2	30 à 40	inhomogeneous	150	about 250
Etch channels /cm^2	40	50	1	> 100

The Al impurity concentration is the lowest in block A (0.96 instead of 2.7 for the others).

From the point of view of etch channels and dislocations, Block D is the worst.

Typical values of the hysteresis are given in the table 3:

The f-T characteristics for an up-cycle and a down-cycle and hysteresis curves are shown on figures 2-3 for the four different manufacturing sources.

According to the hysteresis average, the best units come from B and C lots and the worst from D.

If the difference between the hysteresis values is attributed to the material quality, then, it would suggest that the dislocations or etch channels density are important, for the hysteresis phenomena. We propose to verify using natural quartz resonators.

	B			A			C			D		
	Δf_{down} (mHz)	Δf_{up} (mHz)	hyst (mHz)	Δf_{down} (mHz)	Δf_{up} (mHz)	hyst (mHz)	Δf_{down} (mHz)	Δf_{up} (mHz)	hyst (mHz)	Δf_{down} (mHz)	Δf_{up} (mHz)	hyst (mHz)
1	910	3960	3709	HS	HS	HS	1392	3456	2756	912	2364	2496
2	638	3234	3664	448	2294	2304	HS	HS	HS	2606	14772	14873
3	452	890	1104	2538	6684	7331	1674	4724	3885	4008	9648	8670
4	736	6312	6416	2926	7070	8117	1344	2208	3196	1150	2892	3241
5	702	4720	4468	890	3124	2767	1660	4690	4376	746	9078	9597
6	932	2756	2534	2330	8456	7720	1334	5528	5269	558	5838	5548
7	806	3618	3219	1574	3872	3482				8196	6172	8006
8	446	6632	6865	528	2588	2598				776	9906	9840
9	538	1740	2100	2286	3222	4368				1676	11376	11317
10	260	3856	3894	2648	7242	6804				1812	1822	1122
Hyst average			3797			5054			3896			7471

Table 3

Figure 3. f-T characteristics and hysteresis curves for the source B.

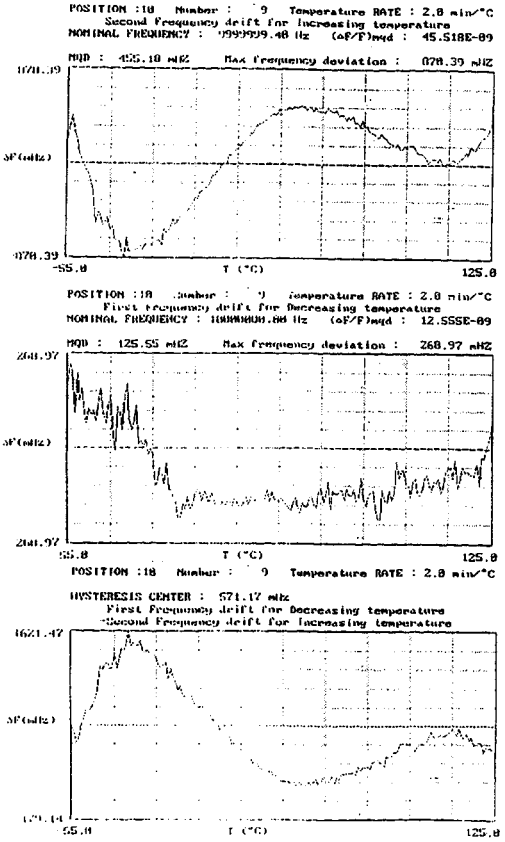


Figure 2. f-T characteristics and hysteresis curves for the source A.

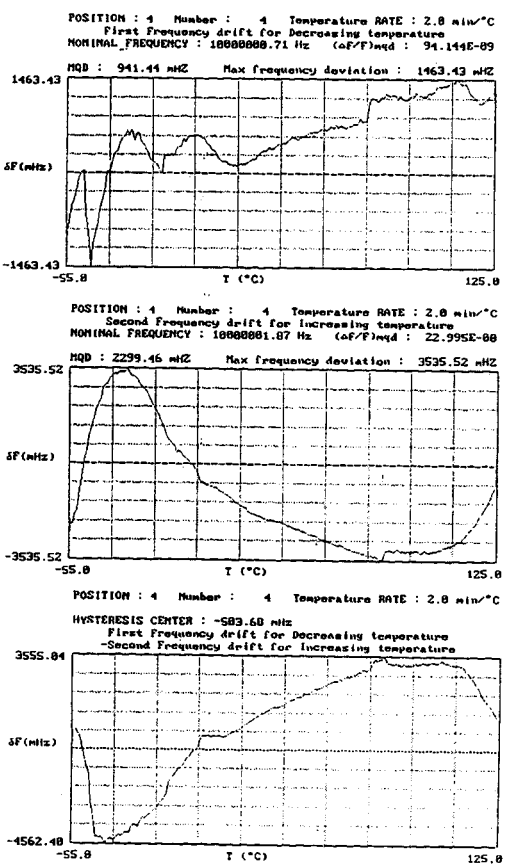


Figure 5. f-T characteristics and hysteresis curves for the source D.

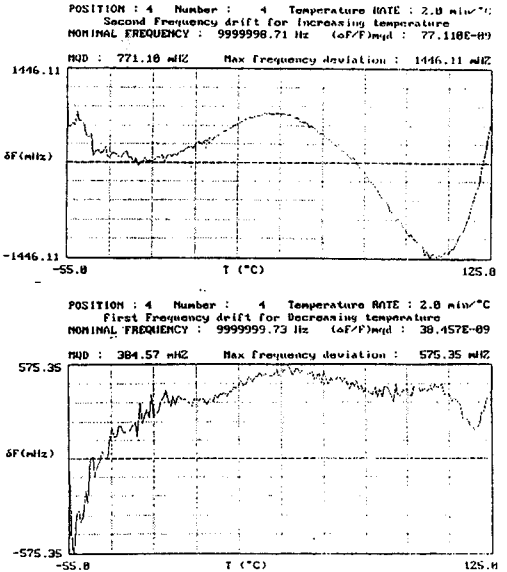
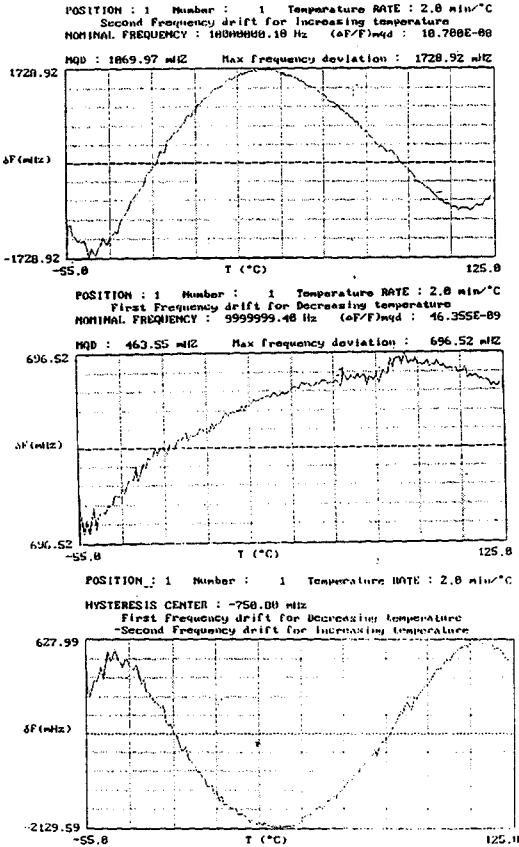


Figure 4. f-T characteristics and hysteresis curves for the source C.



COMPARISON BETWEEN THE DIFFERENT ENCLOSURES AND THE HYSTERESIS PHENOMENA

For the most part of units, we used ceramic flatpacks to obtain the smallest volume. However, to investigate the influence of the mounting system and the conducting paste, we made a batch enclosed in ceramic flatpacks. The batch was then reproccessed and mounted in metal holder.

Table 4 gives the results of the experiments:

	Ceramic flatpack			Metal HC-37/U		
n°osc	Δf_{down} (mHz)	Δf_{up} (mHz)	hyst (mHz)	Δf_{down} (mHz)	Δf_{up} (mHz)	hyst (mHz)
1	688	5516	5684			
2	916	6116	6437	906	2464	2137
3	946	5636	5653	1154	2238	2757
4	1830	3804	3688	1296	3746	2578
5	1836	5762	5612	3864	3808	3465
6	1054	8710	9380	854	4084	3987
7	802	9380	9228	900	3690	3452
8	992	3292	3335	2402	4720	2856
9	980	15184	15236	848	1864	2452
10	750	8586	8769	1616	4322	3336
Hysteresis average			6420			3002

Table 4

With the exception of the enclosure, there are major two parameter changes: the conducting paste and the mounting system.

The high sealing temperature of the ceramic flatpack does not permit the use of the same paste employed in the metal holder due to its low polymerisation temperature.

The results shown in table 4 indicate the importance of the mounting structure and the type of paste, we cannot as yet determine the dominant factor. Further experiments are under way.

INFLUENCE OF ANNEALING

Annealing of quartz prior to electrode deposition (480°C during 72 hours) does not modify the hysteresis values.
On the other hand, very fast temperature cycles between -55°C and 185°C applied on the resonators appear to have a signicant influence (Table 5).

	Normal procedure			With rapid temperature cycling		
n°osc	Δf_{down} (mHz)	Δf_{UD} (mHz)	hyst (mHz)	Δf_{down} (mHz)	Δf_{UD} (mHz)	hyst (mHz)
1	896	2596	3406	HS	HS	HS
2	1890	4450	3143	1074	2946	2535
3	970	3300	2877	716	1672	1654
4	580	5778	5837	664	2890	2847
5	1312	2496	2148	786	2272	2023
6	1768	6788	7051	1694	6534	5818
7	1402	3158	3872	810	2116	2558
8	992	2336	2924	HS	HS	HS
9	658	4622	4634	674	2356	2196
10	1732	4932	3508	944	2570	2292
Hysteresis average			3940			2740

Table 5

The hysteresis curve of one of the pieces before and after rapid temperature cycling is given in the figures 6 and 7.

The presentation of the hysteresis curve in the temperature range: -55°C; +125°C is the same, it also appears to depend on the material quality.

Rapid temperature cycle improves the hysteresis characteristics approximately by a factor of two. We can therefore deduce that the mounting system or the silver paste have a very important influence on the thermal hysteresis.

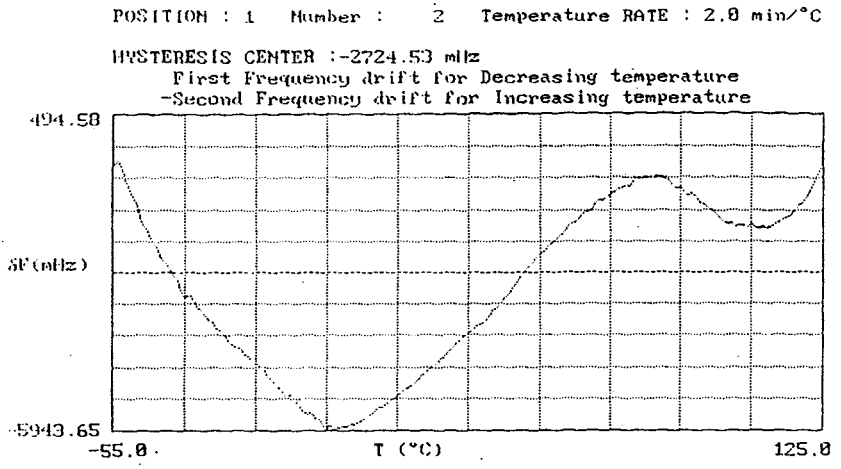


Figure 6. Hysteresis before annealing.

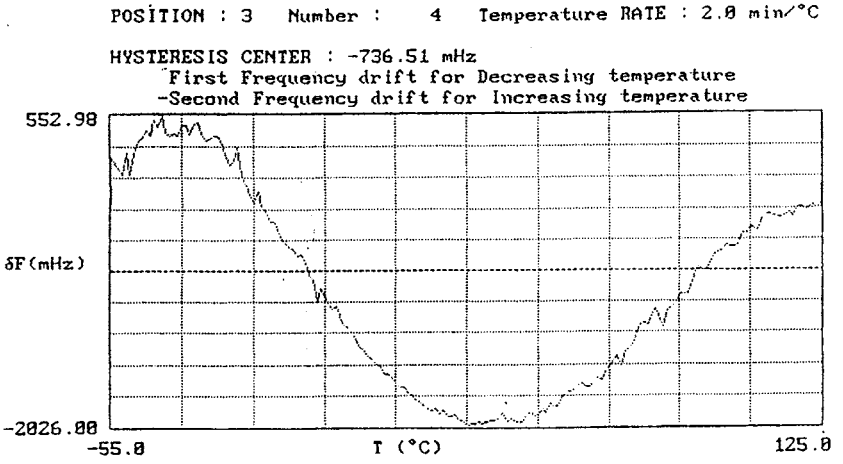


Figure 7. Hysteresis after annealing.

CONCLUSION

For R.L. Filler (2), it is quite evident that each manufacturing lot has a characteristic. Due to the over whelming influence of the mounting system and paste, it is not clear us to how much the hysteresis can be attributed to the material. Nevertheless, for resonators using the same manufacturing techniques, it is true to say that the best results are obtained using quartz with a low dislocation and etch channel density.

Further experiments will be performed to improve the mounting system in order to determine the ideal manufacturing process to decrease thermal hysteresis.

ACKNOWLEDGE

The autors thank R. Besson and J.J. Boy of LCEP and J.P. Zecchini and K. Yamniof LCCM for making the etch channel and dislocation density and chemical analysis measurements.
We acknowledge Mike Hodgkins for helpfull discussions.
The autors wish to acknoledge the support of the french army.

REFERENCES

- (1) P. COLLET, G.PUTOIS and F. VIGNAUD, Proceedings of the 6th EFTF, pp 159-163, (1992).
"A microcomputer compensated crystal oscillator MCXO".

- (2) R.L. FILLER, 44th ASFC, pp 176-184, (1990).
"Thermal hysteresis in quartz crystal resonators and oscillators".

TORSIONAL QUARTZ RESONATORS

Analytical and numerical models

P. TRUCHOT, R. BOURQUIN

Laboratoire de Chronométrie, Electronique et Piézoélectricité
Ecole Nationale Supérieure de Mécanique et des Microtechniques
Route de Gray - 25030 BESANCON CEDEX - FRANCE

SUMMARY

Analytical models of torsional quartz resonator are compared with its predicted behaviour given by finite element method, particularly the constant of torsion and the thermal sensitivity of the resonant frequency that is directly related to the constant of torsion. The first and second order coefficient of temperature of the frequency are calculated by using the usual way, or by using the material (Lagrangian) description in the reference state at reference temperature. The calculus of the electrical current of the resonator shows that it is necessary to consider the St Venant's function for the determination of motional parameters.

1/ INTRODUCTION

The basis of the analytical study of torsion of beam is the theory of St Venant that has been extended to the case of anisotropic material by Love [1] and Green [2]. Depending of the used method, the constant of torsion may be expressed by using the elastic compliances s_{ij} or the elastic stiffness c_{ij}. The obtained constants of torsion are not fully equivalent and the predicted resonant frequencies of resonators and their thermal sensitivities may differ significantly. Torsional quartz resonators have been studied by several authors [3-5].

The treatment of the problem, in the case of an anisotropic solid is less difficult if some elastic coefficients, namely c_{46} or s_{46}, are neglectable. In this last case, the obtained results are closer to the experience than for a cut where c_{46} (or s_{46}) is high.

The study of torsional resonator can be also achieved by using finite element method, in the most general case. An advantage of this method is the possibility to include the attachment of the resonator. This last point is important when one consider very small devices like those that can be manufactured by using chemical etching. When this process is used, the length of the bar is generally oriented in the YZ cristallographic plane.For this reason, we consider more particularly single rotated resonators.

The thermal sensitivity of the resonant frequency can be calculated in the usual manner, which involve the description of the resonator in the deformed state after heating. A less tedious way is to use the material (Lagrangian) description of the resonator in the reference state at reference temperature [6]. In this case, only the effective elastic coefficients, employed in this description, vary when the temperature change.

2/ CONSTANT OF TORSION

2.1/ Analytical models

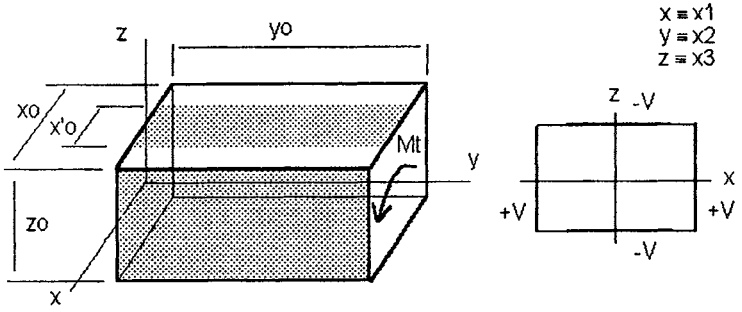


Fig. 1

The geometry of the rectangular beam is depicted on the Figure 1. According to St Venant's theory, the mechanical displacements u, v, w along the three axes x, y, z are

$$\begin{aligned} u &= -\alpha z \\ v &= \tau\, \varnothing \\ w &= \alpha x \end{aligned} \qquad (1)$$

where $\tau = \partial\, \alpha\, / \partial y$ is the torsional angle per unit length and $\varnothing$ (x, z) is the St Venant's torsion function. The deformations in the bar are S_4 and S_6 and the corresponding stresses, assuming τ constant, are :

$$\begin{aligned} T_4 &= \tau \left[c_{44} (\partial\, \varnothing / \partial z + x) + c_{46}\, (\partial\, \varnothing / \partial x - z) \right] \\ T_6 &= \tau \left[c_{46} (\partial\, \varnothing / \partial z + x) + c_{66}\, (\partial\, \varnothing / \partial x - z) \right] \end{aligned} \qquad (2)$$

All elastic coefficients are expressed in the rotated frame (x, y, z).

In the case where the terms containing c_{46} may be neglected, the function $\varnothing$ is given by [1] :

$$\varnothing (x, z) = -xz + \frac{8 z_o^2}{\pi^3} \frac{1}{K_c} \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)^3} \frac{sh\,(2n+1)\,\pi\, K_c\, x / z_o}{ch\,(2n+1)\,\pi\, K_c\, x_o / 2 z_o} \sin \frac{(2n+1)\,\pi\, z}{z_o} \qquad (3)$$

where $K_c = (c_{44}/c_{66})^{1/2}$

The torque M_t is related to the torsional angle τ by the torsional constant C_t :

$$M_t = C_t\ \tau = \tau \iint \left(C_{44}\ x^2 + C_{66}\ z^2 + C_{44}\ x \frac{\partial \varnothing}{\partial z} - C_{66}\ z \frac{\partial \varnothing}{\partial x} \right) dx\ dz \qquad (4)$$

using (3), C_t is then expressed as :

$$C_t^c = \frac{c_{66}}{3} \frac{z_o^4}{r} \left[1 - \frac{r}{K_c} \frac{192}{\pi^5} \sum_{n=0}^{\infty} \frac{1}{(2n+1)^5}\ th \frac{(2n+1)\ \pi\ K_c}{2r} \right] \qquad (5)$$

where $r = z_0/x_0$ labels the width to thickness ratio.

If one consider that the main stress in the bar are T_4 and T_6 (solid with a neutral fiber), the induced deformations are :

$$S_4 = s_{44}\ T_4 + s_{46}\ T_6 \qquad S_6 = s_{46}\ T_4 + s_{66}\ T_6 \qquad (6)$$

or

$$T_4 = \frac{s_{66}}{s_{44}\ s_{66} - s_{46}^2}\ S_4 + \frac{s_{46}}{s_{45}^2 - s_{44}\ s_{66}}\ S_6 \qquad (7)$$

$$T_6 = \frac{s_{46}}{s_{46}^2 - s_{44}\ s_{66}}\ S_4 + \frac{s_{44}}{s_{44}\ s_{66} - s_{46}^2}\ S_6$$

and in the case where $s_{46} = 0$, instead of (5), the constant of torsion is given by :

$$C_t^s = \frac{1}{3\ s_{66}} \frac{z_o^4}{r} \left[1 - \frac{r}{K_s} \frac{192}{\pi^5} \sum_{n=0}^{\infty} \frac{1}{(2n+1)^5}\ th \frac{(2n+1)\ \pi\ K_s}{2r} \right] \qquad (8)$$

with $K_s = (s_{66}/s_{44})^{1/2}$

The Figure 2 shows the orientation of the doubly rotated bar given by the angle Ψ and θ, the bar corresponding to a zytw (Ψ/θ) cut following the IEEE 1978 convention .

The Figure 3 depicts the values of C_t^c / z_o^4 in function of the θ angle and the ratio r, for an angle Ψ=0, in the case of the so called C model (eq 5). For this Ψ angle, the coefficient c_{46} relative to the rotated frame (x y z) is nul.

In the case of the use of the s coefficients, the corresponding curves for the constant C_t^s / z_o^4 are given by the Figure 4.

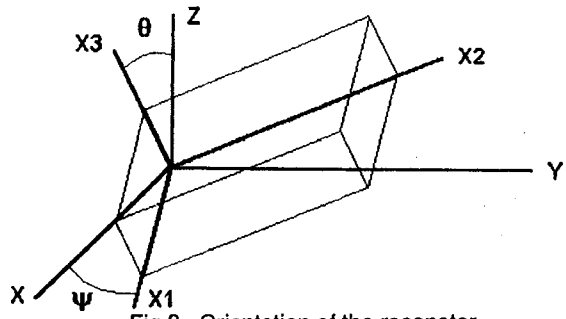


Fig 2 Orientation of the resonator

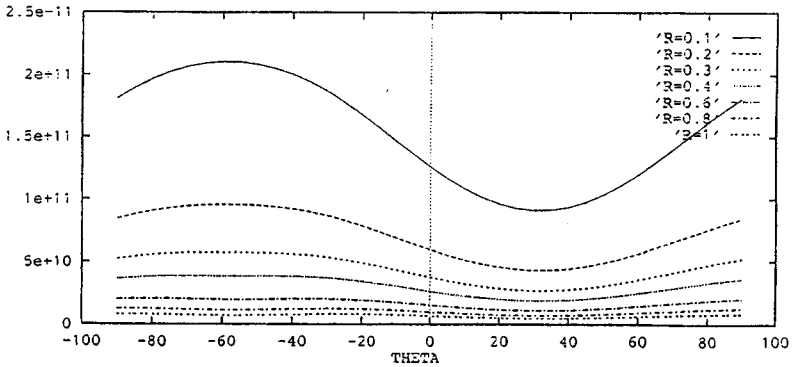


Fig 3 C_T^c / z_o^4 function of orientation θ and aspect ratio r= z_o / x_o

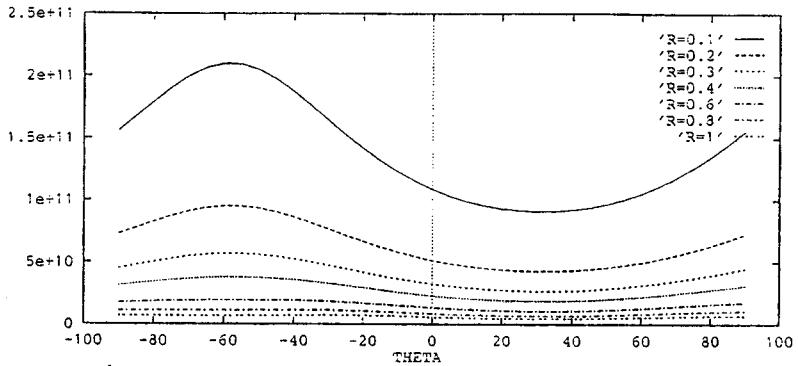


Fig 4 C_T^s / z_o^4 function of orientation θ and aspect ratio r= z_o / x_o

2.2/ Use of finite element method

Numerical calculation of the C_t constant has been performed by using Systus finite element software.The mesh of the bar is shown on Figure 5a. It use elements with quadratic interpolation. The bar is rigidly fixed at one extremity and the torque is exterted by two forces applied on two opposite sides at the other extremity, the neutral fiber having its displacements u and w forced to zero. The rotation of the sections of the bar is shown on the Figure 5b. In order to have a good accuracy in the determination of the angle of torsion per unit length, the sections at the extremities of the bar are not used, because theirs displacements may be affected by edge effects.

The Table I and II gives the constants of torsion calculated from the formulas (5) and (8) and the constant C_t^N given by the FE method. In the case of the Table I, the orientation of the bar ($\Psi = 0$, $\theta = 0$) is choosen so that the coefficient c_{46} (or s_{46}) is zero, while the Table II correspond to an orientation ($\Psi = 90°$, $\theta = 60°$) for which the coefficients c_{46} (or s_{46}) are maximum.

We note a better agreement between C_t^s and C_t^N than between C_t^c and C_t^N, particularly in the case where $c_{46} = 0$. The fact to neglect this coefficient cause an error that is, in the worst case, of the order of 5% for the C_t^s constant. In all cases, the C_t^c constant exhibit a deviation of an order of 10%.

Table I
Constants of torsion : $\Psi = 0$, $\theta = 0$

r, z_0	0.19, 0.12	0.5, 0.3	1, 0.64	3, 0.6	5, 0.6
C_t^c	$1.31.10^{-5}$	$1.59.10^{-4}$	$11.15.10^{-4}$	$6.92.10^{-5}$	$1.70.10^{-5}$
C_t^s	$1.14.10^{-5}$	$1.37.10^{-4}$	$9.61.10^{-4}$	$5.96.10^{-5}$	$1.46.10^{-5}$
C_t^N	$1.15.10^{-5}$	$1.40.10^{-4}$	$9.94.10^{-4}$	$6.40.10^{-5}$	$1.58.10^{-5}$

Table II
Constants of torsion : $\Psi = 90°$ $\theta = 60°$

r, z_0	0.19, 0.12	0.5, 0.3	1, 0.64	3, 0.6	5, 0.6
C_t^c	$1.72.10^{-5}$	$1.91.10^{-4}$	$11.71.10^{-4}$	$5.97.10^{-5}$	$1.41.10^{-5}$
C_t^s	$1.45.10^{-5}$	$1.64.10^{-4}$	$10.30.10^{-4}$	$5.36.10^{-5}$	$1.28.10^{-5}$
C_t^N	$1.52.10^{-5}$	$1.71.10^{-4}$	$10.60.10^{-4}$	$5.46.10^{-5}$	$1.29.10^{-5}$

3/ THERMAL DEPENDENCE OF THE RESONANT FREQUENCY

The motion of a torsional resonator is governed by the equation :

$$\rho I_y \partial^2 \alpha / \partial t^2 - C_t \partial^2 \alpha / \partial y^2 = 0 \qquad (9)$$

where $I_y = x_o\, z_o \left(x_o^2 + z_o^2\right) / 12$ is the moment of inertia of a section of the rectangular bar and ρ the mass density. For a resonator clamped at one extremity while the other is free, the angle of rotation is :

$$\alpha = A \sin a y\, e^{j\omega t} \tag{10}$$

and the resonant frequency is then :

$$f = \frac{n}{4 y_o} \left(\frac{C_t}{\rho I_y}\right)^{1/2} \quad , n \text{ odd} \tag{11}$$

The first order frequency coefficient $T^{(1)} f = 1/f\, \partial f/\partial T$ and the second order $T^{(2)} f = 1/2f\, \partial^2 f/\partial T^2$ can be computed from (11) by using as C_t, the constant C_t^c (eq.5) or the constant C_t^s (eq 8). In both cases, in performing the calculation, we have neglected, as usual, the variation of orientation of the bar when the temperature change. The numerical values of the thermal derivaties of the elastic constants are those given by Bechmann and all [7].

This way is straightforward but tedious. A more powerful method consist to use a Lagrangian description. In this case, the involved quantities are expressed in terms of the coordinates X_j in the undeformed state at the reference temperature T_o. The Piola stress tensor K_{ij} is used instead the usual stress tensor T_{ij} [6]. K_{ij} is related to the deformation gradients $\partial u_k/\partial X_l$ through the material coefficients G_{ijlk} [6, eq 22].

$$K_{ij} = G_{ijlk}\, \partial u_k \,/\, \partial X_l \tag{12}$$

The law of motion and the boundary conditions take the form :

$$\partial K_{ij} \,/\, \partial X_i = \rho_o\, \partial^2 u_j \,/\, \partial\, t^2 \tag{13}$$

$$K_{ij} N_j = 0 \qquad \text{on} \quad S_o (T_o)$$

where N is the normal of the boundary surface S_o and ρ_o the density at the reference temperature T_o.The coefficients G_{ijlk} are related to the usual coefficients C_{ijkl} and the expansion coefficients of the material [6]. The constant of torsion C_t^c become C_t^G by substituting G_{96} for C_{66} and G_{47} for C_{44} in matrix notation [9].

In eq 13, the G coefficients are the only temperature depending quantities and, further, the variation of orientation of the bar with the temperature is automatically taken into account.

In the case of singly rotated cut ($\Psi = 0$), the Figures 6 and 7 give the ratio of lateral dimensions of the resonator r, as a function of the θ angle, for which the first and second order temperature coefficient are null.

4/ MOTIONAL INDUCTANCE

We restrict our attention to a single rotated bar ($\psi = 0$) for which the coefficients c_{46} and e_{16}, e_{32}, e_{34} are null.

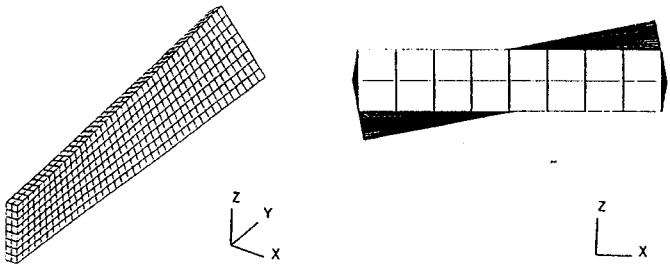


Fig 5 FE computation of a bar in torsion

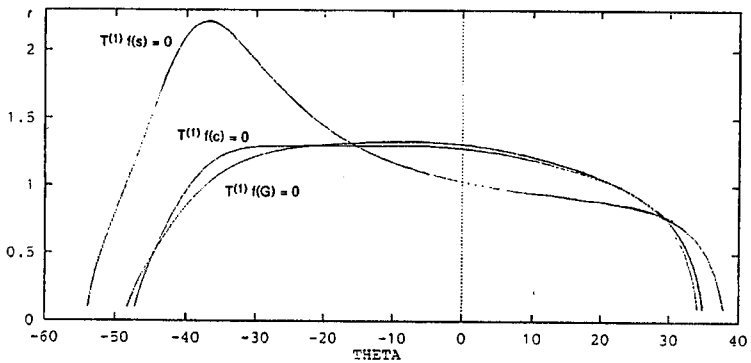


Fig 6 aspect ratio r vs. θ for $T^{(1)}$ f = 0 for the three models

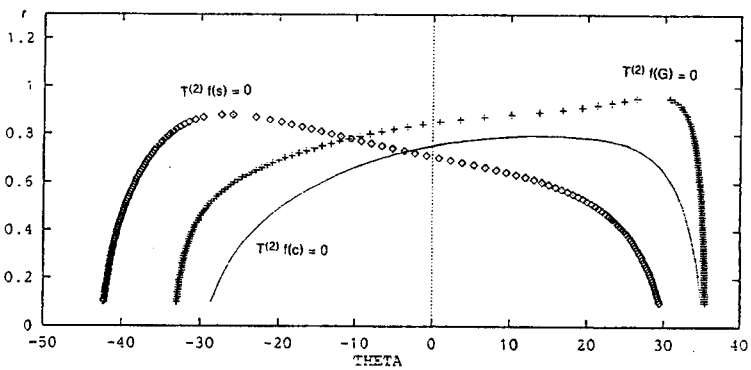


Fig 7 aspect ratio r vs. θ for $T^{(2)}$ f = 0 for the three models

The motional inductance L_1 can be calculated with a good accuracy by considering that the v displacement discribed by the St Venant's function has a noticeable value only in the corner regions of the cross section of the resonator and that the electrical field has a negligible component along the length of the resonator.

The deformations in the resonator are :

$$S_4 = \partial\alpha / \partial y(x + \partial\varnothing / \partial z) \; ; \; S_6 = \partial\alpha / \partial y \, (-z + \partial\varnothing / \partial x) \; ; \; S_2 = \varnothing \, \partial^2 \alpha / \partial y^2 \tag{14}$$

where the rotation angle α is given by (10)

The inductance L_1 may be computed from the stored energy in the resonator by writing that :

$$\frac{1}{2} L_1 |I_1|^2 \equiv \frac{1}{2} \iiint_{réson.} \left(c_{44} |S_4|^2 + c_{66} |S_6|^2 \right)_{S\max} dx \, dy \, dz \tag{15}$$

where we have neglected the contribution of S_2 when evaluating the mechanical energy. I_1 represent the amplitude of the current in the motional branch (L_1, C_1, R_1) of the equivalent electrical circuit of a piezoelectric resonator. The total current through a pair of electrodes is :

$$I_Q = (I_1 + I_o) e^{j\omega t} = 2 \int_{-z_o/2}^{z_o/2} \int_{o}^{y_o} \left(- \frac{\partial D_1}{\partial t} \right)_{x = x_o/2} dy \, dz \tag{16}$$

I_0 being the current through the static capacitance C_0 of the electrical circuit, and :

$$D_1 = - \varepsilon_{11} \, \partial\varphi / \partial x + e_{12} \, S_2 + e_{14} \, S_4 \equiv - \varepsilon_{11} \, \partial\varphi / \partial x + A \left[- e_{12} \, a^2 \, \varnothing \sin ay + e_{14} \, a \, \frac{8 z_o}{\pi^2 \, K_c} \, \frac{sh \, \pi \, K_c \, x / z_o}{ch \, \pi \, K_c \, x_o / 2 z_o} \cos \frac{\pi z}{z_o} \cos ay \right] \tag{17}$$

where $\varnothing$ is reduced to its first term in the sommation.
The electric potentiel φ can be separated into an electrostatic part φ_0, obeing the Laplace's equation, and a piezoelectric part φ_p : $\varphi = \varphi_0 + \varphi_p$. The equation of Poisson div **D** = 0 gives:

$$\varepsilon_{11} \frac{\partial^2 \varphi_o}{\partial x^2} + \varepsilon_{33} \frac{\partial^2 \varphi_o}{\partial z^2} = 0$$

$$\varepsilon_{11} \frac{\partial^2 \varphi_p}{\partial x^2} + \varepsilon_{33} \frac{\partial^2 \varphi_p}{\partial z^2} = e_{12} \frac{\partial S_2}{\partial x} + e_{14} \frac{\partial S_4}{\partial x} + e_{36} \frac{\partial S_6}{\partial z} \tag{18}$$

Since we need the eigenmode of resonance, in solving (18-2) the electrodes are electrically short circuited [9]. These equations are solved by a numerical method [10], the normal component of the electric displacement being equated to zero at the boundary surfaces not covered by an electrode. The currents are then computed from (16) with the help of (17) (18) (14) and (10) :

$$I_1 = AI_m = -2\,j\omega \left[16\, e_{14}\, A\, \frac{z_o^2}{\pi^3\, K_c}\, th\, \frac{\pi\, K_c\, x_o}{2 z_o} - \varepsilon_{11} \int_{-z_o/2}^{z_o/2} \int_{o}^{y_o} \left(\frac{\partial \varphi_p}{\partial x} \right)_{x = x_o/2} dy\, dz \right] \tag{19}$$

$$I_o = 2\, j\omega\, \varepsilon_{11} \int_{-z_o/2}^{z_o/2} \int_{o}^{y_o} \left(\frac{\partial \varphi_o}{\partial x} \right)_{x = x_o/2} dy\, dz$$

The integral are numerically computed. The inductance L_1 is finally given, in the case of the fundamental mode ($a = \pi/2y_o$), by :

$$L_1 \cong \frac{\pi^2\, x_o\, z_o}{96\, y_o} \left(C_{66}\, z_o^2 + C_{44}\, x_o^2 \right) / \left| I_m \right|^2 \tag{20}$$

The plot of the inductance L_1, as a function of the relative width $e = x'_o/x_o$ of the electrodes on the faces normal to z, for several values of the aspect ratio $r = z_o/x_o$, is shown on the figure 8. It exhibit a strong dependence of L_1 with respect of these two parameters.

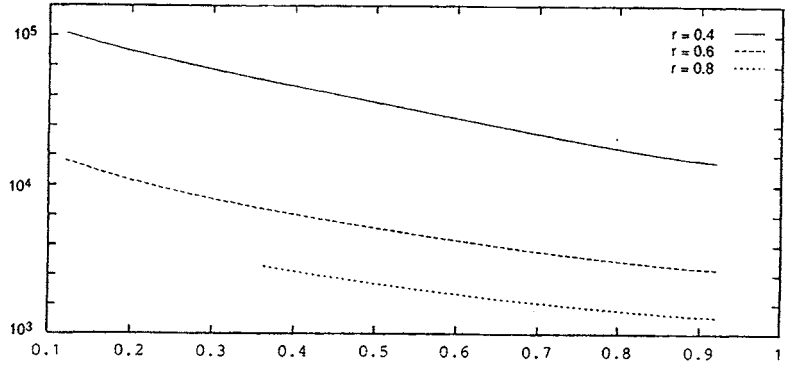


Fig. 8 log(L_1) vs. relative width e of electrodes.
See note under table III for sizes of resonators.

The table III below gives some experimental results and the corresponding theoretical values. f_r^c, f_r^s are respectively the resonant frequency calculated by using the C_t^c or C_t^s torsional constant, while f_r^n is the resonant frequency obtained by modelizing the resonant device with the use of FE method.

Param.	Exp. values		Theoretical values			
	f_r	L_1	f_r^c	f_r^s	f_r^n	L_1
θ = 38 r = 0.4 e = 0.72	51720	44780	54600	54250	53667	21140
θ = 34 r = 0.6 e =0.5	80004	9460	87192	86651	85884	5200
θ = 34 r = 0.6 e = 0.56	80008	8076	87192	86651	85884	4600
θ = 34 r = 0.8 e = 0.8	94534	4010	98358	97309	95323	1440
θ = 34 r = 0.8 e = 0.72	94516	5250	98358	97309	95323	1570

Note : The sizes in mm of the resonators are the following : [θ = 38, r = 0.4] : xo = 1 yo = 10 [θ = 34, r = 0.6] : xo = 2 yo = 8 ; [θ = 34, r = 0.8] : xo = 1 yo = 8

Acknowledgements :

The authors would like to thank Dr B. DULMET for his help in performing FE computation.

REFERENCES

[1] A.E.H. Love - A treatise on the mathematical theory of elasticity - Dover - New York (1944).

[2] A.E. Green - The flexure and torsion of aeolotropic beams - Proc. Cambridge Phil. Soc., (Jan. 1947), p 68-74.

[3] R.G. Pozdnyakov - Quartz crystal resonating in torsion - Sov. Phys. - Crystallography - Vol. 15, N° 1 (Jul-Aug. 1970), p 63-67.

[4] J. Hermann - DT-cut torsional resonators - Proc. of the 31st Ann. Symp. on Freq. Cont. - Atlantic City - (1977), p 55-61.

[5] H. Kawashima, M. Nakazato - TT-cut torsional quartz crystal resonator - Proc. of the 45th Ann. Symp. on Freq. Cont. - Los Angeles - (1991), p130-136.

[6] B. Dulmet, R. Bourquin - Influence of the definition of material constants in the study of crystal plates thermal behaviour - Proc. of the 3rd EFTF - Besançon (1989) p 305-312.

[7] R. Bechmann, A. Ballato, T. Lukaszek - High order temperature coefficients of the elastic stiffnesses and compliances of quartz - P.I.R.E. (1962), p1812-

[8] The symetry with respect to the indice of the G coefficients is lower than for the c coefficients. The G coefficients are arranged in an 9x9 matrix. Both family of coefficients are identical at the reference temperature.

[9] P. Lloyd - Equations governing the electrical behavior of an arbitrary piezoelectric resonator having n electrodes - B.S.T.J. Vol. 46, (1967), p 1881-1900.

[10] E. Durand - Electrostatique T II et T III - Masson - Paris (1964).

Compact Ovenized Crystal Oscillators with Internal Synchronization to Standard Frequency Transmitters

Bernd W. Neubig / Oliver Leibfried / Jochen Dugeorge

TELE QUARZ GROUP

1. Introduction

Ovenized crystal oscillators (OCXO) show a very low phase noise and excellent short-term stability. The long-term behaviour is determined mainly by the aging of the crystal oscillator, for AT-cut and SC-cut that is between 10^{-8} and 10^{-6} per year. Most application therefore require to re-adjust the oscillator after a certain time to compensate for the aging.

Many OCXOs produced today are used in base stations and other remote locations where the readjustment is extremely difficult. Synchronization to a standard frequency transmission of one of the VLF transmitters is favourable and allows to achieve a long-term stability of 10^{-9}, while the short-term stability still is determined by the OCXO.

This table presents a solution in which the complete OCXO plus the VLF receiver and the frequency synchronization circuit are included in a very compact enclosure of the size 67 x 60 x 40 mm^3 (2,6" x 2,4" x 1,6"). The whole unit is not larger than a conventional OCXO. For interference-free reception an external antenna can be connected to a coaxial cable.

2. Synchronization requirements

VLF transmitters for standard frequencies are used throughout the world. In Germany the transmitter DCF-77 at 77,5 KHz, located near Frankfurt/Main, Germany, is the legal distributor of time and frequency. Its output frequency is based on three high accuracy primary cesium standards of the German standards authority PTB (German equivalent of NIST). Similar transmitters in Central Europe are HBG (Switzerland) at 75 KHz, OMA (Czechia) at 50 KHz , MSF and GBR (UK) at 60 KHz resp. 16 KHz. In the US OMEGA and LORAN C transmitters can be used as a reference.

The frequency accuracy which can be recovered by a VLF receiver is degraded by variations in the propagation conditions (delay, reflexions, etc.) and also by a pseudo-random phase keying of the DCF-77 carrier which was introduced in 1983. Therefore, the inherent stability of the Cs sources can only be approached for averaging times longer than one day. Figure 1 shows the frequency stability vs. averaging time (Allan-Variance) for different sources. The dotted line marked 'DCF' is the useable accuracy of a DCF signal received approximately 300 km from the transmitter. The shaded area marked 'Q' indicates the typical range for crystal oscillators.

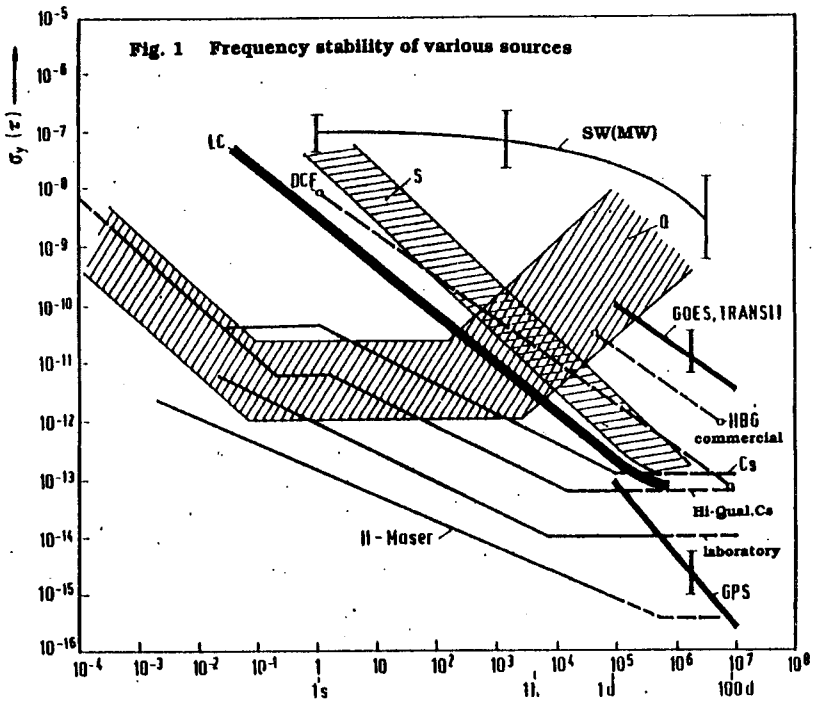
Synchronization of an OCXO to a VLF reference transmitter therefore imposes the following requirements:

- the good short-term stability of the OCXO shall not be degraded by the synchronization
- very short distortion and interferences of the reception is caused by man made noise, ignitions, thunderstorms, etc. must be isolated from the OCXO
- at interruptions of the transmission or other loss of synchroni-zation the OCXO frequency shall not float
- the stabilization characteristic of the OCXO after power-on shall not be degraded by the required synchronization

The aim of the presented development was:

- the whole oscillator including the VLF receiver and the synchronization has to be included in a conventional OCXO
- the current consumption shall be approximately the same as for already existing OCXOs
- the efforts for tuning and adjustment of the whole unit should be minimized.

All these requirements could be fulfilled by the presented solution



3. Description of the device

Because of the small available space, concept using complicated synthesizers and multiple PLL loops could not be used. The required high time constants in excess of one day in connection with the fast warm-up characteristic required are not easily realizable with conventional PLL circuits. Because of space restrictions and to avoid adjustment the whole receiver circuit in the unit does not use any inductors. The method described below does not deteriorate the characteristics of the precision OCXO and nevertheless is realized with relatively low parts count. The aim of the method was to reach a frequency stability in the order of $\pm 1 \cdot 10^{-9}$ by a suitable frequency locking principle, not a phase locking loop. Thus, the OCXO is not locked strictly to the phase of the VLF standard frequency transmitter which is the cause of the complexitiy of usually applied PLL circuits. The frequency control loop is realized by a low-cost micro-controller with peripheral circuits (counter, D/A converter etc.). The main design work therefore is concentrated in the software of the system, while hardware requirements are reduced drastically. Figures 2 and 3 shows the working principle of the circuit.

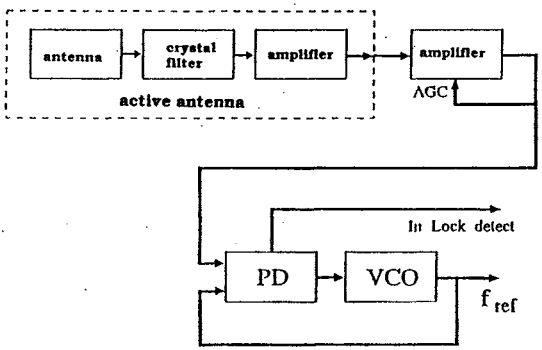


Fig. 2 Building blocks of the frequency synthesis

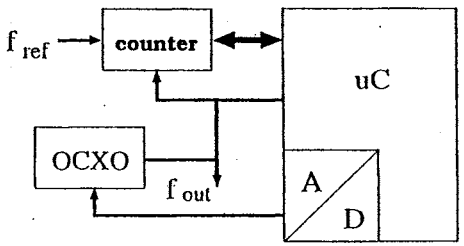


Fig. 3 Building blocks of the digital part

Four functional blocks are shown:
- receiver
- tracking oscillator (used as tracking filter)
- precision OCXO
- micro-controller

The receiver consists of an active antenna, whose DC supply is delivered through the coaxial connection to the OCXO unit. The transmitted standard frequency is picked up by a tuned ferrite antenna. It is combined with a single-pole crystal filter which filters the incoming signal with a bandwidth of approximately 2 to 3 Hz. This small bandwidth already suppresses a large amount of the amplitude modulation of the transmitter (time and date coding). The circuit is given in figure 4. The capacitor C3 in the opposite lattice branch the filter can be adjusted either to a symetrical response or to a response with a finite attenuation pole at a given interference frequency. Because of the good filtering of a received signal direct at the front end of the receiver, disturbing and interfering signals and frequencies in the vicinity are effectively filtered out, so that they cannot cause any overdrive and limiting actions of the subsequent stages or even blocking of them. The inclusion of the filter in the antenna preamplifier leads to a significant space reduction for the receiver itself. The following preamplifier has a high input impedance and a gain of approximately 6. An emitter follower matches the amplifier to the cable impedance of 50 Ω. The total current consumption of the preamplifier is 1.7 mA. For a sufficient reception quality the antenna unit should be mounted away from the system at a high position. The ferrite stick should be oriented for maximum sensitivity in the direction of the transmitter.

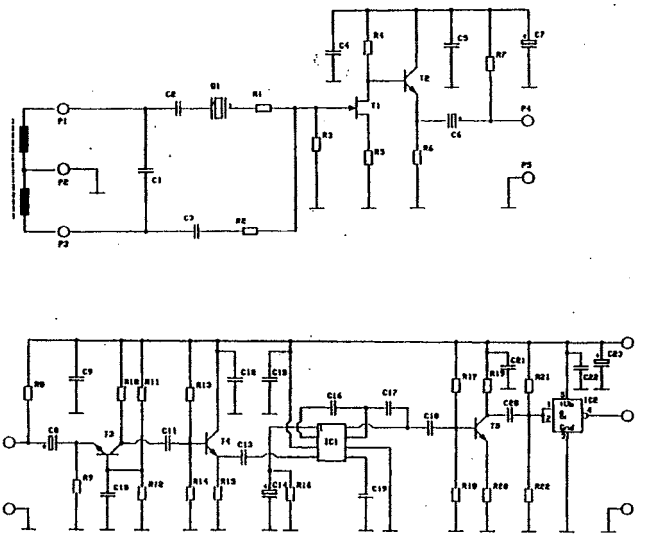


Fig. 4 Circuit diagram of the receiver

The signal from the antenna preamplifier goes into the receiver which is included in the OCXO unit. The input impedance of T3 is matched to 50 Ω. The voltage gain is approx. 80. The following transistor T4 couples the amplified AC-voltage into IC1 of a low impedance. IC1 contains 2 amplifiers with controllable gain and an AGC circuit which maintains the output signal at pin 8 at a constant level of 90 mV_{rms}. The time constant of the AGC is such that the residual amplitude modulation of the DCF-77 signal is suppressed completely. At the collector of T5 we get a sinusoidal AC-voltage which is shaped into a clean rectangular signal used for the processing with the following digital components by IC2. The maximum overall gain of the receiver circuit is 132 dB. That is sufficient for application throughout Germany and in most parts of Central Europe. The dynamic range of the AGC circuits is approx. 55 dB. In areas with extremely high field strength, e. g. close to the transmitter, the signal is already limited in the amplifier stages in front of the AGC control. This, however, has no negative influence on the subsequent stages because the above-mentioned modulation of the DCF-77 signal is not desired anyhow. In those cases the AGC control ist not effective.

The squared signal at the output of the receiver has a very poor short-term stablity and high jitter. This is improved by the following tracking filter, which can be operated with very narrow bandwidth by a crystal stable oscillator. With a loop time-constant of 3 ... 5 seconds the phase jitter of the signal can be improved significantly. Because of the fact that PLL circuits work like ideal bandpass filters beyond their open loop bandwidth, interferences with very short time constants are suppressed effectively by the circuit. The output of this VCXO delivers a square wave signal which represents the received carrier with moderate short-term stability. In case of carrier loss or of longer lasting interferences an integrated in lock detector delivers an "out-of-lock signal" to the micro-controller. Thus, erraneous synchronizations are avoided effectively. For short-term carrier interuptions the PLL is able to deliver a carrier signal for some time with approximately correct phase. Thus, short interferences by thunderstorms, ignitions etc. have no influence on the function of the circuit. The PLL circuit consists of a standard CMOS IC 4046, from which only the phase detector and the lock detector are used. Despite the only partial use of this IC this solution was found to be the optimum between current consumption and space requirement.

The cleaned output signal of the described circuit is now fed into the digital portion of the unit (figure 5). This part performs the synchronization of the OCXO to the received standard frequency.

The received and processed carrier signal goes to a flip-flop (IC1) which can reset the counter (IC2) through the micro-controller. The beginning of the count is synchronized with the falling edge of the standard frequency signal. At the same time the prescaler, which is part of the IC 1, is activated. The micro-controller divides the input frequency further down by software. By this, very long time constants can be realized. The output frequency of the OCXO is continuously measured in IC2. Because of the high frequency stability of the OCXO (in short terms) it is sufficient that the counter of IC 2 is read only once a second. As long as the frequency of the OCXO does not move more than 64 Hz (at a nominal frequency of 10 MHz), no information is lost by a probable counter overflow, which occurs every 12,8 µsec. This condition is fulfilled for all OCXOs without problems.

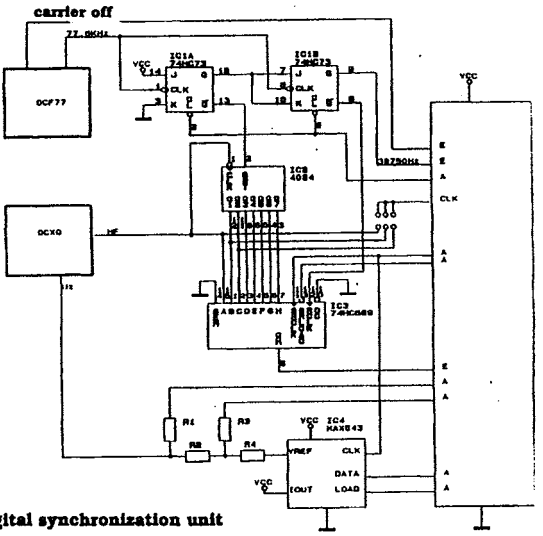


Fig. 5 Digital synchronization unit

The counter is read out by the shift register IC3. The inputs of the shift register are latched, so that the processor has sufficient time to read the counter serially. IC4 computes the required pulling voltage for the OCXO by a numerical method. This voltage is output serially via a 12 bit D/A-converter which was extended to 14 bit resolution by additional resistors. For a pulling range of $\pm$ 2 ppm (10^{-6}) this leads to a resolution of 0,25 ppb (10^{-9}). The clock of the micro-controller is derived from the output frequency in IC2. Additional clock generators are not necessary. The whole microprocessor control circuit is located on a small double-sided SMT pc-board which is mounted vertically on the OCXO board.

4. Software

The flow chart of the software routines are given in figure 6. After the switch-on of the system, a pulling voltage for the OCXO is generated by the use of a parameter stored internally in the EEPROM of the micro-controller. This parameter is individually derived from the measurement during production of the unit. Therefore, the OCXO skips nearly immediately to nominal frequency even at the first operation. The value of the pulling voltage is continuously renewed, so that the output frequency is also recovered very fast after a power-down. Also after loss of the standard frequency carrier the stored value is used. After a suitable warm-up time, depending on the used oscillator type (5 to 15 minutes), the controller starts with the synchronization. From the standard frequency signal interrupts are derived which decrements a software based counter. Each second the counter overflows, which is used to read out the counter of IC2. My means of a digital controlled algorithm the correction values for the D/A-converter are computed from the measured values. The frequency error of the OCXO therefore is well below 1 ppb as long as the standard frequency carrier can be received. If no carrier is present the last correction value in the EEPROM is frozen. This means that the aging of the oscillator follows the regular aging curve of the free running OCXO for this short period.

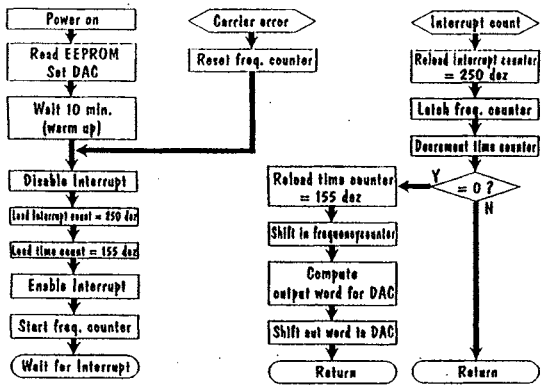
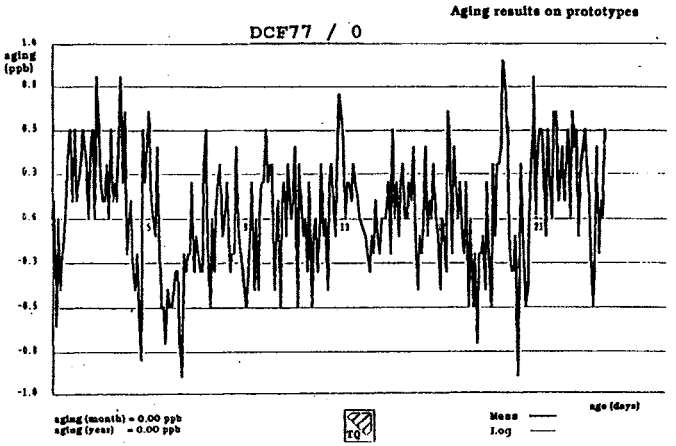


Fig. 6 Flow chart of the synchronization algorithm

Figure 7 shows a typical example of a practically measured "aging curve" of a OCXO unit synchronized to DCF-77 by the described method.



5. Summary

The paper describes a compact ovenized crystal oscillator which is internally synchronized to a standard frequency transmitter in the VLF range. The received signal is cleaned up by analogue and digital circuitry. The synchronization is realized by software which is run in a micro-controller. All components, receiver, clean-up circuitry micro-controller, OCXO (except the antenna) are included in an conventional OCXO enclosure. The current consumption of the synchronization unit is only a small fraction of the overall current consumption of the OCXO circuit itself.

The method allows the aging of an OCXO to be kept below $1 \cdot 10^{-9}$ for the whole life time of the OCXO. The sensitivity to external interferences, carrier loss, etc. are minimized.

6. References

(1) Benjamin Parzen with Arthur Ballato
Design of Crystal- and other harmonic oscillators
Wiley - Interscience publication 1983

(2) Tagungsdokumentation des "ZVEI" - Symposium Schwingquarze 1989
Fachverband Bauelemente der Elektronik im Zentralverband
Elektrotechnik- und Elektronikindustrie e. V.

(3) W. Hilberg
Funkuhrtechnik
Oldenbourg Verlag, München

(4) Michael Arnold
Zeitzeichen- und Normalfrequenzempfang
Franzis Verlag, München

(5) Roland Best
Theorie und Anwendung des Phase Locked Loops
AT Verlag Aarau, Stuttgart

BALANCED BRIDGE IMPEDANCE ELEMENT SAW FILTERS

James Heighway, Sergei N. Kondratyev, and Victor P. Plessky

Advanced SAW products SA, CH-2022 Bevaix, Switzerland

1. ABSTRACT

*Over the last few years a new generation of Surface Acoustic Wave (SAW) devices with low loss and operating at high frequency have been invented to meet the demand of mobile communication systems. Besides outstanding electrical characteristics, these filters need to have small size and weight, must be highly reproducible, need no tuning, must be very cheap and if necessary, able to handle power levels up to a few Watts. In this work we introduce a new class of SAW filters - **Balanced Bridge Impedance Element Filters**, - which have all these advantages; low insertion loss, extremely small size and weight, a wide variety of centre frequencies and passbands, excellent suppression in the stopband and require no external matching circuits .*

2. BASIC CONCEPTS

The impedance of an interdigital transducer (IDT) having a large number of electrodes N ($N * K^2 \geq 1$, where K^2 is electromechanical coupling coefficient) strongly changes in the frequency range corresponding to effective excitation of surface acoustic waves (SAW's).

Figure 1 illustrates the absolute value of the impedance dependence on frequency for a long interdigital transducer on a 64-LiNbO3 substrate (the structure contains 100 pairs of electrodes with period p≈2.35μ and aperture W=65μ) The frequency is in MHz, the impedance magnitudes in Ohms. The frequency dependence of the impedance is similar to that of classic L-C resonators with low resistivity at the resonance frequency and orders of magnitude higher resistivity at the anti-resonance frequency. The ratio of the max. to min. impedance can be increased by adding reflectors at both sides of the IDT, that is by creating a one port synchronous resonator. The solid curve of Fig. 1 illustrates this case. The resonator contains 70 pairs of electrodes in an IDT (p≈2.26 μ, and short reflectors on both sides of 16 shorted fingers). Both curves are simulated using the COM approach [1] . Generally speaking, this type of one port SAW device can be treated as ***a SAW impedance element*** **(IE)**, the frequency characteristic of which can be manipulated over a wide range using different substrate materials, types of waves, or designs of the IE itself.

The main idea of this work is to use these impedance elements as a building block for filters designed using classic network theory. Some lattice type filters have been developed by Japanese authors [2]-[3] based on this approach, but the full potential of this idea and some revolutionary and significant improvements can occur in SAW device design which have not yet been wholly understood. At some stage in all SAW devices , a signal is transformed from electrical into acoustic form and later reconverted. This fundamental feature results in several well known difficulties: loss, low power handling ability, and sensitivity to the accuracy of positioning of electrodes. All SAW device design is a struggle to overcome these difficulties.

With IE designs these difficulties can be overcome easily. Imagine that we have a simple Γ - type scheme including two IE's : the first (#1, solid curve on Fig. 1) in series between the input and the output (50/50 ohm) and the second (#2, dotted curve, Fig. 1) in parallel with the output. The first IE has low impedance , say 3 to 5 Ohms, at frequency $f1 \approx 950\ MHz$ and high impedance , more than 1000 Ohms at a higher frequency $f2 = f1 + \Delta f \approx 990\ MHz$ (numbers which can be realized without any difficulty). The parallel connected element can be

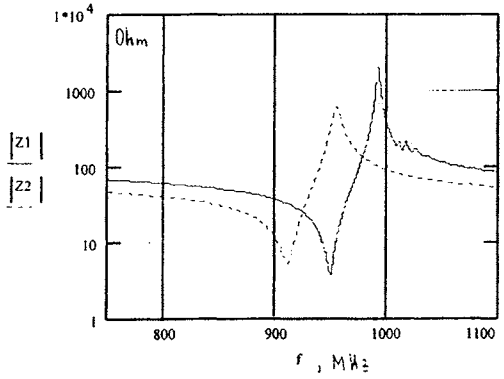


Fig. 1 Impedance of long IDT (dotted curve) and synchronous resonator (solid curve) on 64-LiNbO3 substrate.

chosen to have similar characteristics but shifted to a lower frequency: min. impedance at $f3 = f1 - \Delta f$ and maximum at $f = f1$ (dotted curve, Fig. 1) . It is obvious that this device will work as a passband filter (See Fig. 2, simulated filter performance) with a passband around frequency $f1$ and deep notches at $f2$, $f3$. The low loss at frequency $f1$ is due to the fact that the series connected element has, at this frequency, a very low impedance, while the parallel connected one has at this frequency a very high impedance.

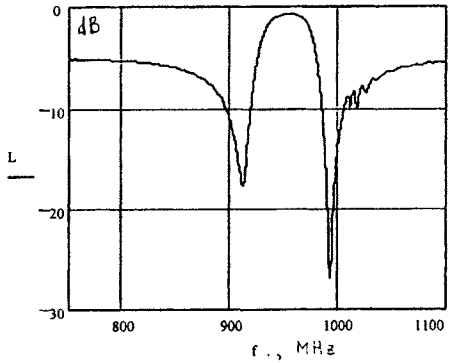


Fig. 2 Two IE Γ-type filter.

The out-of-band suppression of the filter is determined by the static capacitance ratio of the IDTs. In practice this out-of-band suppression is not adequate in the one section case discussed above and a few sections must be connected in series.

Note that there is no complete electric ↔ acoustic signal energy transformations. Only a small part of the voltage is applied to IE number 1 at the passband frequency. Since the distribution of voltages is approximately proportional to the impedances, practically all the input voltage in

the passband frequency range is transferred to the output. Thus the device can handle high power levels with low insertion loss.

3. LADDER TYPE IE-FILTERS

Figures 3a and 3b show the simulated and measured results for 2- and 4-section ladder type filters wherein the elementary section of two IE's is exactly the same as in an above example.

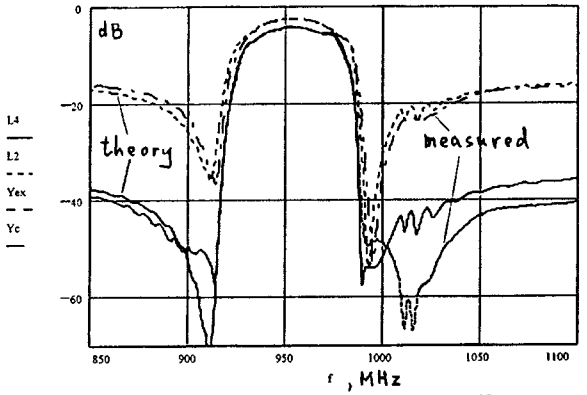


Fig. 3a Two and 4 section ladder type filters

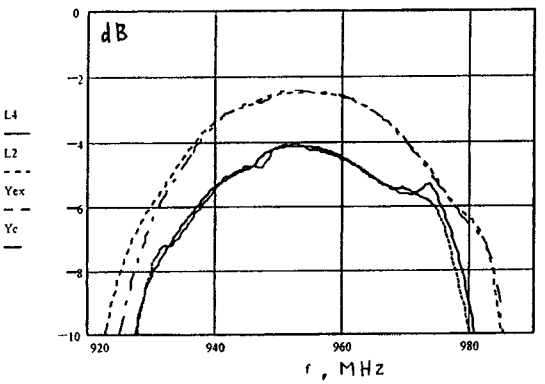


Fig. 3b. Top part of the curves for 2 and 4 section filter.

It can be seen that the simulation gives excellent results in the passband and is quite accurate up to the 50 dB level where the feedthrough must be taken into account. These particular comparisons do not show the best achievable results and only serve to illustrate the principle. The insertion loss can be reduced to a value about 1 dB for 2-section devices and 2 dB for a 4-section ladder type filter.

4. BALANCED BRIDGE FILTERS

The main drawback of ladder type IE filters is the relatively poor suppression in the stopband. At the stopband frequencies the IDTs do not excite the waves and act as simple capacitors. Part of the voltage from the input is inevitably transferred to the output through the ladder scheme of capacitors. This output signal level is basically determined by the capacitance ratios in the parallel and series arms of the scheme and is almost independent of frequency.

We propose a new balanced bridge type of SAW impedance element filter in which this static capacitance is compensated. This filter comprises a first pair and second pair of SAW impedance elements electrically coupled to form a bridge circuit, wherein the centre frequency of each SAW impedance element of the first pair is different to the centre frequency of each SAW impedance element of the second pair , and the product of the static capacitances of the first pair is substantially equivalent to the product of the static capacitances of the second pair. In particular , all static capacitances can be made equal. The advantage of this bridge filter is that the voltages transmitted from the input to the output due to static capacitances are canceled, thereby reducing out of band transmission. For frequencies of an input signal in the passband range, the admittances of the IE in each of the before mentioned pairs are different (due to SAW generation and due to the shift in electrode structure periods), and thus the input signal sees a different admittance sequence in each arm of the bridge. The bridge is strongly non balanced and the signals pass from the input to the output through the low resistivity IE's. The frequency shift can be chosen so that the resonance frequencies of one pair of IE's (where the impedance is low) approximately coincides with the anti-resonance of the second pair of impedance elements (See Fig. 4) . The impedance of these second pair of IE's is high and they do not have important influence in the passband. But in the stopband all impedances become progressively identical , the bridge scheme balanced and the output signals suppressed.

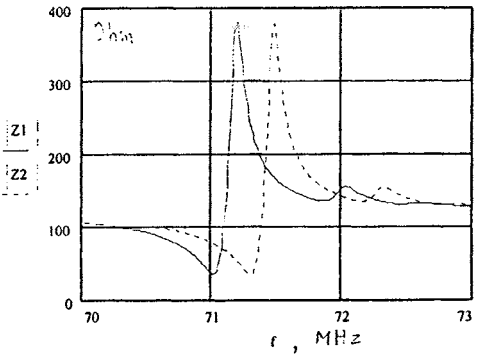


Fig. 4 Simulated impedance curves for two synchronous resonators on 112-LiTaO3 (frequency in MHz, impedance in Ohms).

Each resonator contained 90 pairs of electrodes in the IDT, 46 reflecting electrodes , the periods were approximately 23 and 22.9 microns, the apertures equal to 400 μ in both cases.
In the passband frequency range the first pair of IE practically connects the input to the output with minimal losses, while the signal cannot pass through the second pair of IE's which have high resistivity.

Fig. 5 shows an example of a filter performance designed using this approach with IE's the same as on Fig. 4.
The frequency shift between the arms determines the passband of the filter and to provide low loss, must be chosen comparable to the electro-mechanical coupling coefficient of the substrate. Two or more sections can be connected in series to get better stopband suppression. From a practical design point of view this new type of balanced bridge IE filter has a number of important advantages.

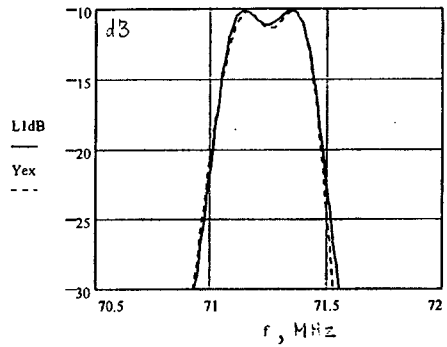


Fig. 5 Two section balanced bridge filter simulated (solid line) and measured (dotted line) frequency characteristics

The impedance elements are just simple long transducers (possibly with synchronous reflectors) having relatively large number of electrodes. The performance of these elements can be accurately modeled using existing models. The separate IE's are not interacting acoustically, they can be placed independently on the substrate. The large number of electrodes makes the resistivity of the electrodes relatively unimportant and parasitic capacitances are normally much smaller than the static capacitances of the transducers. The fact that SAW's do not have to be generated and received makes the overall length of the device shorter than in case of conventional SAW filters of the same passband.
Using different types of surface waves (Rayleigh waves, leaky SAW, STW) and different substrate materials, balanced bridge filters with passband ranges of 0.05% - 5% have been realized. At present, balanced bridge IEF's have been operated at frequencies up to 1.5 GHz (See parallel paper [4] at this conference). In this 1.5 GHz device the insertion loss was about 3 dB, the passband close to 1 MHz, sidelobe suppression better than 40 dB. It is clear that device performances can be superior to conventional SAW device characteristics.

5. CONCLUSIONS

A new class of SAW devices has been proposed - balanced bridge SAW impedance element filters - which have significant advantages compared both with conventional SAW filters and with recently developed ladder type SAW impedance element filters.

6. ACKNOWLEDGMENTS

The authors are grateful to C. S. Hartmann, Prof. D. P. Chen, Dr. M. A. Sharif, Dr. T. Thorvaldsson, and P. Durrant for fruitful discussions.

7. REFERENCES

[1] V. P. Plessky, C. S. Hartmann, " Characteristics of leaky SAWs on 36-LiTaO3 in periodic structures of heavy electrodes", 1993 Ultrasonics Symposium, pp. 1239-1242.

[2] M. Hikita, N. Shibagaki, T. Akagi, K. Sakiyama, "Design methodology and synthesis techniques for ladder-type SAW resonator coupled filters", 1993 Ultrasonics Symp., pp. 15-24.

[3] O. Ikata, T. Miyashita, T. Matsuda, T. Nishihara, and Y. Satoh , "Development of low-loss band-pass filters using resonators for portable telephones", 1992 Ultrasonics Symposium, pp. 111-114

[4] V. P. Plessky, T. Thorvaldsson, "GHz rang STW resonators and narrow band filters", These Proceedings.

Full–Wave Analysis of Surface Acoustic Wave Propagation in Planar ZnO–Diamond–Si Structures

U. Rösler[1], R. Weigel[1], H. Meier[2], P. Russer[1*3]

[1] Technische Universität München, Lehrstuhl für Hochfrequenztechnik,
Arcisstr. 21, D–80333 München, Germany
[2] Siemens Matsushita Components, OFW E NT,
Balanstr. 73, D–81617 München, Germany
[3] Ferdinand Braun Institut für Höchstfrequenztechnik,
Rudower Chaussee 5, D–12489 Berlin, Germany

Summary – A full–wave analysis of the propagation of surface acoustic waves (SAW's) in a ZnO(001)–diamond(001)[100]–Si(001)[100] structure is presented. For computation, we used the latest available material constants. We compute the modal dispersion, i.e. the phase velocities for all existing modes, as a function of the ZnO thickness for various thicknesses of the diamond film. For thin ZnO films, the velocity values for all modes are higher than those for SAW's on conventional substrates. The attractive modes and layer thicknesses have been identified, and modes with sufficiently large electromechanical coupling constants and velocities of more than 10000 m/s have been found.

1 Introduction

Recently, high–quality polycrystalline diamond–like carbon thin–films have been successfully deposited on silicon (Si) substrates [1]. Among the many promising applications foreseen for such thin diamond–like carbon films is their use in the surface acoustic wave (SAW) area [2,3], since diamond exhibits the highest acoustic wave velocity of all materials due to its high Young's modulus [4]. Velocities of modes propagating in planar thin–film structures are expected to be more than twice as high as with common SAW materials (such as quartz, $LiNbO_3$ and $LiTaO_3$). If we assume, for example, a wave velocity of 10000 m/s, a transducer with quarter wavelength–width electrodes in fundamental mode operation, and a line width of, say, 0.5 μm, we get a frequency of 5 GHz. For comparison, the leaky surface acoustic wave (LSAW) on 36°rot–YX $LiTaO_3$

would result in a frequency of only 2.1 GHz. Since both diamond and silicon are non–piezoelectric materials, usually ZnO (which is an excellent piezoelectric medium [5]) is applied as a superstrate. Thus, the use of thin–film diamond for SAW applications demands for a thorough theoretical study of the SAW propagation in three–layered media composed of a piezoelectric superstrate, a diamond film, and a substrate. In what follows, we will investigate ZnO–diamond–Si thin–film structures using an accurate modal analysis which fully includes anisotropy. The method is based upon a combination of matrix–, partial wave– and simplex–analysis.

2 Theory

Let us consider a half–space of a Si substrate layered with a diamond film, of thickness h_1, and a piezoelectric layer of ZnO, of thickness h_2, on top, as shown in Fig. 1.

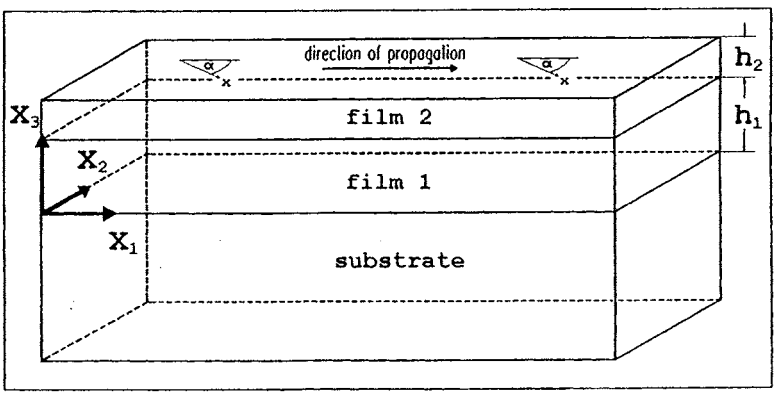


Fig. 1: Si substrate coated with diamond film (1) and zinc oxide film (2)

A Cartesian coordinate system is introduced so that the SAW's travel in the x_1 direction along the surface whose normal is in the x_3 direction with the substrate being in the lower halfspace $x_3 < 0$. This coordinate system may have an arbitrary orientation with respect to the crystal axes provided the tensors describing the media are properly transformed to the (x_1,x_2,x_3) coordinates. Formulas which govern this transformation may be found in Reference [6].

Elastic wave propagation in an anisotropic medium is generally governed by the mechanical equations of motion and (if the material is piezoelectric) by Maxwell's equations. Since the velocities of the surface acoustic waves are many magnitudes less than those of electromagnetic waves, the quasistatic approximation of Maxwell's equations can be used. Each region of the sandwiched structure shown in Fig. 1 is characterized by its own set of constants which are as follows: elastic stiffness, piezoelectric and dielectric tensors c_{ijkl}, e_{kij} and ϵ_{ik}, respectively, and mass density ρ. In general, the three particle displacements u_i and the potential ϕ must satisfy the following set of four coupled wave equations in each medium:

$$\sum_i \sum_k \left[\sum_l c_{ijkl} \frac{\partial^2 u_k}{\partial x_i \partial x_l} + e_{kij} \frac{\partial^2 \phi}{\partial x_k \partial x_i} \right] = \rho \frac{\partial^2 u_j}{\partial t^2} \tag{1}$$

$$\sum_k \sum_i \left[\sum_l e_{ikl} \frac{\partial^2 u_k}{\partial x_i \partial x_l} - \epsilon_{ik} \frac{\partial^2 \phi}{\partial x_k \partial x_i} \right] = 0 . \tag{2}$$

The indices in equations (1) and (2) change from 1 to 3. Assuming lossless media, for the wave solutions in the substrate s, the films f_1 and f_2, and the vacuum v the following partial wave ansatz is made:

$$u_{in}^q = A_{in}^q \exp(jk\alpha_n^q x_3) \exp[jk(x_1 - vt)] \quad i=1,2,3 \tag{3}$$

$$\phi^q = B_n^q \exp(jk\alpha_n^q x_3) \exp[jk(x_1 - vt)] \tag{4}$$

We have:

$q = s$	in the substrate $x_3 < 0$,
$q = f_1$	in the diamond film $0 \leq x_3 < h_1$,
$q = f_2$	in the zinc oxide film $h_1 \leq x_3 \leq h_1+h_2$, and
$q = v$	in the vacuum $x_3 > h_1+h_2$
	with $A_{in}^v = 0$ (i=1,2,3), and
	$\alpha_n^v = +j$.

Here, A_{in}^q and B_n^q are the normalized partial amplitudes, and α_n^q are the transverse decay constants in x_3–direction. Wavenumber and velocity are denoted k and v, respectively. Substituting the formulation of the solution into equations (1) and (2) we obtain for each medium (in general) a set of four homogeneous equations for the relative amplitudes A_{in}^q and B_n^q, i.e each region is described by a 4x4 system matrix. The system matrices have a nontrivial solution if

their determinants of coefficients vanish. This condition leads to polynomial equations of the 8th order in the transverse decay constants. For a given value of the velocity v we obtain in general four eigenvalues α_n^q for a physical substrate solution (α–solutions with a negative imaginary part only) and eight eigenvalues α_n^q for a physical film solution. For every set of eigenvalues we find the eigenvector of normalized amplitudes.

The superposition of the partial waves (four in the substrate and eight in each layer) leads to the general physical solution of equations (1) and (2):

(i): substrate $x_3 < 0$:

$$u_{in}^s = \sum_{n=1}^{4} W_n A_{in}^s \exp(jk\alpha_n^s x_3) \exp[jk(x_1 - vt)] \tag{5}$$

$$\phi^s = \sum_{n=1}^{4} W_n B_n^s \exp(jk\alpha_n^s x_3) \exp[jk(x_1 - vt)] \tag{6}$$

(ii): diamond film $0 \leq x_3 < h_1$:

$$u_{in}^{f1} = \sum_{n=5}^{12} W_n A_{in}^{f1} \exp(jk\alpha_n^{f1} x_3) \exp[j(x_1 - vt)] \tag{7}$$

$$\phi^{f1} = \sum_{n=5}^{12} W_n B_n^{f1} \exp(jk\alpha_n^{f1} x_3) \exp[jk(x_1 - vt)] \tag{8}$$

(iii): zinc oxide film $h_1 \leq x_3 \leq h_1+h_2$:

$$u_{in}^{f2} = \sum_{n=13}^{20} W_n A_{in}^{f2} \exp(jk\alpha_n^{f2} x_3) \exp[j(x_1 - vt)] \tag{9}$$

$$\phi^{f2} = \sum_{n=13}^{20} W_n B_n^{f2} \exp(jk\alpha_n^{f2} x_3) \exp[jk(x_1 - vt)] \tag{10}$$

(iv): vacuum $x_3 > h_1+h_2$:

$$\phi^v = B^v \exp(-kx_3) \exp[jk(x_1 - vt)] \tag{11}$$

with W_n being weighting factors, i.e. normalized amplitudes. These equations have to fulfill the mechanical and electrical boundary conditions at the interfaces

$x_3 = 0$, $x_3 = h_1$, and $x_3 = h_1+h_2$ which are as follows (i = 1,2,3):

(i): at the interface $x_3 = 0$:

$$u_i^s = u_i^{f_1} \tag{12}$$

$$T_{3i}^s = T_{3i}^{f_1} \tag{13}$$

$$(\vec{E}^s - \vec{E}^{f_1}) \times \vec{n} = \vec{0} \tag{14}$$

$$(\vec{D}^s - \vec{D}^{f_1}) \cdot \vec{n} = 0 \tag{15}$$

(ii): at the interface $x_3 = h_1$:

$$u_i^{f_1} = u_i^{f_2} \tag{16}$$

$$T_{3i}^{f_1} = T_{3i}^{f_2} \tag{17}$$

$$(\vec{E}^{f_1} - \vec{E}^{f_2}) \times \vec{n} = \vec{0} \tag{18}$$

$$(\vec{D}^{f_1} - \vec{D}^{f_2}) \cdot \vec{n} = 0 \tag{19}$$

(iii): at the interface $x_3 = h_1+h_2$:

$$T_{3i}^{f_2} = 0 \tag{20}$$

$$(\vec{E}^{f_2} - \vec{E}^{v}) \times \vec{n} = \vec{0} \tag{21}$$

$$(\vec{D}^{f_2} - \vec{D}^{v}) \cdot \vec{n} = 0 \tag{22}$$

Herein, T, $\vec{E}$, and $\vec{D}$ denote mechanical stress, electrical field strength, and electrical displacement, respectively. These conditions can be combined to form a 20x20 matrix – the boundary matrix. For a physical solution, the determinant of the boundary matrix must also vanish. The general solution must cause the determinants of the system matrices as well as the determinant of the boundary matrix to vanish. Because of their complicated structures these matrices cannot be transformed into equations for the velocity v and formulas for the decay constants α. Thus, the following numerical procedure has been used. For a chosen value for v the eigenvalues and eigenvectors from the set of equations for the relative amplitudes A_{in}^q and B_n^q are found and substituted into the determinant of the boundary matrix. If this determinant is zero with a prescribed accuracy the chosen and computed parameters are correct. Then the weighting factors W_n are computed and after the substitution in equations (5) to (11) we obtain the solution. The absolute amplitudes can be determined from the power flow. In the second case, if the determinant of the boundary matrix has a nonzero value, we vary the phase velocity v using a simplex algorithm, and the procedure must be repeated unless the determinant of the boundary conditions is zero.

3 Numerical Results

We will present results for the SAW propagation in a ZnO(001)–diamond(001)[100]–Si(001)[100] structure. The material data were as follows [7,8]:

Si: $c_{11} = 1.657 \cdot 10^{11}\ \mathrm{Nm^{-2}}$, $c_{12} = 0.639 \cdot 10^{11}\ \mathrm{Nm^{-2}}$,
$c_{44} = 0.7956 \cdot 10^{11}\ \mathrm{Nm^{-2}}$, $\epsilon_{11} = \epsilon_{33} = 11.7\ \epsilon_0$, $\rho = 2332\ \mathrm{kgm^{-3}}$

Diamond: $c_{11} = 10.8 \cdot 10^{11}\ \mathrm{Nm^{-2}}$, $c_{12} = 1.25 \cdot 10^{11}\ \mathrm{Nm^{-2}}$, $c_{44} = 5.76 \cdot 10^{11}\ \mathrm{Nm^{-2}}$,
$\epsilon_{11} = \epsilon_{33} = 5.67\ \epsilon_0$, $\rho = 3512\ \mathrm{kgm^{-3}}$

ZnO: $c_{11} = 2.097 \cdot 10^{11}\ \mathrm{Nm^{-2}}$, $c_{12} = 1.211 \cdot 10^{11}\ \mathrm{Nm^{-2}}$,
$c_{13} = 1.05 \cdot 10^{11}\ \mathrm{Nm^{-2}}$, $c_{33} = 2.109 \cdot 10^{11}\ \mathrm{Nm^{-2}}$,
$c_{44} = 0.4247 \cdot 10^{11}\ \mathrm{Nm^{-2}}$, $e_{15} = -0.480\ \mathrm{Asm^{-2}}$, $e_{31} = -0.573\ \mathrm{Asm^{-2}}$,
$e_{33} = 1.32\ \mathrm{Asm^{-2}}$, $\epsilon_{11} = 8.55\ \epsilon_0$, $\epsilon_{33} = 10.2\ \epsilon_0$, $\rho = 5680\ \mathrm{kgm^{-3}}$

3.1 Dispersion

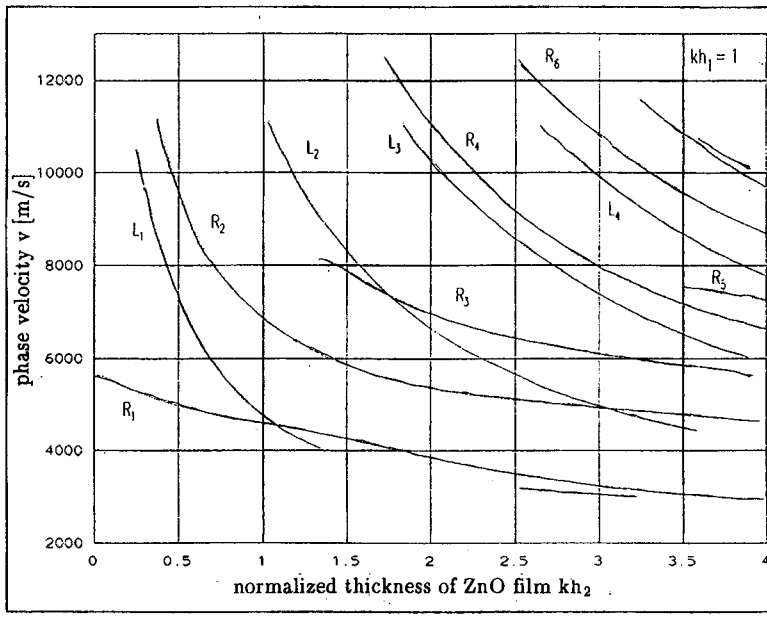


Fig. 2: Computed phase velocity as a function of normalized ZnO film thickness kh_2 for a normalized diamond film thickness $kh_1 = 1$

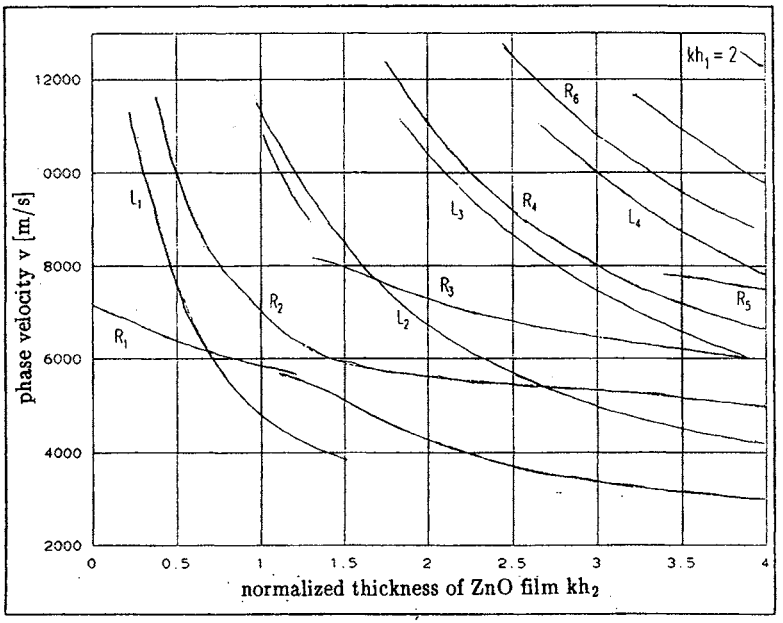


Fig. 3: Computed phase velocity as a function of normalized ZnO film thickness kh_2 for a normalized diamond film thickness $kh_1 = 2$

Figs. 2 to 4 show the computed phase velocities in dependence of the normalized ZnO thickness for three normalized thicknesses of the diamond film, $kh_1 = 1$, $kh_1 = 2$, and $kh_1 = 4$. The wavetypes of up to the sixth mode are given. For thin ZnO films, the phase velocities for all modes are higher than those for SAW's on conventional substrates. We found that the velocities increase with increasing diamond thickness, but are nearly constant for kh_1–values higher than about $kh_1 = 4$. Thus, thin diamond films can be used for a practical substrate .

3.2 Electromechanical Coupling Constant

The electromechanical coupling constant k^2 is defined as

$$k^2 = \frac{2(v_0 - v_s)}{v_0} \tag{23}$$

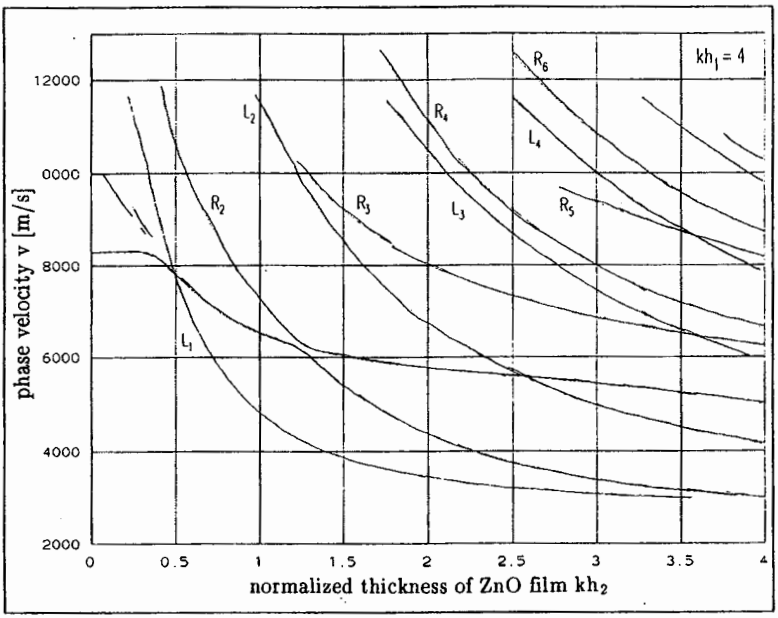


Fig. 4: Computed phase velocity as a function of normalized ZnO film thickness kh_2 for a normalized diamond film thickness $kh_1 = 4$

with v_o and v_s being the phase velocity of the free and metalized surface, respectively [2]. For a high k^2, let the interdigital transducer (IDT) be arranged at the interface $x_3 = h_1$, i.e. between ZnO film and diamond film. Computing v_o, the influence of the IDT can be neglected. Computing v_s, the IDT is modeled as a metal film with zero thickness and infinite conductivity. Thus, the mechanical boundary conditions formulated above apply without any change. The electrical boundary conditions equations. (18) and (19) however have to be modified such that

$$\vec{E}^{f_1} \times \vec{n} = \vec{E}^{f_2} \times \vec{n} = \vec{0} \tag{24}$$

$$(\vec{D}^{f_1} - \vec{D}^{f_2}) \cdot \vec{n} = \rho_s \tag{25}$$

where ρ_s is the surface charge density at $x_3 = h_1$. For given weighting factors W_n, ρ_s can be calculated from equation (25). Fig. 5 shows the electromechanical coupling constant in dependence of the ZnO thickness for $kh_1 = 4$. As is seen,

k^2 values higher than 5% occur for the R_1–mode near $kh_2 = 2.5$. The R_1–mode is somewhat slow compared to the fast R_2–mode, for which k^2 values near 1% at $kh_2 = 1.5$ can be achieved.

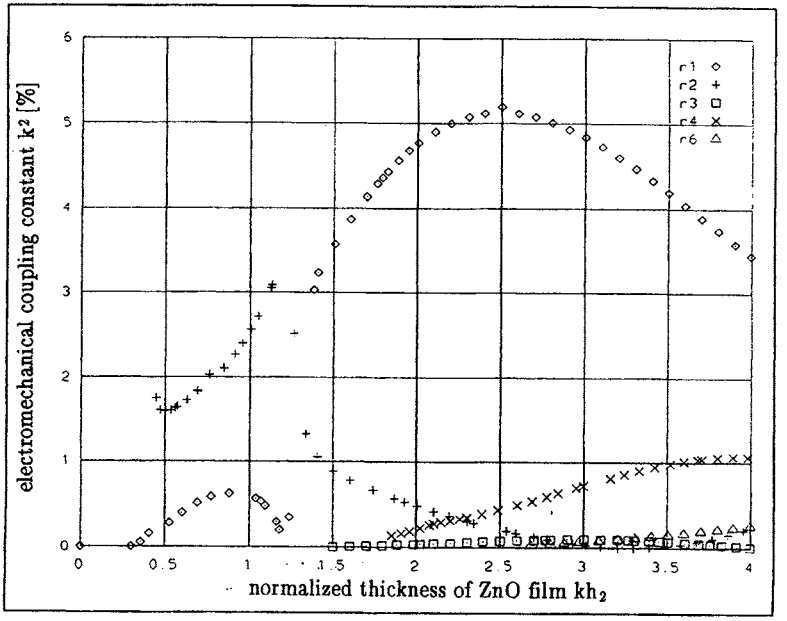


Fig. 5: Electromechanical coupling constant k^2 as a function of normalized ZnO film thickness kh_2 for a normalized diamond film thickness of $kh_1 = 4$

4. Conclusion

A full wave analysis tool for computing the phase velocity and electromechanical coupling constant of SAW propagation in multilayer structures has been demonstrated. Numerical results have been given for a ZnO(001)–diamond(001)[100]–Si(001)[100] structure. As wavetypes for practical use in particular the lower modes such as the zeroth to second modes have been identified. Phase velocities higher than 10000 ms^{-1} with moderate electromechanical coupling constants are predicted. The theoretical results indicate the possibility that polycrystalline diamond–like carbon thin–films may be used as a substrate for future SAW com–

ponents operating in the frequency range of up to 5 GHz. The recent progress in the polycrystalline diamond growth techniques are very encouraging. From the mass–production point of view however, many technological engineering issues must still be addressed before technical applications can become possible.

Acknowledgement

The authors thank W. Ruile from Siemens AG München, Germany for stimulating this work. They are gratefully to G. Jacoby from the Technische Universität München, Germany for providing basic software tools.

References

[1] D. Schneider, H.–J. Scheibe, T. Schwarz, P. Hess, "Characterization of thin diamond–like carbon films by ultrasonic surface waves", Diamond and Related Materials 2, pp. 1396–1401, 1993

[2] C. Campbell, *Surface Acoustic Waves and their Signal Processing Application.* San Diego: Academic Press, 1989

[3] C.C.W. Ruppel, R. Dill, A. Fischerauer, G. Fischerauer, W. Gawlik, J. Machui, F. Müller, L. Reindl, W. Ruile, G. Scholl, I. Schropp, K.C. Wagner, "SAW devices for comsumer applications", IEEE Trans. Ultrason., Ferroelec., Freq.Contr. UFFC–40, S. 438–452, 1993

[4] K. Yamanouchi, N. Sakurai, T. Satoh, "SAW propagation characteristics of piezoelectric thin film/diamond structure", IEEE Ultrasonics Symposium Proceedings, pp. 351–354, 1989

[5] F.S. Hickernell, "The microstructural characteristics of thin–film zinc oxide for SAW transducers", IEEE Ultrasonics Symposium Proceedings, pp. 239–242, 1984

[6] V.M. Ristic, *Principles of Acoustic Devices.* New York: Wiley, ch. 6, 1983

[7] B.A. Auld, *Acoustic Fields and Waves in Solids.* New York: Wiley, 1973

[8] H. Nakahata, A. Hachigo, S. Shikata, N. Fujimori, "High frequency surface acoustic wave filter using ZnO/diamond/Si structure", IEEE Ultrasonics Symposium Proceedings, pp. 377–380, 1992

Selfadjusting SAW IF Channel Filtering For Mobile Communication Systems

C. Rheinfelder[1], G. Scholl[2], R. Weigel[1], P. Russer[1+3]

[1] Technische Universität München, Lehrstuhl für Hochfrequenztechnik,
Arcisstr. 21, D–80333 München, Germany
[2] Siemens AG, ZFE ST KM 11,
Otto–Hahn–Ring 6, D–81739 München, Germany
[3] Ferdinand Braun Institut für Höchstfrequenztechnik,
Rudower Chaussee 5, D–12489 Berlin, Germany

Summary – Surface acoustic wave (SAW) filters with high sidelobe suppression, steep passband skirts and small chip sizes are key components for intermediate frequency (IF) channel selection in modern mobile communication systems. Usually, these IF filters are fabricated on quartz substrates with precise photolithographic techniques. However, unavoidable fabrication tolerances, temperature drift and aging require some design reserve for achieving a specific filter characteristic, thus degrading the neighbour–channel selectivity. We report on a novel receiver design which does not demand for a design reserve, thus improving the filter performance drastically. As an IF bandpass filter, a new six–pole transverse–mode coupled SAW resonator filter (TCF) has been incorporated in the receiver path.

1 Introduction

In recent years, surface acoustic wave (SAW) resonator filters have found widespread application in mobile communication systems [1,2]. They offer very low insertion loss over the passband, steep passband skirts, high close–in selectivity, and small chip size. Due to their superior amenability to mass–production, SAW filters can be fabricated at a very low price. In the intermediate frequency (IF) range of the upcoming European GSM standard and the PCN system, SAW filters are currently used in the receiver (Rx) path. In order to obtain miniaturized SAW filters, IF frequencies as high as possible are attractive. However, the higher the IF frequency, the tighter the required tolerances for the center fre–

quency. Up to now, some design reserve must be provided to account for temperature and manufacturing parameter variations, thus degrading the neighbour–channel selectivity. We report on a design principle which overcomes these drawbacks and does not demand for a passband reserve anymore. The filter can now be designed to cover the given channel accurately and for optimum selectivity. Thereby, the required manufacturing tolerances can be relaxed drastically.

2 Receiver Schematic

Fig. 1 shows the basic block diagram schematic of the new receiver circuit together with the frequency levels [3]. The block diagram can be subdivided in the following four building blocks: (i) the RF input; (ii) the self–aligning IF filter configuration; (iii) the baseband output; and (iv) the external local oscillator LOE. In Fig. 1 BF denotes the baseband frequency. The RF signal is heterodyned to the IF frequency by a first mixer MX1. The channel filter in the IF stage is a narrowband SAW resonator filter. This filter shares the quartz substrate with a SAW resonator which is used as the frequency controlling element in the feedback loop of the oscillator LOS. The operation is based on the fact that all parameter fluctuations (manufacturing tolerances, temperature drift, and aging) cause the same variation of the resonance frequency df of the SAW resonator and the SAW resonator filter.

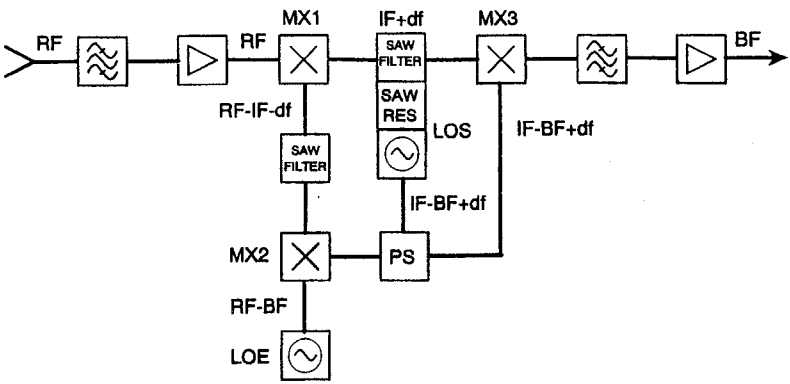


Fig. 1: Block diagram schematic of the receiver circuit

Let us assume a small shift in the center frequency df of the channel filter. Then LOS will have the same frequency shift df and will oscillate at the frequency IF–BF+df, converting the incoming signal of LOE to the frequency RF–IF–df. This signal is the local oscillator signal for mixer MX1, converting the RF signal directly to the center frequency IF+df of the channel filter. With the help of mixer MX3 the IF signal is heterodyned to the baseband frequency BF. An additional SAW low–loss filter is inserted between MX1 and MX2 to suppress unwanted spurious harmonics. The RF and IF frequency of the present prototype circuit are at 900 MHz and 475 MHz, respectively. PS denotes a power splitter.

3 Single Substrate SAW Device Incorporating TCF and Resonator

Featuring very low insertion loss across the passband, steep passband skirts, and a high close–in selectivity on smallest chip size, transverse–mode coupled resonator filters (TCF's) are typically used as narrowband filters in mobile communication systems [4–7]. The schematic layout of a two–pole resonator filter is given by Fig. 2. It consists of two identical tracks. Each track can be viewed as a one–port resonator. Both tracks together form an acoustical waveguide supporting the fundamental mode M_0 and the first antisymmetric mode M_1. In the upper track both waveguide modes are excited. The interdigital transducer (IDT) in the lower track couples to the propagating modes and connects the filter to an external load.

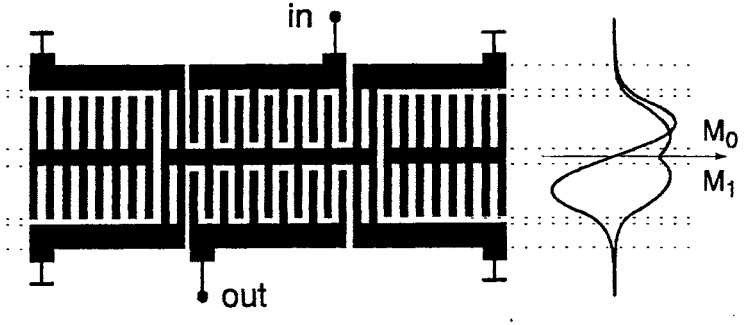


Fig. 2: Two–pole transverse–mode coupled resonator filter (TCF)

To improve the skirt steepness several two–pole sections are cascaded. We de–

veloped a six–pole filter which is composed of three two–pole sections [8]. The 3 dB–bandwidth of the resonator filter was approximately 190 kHz. At 1200 kHz offset from the center frequency a selectivity of more than 30 dB was achieved. A two–port resonator with the same sensitivity to fabrication tolerances and the same temperature behavior as the TCF was also designed. Both TCF and resonator were integrated on a single rotated Y–cut quartz substrate and mounted in a TO–8 package.

4 Oscillator Design

Fig. 3 gives a picture of packaged receiver circuit the diagram of which is shown in Fig. 4. A detailed description of the individual circuit parts may be found in Reference [9]. The SAW chip is clearly seen with the TCF to the left, and the resonator to the right. The framed part in Fig. 4 surrounds the SAW resonator–stabilized oscillator LOS followed by a buffer amplifier. The nonlinear simulation and optimization of the oscillator circuit [10] together with the buffer amplifier (i.e. the framed part of Fig. 4) were carried out by use of the commercial software tool Microcap IV [11].

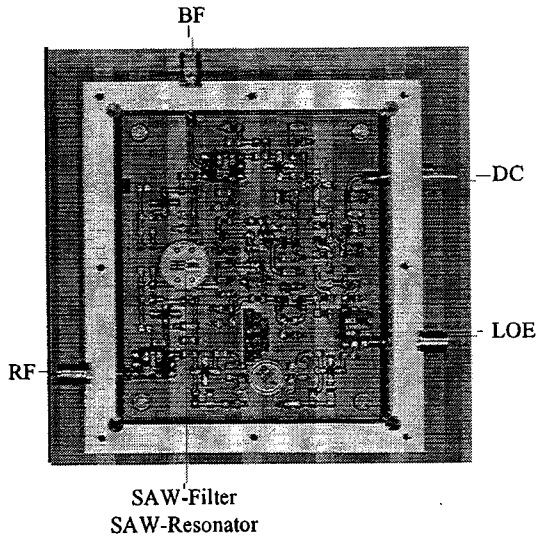


Fig. 3: Packaged receiver circuit

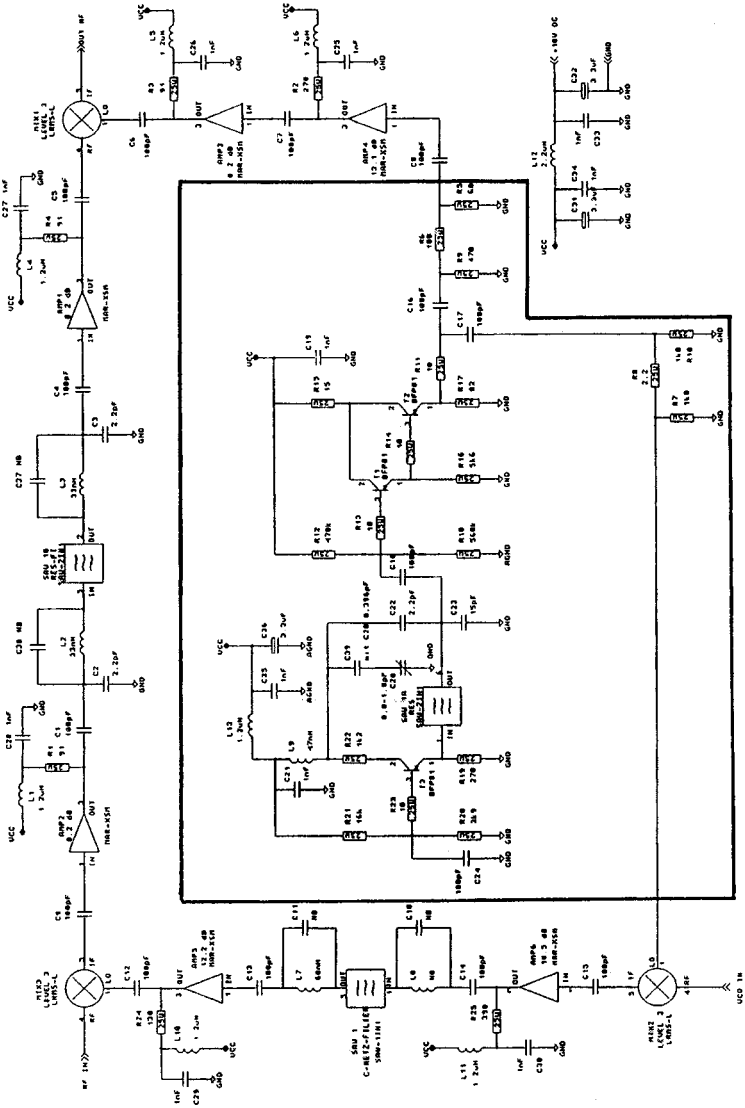


Fig. 4: Circuit diagram of complete receiver

Using Monte–Carlo analysis, the circuit elements were chosen in such a way that

parameter variations should have had no influence on the oscillator performance. The oscillator elements, in particular the resistor connected to the emitter which provides negative feedback, and the resistor connected to the collector were optimized for effective suppression of spurious higher harmonics. Employing a modified Darlington configuration we achieved a very high input impedance and a low output impedance for the buffer amplifier. The transient analysis outcome of the optimized oscillator simulation is shown in Fig. 5.

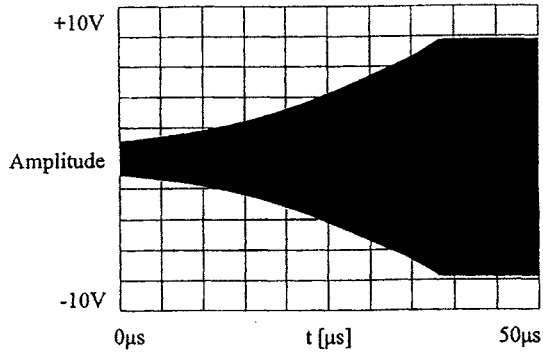


Fig. 5: Transient analysis of the local oscillator

5 Measurement and Experimental Results

The experimental test setup is shown in Fig. 6. We converted the baseband signal back to the RF frequency. Thus, the behavior of the IF path could easily be measured with a network analyzer. Fig. 7 shows the power flow chart of the modified receiver prototype. Additionally, we substituted the IF bandpass filter by a second resonator with the same sensitivity to fabrication tolerances, which simplifies the detection of center frequency variations. For an optimum performance of the radio receiver the power levels of the oscillator signals at the mixers MX1 and MX3 in Fig. 1 must be stable against temperature drift and variations of the circuit elements. In Fig. 8 we see two resonator pairs with different metalization heights. The resonators to the left operate as the frequency controlling elements in the oscillator circuit LOS shown in Fig. 1. The resonators to the right represent the SAW IF filters. We plugged each resonator pair in our receiver

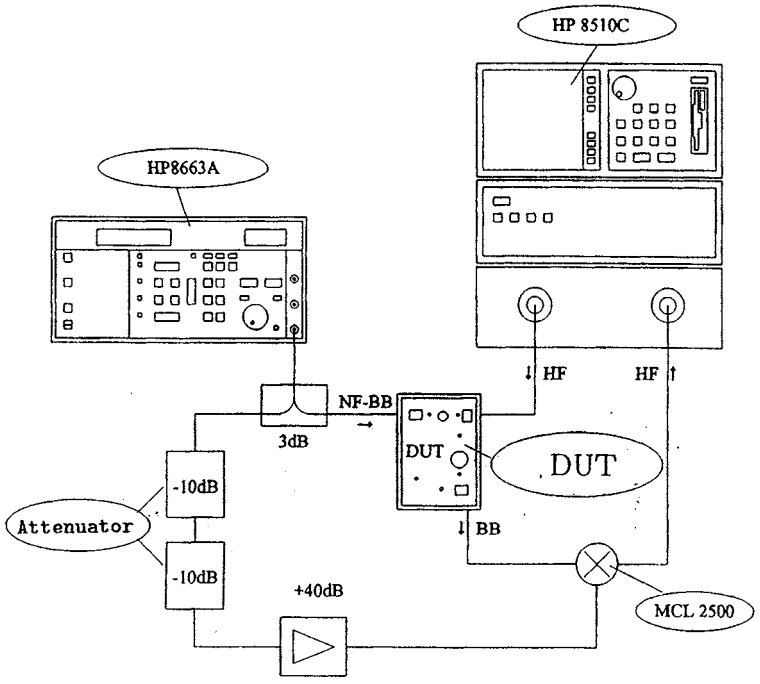


Fig. 6: Measurement test setup

circuit and measured the frequency response in the way described above. We found that a mean frequency offset of about 76 kHz between the resonator pairs caused by the different metalization heights was reduced to approximately 9 kHz. Thus, the frequency stability was improved by a factor of 8. Parameter variations in the photolithography result in a 12 kHz offset of the resonance frequencies. We also measured the frequency response of a resonator pair at two different temperatures (20 ^{0}C and 75 ^{0}C). Usually this temperature difference would result in a frequency shift of about 45 kHz. Here however, the difference in the resonance frequency of the receiver transfer functions was only 1.25 kHz. Thus, the temperature stability was improved by a factor of 36. Fig. 9 shows the frequency responses of the receiver circuit with a six–pole resonator filter at two different temperatures in the IF stage. A temperature difference of approximately 55 ^{0}C caused a frequency shift of the upper passband skirt of 2.9 kHz to higher frequen–

cies or a broadening of the relative bandwith of the filter of approximately 1.5 %.

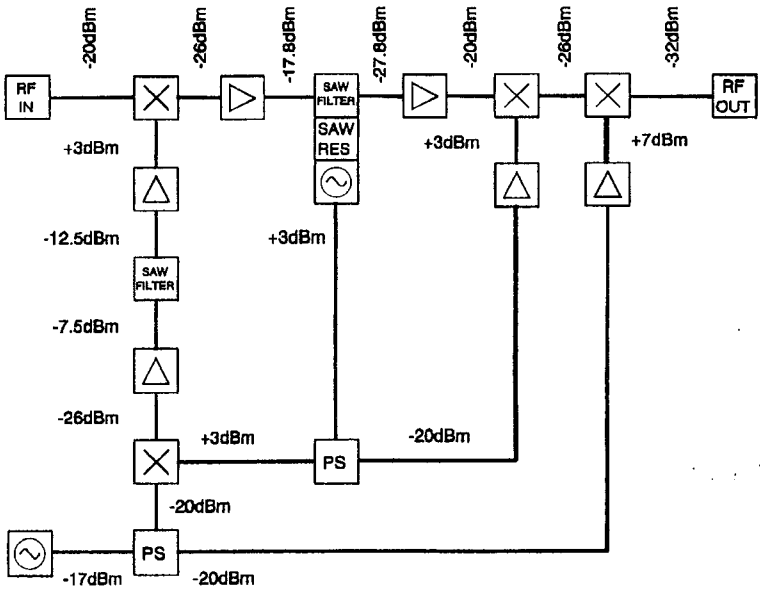


Fig. 7: Power flow chart of the modified receiver circuit

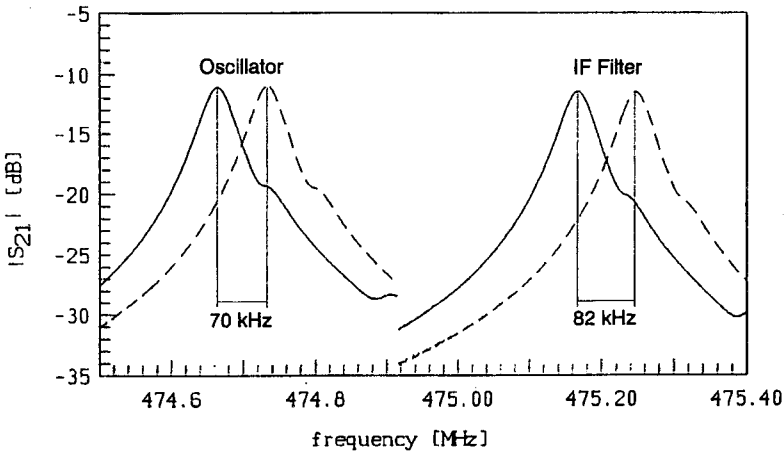


Fig. 8: Responses of two resonator pairs with different metalization heights

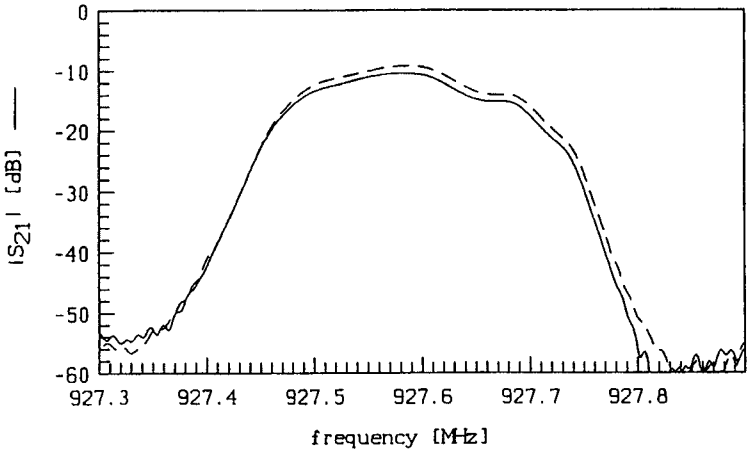


Fig. 9: Frequency responses of the receiver circuit with a TCF at 20 °C (——) and at 75 °C (– – –) in the IF stage

6 Conclusion

The present work demonstrates the feasibility of a new SAW receiver design for the compensation of frequency variations caused by parameter fluctuations of the SAW IF filter. In contrast to the conventional receiver design with a fixed IF we converted the RF signal exactly to the varying center frequency of the SAW channel filter in the IF stage. By doing this we were able to significantly enhance the frequency stability of our prototype receiver by more than one order. As an essential part of the new receiver architecture we integrated a SAW resonator and a SAW resonator filter with the same sensitivities to fabrication and temperature variations on a single chip.

Acknowledgement

The authors thank S. Berek, H. Zottl, and B. Bienert from Siemens AG München for the ready fabrication of the SAW devices. They are also gratefully to P. Babel and J. Geisel for the many helpful discussions.

References

[1] C.C.W. Ruppel, R. Dill, A. Fischerauer, G. Fischerauer, W. Gawlik, J. Machui, F. Müller, L. Reindl, W. Ruile, G. Scholl, I. Schropp, K.C. Wagner, "SAW devices for comsumer applications", IEEE Trans. Ultrason., Ferroelec., Freq.Contr. UFFC–40, S. 438–452, 1993

[2] K. Shibayama, K. Yamanouchi (Eds.), *Proceedings of International Symposium on Surface Acoustic Wave for Mobile Communications.* (Tohoku University, Sendai, Japan) 150th Committee in Japan Society for the Promotion of Science, Tokyo Press, Tokio, 1992

[3] K. Yajima, *European Patent Application No. 91111361.1.* 1991

[4] H.F. Tiersten, R.C. Smythe, "Guided acoustic surface wave filters", IEEE Ultrasonics Symposium Proceedings, pp. 293–294, 1975

[5] M. Tanaka, T. Morita, K. Ono, Y. Nakazawa, "Narrow bandpass filter using double–mode SAW resonators on quartz", Proceedings 38th Annual Frequency Control Symposium, pp. 286–293, 1984

[6] C.S. Hartmann, D.P. Chen, J.Heighway, "Modeling of SAW transversely coupled resonator filters using coupling–of–modes modeling technique", IEEE Ultrasonics Symposium Proceedings, pp. 39–43, 1992

[7] G. Scholl, W. Ruile, P. Russer, "P–Matrix modeling of transverse–mode coupled resonator filters," IEEE Ultrasonics Symposium Proceedings, 1993, *to be published*

[8] C. Rheinfelder, G. Scholl, R. Weigel, C.C.W. Ruppel, P. Russer, "Feedback–controlled compensation of parameter fluctuations of mobile radio SAW IF channel filters", Proceedings IEEE MTT–S Symposium, 1994, *to be published*

[9] C. Rheinfelder, *Aufbau und Untersuchung einer selbstjustierenden ZF–Anordnung zur Kanalfilterung in Mobilfunksystemen.* Diploma Thesis, Technische Universität München, 1993, *in German*

[10] M.M. Driscoll, "Low–noise microwave signal generation using bulk– and surface–acoustic–wave resonators", IEEE Transactions on Sonics and Ultrasonics, vol. SU–35, no. 3, pp. 426–434, 1988

[11] M.S. Roden, *Microcap IV. An Electronic Circuit Analysis Program.* The Benjamin–/Cummings Publishing Company Inc., 1993

SCATTERING OF SURFACE ACOUSTIC WAVES AT A FINITE NUMBER OF MASSIVE ELECTRODES

H.P. REICHINGER and A.R. BAGHAI-WADJI

Department of Applied Electronics, IAEE 3592,
Vienna University of Technology,
Gußhausstraße 27–29, A-1040 Vienna, Austria

Abstract

In recent works we have shown that the combination of Finite-Element-Method (FEM) and Boundary-Element-Method (BEM) makes possible the analysis of **periodic** Surface Acoustic Wave (SAW) structures with massive electrodes. In this hybrid-technique we apply FEM to electrode regions only and use BEM in the semi-infinite substrate as well as in the adjacent free space. This solution strategy allows the analysis of arbitrarily complicated electrode geometries with a minimum amount of computer resources. In the present work we apply this method to structures consisting of **a finite number** of massive electrodes. The details of the underlying theory are illuminated by discussing the scattering of surface waves at a single electrode. Numerical results concerning reflection- and transmission-coefficients of the incident SAW and the conversion of surface into bulk waves are presented. Power reflection-coefficients of magnitudes up to 15%, as a result of SAW-interaction with a single massive electrode have been calculated. Results concerning electrode's vibrational states and stress distributions generated at the electrode/substrate interfaces conclude the paper.

1 Introduction

Recently we presented a method to analyze the effect of massloading of a periodic array of electrodes on a piezoelectric substrate [1], [2]. In this technique the electrodes may have arbitrary cross-section and the elastic material can be inhomogeneous

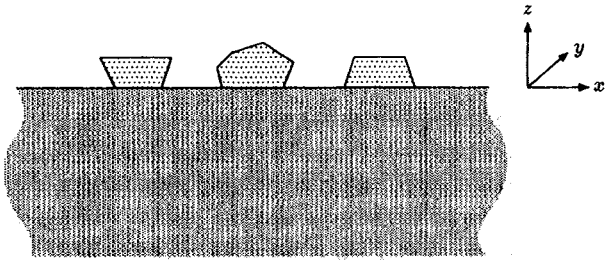


Figure 1: 2D-Model of a SAW-device with a finite number of electrodes.

and anisotropic. This rigorous (full-wave) approach is easy-to-implement and, at the same time, efficient in terms of computation time and memory requirements. This results from the following facts: (a) we restrict the flexible, but numerically expensive Finite Element Method (FEM) to the comparatively small electrode regions. (b) By definition, the application of Boundary Element Method (BEM) reduces the number of dimensions of the problem by one. Using proper Green's functions the stress-free condition on the unloaded portion of the substrate surface will be automatically fulfilled.

This paper extends the above-mentioned works to investigate the massloading effect of a finite number of electrodes in SAW-devices. Assuming a harmonic time-dependence according to $e^{j\omega t}$, we determine the frequency dependence of the power reflection-coefficient r_P and transmission-coefficient of an incident SAW at elastic electrodes. Our analysis shows magnitudes of r_P up to 15% of a SAW scattered at a single electrode. Furthermore, results concerning power conversion into bulk waves of an impinging SAW is presented. In addition, examples of the vibrational state of the electrodes and the stress-distribution on the electrode/substrate interfaces are given.

2 Statement of the Problem

Assuming long apertures, we define a two-dimensional model ($\partial/\partial y \equiv 0$) of a SAW-transducer. Consider the structure sketched in Fig.1, in which the used co-ordinate system also is shown. The lower semi-space ($z < 0$) is filled with a piezoelectric material while the upper semi-space ($z > 0$), outside the electrode regions, is free space. On the plane interface ($z = 0$) we assume a finite number of metallic electrodes.

3 Analysis of the Electrodes

For the sake of simplicity we restrict our discussion to one electrode only. Assuming a non-piezoelectric electrode-material, the equations of the mechanic and electric field within the electrode region decouple. The electrical properties of the electrode are described by the values of their conductivity κ and permittivity ε. Mechanically, the electrode is characterized by the mass density ρ and the elastic stiffness tensor c. In our model the site of coupling is the electrode/substrate interface. This property significantly simplifies the analysis by allowing the individual treatment of the mechanic and electric field in the upper semi-space.

Electrical interaction: For frequencies up to a few hundred MHz, the assumption of quasi-static electric field for wave propagation in piezoelectric materials provides sufficiently accurate results. This implies that the electric charges are pushed to the bounding surfaces of the electrode. (A conductive material for the electrode is assumed.) On the other hand, comparatively large values of relative permittivity-constants of the substrate imply that the charges are predominantly concentrated on the electrode/substrate interface. Keeping this fact in mind, we neglect the charges on the free-parts of the electrode boundary, and obtain a model in which the semi-space ($z > 0$) is filled with ε_0 and the charges exist only on the electrode/substrate interface. The reader recognizes that this assumption is not essential, and that a minor addition of routine work allows the inclusion of the charges on the free parts of the electrode's bounding surface.

On metallized regions the charges are regarded as our unknowns. We distinguish two types of boundary conditions: a) for a driven electrode the given constant potential-value φ has to be imposed in metallized regions for the determination of the charge values in this area. b) In the case of a floating electrode the charges and the *a priori* unknown potential-value have to be determined in such a way that the charge integral on the electrode vanishes.

Mechanical interaction: The FEM is best suited for the analysis of the electrode. It can be straightforwardly extended to cover anisotropic and inhomogeneous materials. Furthermore, there are virtually no restrictions to the cross-section geometry of the electrode. A disadvantage of FEM is its rather high consumption of computation time and memory. However, in our method we overcome this disadvantage by limiting the FEM-application to the small electrode-region.

The starting point of our FEM-application results from a variational formulation [4, 5] according to Hamilton's principle in two dimensions

$$\delta \int_{t_1}^{t_2} \left(\int_A \rho \dot{u}_i \dot{u}_i \, dA - \int_A T_{ij} S_{ij} \, dA + \int_s f_i u_i \, ds \right) dt = 0 \qquad (1)$$

which is as fundamental as Newton's equation of motion. Here δ indicates variation with respect to time, u_i and f_i are the vectors of the mechanical displacement and the surface force density. T_{ij} and S_{ij} denote the stress and strain tensors, respectively. Furthermore, A and s, respectively, stand for the area and the surface line of the considered two-dimensional body. Note that Einstein's summation convention is adopted in (1). A dotted variable means its differentiation with respect to time. The first and second integral represent the kinetic and potential energy, respectively. The third integral is an expression for the work performed by f_i.

Restricting to time-harmonic solutions, the application of FEM to (1) yields a set of algebraic equations which in matrix form gives

$$\left(\underline{\underline{M}}_p - \omega^2 \underline{\underline{M}}_k\right) \vec{U} = \underline{\underline{M}}_f \vec{F}. \tag{2}$$

In Eq. (2) $\vec{U}$ and $\vec{F}$, respectively, denote the mechanical displacements in the nodes of the finite elements and the unknown coefficients resulting from the expansion of f_i into (piecewise-constant) basis functions. In this work quadratic ansatz functions within each (triangular) finite element have been used. As the reader will recognize the knowledge of the displacements $\vec{U}^{(I)}$ in the electrode/substrate interface is essential in our method. Thus we rewrite (2) as

$$\vec{U}^{(I)} = \underline{\underline{M}}_{bc} \vec{F}. \tag{3}$$

Eq. (3) is the mechanical boundary condition in discretized form underneath the electrode.

3.1 BEM-Analysis of the Unbounded Regions

So far we have reduced the influence of the electrode to mechanical and electrical boundary conditions on the electrode/substrate interface. The governing electrical equation in the upper half-space, and the electro-mechanical equations in the substrate will be treated via a BEM application.

In free space the electric potential φ obeys Laplace's equation

$$\varphi_{,ii} = 0, \tag{4}$$

whereas in the substrate the four coupled partial differential equations

$$\begin{aligned} c_{ijkl}\, u_{k,li} + e_{kij}\, \varphi_{,ki} &= \rho\, \ddot{u}_j \\ e_{ikl}\, u_{k,li} - \varepsilon_{ik}\, \varphi_{,ki} &= 0 \end{aligned} \tag{5}$$

are valid. Subscripts preceded by a comma denote spatial partial derivation. The tensor of piezoelectric coupling is denoted by e.

A variable-separated solution ansatz can be obtained in the form

$$\begin{aligned} u_i &\propto e^{-jkx} e^{\lambda_i z} e^{j\omega t} \\ \varphi_i &\propto e^{-jkx} e^{\lambda_4 z} e^{j\omega t}. \end{aligned} \tag{6}$$

Using (6) in (4) and (5) and imposing the boundary conditions, we obtain an equation which involves the problem-characterizing 4x4 Green's function matrix $\underline{\underline{G}}$:

$$\begin{pmatrix} u_1(k, z=0) \\ u_2(k, z=0) \\ u_3(k, z=0) \\ \varphi(k, z=0) \end{pmatrix} = \underline{\underline{G}} \cdot \begin{pmatrix} f_1(k, z=0) \\ f_2(k, z=0) \\ f_3(k, z=0) \\ \sigma(k, z=0) \end{pmatrix}. \tag{7}$$

As is evident $\underline{\underline{G}}$ establishes a relation between u_i and φ, and f_i and the charge density σ.

The following convention is used

$$u(x,z) = \int_{-\infty}^{\infty} \frac{dk}{2\pi}\, u(k,z) e^{-jkx} \tag{8}$$

$$u(k,z) = \int_{-\infty}^{\infty} dx\; u(x,z) e^{jkx}. \tag{9}$$

Note that $\underline{\underline{G}}$ satisfies the following boundary- and interface conditions: φ is continuous across the interface-plane (z=0) and vanishes for z tending to infinity in the upper half-space. Sommerfeld's radiation condition is required for z tending to infinity in the lower half-space. The jump discontinuity of the dielectric displacement across the interface equals the charge density, and the stress distribution normal to the interface is equal to the force density f_i.

The boundary conditions which have not been addressed so far are the following: a) φ in metallized regions is known, b) u_i and T_{i3} are continuous across the interface in the electrode regions.

The unknown charges σ and forces f_i (equivalent to stresses T_{i3}) are expanded into basis functions with unknown coefficients and are determined in conjunction with (3) by the Method of Moments [7].

4 Numerical Results

The results in Figs. 2–12 have been obtained under the following conditions: i) a single electrode is assumed on a YZ-$LiNbO_3$ substrate. ii) The electrode is modeled as being purely elastic. (Performing the calculations in Figs. 6 and 7, the electrodes'

electrical conductivity is also accounted for.) iii) The electrode has the mass density ρ and the stiffness c of polycrystalline gold in Fig. 7, and the properties of polycrystalline aluminum in the remaining figures. iv) In all calculations presented here the width of the electrode/substrate interface is $4.3595\mu m$. This particular value is equal to a quarter of the wavelength of a wave propagating along the free substrate surface at the frequency 200 MHz. v) The cross-section of the electrode is rectangular for the calculations presented in Figs. 2–7. The electrode's cross-section is trapezoidal for the results in Fig. 8. The assumed cross-section geometry is shown in Fig. 9 for the calculations provided in Figs. 10–12.

In Figs. 2–4 the percentage of the incident SAW-power reflected (r_P), transmitted, and converted into bulk waves by a rectangular aluminum electrode is shown. The information in Figs. 5 and 6 is provided for comparison of reflection assuming both a conductive and a non-conductive electrode. For a conductive gold-electrode r_P is shown in Fig. 7. The dependence of r_P on the corner angle of a trapezoidal electrode can be found in Fig. 8.

The shape and FEM-mesh [6] of a non-symmetric electrode is sketched in Fig. 9. For this electrode r_P is given in Fig. 10 for an incident SAW from the left- and the right-hand side, respectively. (For comparison purposes, the dashed line shows r_P due to a rectangular electrode without the indentation of the non-symmetric electrode.) As can be seen, for frequencies above $\approx$ 300 MHz, r_P strongly depends on the direction of the incident SAW. The following example concludes the paper: We investigate the vibrational state of the electrode at two 'critical' frequencies f_1 and f_2 ($f_1 \approx$ 300 MHz: the first valley of r_P and $f_2 \approx$ 400 MHz: the second peak of r_P). To this end an impinging SAW from the left-hand side with a unit magnitude of the accompanying φ has been assumed. It turns out that the stress in the electrode/substrate interface has the form given in Fig. 11. The time-harmonic electrode-vibration at consecutive time-instants is illustrated in Fig. 12.

5 Conclusion

We have presented a rigorous, efficient, and easy-to-implement technique for modeling SAW-transducers with a finite number of massive electrodes. The reflection-coefficients of a SAW impinging on an isolated massive electrode have been computed, giving calculated coefficients with an amount of up to 15 %.

Acknowledgements

This work was supported by Siemens AG, Corporate Research and Development, Munich, Germany. The authors thank F. Seifert and C. Ruppel for valuable discussions.

References

[1] H.P. Reichinger and A.R. Baghai-Wadji, "Dynamic 2D Analysis of SAW–Devices Including Massloading," *Proc. IEEE Ultrasonics Symposium*, Vol.1, Oct.1992, pp. 7–10.

[2] H.P. Reichinger and A.R. Baghai-Wadji, "Stress Pattern on the Electrode/Substrate Interfaces in SAW-Devices," *Proc. IEEE Ultrasonics Symposium*, Vol.1, Nov.1993, pp. 153–156.

[3] B.A. Auld, *Acoustic Fields and Waves in Solids,* Vol. 1, New York: Wiley, 1973.

[4] O.C. Zienkiewicz. *The Finite Element Method,* 3rd ed. McGraw–Hill, 1977.

[5] H.R. Schwarz, *Methode der finiten Elemente,* Stuttgart: B.B. Teubner 1980.

[6] R.E. Bank, *PLTMG: User's Guide 6.0,* Vol. 7, Philadelphia: Siam, 1990.

[7] R.F. Harrington, *Field Computation by Moment Methods.* New York: Macmillan, 1968.

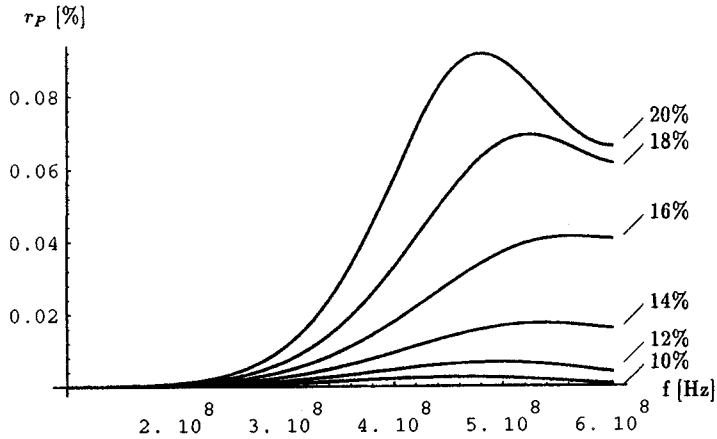


Figure 2: Power ratio reflected/incident SAW. The parameter is the rectangular aluminum electrode's height-to-width ratio. (The electrode is modeled as being purely elastic.)

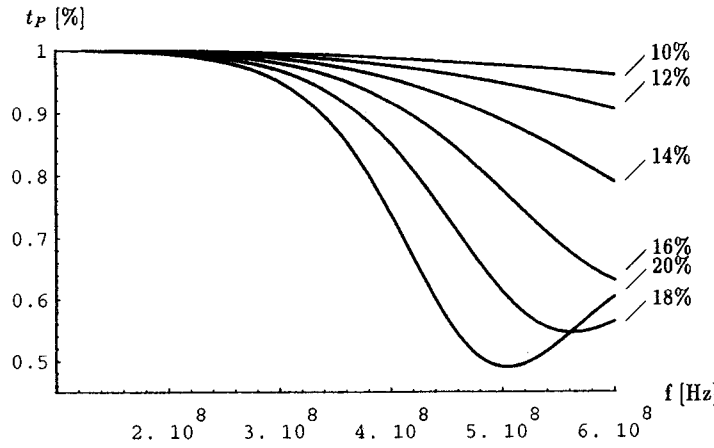


Figure 3: Power ratio transmitted/incident SAW. The parameter is the rectangular aluminum electrode's height-to-width ratio. (The electrode is modeled as being purely elastic.)

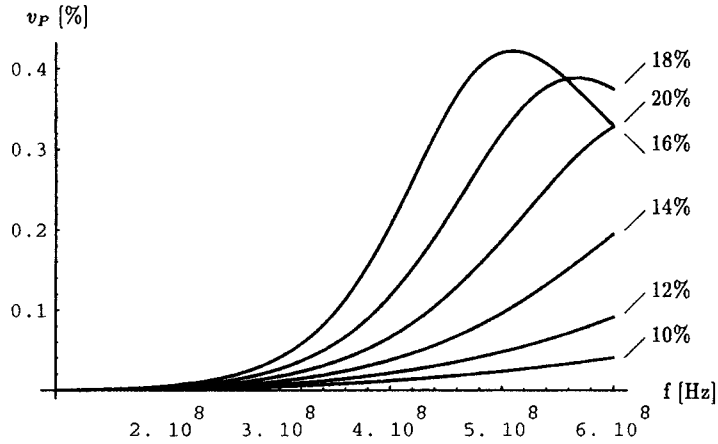


Figure 4: Power ratio of generated bulk wave to incident SAW. The parameter is the rectangular aluminum electrode's height-to-width ratio. (The electrode is modeled as being purely elastic.)

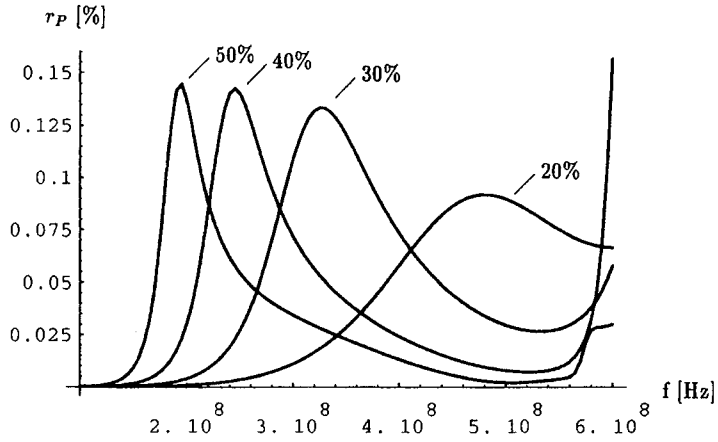


Figure 5: Power ratio reflected/incident SAW. The parameter is the rectangular aluminum electrode's height-to-width ratio. (The electrode is modeled as being purely elastic.)

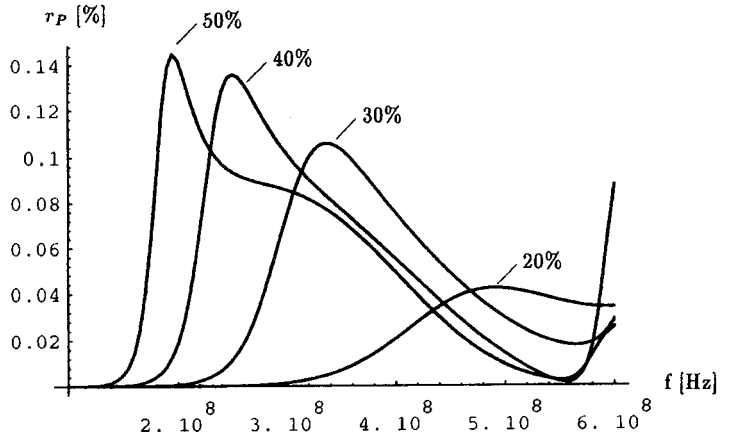


Figure 6: Power ratio reflected/incident SAW. The parameter is the rectangular aluminum electrode's height-to-width ratio. (The electrode is modeled as being elastic as well as conductive.)

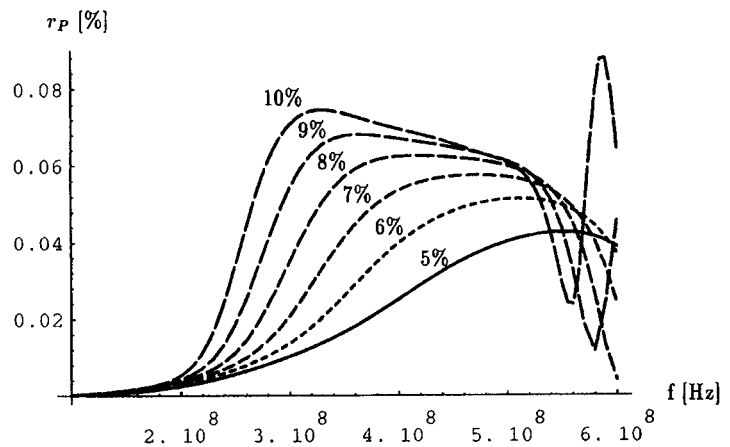


Figure 7: Power ratio reflected/incident SAW. The parameter is the rectangular gold electrode's height-to-width ratio. (The electrode is modeled as being elastic as well as conductive.)

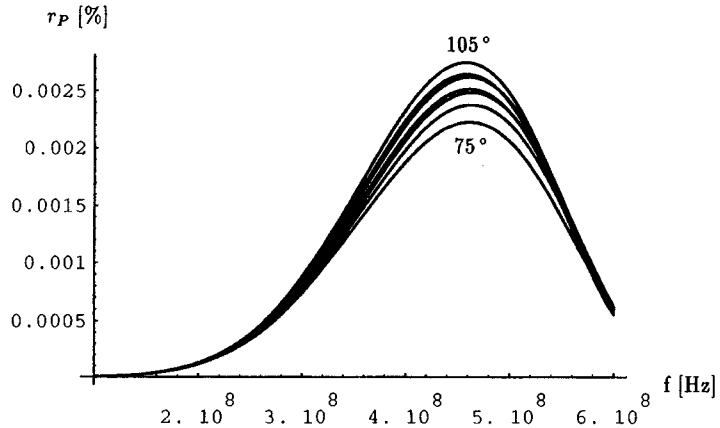


Figure 8: Power ratio reflected/incident SAW. The parameter is the angle of the trapezoidal aluminum electrode, increasing from 75° to 105° in units of 5°. (The electrode is modeled as being purely elastic.)

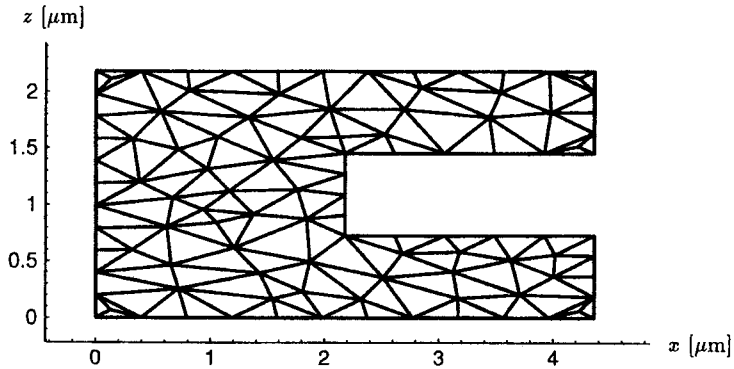


Figure 9: Cross-section geometry and FEM-mesh of the non-symmetric aluminum electrode assumed for the calculations in Figs. 10–12. (The electrode is modeled as being purely elastic.)

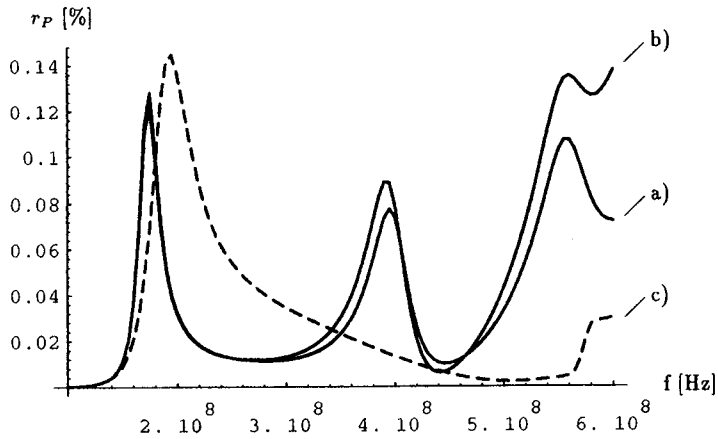


Figure 10: Power ratio reflected/incident SAW. SAW incident from a) the left- b) the right-hand side. The dashed line c) shows the result for a rectangular electrode without the indentation.

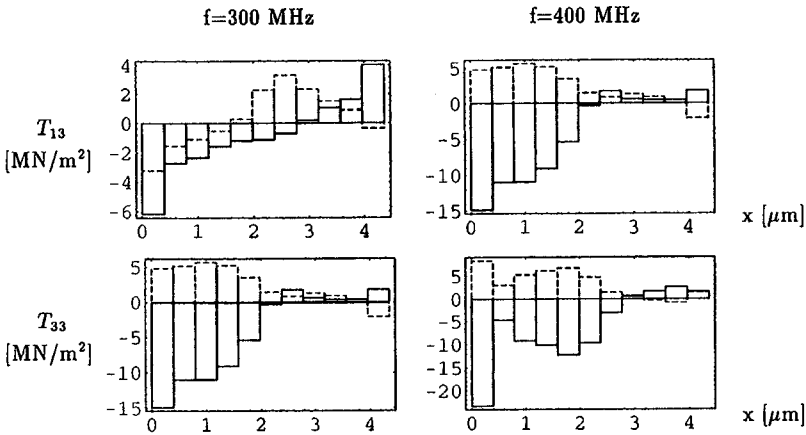


Figure 11: Time-harmonic stress in electrode/substrate interface due to incident SAW (unit magnitude of φ) with 300 MHz (left column) and 400 MHz (right column). Upper and lower row show T_{13} and T_{33}, respectively. T_{23} vanishes because of pure rayleigh cut. Solid and dashed lines indicate real and imaginary part, respectively.

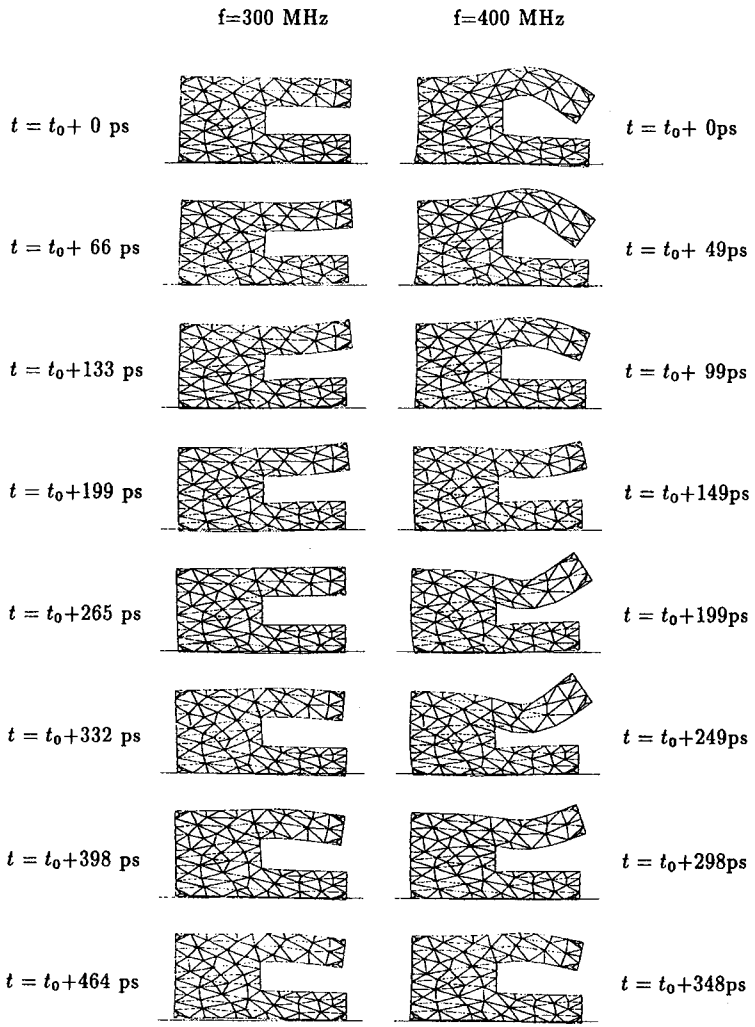


Figure 12: Time-harmonic vibrations due to an incident SAW with 300 MHz (left column) and 400 MHz (right column). The vibrations are shown at discrete time-instants according to values of ωt increasing in 45° units from the top to the bottom. (The displacements are magnified by a factor of 300.)

2D ANALYSIS OF SAW AND PSAW PERIODIC TRANSDUCERS USING FEM AND BEM*

Pascal Ventura, Jean Desbois and Laurent Boyer

Thomson Microsonics, 399 Route des Crêtes, 06904 Sophia-Antipolis, France

* This work has been supported by DRET

ABSTRACT

Emission and propagation of True SAW (TSAW) and Pseudo SAW (PSAW) under a metallic periodic grating are the basic phenomena for the SAW filter low loss technology. Their precise modelling is becoming a key element in the design of more and more compact filters.

In the present paper, we present a 2D model to compute the dispersion curve as well as the coupling efficiency.

In order to take into account complex electrode profiles, FEM is applied only to the electrode and is used to compute relationships between mechanical stresses and displacements at the electrode-substrate interface.

Thus, Floquet's analysis and a Green's function formalism are used to derive boundary relationships for the semi-infinite piezoelectrical propagation medium.

Both theoretical explanations, calculation results and comparisons with experimental data will be shown for common substrates.

1. INTRODUCTION

For the first SAW bandpass filters, the mechanical part of the reflection coefficient on metallic electrodes was considered as a parasitic phenomenum and had to be minimized. On the contrary, for current SAW low-loss bandpass filters, such as SPUDT-DART based filters or resonator filters, the reflection coefficient under a metallic electrode is a key design parameter. It is used for both insertion loss minimization and acoustic path folding (/1/, /2/), resulting in more compact low loss structures. On quartz, commonly used for narrowband low loss applications because of its good temperature stability, the main part of the reflection is due to pure mechanical effects. For high piezoelectrically coupled substrates such as 36° lithium tantalate or 41° lithium niobate,

both mechanical and electrical boundary perturbations have a strong effect on the propagation of TSAW or PSAW and therefore have to be taken simultaneously into account in a numerical model.

Even for simple electrode shapes such as rectangular ones, relationships between the stress vector and the displacement field at one of the faces of the electrode can't be computed with analytical expressions in the non pure shear horizontal case. Therefore, only numerical methods can be used. During the last ten years, there has been an important activity in computing by perturbation theory or variational methods the dispersion curve in the case of SAW propagation under a periodic grating (/3/ to /8/). For these methods, simplifying assumptions (such as equivalent isotropic substrate, ...) are needed and lead to inaccurate results for large metal thickness to wavelength ratios.

Only numerical methods such as Finite Elements or Finite Differences can be efficient enough for so complex a problem. The first published FEM model of the mechanical perturbation on SAW propagation (/9/ and /10/) requires meshing the electrode and a grating period of the substrate (over a few SAW wavelengths) leading to a great amount of computer time. Recently /11/, a combined FEM / BEM method was presented where the piezoelectric substrate is taken into account by its periodic Green's function while FEM is only used to compute acoustical behaviour of the electrode.

In this paper, we present a mixed FEM / BEM model of the electrode perturbation on TSAW and PSAW propagation using the recently developed concept of harmonic admittance /12/. This method keeps the advantages of restricting FEM discretization to only the electrode and allows the use of theoretical results and the algorithms already published /12/.

2. THEORY

Let us consider a piezoelectric semi-infinite medium the interface of which is covered by an infinite p-periodic array of metallic electrodes with infinite length and finite other dimensions (width : a, height : h) (figure 1). An electrical current I_n per unit electrode length is entering the n^{th} electrode at the potential V_n.

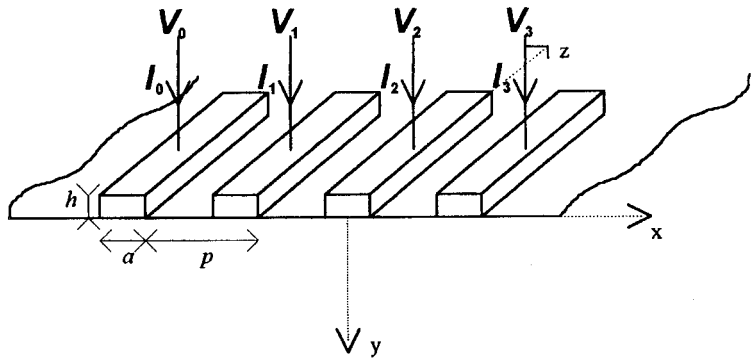


figure 1 : transducer structure schematic.

In what follows, we assume a $\exp(j\omega t)$ temporal harmonic motion. The plane strain assumption will be made due to the infinite length of the electrode. Finally, only on an electrical point of view, the height of the electrode will be considered negligible.

It is well known that the electroacoustic behaviour of the transducer is completely described by the mutual admittance $Y_{m,n}$ between electrodes. Due to the translation symmetry of the structure, $Y_{m,n}$ only depends on the difference $(m-n)$ and we will define : $Y_{m-n} = Y_{m,n}$.

In order to compute the SAW or PSAW contribution to the mutual admittance, we will adopt the same method as Zhang et al /12/. Let us consider a spatial harmonic excitation of the array :

$$V_m(\gamma) = V_0 \exp(-j2\pi m\gamma) \tag{1}$$

leading to a corresponding spatial harmonic current distribution :

$$I_m(\gamma) = I_0 \exp(-j2\pi m\gamma)$$

Due to the translation symmetry, the equivalent admittance of the n^{th} electrode is independent on n. Let us define it as the harmonic admittance of the array :

$$Y(\gamma) = I_n(\gamma)/V_n(\gamma) \tag{2}$$

As was shown in /12/, $Y(\gamma)$ is the Fourier serie of the mutual admittance :

$$\begin{cases} Y(\gamma) = \sum_{m=-\infty}^{+\infty} Y_m \exp(-j2\pi m\gamma) \\ Y_m = \int_0^1 Y(\gamma)\exp(j2\pi m\gamma)d\gamma \end{cases} \tag{3}$$

2.1. HARMONIC ADMITTANCE

Because of the periodicity of the problem, it is possible to compute the harmonic admittance by restricting the analysis to a single period. Due to the γ-harmonic excitation of the array, and using the Floquet's theorem it is possible to expand the mechanical displacements $\boldsymbol{u}_\gamma$, the stress vector $\boldsymbol{t}_\gamma$, the electrical potential ϕ_γ and charges σ_γ at the substrate interface into spatial harmonics :

$$\begin{aligned} \begin{pmatrix} \boldsymbol{u}_\gamma(x) \\ \phi_\gamma(x) \end{pmatrix} &= \sum_{m=-\infty}^{+\infty} \begin{pmatrix} \boldsymbol{u}_{\gamma,m} \\ \phi_{\gamma,m} \end{pmatrix} \exp(-j2\pi(\gamma+m)x/p) \\ \begin{pmatrix} \boldsymbol{t}_\gamma(x) \\ \sigma_\gamma(x) \end{pmatrix} &= \sum_{m=-\infty}^{+\infty} \begin{pmatrix} \boldsymbol{t}_{\gamma,m} \\ \sigma_{\gamma,m} \end{pmatrix} \exp(-j2\pi(\gamma+m)x/p) \end{aligned} \tag{4}$$

Let us now consider $\boldsymbol{G}_{\gamma,m}$ the wave number dependent Green's function (Fourier transform of the Green's function) of the homogeneous semi-infinite anisotropic substrate for a slowness $(\gamma+m)/fp$ in the propagation direction. $\boldsymbol{G}_{\gamma,m}$ can be computed from the solutions of the Christoffel's equation and governing laws :

$$\begin{pmatrix} \boldsymbol{u}_{\gamma,m} \\ \phi_{\gamma,m} \end{pmatrix} = \boldsymbol{G}_{\gamma,m} \begin{pmatrix} \boldsymbol{t}_{\gamma,m} \\ \sigma_{\gamma,m} \end{pmatrix} \tag{5}$$

Finally, an integral equation relating displacement, electrical potential, stress fields and electrical charges at the interface of one period can be derived :

$$\begin{pmatrix} \boldsymbol{u}_\gamma(x) \\ \phi_\gamma(x) \end{pmatrix} = \frac{1}{p}\int_0^p \boldsymbol{G}_\gamma(x-x') \begin{pmatrix} \boldsymbol{t}_\gamma(x') \\ \sigma_\gamma(x') \end{pmatrix} dx' \tag{6}$$

where, $\boldsymbol{G}_\gamma(x) = \sum_{m=-\infty}^{+\infty} \boldsymbol{G}_{\gamma,m} \exp(-j2\pi(\gamma+m)x/p)$

Assuming stress free boundary conditions between electrodes (no mechanical stress or electrical surface charge density), the integral in (6) can be restricted to $-a/2$ and $a/2$ lower and upper boundaries.

The presence of the electrode on the piezoelectric substrate interface generates both mechanical shear and compression stress and a surface electrical charge density. In order to compute mechanical impedance relationships, we decided to use a Finite Element Method. The section of the electrode is discretized into six node triangle finite elements. Once the mass and stiffness matrices (**M** and **K**) are computed, we obtained linear relationships between mechanical displacements and forces.

$$[K - \omega^2 M]U = F \tag{7}$$

Only surface forces exist at the interface Γ_e of the electrode connected to the substrate, therefore it is possible to condense the linear system (7) to the Γ_e nodal values of U and F (U_e and F_e) :

$$[Z_{FEM}]U_e = F_e \tag{8}$$

In the following the indice i will correspond to the i^{th} finite element node of Γ_e while j will be chosen for the j^{th} finite element.

If $w_i(x)$ is the basis function corresponding to the i^{th} node of Γ_e, F_e is defined by

$$\boldsymbol{F}_{e,i} = \int_{\Gamma_e} \boldsymbol{t}_\gamma(x) w_i(x) dx \tag{9}$$

and the finite element interpolation of $\boldsymbol{u}_\gamma$ under the electrode can be written :

$$\boldsymbol{u}_\gamma(x) = \sum_i \boldsymbol{U}_{e,i} w_i(x) \tag{10}$$

At this level of the analysis, it is necessary to choose basis functions to interpolate mechanical stresses and electrical charges components. For this purpose, we used a

collocation interpolation technique (on each finite element, the mechanical stress field is supposed to be constant) :

$$\begin{pmatrix} t_\gamma(x) \\ \sigma_\gamma(x) \end{pmatrix} = \sum_j \begin{pmatrix} T_{e,j} \\ \Pi_{e,j} \end{pmatrix} P_j(x) \tag{11}$$

where, P_j is the pulse function (equal to 1 on the finite element j, and 0 elsewhere).

Using the same finite element interpolation for the electrical potential as for the mechanical displacement, and applying the moment method to the integral equation (6), we obtained the global linear system :

$$\begin{cases} [Z_{FEM}]U_e = [L_1]T_e \\ [L_2]\begin{pmatrix} U_e \\ \Phi_e \end{pmatrix} = [R]\begin{pmatrix} T_e \\ \Pi_e \end{pmatrix} \end{cases} \tag{12}$$

where, $[L_1]$ and $[L_2]$ allow the coupling of collocation interpolation and quadratic finite element interpolation, and :

$$[R]_{j_1,j_2} = \frac{1}{p^2}\int_{\Delta_{j_1}}\int_{\Delta_{j_2}} G_\gamma(x-x')dxdx' \tag{13}$$

Assuming negligible electrode resistivity, the harmonic admittance can be easily derived from system (12).

Figure 2 shows an example for a Y+36° lithium tantalate substrate of normalized harmonic admittance $Y_v(\gamma)$ defined by : $Y(\gamma)/(2L\omega\varepsilon_\infty)$ where L is the acoustical aperture of the transducer and ε_∞ the effective permittivity of the substrate in the case of an infinite slowness. It is easy to notice the presence of both PSAW and TSAW by the complex and the real pole of the harmonic admittance, while the contribution of SSBW in the mutual admittance creates a discontinuity of the first derivative of $Y_v(\gamma)$.

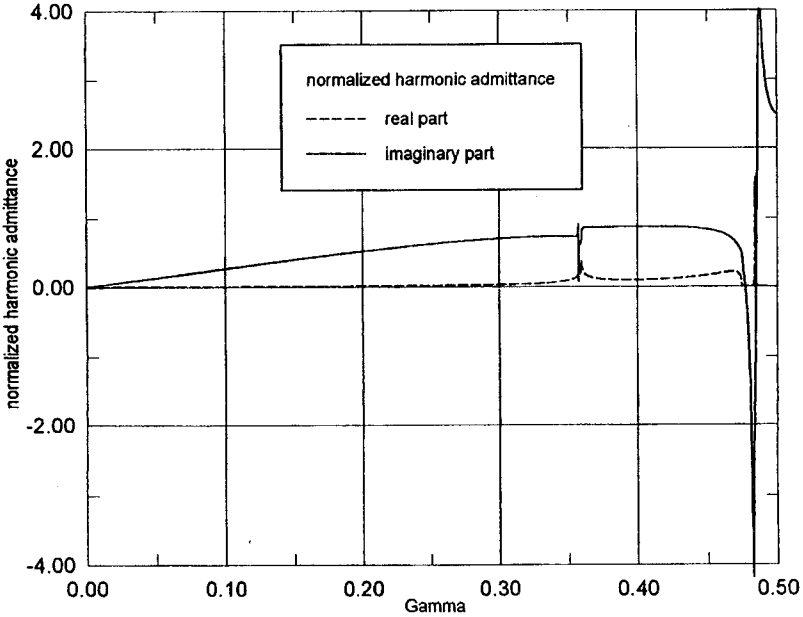


figure 2 : normalized harmonic admittance for the LiTaO3 Y+36° substrate (fp=2007.8)

2.2. SAW AND PSAW CONTRIBUTIONS

The harmonic admittance can be expanded into three different contributions : the dielectric contribution, the surface skimming bulk waves contribution, and the TSAW and PSAW contributions :

$$Y(\gamma) = Y_d(\gamma) + Y_b(\gamma) + Y_s(\gamma) \tag{14}$$

Using the asymptotic expression of the contribution of the TSAW and PSAW to the mutual admittance, it is possible to find a model for $Y_s(\gamma)$ depending on two complex parameters Y_{sc} and γ_{sc} /12/:

$$Y_s(\gamma) = \frac{jY_{sc}}{\tan(\pi\gamma_{sc})} \frac{\sin(\pi\gamma)^2}{\left(\sin(\pi\gamma)^2 - \sin(\pi\gamma_{sc})^2\right)} \quad (15)$$

Physically, γ_{sc} characterizes the phase velocity of the short circuited eigenmode while Y_{sc} is related to the amplitude of the coupling between two electrodes. In the case of a SAW contribution outside the stop band, γ_{sc} and Y_{sc} are respectively real and purely imaginary, as a contrary, for a PSAW, γ_{sc} and Y_{sc} present an imaginary part which is the consequence of the energy waste in bulk waves.

In order to compute these two parameters we assume that close to the pole the variations of the residual part of the harmonic admittance are negligible in front of those of $Y_s(\gamma)$ and we develop an iterative algorithm. In figure 3 the derived PSAW contribution was removed , it illustrates the very good efficiency of this algorithm as well as the accuracy of the model of the PSAW contribution.

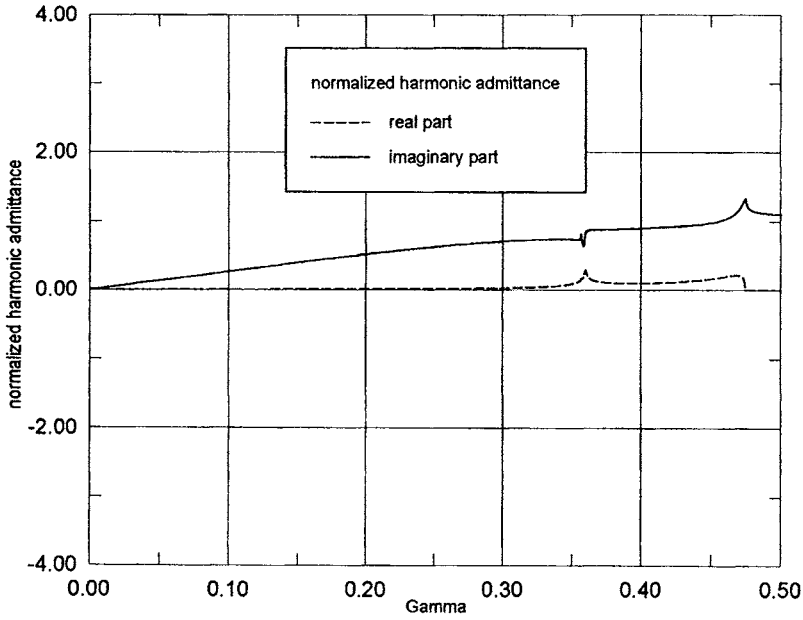


figure 3 : normalized harmonic admittance after removing PSAW contribution for the LiTaO3 Y+36° substrate (fp=2007.81)

III. NUMERICAL RESULTS

In order to illustrate the importance of a numerical simulation, we plot in figures 4, 5 and 6 the mechanical stress distributions $t_{Y,x}$ and $t_{Y,y}$, and the electrical charge distribution σ_Y (normalized at their respective maximum value) for the Y+38° quartz and near the TSAW pole.

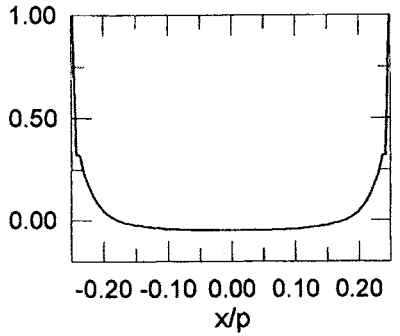


Figure 4 : normalized $t_{Y,x}$

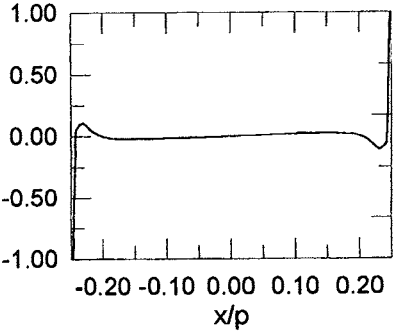


Figure 5 : normalized $t_{Y,y}$

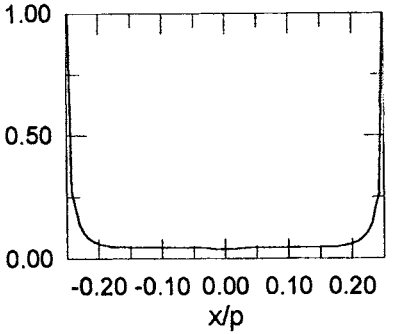


Figure 6 : normalized charge distribution σ_Y

Rectangular aluminium electrodes and the Y+128° lithium niobate piezoelectric substrate were first used to validate our computations. The ratio h/2p is 1%. Figure 5 shows the amplitude of the reflection coefficient for shorted electrodes computed with our model and previously published /13/. The similarity of both results is excellent. Figure 7 also shows the mechanical and the piezoelectric parts of the reflection

coefficient. The opposite sign of the two parts leads to the zero amplitude of the reflection coefficient for $a/p \approx 0.7$.

In order to simulate and investigate low loss bandpass filters this numerical model has been extensively used to compute strip reflectivity and phase velocity for the Y+38° quartz piezoelectrical substrate. First of all, we present in figure 8 the dependence of the mechanical part of the strip reflectivity (in amplitude) versus the metallization height for $a/p = 0.5$. For $h/2p$ less than 3% the curve can be very well fitted by a linear function :

$$R \approx \alpha \frac{h}{2p} \quad \text{where } \alpha = 0.588$$

On the other hand, figure 9 shows the variation of the phase velocity versus the metallization height for two different a/p (0.5 and 0.7).

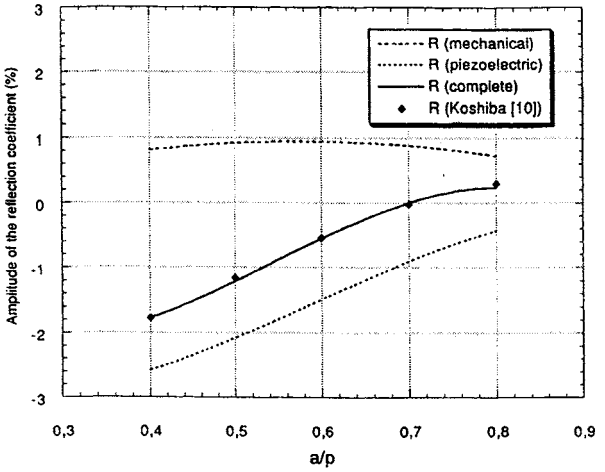


figure 7 : Y+128° $LiNbO_3$ cut. Amplitude of the reflection coefficient per period for shorted aluminium electrodes.

(16)

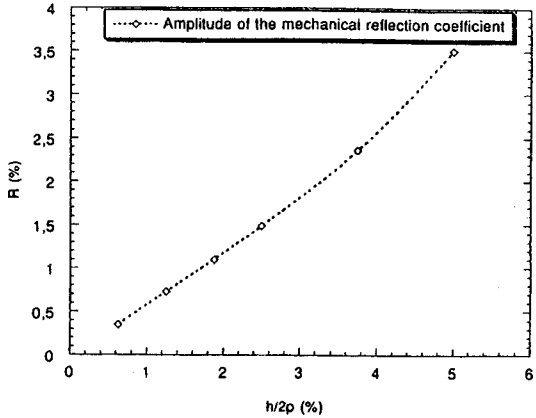


figure 8 : Y+38° quartz cut. Amplitude of the mechanical reflection coefficient per period for aluminium electrodes versus metallization height.

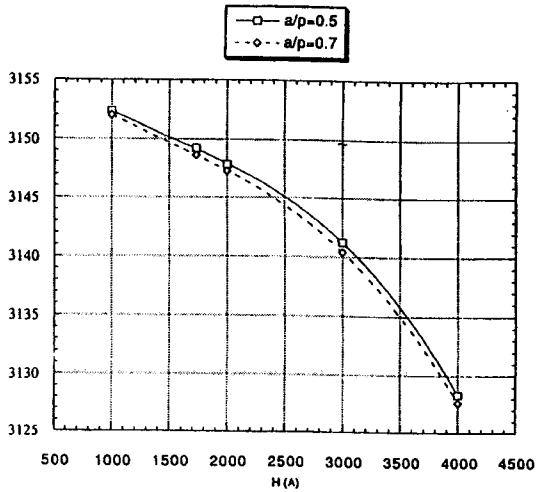


figure 9 : Y+38° quartz cut. phase velocity for a/p=0.5 and 0.7 for aluminium electrodes versus metallization height.

CONCLUSION

A new model based on the harmonic admittance concept has been developed to compute the emission and propagation of TSAW or PSAW propagation under a periodic array of metallic electrodes. FEM has been used to take into account any electrode shape while the piezoelectric substrate is completely characterized by a Green's function formalism. Up to now, this numerical model has been extensively used to characterize low loss bandpass filters on quartz and we are currently improving our simulation of SAW resonator filters on more piezoelectrically coupled substrate /14/. There is an excellent agreement found between our new method and experimental measurements, making this new 2D model a powerful tool for exhaustive characterization of TSAW and PSAW.

REFERENCES

/1/P. Ventura, M. Solal, P. Dufilié and S. Chamaly, "A Global Optimization Procedure for SPUDT Filters" , in *IEEE Ultrason. Proc. Symp.*, 1993.

/2/ M. Solal and J. M. Hodé, "A new compact SAW Low Loss Filter for Mobile Radio", in *IEEE Ultrason. Proc. Symp.*, 1993.

/3/ D. A. Simons, "Reflection of Rayleigh waves by strips, grooves, and periodic arrays of strips and grooves" , *J. Acoust. Soc. Am. 63(5)*, May 1978.

/4/S. Datta and B. J. Hunsinger, "First-order reflection coefficient of surface acoustic waves from thin strip overlays" , *J. Appl. Phys. 50(9)*, Sept. 1979.

/5]/S. Datta and B. J. Hunsinger, "Analysis of Energy Storage Effects on SAW Propagation in Periodic Arrays" , *IEEE Trans. Sonics Ultrason.*, vol. SU-27, Nov. 1980.

/6/H. Robinson, Y. Hahn and J. N. Gau, "Analysis of SAW reflections" , *IEEE Ultrason. Symp. Proc.*, pp. 131-134, 1988.

/7/H. Robinson, Y. Hahn and J. N. Gau, "A comprehensive analysis of surface acoustic waves reflections", *J. Appl. Phys. 65(12)*, June 1989.

/8/H. Robinson, Y. Hahn and J. N. Gau, "A comprehensive analysis of surface-acoustic-waves reflections. II. Second-order effects" , *J. Appl. Phys. 68(9)*, Nov. 1990.

/9/P. Bauerschmidt, R. Lerch, J. Machui, W. Ruile and G. Visintini, "Reflection and Transmission Coefficients of SAW in a periodic Grating Computed by Finite Element Analysis" , *IEEE Ultrason. Symp. Proc.*, 1990.

/10/M. Buchner, W. Ruile, A. Dietz and R. Dill, "FEM Analysis of the Reflection Coefficient of SAWs in an Infinite periodic Array" , *IEEE Ultrason. Symp. Proc.*, 1991.

/11/ H. P. Reichinger and A. R. Baghai-Wadji, "Dynamic 2D Analysis of SAW devices including Massloading" , *IEEE Ultrason. Symp. Proc.*, 1992.

/12/ Y. Zhang, J.Desbois, and L. Boyer, "Characteristic Parameters of Surface Acoustic Waves in a Periodic Metal Grating on a Piezoelectrical Substrate" , *IEEE Trans. Ultrason. Ferroelec., Freq. Contr., vol. 40, No. 3*, May 1993.

/13/ M. Koshiba and S. Mitobe, "Equivalent Networks for SAW Gratings" , *IEEE Trans.Ultrason. Ferroelec., Freq. Contr., vol. 35, No. 5*, Sept. 1988.

/14/ M.Solal and P. Ventura, "Analysis and Design of a Very Low Loss SAW Resonator Filter on 36° Lithium Tantalate", *8th EFTF Proc.*, 1994.

Performance Evaluations of Frequency Multipliers for Atomic Frequency Standards

U. Grimaldi, P. Schumacher, P. Thomann

Observatoire Cantonal - Rue de l'Observatoire, 58 - CH-2000 Neuchâtel - SWITZERLAND

Abstract

The first purpose of the paper is the evaluation, in terms of phase noise, of the performance of some frequency multipliers used in Cesium frequency standards. Then a tentative comparison of the measurement results with the requirements of existing and future atomic clocks is performed.
The contribution of the multipliers to the $\sigma_y(1s)$ of an atomic clock is shown to be lower than 10^{-13}. Such a performance is definitely satisfactory for the present atomic frequency standards and, moreover, appears to be suitable for atomic clocks in the near future. Secondly, it is suggested that this performance may be improved to the level of 10^{-14} at 1 s, by substituting the first multiplying stage with one of the others measured. Preliminary measurements of the output phase sensitivity to temperature and input level are also discussed.

Keywords: Frequency multipliers, Phase noise measurement, Frequency standards.

1. Introduction

Frequency multipliers are one of the main elements in the chain generating the high frequency signal used to interrogate the resonator of atomic frequency standards. As a consequence, such frequency multipliers must fulfil the basic requirement of providing low noise signals, so that high performance atomic frequency standards can be built.

Of course the overall noise is the result of the contributions of all the elements of the chain; so the frequency multiplier performance has to be evaluated considering the capabilities of the local oscillator to produce a clean signal /1/.

In the following, the results of an evaluation of the phase noise contribution of some frequency multipliers are presented. Firstly the multiplier developed at the PTB and used in several primary atomic frequency standards is examined. Secondly the phase noise performance of a frequency multiplier from 5 MHz to 100 MHz, developed at ON, is evaluated. Then the same analysis is performed for a commercial (OSA) frequency multiplier from 10 MHz to 9.2 GHz. Finally, an evaluation of the measurement results is performed with the purpose of comparing the performance of the examined frequency multipliers with the requirements of existing and future atomic frequency standards.

A preliminary evaluation of the phase sensitivity of the PTB multiplier to ambient temperature and input signal amplitude allows one to estimate within which operational limits the multiplier will not affect the clocks performance.

2. The frequency multipliers

The PTB multipliers

The frequency multipliers developed at the PTB generate a 9.2 GHz signal from a 5 MHz input signal. Such a multiplication is performed in two main steps: the first step consists in a direct multiplication from 5 MHz to 100 MHz, based on discrete transistor electronics and tuned resonant filters; the second step (times 92) from 100 MHz to 9.2 GHz, is achieved by a step-recovery diode coupled to a tuned resonant cavity.

The 9.2 GHz signal output from the step-recovery diode steers a phase-locked Voltage Controlled Oscillator (VCO); the PLL also allows the frequency modulation required for the interrogation of the Cesium resonance line; furthermore, it filters out the high frequency noise and sidebands of the 9.2 GHz signal /2/. The block diagram of the whole architecture is sketched in Fig. 1.

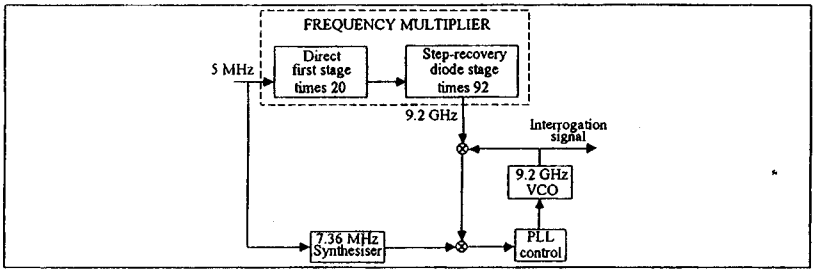


Fig. 1. The block diagram of the PTB multiplier. In the measurements the 7.36 MHz synthesiser was bypassed..

For the phase noise measurements of the 9.2 GHz signal, the PLL output was used in fact the step-recovery diode output level was below the minimum accepted by the measurement system. However this affects the measurement results only for Fourier frequencies higher than the PLL cut-off frequency: this is definitely not critical, since the interesting region spans from 0.1 Hz to some hertz, where the measured phase noise spectrum is actually that of the multiplied signal. Since the 7.36 MHz synthesiser was not needed, it was bypassed and the actual VCO output frequency was 9.195 GHz.

The ON multipliers

A new 100 MHz frequency multiplier was developed at ON in 1990. It is based on the PLL principle, so its structure is basically different from that previously described: it embeds a 100 MHz voltage controlled crystal oscillator (VCXO) whose oscillation frequency is locked to that of a 5 MHz input signal (Fig. 2) through a fast TTL divider. One advantage of the PLL principle is the absence of tuned resonant circuits, which should result in a reduced temperature sensitivity.

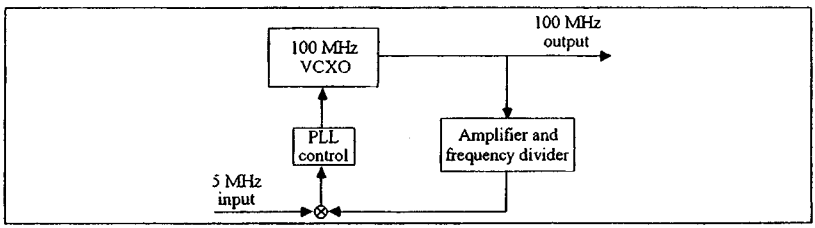


Fig. 2. The block diagram of the 5 MHz to 100 MHz PLL multiplier developed at the Observatoire Cantonal *in Neuchâtel.*

The OSA multipliers

The commercial OSA frequency multiplier was developed in 1987 for the Cesium clock "EUDICS" /3/. This multiplier generates a 9.18 GHz signal starting from a 10 MHz input: it consists of two stages, (x 18 and x 51), conceptually similar to the corresponding ones in the PTB multipliers; the main differences between these two models, apart from the different operating frequencies, are in the way the frequency is modulated to interrogate the atomic resonance (of no concern in the present analysis) and in the lack of a 9.2 GHz VCO in the OSA multiplier.

3. The phase noise measurement scheme

The Frequency Stability Mobile Measuring System (FSMMS) /4/, available at ON, is basically a Hewlett-Packard Model 3048 Phase Noise Measurement System equipped with high-performance BVA quartz oscillators. It allows measurements of the spectral density of the relative phase noise between two signals down to spectral regions not attainable using usual spectrum analysers. In particular it is possible to measure such phase noise spectral density in Fourier frequency intervals below 1 Hz, which are the most interesting for present and future atomic clocks.

The phase noise spectral density is computed via FFT (Fast Fourier Transform); the FSMMS performs every measurement in successive time intervals, during each of which a different Fourier frequency interval is inspected. The measurement result is output by the system in the form of a graph where $L(f) = 0.5\ S_\varphi(f)$ is plotted. Furthermore, from the measured data the FSMMS can also compute the values of $s_y(t)$ for a limited set of t values according to the formula [5]

$$\sigma_y^2 = 2\int_0^{f_h} S_y(f)\,\frac{\sin^4(\pi f\tau)}{(\pi f\tau)^2}\,df \qquad (1)$$

where $S_y(f) = \frac{f^2}{\nu_0^2} S_\varphi(f)$; and ν_0 is the signal nominal frequency.

For this purpose, the FSMMS requires that the values of τ and of the corresponding f_h be input from the user. As it has been shown /5/, if such values are chosen so that $f_h \cdot \tau = 10$, a variance whose behaviour is very close to that of the "modified Allan variance" is obtained.

For all the measurements a common source scheme was adopted, as shown in Fig. 3. This scheme was actually necessary to perform the measurements, since it permitted to get rid of the phase noise of the oscillator used to generate the 5 MHz (10 MHz in the case of OSA multiplier) input signal, which was always the most prominent noise contribution.

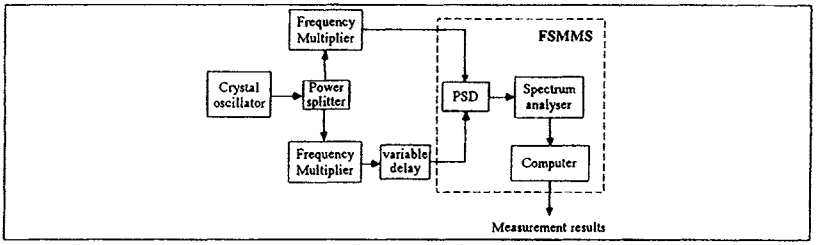


Fig. 3. The common source measurement scheme: a single oscillator was used to generate the input signal for both multipliers under test. The PSD (Phase Sensitive Detector), the spectrum analyser, and the computer are the main parts of the FSMMS.

The FSMMS has a very strict requirement on the quadrature condition of its two input signals: in any time interval in which each measurement is subdivided it allows its input signals to be out of quadrature by no more than some degrees, depending on the peak amplitude of the signals themselves. This can be very critical, especially for higher frequency signals, since the slow thermal drifts of the device under test can severely limit either the minimum Fourier frequency inspected or the measurement accuracy (the available averaging time being limited), or even both. For this reason it was necessary to perform some measurements with both multipliers in a thermally shielded environment, and even so the problem was not completely overcome for all multipliers: in such cases the duration of the acceptable quadrature was exploited to increase the measurement accuracy, for the PTB multipliers, and the inspected spectrum width, for the OSA multipliers. In any case, however, the minimum Fourier frequency inspected was not higher than 100 mHz, as just the interval between 100 mHz and 1 Hz is of main concern in this study.

4. Experimental results

Fig. 4 shows a summary of all the measured phase noise spectra, referred to 5 MHz in order to be easily compared. The phase noise spectral density of the 5 MHz BVA crystal oscillator, used to generate the input signal, is also reported: it was measured with the FSMMS with a measurement scheme similar to that shown in Fig. 3 (curve 4a).

The PTB multipliers

The phase noise spectral density both of the whole multiplier (9.195 GHz signal) and of the first 100 MHz stage were measured. The relevant graphs are reported in Fig. 4.

The 9.195 GHz spectrum (curve 4b) shows a "bump" of noise that reaches its peak at about 30 kHz, which is the cut-off frequency of the amplifying PLL: this "bump" comes from one of the two PLLs for which the cut-off frequency was not optimised with respect to the VCO noise spectrum.

The 100 MHz spectrum (curve 4c) is characterised by white phase noise above 100 Hz and by a sort of flicker phase noise (with a $f^{2/3}$ slope) for frequencies below 100 Hz. Such behaviour is possibly connected with the kind of transistors used or a combination of flicker phase noise (from the electronics) and white frequency noise (from some thermal effects): this phenomenon deserves further investigations.

The two graphs (fig. 4b and c) show the same behaviour for frequencies below 1 Hz: this means that the noise contribution of the second stage of the multiplier is not significant in that frequency interval. On the contrary, above 1 Hz the spectrum of the 100 MHz signal is below the other. A further investigation of this feature will first require two identical multipliers.

In Fig. 5a the plot of the $\sigma_y(\tau)$ points computed from eq (1) for the 9.195 GHz multiplied signal is shown. It is immediately seen that the behaviour of such multipliers meets satisfactorily the most severe requirements of present Cesium clocks /6/, since $\sigma_y(\tau)$ is slightly less than 10^{-13} at $\tau = 1s$.

The ON multipliers

The phase noise spectral density of the 100 MHz PLL multipliers is shown in Fig. 4 (curve d). Below 1500 Hz the behaviour of the multipliers is definitely of the flicker phase noise type, with a f^{1} slope due to mixer noise.

The phase noise introduced by this multiplier is about 10 dB lower than that of the PTB first stage between 100 mHz and 1 Hz. This is due to the PLL design, which is free from the noise normally associated to tuned resonant filters.

The OSA multipliers

The phase noise spectral densities measured for these multipliers are shown in Fig. 4 (curves e, f). The interpretation of such results is quite similar to that relevant to the PTB multipliers: in the spectrum of the 9.18 GHz signal (curve e) it is possible to notice a small bump at about 10 kHz, and the phase noise contribution below 10 Hz is that of the first multiplying stage, whose behaviour is of the flicker phase noise type, with a f^{1} slope.

Besides that, it can be observed that, from about 20 Hz up, the noise of the PTB first stage is almost 10 dB lower than that of both stages of the OSA multiplier. For lower frequencies, instead, the higher level of noise is just due to the higher slope of the PTB multiplier spectrum: this suggests that, if it is possible to remove the cause of such a steeper slope, it will also be possible to improve the overall performance of the PTB multiplier. However, the phase noise contribution of the 100 MHz PLL multiplier is still lower than all the others for frequencies below 20 Hz.

In Fig. 5b the plot of the $\sigma_y(\tau)$ points computed for the 9.18 GHz multiplied signal is reported: at $\tau = 1s$ the value of $\sigma_y(\tau)$ is slightly less than 10^{-13}.

5. Thermal behaviour of the PTB frequency multiplier and phase dependence on the amplitude of the 5 MHz input signal

The two most important phase deviation sources encountered in the normal operation of multipliers are the temperature variation and the input signal AM noise. Some investigations were performed in order to test the sensitivity of the PTB multipliers to such effects: the beat of the two multiplied 9.195 GHz signals was recorded while one multiplier was subjected to a temperature change or an input signal amplitude change.

Effects of temperature

In order to investigate the temperature effect each frequency multiplier was placed in a separate thermal chamber. Several temperature steps having different amplitudes (from 0.5°C to 20°C) were applied to each of them separately. The temperature steps were applied starting from an initial ambient temperature that was varied from about 45°C to about 25°C; however the measurement results were not significantly different.

The phenomenon of the temperature influence on the multipliers is quite complex, probably depending on how the temperature influences the various sub stages of the 100 MHz first multiplying stage. Although a deep understanding of such phenomenon deserves further investigations, it is possible to say that the effects of thermal variations are not higher than 30 ps/K. This means that the ambiant temperature should not vary at a rate higher than 0.1 K/hour if the resultant frequency offset of the clock is not to exceed 10^{-15}. Such a requirement is usually met in a stable laboratory environment.

Effects of the input signal AM noise

The determination of the AM to PM noise coefficient was carried out using an input signal level of 6 dBm, which is the value that the 5 MHz input signal has in normal operation conditions. Anyway, this effect is expected to be strongly non linear, so the present determination of this coefficient is probably valid only in a narrow interval around the 6 dBm level of the input signal.

The measurement was performed by varying the input signal level by 0.05 dB, so that the phenomenon can be considered as linear, while keeping the two multipliers at the same, fixed temperature. The measurement was repeated several times in order to average the resolution error in the reading of the recorded beat signal. The resulting coefficient is about 1 ns/dB. As a consequence, a relative amplitude drift of the 5 MHz input signal of $10^{-6}\,s^{-1}$ will cause a clock frequency offset of 10^{-14}. The relative amplitude stability of the 5 MHz input signal may become a concern in future standards with an expected flicker level well below 10^{-14}; therefore the AM to PM coefficient will deserve further study.

6. Conclusions

Different kinds of multipliers have been tested with respect to their contribution to the phase noise spectrum of the multiplied signal.

It appears that the best performance in terms of introduced phase noise might be obtained by coupling the 100 MHz PLL multiplier with the step-recovery diode second stage from the PTB multiplier.

It has also been verified that in a normal laboratory environment the effects on the PTB multiplier of phase deviation sources such as temperature variations and input signal AM noise can be easily made negligible to the $< 10^{-14}$ level. As a consequence, since the PLL multiplier is expected to be less sensitive to both these phase deviation sources, the sensitivity to them of the "mixed" multiplier is likely to be correspondingly decreased.

However, from the comparative graph shown in Fig. 4 it can be immediately seen that, in the frequency interval of main concern for the atomic frequency standards, i.e. around one hertz, the highest noise contribution is that coming from the quartz oscillator. So it appears that nowadays the strictest limitation in developing a whole interrogation chain for an atomic frequency standard, as far as the phase noise is concerned, is the availability of a local oscillator providing satisfactory performance.

For this reason this is the main field where research and development of new devices can give the strongest improvement to the development of future Cesium clocks providing a significantly improved short term stability.

As far as noise at frequencies up to some kilohertz is concerned, the analysis above shows that the most critical element is the frequency multiplier.

At higher frequencies, instead, if a design concept is adopted in which the actual multiplier is coupled with a VCO via a PLL, an improvement can come from using a VCO providing better performance and from a careful optimisation of the PLL cut-off frequency.

Deeper analysis should be devoted to some phenomena that appeared from the measurement performed in this study. Among them are the effects of the VCO noise on the spectrum of a signal filtered and modulated with a PLL based on such an oscillator, and the effect of temperature and input signal AM noise on the phase of the multiplied signal, in order to possibly reduce the sensitivity of a multiplier to such disturbances.

7. Acknowledgement

The authors want to thank their colleagues at PTB, particularly Mr. Schröder, Dr. Bauch and Mr. Richter, who made possible a large part of this work by making the PTB multipliers available, and whose experience was a valuable help in the interpretation of the results.

The work of U. Grimaldi was supported by a grant of the Human Capital and Mobility Program of the European Community (LEARN: Learning in an atomic clock research network).

8. References

[1] R. J. Douglas and J.-S. Boulanger, 'Local Oscillator Requirements for Timekeeping in the $10^{-14}\ t^{-½}$ Era', *1992 IEEE Frequency Control Symposium*, pp. 6-26.

[2] R. Schröder, 'Frequency Synthesis in Primary Cesium Clocks', *5th European Frequency and Time Forum*, Besançon, France - 12-14 March 1991 - pp. 194-200.

[3] I. Rabian, P. Rochat: "Full digital Processing in a new commercial Cs standard", Proc. 2nd European Frequency and Time Forum, Neuchâtel, Switzerland 16-18 March 1988 pp. 461-468.

[4] 'Frequency Stability Mobile Measuring System', developed under ESTEC contract 9099

[5] Characterization of Clocks and Oscillators (editor: P.B. Sullivan) NIST Technical Note 1337, March 1990

[6] G. Santarelli, P. Laurent, A. Nadir, A. Clairon, " Preliminary Results in a Cesium Atomic Fountain Frequency Standard" These proceedings.

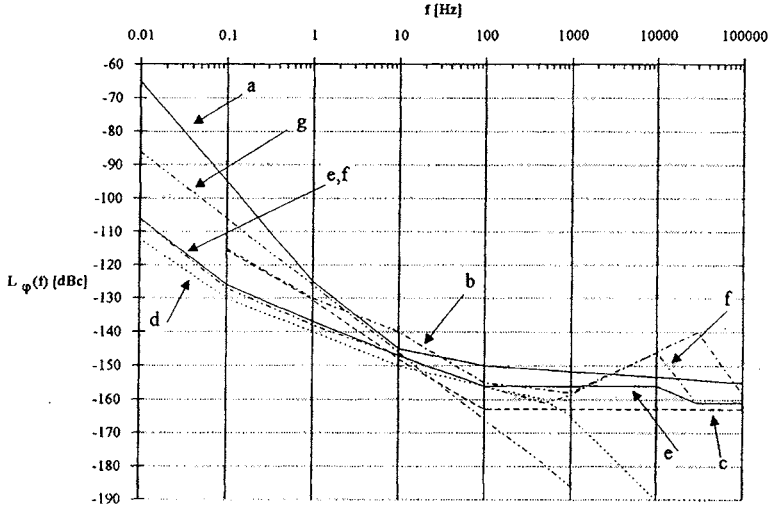


Fig. 4 The summary of all the measured phase noise spectra, referred to 5 MHz: a) 5 MHz BVA oscillator; b) PTB 9.195 GHz (whole multiplier); c) PTB first 100 MHz stage; d) ON 100 MHz PLL; e) OSA 9.18 GHz (whole multiplier); f) OSA first 90 MHz stage; g) hypothetical multiplier with f^2 spectral density of phase noise contributing $10^{-13}\tau^{-1/2}$ to $\sigma_y(\tau)$ of clock.

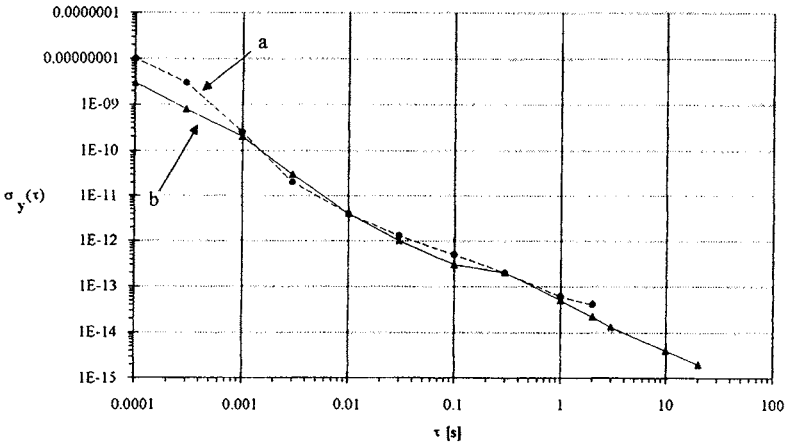


Fig. 5 The plot of the $\sigma_y(\tau)$ points computed from eq. (1): a) PTB 9.195 GHz whole multiplier; b) OSA 9.18 GHz whole multiplier.

A Novel Interpretation of Collisional Cross Sections between Cesium Atoms and Rare Gas Atoms

Masatoshi Kajita
Kansai Advanced Research Center
Communications Research Laboratory
Iwaoka, Nishi-ku, Kobe 651-24 JAPAN

Tel: 81 - 78 - 969 - 2151
Fax: 81 - 78 - 969 - 2154

Summary

Analytical interpretations are given on the collisional broadening and shift cross sections of the Cs D_2 line, which are described with the depth of the van der Waals potential of the perturbers.

1. Introduction

The collisional broadening and frequency shift of an alkali resonance perturbed by rare gas atoms give important information on the interaction potential between the alkali atom and the rare gas atom. The single valence electron of an alkali atom allows a relatively simple theoretical treatment to yield an accurate description of the alkali-rare gas perturber interaction /1-3/. Inoue *et al.* measured the collisional broadening and shift of the Cs D_2 line perturbed by the rare gases He, Ne, Ar, Kr, and Xe /4/. The collisional broadening cross section σ_b and the collisional shift cross section σ_s increase monotonically with both the square of the atomic radius and the polarizibility of the perturbers. Noteworthy are the very small σ_s for the perturbation by helium, and the discontinuity in the σ_b dependence on atomic radius and polarizibility occurring between the interaction with neon and the interaction with argon. Inoue *et al.* suggest that the shift cross section is determined by the physical dimension of the perturber, while the broadening cross section depends on a complex interaction between the

cesium atom and rare gas atoms. In the present paper we address this problem of the discontinuity in the collisional broadening cross section σ_b and give an analysis based on a van der Waals potential parameter.

2. Analysis

Figure 1 shows the correlation between σ_b and $\varepsilon + \varepsilon_c$ where ε is the depth of the Lennard-Jones potential of the perturbers and ε_c is a constant, typically assumed to be $\varepsilon_c / k_B = 65$K as in Ref. /5/. It is shown that σ_b is proportional to $\varepsilon + \varepsilon_c$ for every rare gas perturber. The same result was shown in Fig. 2 for broadening of molecular spectra caused by collisions with isotropic projectiles (rare gas atoms and spherical molecules) /5/. On the other hand σ_s is proportional to $\sqrt{\varepsilon} + \beta$, where β is a constant (as shown in Fig. 3) ($\beta / \sqrt{k_B} = -3.2\sqrt{\mathrm{K}}$). The parameters ε_c and β represent the influence of the exchange force. These proportionalities also hold in the impact approximation calculations of Kielkopf /6/.

3. Interpretation

Generally, collisional broadening is caused by non-radiative transitions to other quantum states. For molecular spectra it is caused by rotational relaxations. As shown in Fig. 4, σ_b is proportional to the square of the matrix element of the Hamiltonian at the distance where the intermolecular interaction is the strongest, given by the parameter D of Ref. /5/. The proportionality between σ_b and D shows that σ_b is determined just by the transition probability at the distance where the interaction is strongest. It is reasonable to give the same interpretation to the broadening of the Cs D_2 line as shown below.

The potential between an alkali atom and a rare gas atom at a long distance is expressed as /1/

$$V(r,\theta) = -\frac{\alpha e^2 r^2}{2R^6}(1 + 3\cos^2\theta) \quad , \tag{1}$$

where α is the polarizability of the perturber, R is the interatomic distance, r is the distance between the alkali core and the valence electron, θ represents the angle between $\boldsymbol{R}$ and $\boldsymbol{r}$, and e is the unit electric charge. By treating the single valence electron and the remaining atomic core of the alkali atom as separate "atoms" we can model the alkali atom as a molecule with a permanent dipole moment $\boldsymbol{er}$. The potential of Eq. (1) coincides with the formula for a dipole-induced dipole interaction which causes rotational relaxation in the molecular case /7/. All transitions which change the

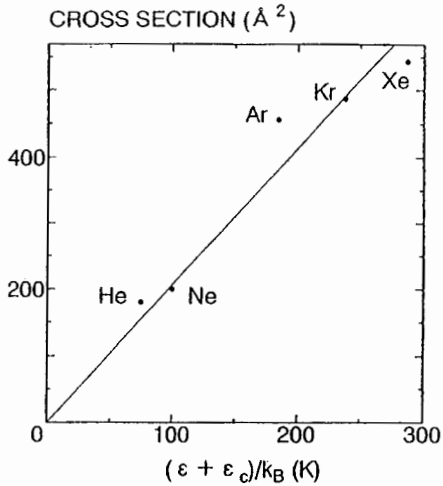


Fig. 1

Collisional broadening cross section as a function of van der Waals potential of perturbers (ε)

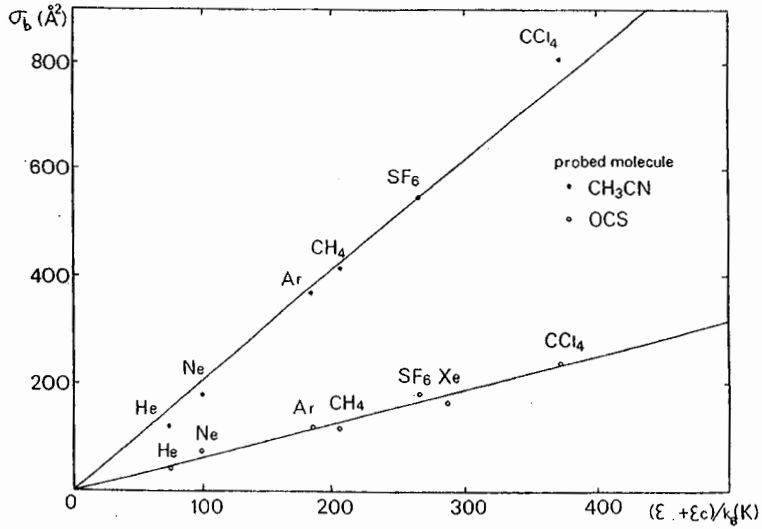


Fig. 2

Collisional broadening cross section σ_b between polar molecules and spherical perturbers as a function of van der Waals potential of perturbers

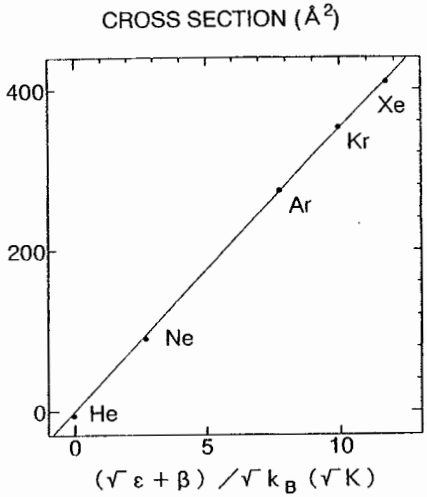


Fig. 3

Collisional shift cross section σ_s as a function of the square root of the depth of the van der Waals potential of perturbers ($\sqrt{\varepsilon}$)

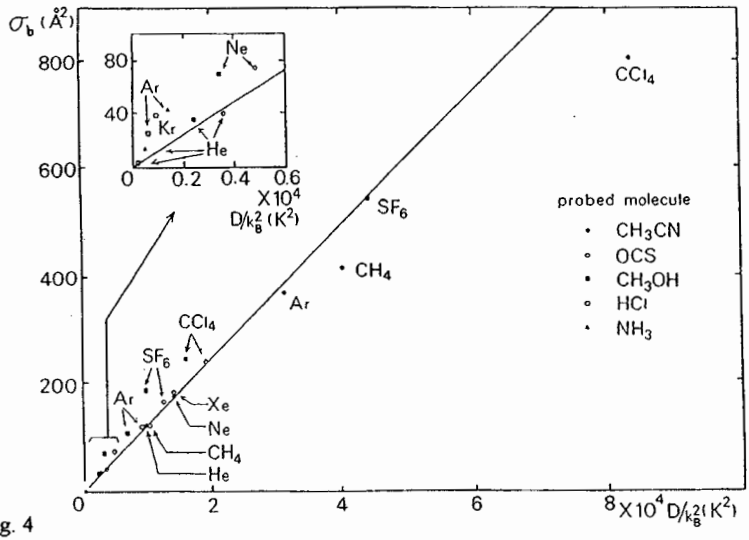


Fig. 4

Collisional broadening cross section σ_b between polar molecules and spherical perturbers as a function of the maximum interaction given by square of the matrix element of the interaction Hamiltonian as in Ref. /5/

state of the valence electron can be interpreted as analogous to rotational transitions. Practically just the θ-dependent term of Eq. (1) has non-diagonal matrix elements which can cause the fine structure changing /2/. It is therefore reasonable that the proportionality between σ_b and $\varepsilon + \varepsilon_c$, found with molecular spectra, holds also with the Cs D_2 line.

The collisional frequency shift is caused by the energy shift of the unperturbed levels as described by the interatomic potential. The frequency shift at the distance where the interaction is the strongest is

$$\Delta\nu_{max} = [\, \varepsilon_p \text{ (upper state)} - \varepsilon_p \text{ (lower state)} \,] / h , \qquad (2)$$

where h denotes the Planck constant, and ε_p represents the depth of the van der Waals potential between the Cs atom and the perturber, which is proportional to $\sqrt{\varepsilon}$ /7/. The linearity between σ_s and $\sqrt{\varepsilon}$ shows the linearity between the cross section and $\Delta\nu_{max}$. It is reasonable to presume that σ_s is determined just by $\Delta\nu_{max}$, considering that even if some other parameter is necessary, it is independent of the perturbers.

4. Conclusion

In analogy with the molecular model of Ref. /5/ the present analysis shows that the collisional broadening and shift cross sections of the Cs D_2 line can be estimated by the depth of the van der Waals potential of the perturbers. It is interpreted that the collisional broadening cross section is determined just by the transition probability at the distance where the interaction is strongest and that the collisional shift cross section is determined just by the frequency shift where the interaction is the strongest.

5. References

/1/ Pascale, J. and Vandeplanque, J.: J. Chem. Phys. **60** (1974) 2278.

/2/ Baylis, W.E.: J. Chem. Phys. **51** (1969) 2665.

/3/ Czuchaj, E. and Sienkiewicz, J.: Z. Naturforsch. **34a** (1979) 694.

/4/ Inoue, Y., Uchida, K., Hori,H. and Sakurai, T.: J. Phys. Soc. Jpn. **59** (1990) 516.

/5/ Kajita, M., Odashima, H., and Shimizu, T.: J. Chem. Phys. **94** (1991) 6014.

/6/ Kielkopf, J.F.: J. Phys. B: Atom. Molec. Phys. **9** (1976) L547.

/7/ Hirschfelder, J.O., Curtiss, C.F., and Bird, R.B.:
Molecular Theory of Gases and Liquids (Wiley, New York, 1954).

Frequency Stabilization of a Rb Maser with an External Feedback Loop by Self-Tuning of Cavity Resonance Frequency

Nobunori Oura, Tsuyoshi Masato, Tomoyuki Doi,
Fumio Asamura, Naimu Kuramochi* and Yasusada Ohta**
Tokyo Institute of Technology, *Toin University of Yokohama,
**Communications Research Laboratory, Ministry of Postal Services
4259 Nagatsuta, Midoriku, Yokohama 227, Japan

1. Introduction

A Rb maser operating in TE_{021} cavity mode pumped by a laser diode light with a wavelength of $780nm$ is reported/1/. The experiment of a Rb maser with a cavity in TE_{011} mode, as small as one-eighth in volume the existing one, with an external feedback loop pumped by a diode laser with a wavelength of $780nm$ has appeared/2 /. The improvements have been performed on the temperature compensation for the cavity resonance frequency, examination of the Q-value of the external feedback loop, divergence of pumping light in the cavity with a pair of divergence lens, and temperature control circuit of the cavity. These improvements did not make the frequency stabler than expected.

To improve the frequency stability, the self-tuning(auto-tuning) of the cavity resonance frequency method devised for hydrogen maser/3/4/ was introduced to the Rb maser. As a result, the frequency stability was improved by one order of magnitude. The adjustment and improvement of the phase locked loop, the bias-Tee circuit with a varactor diode encapsulated in a holder enhanced the frequency stability.

2. Self-tuning method

The oscillating frequency of a maser can be expressed by Eq.(1) /4/5/,

$$f_m - f_{Rb} = P_l(f_c - f_{Rb}) \tag{1}$$

where f_m is the maser frequency, f_{Rb} the transition frequency of Rb atoms, f_c the cavity resonance frequency. The factor P_l is the pulling factor defined as follows,

$$P_l = \frac{Q_c}{Q_l} \tag{2}$$

where Q_l is the line Q of Rb atoms, about 10^9, Q_c the Q-factor of the cavity, in this case the enhanced Q-factor of the maser is about 20×10^4.

Equation (1) means that the frequency stability of a Rb maser derived from the transition frequency of Rb atoms is degraded by a factor of P_l, about 10^{-4}, by the cavity pulling. The degraded frequency stability can be improved as in a hydrogen maser/4/5/ with introduction of the self-tuning of the cavity resonance method, that is, the adjustment of the cavity resonance frequency to the transition frequency by detecting the frequency difference between the cavity resonance frequency and the transition frequency of Rb atoms.

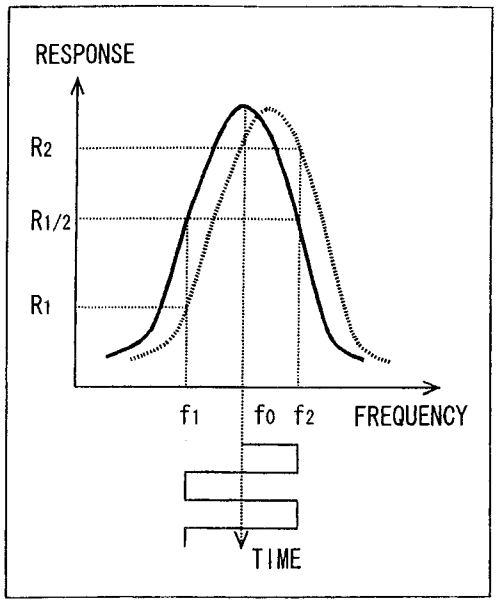


Fig. 1: Principle of the cavity tuning

The transmission resonance curve of the cavity can be considered symmetrical around the resonance frequency as shown in Fig.1. When the test signals with the identical amplitude at frequencies $f_1 = f_0 - \Delta f$ and $f_2 = f_0 + \Delta f$ are injected alternatively into the cavity, the signal amplitudes detected by synchronous-detection from the cavity at each test frequency are the same. Therefore, the output of the detector is null, if the resonance frequency coincides with the transition frequency, otherwise the signal amplitudes are different to each other. The output voltage, namely, $(R_2 - R_1)$ is an error signal, which is delivered as the control signal to the varactor diode to adjust the resonant frequency of the cavity.

3. Experiment of the self-tunig of the cavity

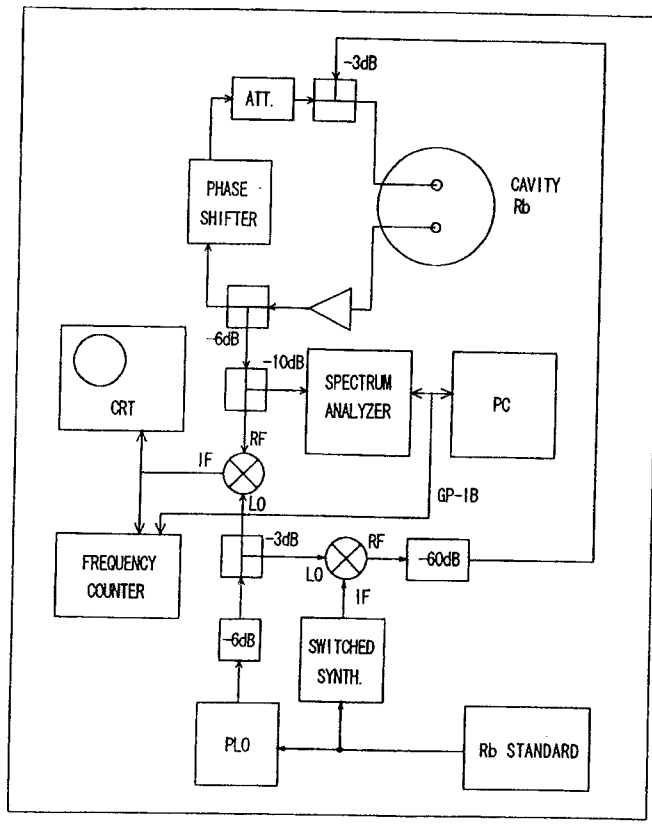


Fig. 2: Schematic diagram for observing the resonance curve affected by the injection signals to the cavity

3.1 Experiment of the test signal injection

A test signal from the switched synthesizer shown in Fig.2 was injected into the cavity in maser operation to investigate whether the resonance curve of the cavity was symmetrical about the resonance frequency. The resonance curve of the cavity was symmetrical and the half value at full width was about 15kHz. The frequency spectrums of the cavity observed on the display of a spectrum analyzer are depicted in Fig.3 at an injection frequency step of 2kHz as a parameter, when the injection frequency was varied from 10kHz below the resonant frequency to 10kHz above the resonance frequency at a 2kHz step. The magnified spectrum around the center frequency is shown in Fig.4.

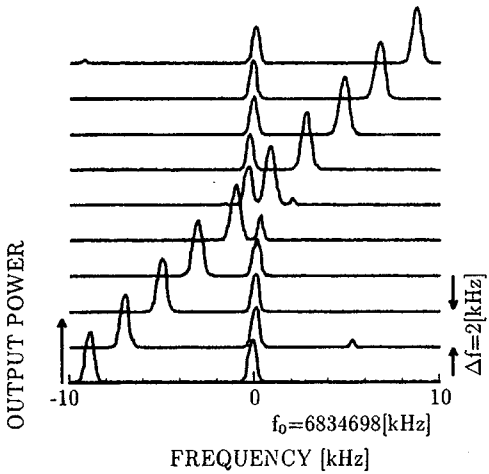


Fig. 3: Observed resonance curves affected by the injection signals

If the frequency of the test signal is close to the center frequency within ±5kHz, it can be seen that the test signal affects the resonance characteristics of the cavity. From Fig.4 it is desirable that the detune frequency of the probing signal from the center frequency should be 5kHz to 20kHz.

3.2 Improvement of the circuits and frequency stability

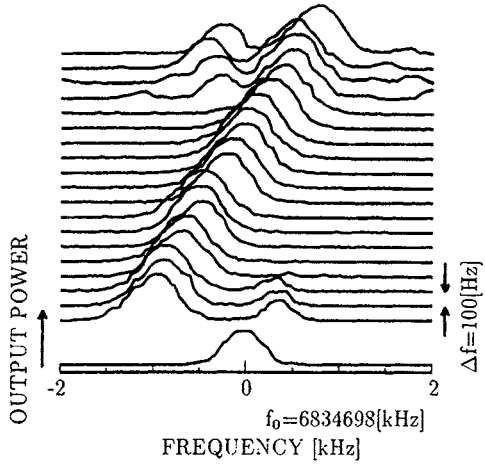


Fig. 4: Magnified profiles of those in Fig.3

A schematic diagram of the maser is shown in Fig.5. Coaxial cables of different length were employed as a phase shifter to adjust the signal phase to the bias-Tee and the varactor in an experimental setup in the early stage. The precise adjustment of the phase could not be achieved by exchanging the different length of the coaxial cable so that the variable phase shifter was employed in place of the coaxial cable as shown in Fig.5. The varactor diode was encapsulated in a holder and connected to the bias-Tee. When the maser with an external feedback loop operates only as a microwave oscillator, the oscillating frequency response without overshoot and hysteresis as shown in Fig.6 was observed by applying a step voltage to the bias-Tee instead of the synchronous detection voltage. Although the applied voltage to the bias-Tee was stepwise, the response was not so owing to the time constant of the recorder. The cavity is shielded magnetically with high permeability substance (μ metal) threefold. The resonance frequency of the cavity is compensated thermally with design by combining of different materials for the end-plate and cylinder. The frequency temperature coefficient of the cavity at 60°C is -0.18kHz/°C. The cavity cylinder is not glued with adhesive cement with the end-plates but the end-plates are placed on the cylinder. The cavity temperature is kept within ±0.05°C by PID control.

The frequency trend of the maser is shown in Fig.7 together with overlapping the trend without the self-tuning method. The frequency stability is improved by one order of magnitude than that without the self-tunig method.

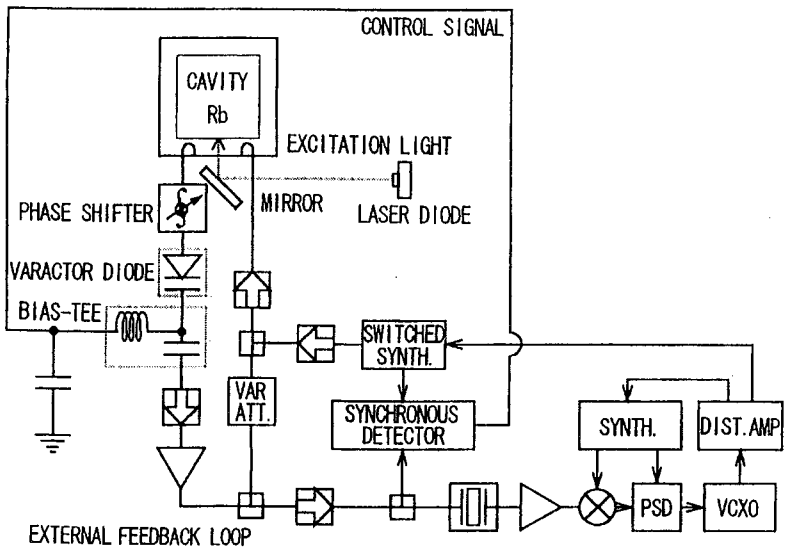


Fig. 5: Schematic diagram of the Rb maser

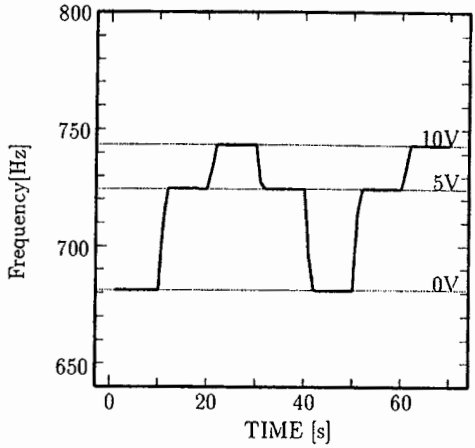


Fig. 6: Step responce of the bias-Tee circuit

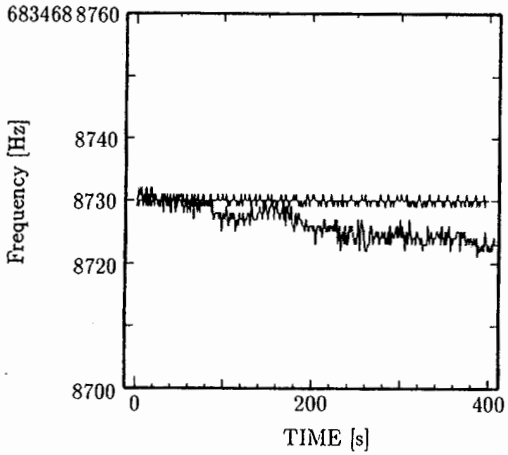


Fig. 7: Oscillating frequency trend of the Rb maser

4. Conclusion

The introduction of the self-tuning method to the Rb maser with an external feedback loop improved the frequency stability by one order of magnitude than that of the maser without the self-tuning method.

The frequency stability of the maser is beyond that expected as a whole, since the frequency stability of the VCXO is approximately in a few parts par 10^{10} and the electronic circuits are not shielded electrically and controlled thermally and the cavity is not glued. The frequency will be stabler with the improvements of these facotrs.

Bibliography

/1/A.Michaud, P.Trembly, and M.Tetu:"Experimantal study of the laser diode pumped rubidium maser,"IEEE Trans. Instrum. Meas.,IM-40(2),p.170, 1991

/2/N.Oura et al.:"Fundamental characteristics of the laser-diode-pumped rubidium maser with an external feedback loop,"Proc. 6th EFTF, p.527, Netherland, March 1992

/3/H.T.M.Wang:"An oscillating compact hydorgen maser,"34th AFCS,p.364,May 1980

/4/Y.Ohota et al.:" Development of a very small hydrogen maser using a loop-gap resonator," Trans.I.E.I.C.E.,C-I, vol.J74-C- I,No.6,pp.222,June 1991 (in Japanese)

/5/P.Kartaschoff;"Frequency and time,"p.53, Academic Press, 1978

Proton Irradiation Test of the Teflon® Wall Coating in the Atomic Hydrogen Maser in Preparation to Space Applications

C.Couplet and L.-G. Bernier

Observatoire Cantonal - Rue de l'Observatoire, 58 - CH-2000 Neuchâtel - SWITZERLAND

Abstract

In order to determine what is the damage caused by radiations in orbit to the Teflon® wall coating of the storage bulb in the hydrogen maser, a proton irradiation test was performed on a standard hydrogen maser built by Neuchâtel Observatory. The results of the test show that we can't see any degradation of the performances of the maser after irradiation.

1. Introduction

High energy protons encountered by spacecrafts in the Earth's Trapped Radiation Belts are known to break polymers, producing radicals. On the other hand, radicals on the wall coating are known to be a cause of relaxation in the hydrogen maser.
We have performed a radiation test on a hydrogen maser precisely to determine if the Teflon® coating in the storage bulb is degraded by space radiations. We adapted the experimental conditions to the standard ground maser in order to have the same radiation dose as in the future space maser.

2. Orbit Simulation

Trapped radiation model

An orbit simulation was performed at ESTEC assuming the Meteor 3M Russian satellite orbit. The parameters of the Meteor 3M nearly circular orbit are 98.5° inclination and 950 km altitude.
The trapped radiation models employed were those developed by Vette and colleagues [1]. The flux along the 14-orbit trajectory were averaged to produce average particle spectra. The three year mission considered for Meteor 3M, is entirely during a period of solar minimum activity (1993 to 1999), so the solar flare fluencies were not considered.
The mission averaged spectra were used to calculate the cumulative dose deposition over the three year mission as a function of the aluminium shielding thickness (see shieldose code [2]) as seen on figure 1.

The shielding of the space maser

We have calculated the equivalent shielding of the magnetic shields layers and of the sapphire loaded microwave cavity that surround the Teflon® coating in the compact space maser designed by Neuchâtel Observatory. The shielding assumptions are shown in table 1.

Effective radiation fluxes on the Teflon® coating

Assuming a shield of 5 g/cm^2 corresponding to 20 mm of aluminium, it was determined that a negligible amount of electrons and Bremsstrahlung get to the Teflon® wall coating.

As shown on figure 1, the only significant dose is produced by the protons. The dose of protons after 20 mm of aluminium is equivalent to 1500 rad for three year.

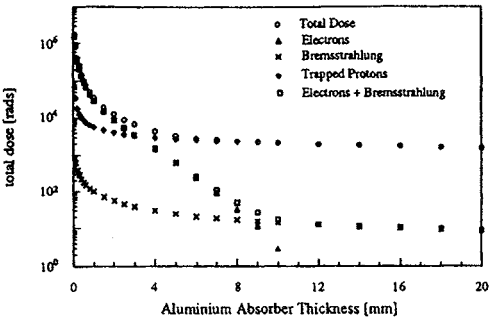


Fig. 1: Dose at center of aluminium spheres for a period of 3 years

layer number	thickness [mm]	density [g/cm^3]	material	shield [g/cm^2]
4	0.5	8.7	magnetic shields	1.74
1	5	3.97	sapphire cylinder	1.99
1	5	2.7	Al microwave cavity	1.35
			Total	5

Table 1: Shielding Assumptions for Compact Space Maser

Energies of the radiations

We must also consider the energies of the radiation: only the protons of energies > 67.5 Mev, corresponding to a range of 5 g/cm^2 aluminium, come to the Teflon®.
We calculated the orbit average flux spectrum as shown in figure 2.
We know the integral flux for 67.5 Mev (corresponding to a range of 5 g/cm^2 aluminium) and we can calculate that the fluency on the Teflon® wall coating over the three year mission (9.4673 10^7 sec) is equal to 1.42 10^{10} cm^{-2}.
For the test, it was also important to calculate the maximum flux. The peak proton flux over the orbit is 3'000 cm^{-2} s^{-1} for proton of energies > 67.5 Mev.

3. Effect of Radiation on Teflon

According to Dupont documentation [3] and [4], the Teflon® FEP (used in hydrogen maser) has a threshold of mechanical damage in the range 2-7 10^5 rad in air and 2-7 10^6 rad in vacuum. It retains 50% of initial tensile strength after a dose of 10^7 rad in air and 10^8 rad in vacuum. Teflon® FEP will tolerate a dose 10-100 times larger than TFE. The mechanical damages are larger on TFE than on FEP

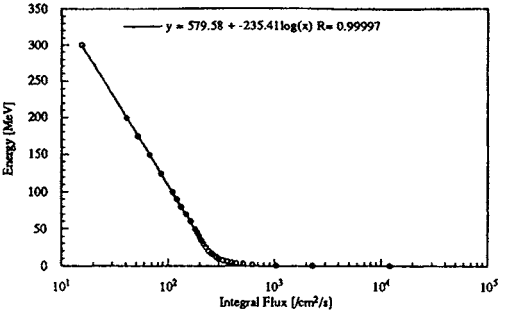


Fig. 2: Orbit average spectra of trapped protons at solar minimum

since the mechanical properties depend directly on the length of the polymer chains. A break of the chain by the radiations reduce more the molecular weight in TFE than in FEP even if the number of radicals created is the same. We expect a lot of radicals to be formed in irradiated Teflon® before the mechanical properties are degraded. The atomic hydrogen stored in the Teflon® coated storage bulb is likely to react with the radicals created by the radiation eventually producing new radicals. In any case, the radicals on the surface of the Teflon® coating are a source of atomic relaxation. In air, the oxygen molecules can react with the radicals on the Teflon® coated surface and stabilize them. The effect of radiation is then more important but we are not directly interested in the reactions in air while we work only in vacuum. Therefore, knowing the radiation dose on Teflon® in space hydrogen maser and the importance of radicals at the surface, we thought that it was important to make this experiment.

4. Experiment

Test Maser

For this test, we irradiate a functioning standard atomic hydrogen maser. It is important to operate the maser continuously through the entire duration of the irradiation test because if the hydrogen flux is stopped or the maser dismounted, the relaxation properties of the Teflon® coated surface will change. In fact, it is better if the bulb is not in contact with air and the hydrogen flux is constant.
For the experiment, we must take into account that the standard maser used in the test has a different physical package than the space maser design. The shielding characteristics of the standard maser are shown in table 2.

layer number	thickness [mm]	density [g/cm^3]	material	shield [g/cm^2]
1	1	8.2	supra 50	0.82
3	0.8	8.7	shielding (mumetal)	2.088
	total: 19.44	2.7	aluminium	5.2488
1	1.5		quartz	negligible
			Total	**8.16**

Table 2: Shielding characteristics for Standard Maser

We have corrected for the energy degradation and for the flux dilution produced by the 8 g/cm^2 shielding of the standard test maser in order to properly simulate the 5 g/cm^2 shielding of the space maser.

Proton Irradiation Facility (PIF) at PSI

At PSI (Villigen - Switzerland), we made a ground simulation of space protons irradiation. PIF delivers protons with energies and flux comparable with those found in space. The PSI cyclotron delivers a proton beam with energy of 590 Mev and a beam splitter directs a few percent of the beam into the PIF beam line. Two different degraders reduce the original proton energy to the desired value: the first one, made of carbon and copper, define the initial energy: 300,254,150,102 and 60 Mev and the second one, made of aluminium slabs, with total thickness between 0 and 255 mm can reduce the beam energy to the energy desired. Beam collimators are used to obtain the desired parameters of the beam. A detector and a radiation counter, mounted directly in front of the irradiated sample normalize the proton flux. [5]

Geometry of Maser Irradiation

Projected on the horizontal plane, the storage bulb has a 140 mm diameter and is placed at the center of the 500 mm x 500 mm physics package. The storage bulb is cylindrical with hemispherical ends. The vertical length of the storage bulb is 30 cm.
The proton beam of the PIF is directed at the center of the bulb. It was impossible to move the maser up and down because of space limitation in the lab. It would have been impossible to irradiate the lower half of the storage bulb. Moreover the largest proton beam that can be produced is 100 mm x 100 mm. Therefore only a limited portion of the storage bulb could be irradiated at a time.
It was decided to irradiate only a 100 mm x 100 mm area in the middle of the top half of the storage bulb. Taking into account the curvature of the cylindrical storage bulb, the area of the wall coating that lies across the 100 mm x 100 mm beam is 2 x 112 cm^2. The comparison with the 1300 cm^2 total surface of the storage bulb shows that more than 17 % (more because of the diffraction of the beam) of the storage bulb wall coating was irradiated.
The electronic package in the standard maser is below and above the storage bulb and is not irradiated .

Test Plan

For the test, it was important to consider on the one hand the peak proton flux over the orbit in space and, on the other hand, the total fluency on the Teflon® wall coating over three years. The test plan is shown in table 3. The indicated fluxes are those seen by the Teflon storage bulb.

Run N°	Purpose	duration [hours]	mean flux on Teflon® [cm^{-2} s^{-1}]	fluency on Teflon® [cm^{-2}]
1	test peak flux	1	3'000	1.08E+07
2	test about 10 times peak flux	1	40'000	1.44E+08
3	3 years mission fluency in 10 hours	5	400'000	7.2E+09
4	15 years fluency	5	4'000'000	7.2E+10

Table 3 : Test Plan

It was also decided to irradiate with 3 different proton energies in order to simulate the energy spectrum encountered in orbit. We consider the energy at center of maser, on the Teflon® wall coating and we used aluminium degraders for the 300 Mev proton beam of PIF to achieve the desired energies. The degrader settings are shown in table 4

aluminium degrader [mm]	energy incident on maser [Mev]	energy at center of maser [Mev]	energy spectrum weight factor
89	227.8	200.3	0.188
175	140.7	100.3	0.375
201	106.3	50.6	0.438

Table 4: Degrader Settings for Specified Proton Energy at Center of Maser

The incident energy is attenuated because of the shielding of the physics package around the Teflon® coating (30 mm or 8 g/cm^2 aluminium equivalent). We must also take into account the weight factors for the different energies. In table 5 are shown the detailed dosages for the four runs.

run N°	energy at the center of the maser [Mev]	energy spectrum weight factor at center	fluency incident [cm^{-2}]	flux attenuation by maser shielding	fluency at center of the maser [cm^{-2}]	total fluency
1	200.3	0.188	3.22E+06	1.589	2.03E+06	
	100.3	0.375	7.53E+06	1.859	4.05E+06	
	50.6	0.438	9.88E+06	2.091	4.73E+06	1.08E+07
2	200.3	0.188	4.29E+07	1.589	2.7E+07	
	100.3	0.375	1.00E+08	1.859	5.38E+07	
	50.6	0.438	1.32E+08	2.091	6.31E+07	1.44E+08
3	200.3	0.188	2.15E+09	1.589	1.35E+09	
	100.3	0.375	5.02E+09	1.859	2.7E+09	
	50.6	0.438	6.59E+09	2.091	3.15E+09	7.2E+09
4	200.3	0.188	2.15E+10	1.589	1.35E+10	
	100.3	0.375	5.02E+10	1.859	2.7E+10	
	50.6	0.438	6.59E+10	2.091	3.15E+10	7.2E+10

Table 5: Detailed dose for the four runs of the plan

5. Realization and results

Amplitude measurement

The test was performed at PIF on July 6 and 7, 1993. The 34 monitoring channels of the maser were recorded in real time from the control room. In addition the analog maser signal amplitude output was recorded. Table 6 show the observations made during the test.
A supplementary run with low energy protons was also performed. Low energy protons are expected to cause maximum damage if the degradation mechanism is linked to energy deposition. The aluminium degrader was set to 21.1 cm so that the total shield including the maser was 24.1 cm which is the critical length for 300 Mev protons. With this setting the protons at the center of the maser were almost stopped and had the maximum probability of losing the rest of their energy in the wall coating.

We observed that the amplitude did not change during the test.

run N°	effective experiment	observations
1	as planned	no signal degradation
2	as planned	no signal degradation
3	interrupted during the third period at energy 50 Mev, the accumulated fluency is 80% of the planned value and corresponds to 1.2 years in orbit.	no signal degradation
4	interrupted during the third period at energy 50 Mev. The accumulated fluency is 76 % of the planned value and corresponds to 11.4 years in orbit.	no signal degradation
5 (supplementary)	Low energy protons. Energy null at center of maser. The protons are stopped in the maser bulb. The total fluency is 1.0 10^9 cm^{-2}.	no signal degradation

Table 6 : Test realization on standard maser

Quality Factor Measurement

Figure 3 shows the atomic quality factor without the spin exchange broadening contribution, i.e. extrapolated to zero hydrogen pressure. The quality factor was measured three times before the radiation test and several times after, up to 25.2.94 (figure 3). Within the measurement error a degradation of the Teflon performance cannot be demonstrated. There is even a slight trend towards an improvement of the quality factor with time.

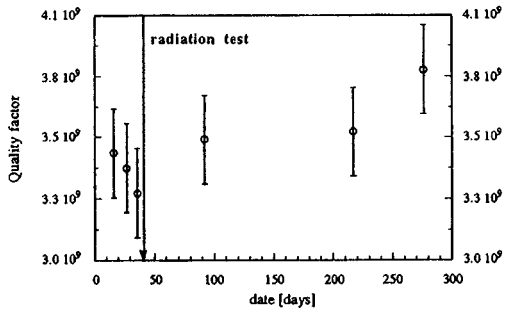


Fig. 3 :Atomic Quality factor of the maser before and after the test

6. Conclusion

The total proton irradiation fluency used in the test corresponds to more than 12 years of maser operation in orbit and no signal degradation of the maser signal was observed. A cross check by the measurement of the atomic quality factor before and after irradiation does not show a degradation of the Teflon coating performance either.
The test was very realistic because it was performed under real operating conditions.

The main conclusion is that the shielding effect of the hydrogen maser physics package combined with the very small thickness of the Teflon coating are such that the probability of interaction of space protons with the Teflon coating is very low and the radicals formed does not damage the Teflon® in its interaction with hydrogen.
Therefore the radiation environment encountered in orbit seem to be safe regarding the operation of the hydrogen maser physics package.

7. Acknowledgements

We want to thank Wojtek Hajdas and Felix Mattenberger of the PIF facility as well as Eamonn Daly and Hugh Evans of ESTEC for their contributions to this experiment.

This work was perform under ESTEC contract N° 9099 / 90 / NL / US (SC), W.O.6.

8. References

[1] J.I. Vette, "Trapped radiation models in Development of Improved Models of the Earth's Radiation Environment",
Technical Note 1 (Chapter 4), 1989.

[2] S.M. Seltzer, "SHIELDOSE: a computer code for space shielding radiaiton dose calculations",
TN-1116 (NBS), 1980

[3] "Radiation Tolerance of "Teflon" Resins", The journal of Teflon® Reprint N° 37

[4] "Radiation Effects on Teflon®: fluorocarbon resin" by Frank L. Bouquet, Graham,
WA : Systems Company, cop. 1989

[5] Proton Irradiation Facility PIF, PSI Paul Scherrer Institute, Space Technology,
Summary Report - 1992

Development of a Cryogenic Hydrogen Maser as a Flywheel Frequency Standard

R. Mossavati, J. C. Gallop, C. F. Wall
Centre for Basic Metrology, National Physical Laboratory
Teddington, TW11 0LW, UK.

Summary

We have begun to develop a prototype Cold Hydrogen Maser (CHM). The main features of this CHM, which is designed to operate in the pulsed mode at 4.2 K, is the use of low loss alumina for the fabrication of the microwave cavity in order to reduce the overall size. A numerical simulation was performed to find the dimensions of the microwave cavity required for the TE_{011} mode to be at 1.420 GHz, and the model has been confirmed experimentally. The CHM is intended to be used as a flywheel frequency standard at the NPL with an expected medium-term stability of one part in 10^{16} or better.

1. The CHM

The hydrogen maser was devised by Ramsey and coworkers /1/ in the early 1960's. The cold hydrogen maser is expected to improve, by up to three orders of magnitude /2/, on the performance of the best frequency standards currently available (Fig.1). In our CHM (Fig.2) the hydrogen molecules are dissociated at room temperature in a discharge bulb energised by an rf circuit which resonates at about 70 MHz. They are then formed into a beam of atoms. After flowing down a PTFE lined transport tube into the cryostat, they are state selected by a hexapole magnet. Atoms in the (F=1) states travel into the maser bulb where they decay into the (F=0) state and radiate power at about 1420 MHz. The storage cell is located within a sufficiently high-Q microwave cavity so that an oscillation at the resonant frequency builds up until an equilibrium value is reached. The emitted signal, which is of the order of 10^{-12}W, is picked up by a probe positioned near one end of the cavity, amplified and transmitted out of the cryostat through a low loss microwave transmission line. Following a further stage of amplification, the signal is mixed down to audio frequencies and detected.

The frequency of the microwave cavity mode has to coincide with the hyperfine frequency of the hydrogen, which is 1 420 405 751 Hz in a flux density of 0.1 μT . The requirements of the hydrogen maser determine the ideal cavity mode to be the TE_{011}. A numerical simulation was performed to find the dimensions of the cavity required for obtaining the TE_{011} mode at the hyperfine transition frequency of the hydrogen atom. The simulation assumes an outer diameter of 115mm and solves Maxwell's equations with the appropriate boundary conditions to obtain the inner diameter and length of the microwave cavity that will realise the TE_{011} mode at the hyperfine transition frequency. The outcome is a cavity that consists of a tube of alumina 130 mm long, 115 mm outer diameter, and 68

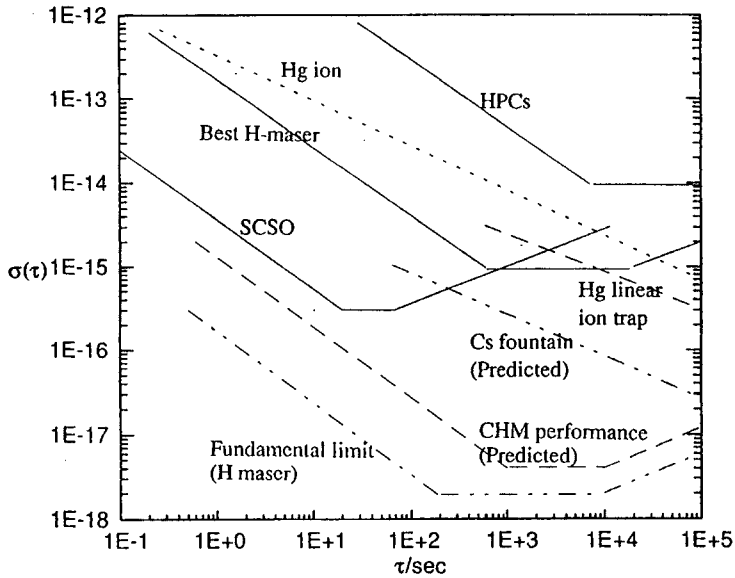


Fig. 1. Comparison of the CHM predicted stability against other leading frequency discriminators

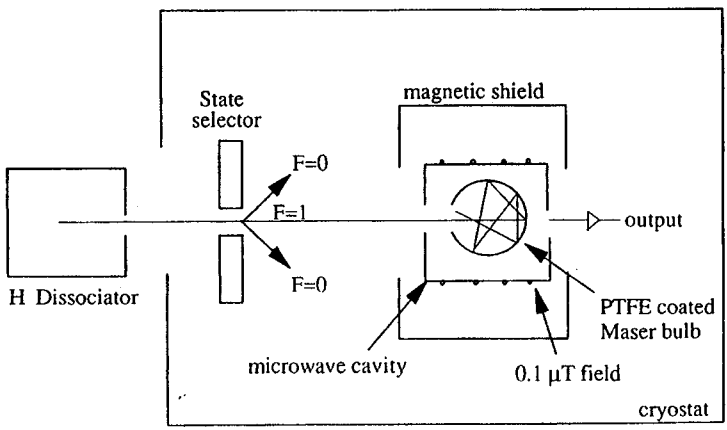


Fig. 2. Schematic diagram of the CHM

mm inner diameter which is bounded by OFHC copper plates at either end. Cavity tuning is accomplished by reducing the length of the inner region and using a screw inserted parallel to the axis half a radius away from the centre. The outer surface of the cavity is silvered to improve its Q factor, which varies from 5000 at room temperature to about 30,000 at 4.2 K. The resonant frequency of the cavity increases by 12.5 MHz on cooldown (Fig.3).

The state selector hexapole magnet is of conventional design /3/ but is adapted for low temperatures by reducing the length to 13mm, and keeping the aperture diameter 3mm. It is composed of high coercivity NdFeB magnet pieces inserted into a soft magnet frame made out of vanadium permendur. There is a uniform positive field gradient of 1.1 T from the centre, maximising at the pole tips.

Wall coating of the masing volume is another important factor influencing the overall performance of the CHM. We have performed extensive PTFE coating trials using a solution of Teflon AF1600 dissolved in Fluorinert FC77, applied to a 5cm diameter quartz bulb with a narrow stem. This yields a uniform, transparent coating without any detectible surface defects. We are about to investigate its effectiveness as a buffer layer on the maser bulb at 4.2 K, and will then test a solution of FEP 120. We then hope to use frozen hydrogen on top of the PTFE to confine the hydrogen atoms at 4.2K.

We have calculated the type and number of shields needed to keep out stray magnetic fields and the earth's field. This suggested that a three layer nested arrangement of mu-metal shields will give the required shielding factor of approximately 10^4. Additional low temperature shielding is provided by a Cryoperm /4/ shield which surrounds the experiment and provides a shielding factor of 1500 at 4.2K.

2. Pulsed operation of the CHM

At low fields the frequency of the (F=1, m_F=0) to (F=0, m_F=0) hyperfine transition of the ground electronic state of the H atom can be written (to 2nd order in H_0) :

$$\nu(H_0) = a/h + h^2(\gamma_e+\gamma_p)^2 H_0^2/(2a/h) \qquad (1)$$

where γ_e and γ_p are the absolute values of the gyromagnetic ratios for the electron and proton, H_0 is the applied magnetic field and a is the hyperfine coupling constant (a/h = 1 420 405 751 Hz). Thus one can have narrow lines, whose width is limited only by the interactions of the atoms with themselves and the walls. When separated the two levels can be treated as an effective spin 1/2 system, with the z component of the applied field couples to the x component of the fictitious spin with effective gyromagnetic ratio $\gamma_e + \gamma_p$. So the system behaves identically to a standard magnetic resonance system, and the FID signal from a pulsed input can be treated in the same way.

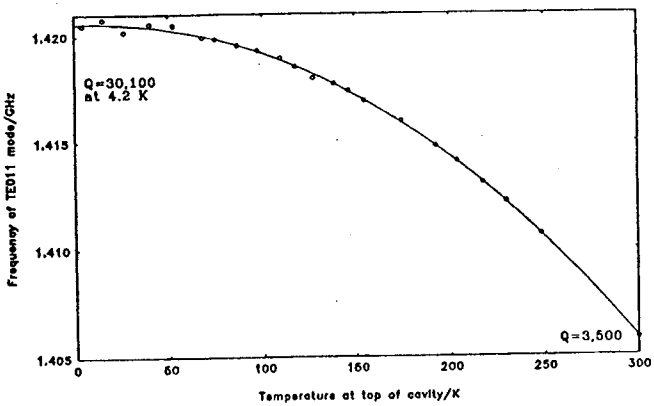


Fig.3. Temperature dependence of the TE011 mode of the microwave cavity

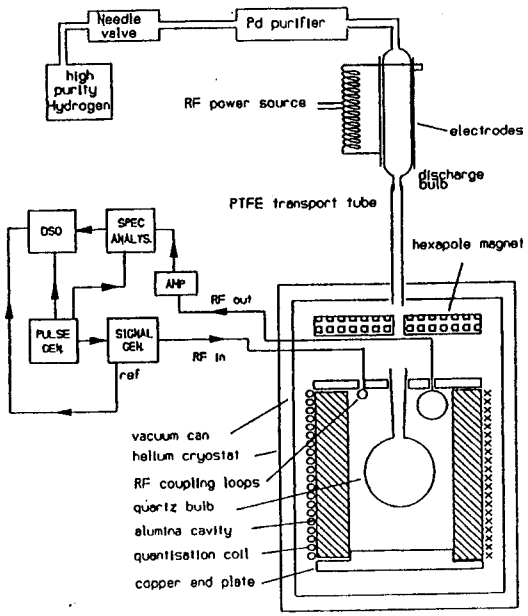


Fig.4. Pulsed operation of the CHM

The receiver system for the pulsed operation of the CHM is displayed in (Fig.4). The sensitivity of the receiver is 10^{-14}W which is sufficiently high for an expected output signal of 10^{-12}W. During the recent run with the CHM assembly, we attempted a range of pulse lengths ($1\mu s < \tau < 1ms$) and powers ($10^{-6}W < P < 10^{-2}W$) but failed to observe an output signal. This could be due to a number of causes, the likeliest of which are the low H atom intensity and poor focusing of the hydrogen atom beam into the masing volume. Investigation of these characteristics necessitated means of detecting H beam intensity and profile.

3. H atom detection

We have successfully detected atomic hydrogen at room temperature using molybdenum trioxide (MoO_3) film evaporated onto a glass substrate, which change from milky white to blue on exposure to atomic hydrogen /5/. Although it is a useful means of obtaining qualitative information on beam geometry, the technique is relatively insensitive, and remote observations of the colour change are inconvenient.

So in order to obtain a quantitative measure of the beam intensity and a more precise geometrical resolution at low temperatures, we have resorted to using a sensitive microcalorimeter made out of phosphorus doped silicon which is capable of measuring 0.5 pW at 4.2 K over an area of about 0.2 mm^2 (Fig.5). Resistance of the bolometric element as a function of temperature is shown in (Fig.6), which also illustrates the increased sensitivity at low temperatures. The detection threshold corresponds to a flux of 10^6 hydrogen atoms per second impinging and recombining on the device. Since we expect a flux of 10^{12} to 10^{13} atoms per second entering the maser bulb, this should give us a very sensitive hydrogen atom sensor.

Acknowledgements

The support of the members of the Centre for Basic Metrology is gratefully acknowledged. The microcalorimeter was supplied by Dr. de Podesta of Birkbeck College.

References

/1/ Kleppner, D.; Goldenberg, H. M.; and Ramsey, N. F. : *Theory of the hydrogen maser*, in Phys. Rev., Vol. 126 (1962), pp 603-615.

/2/ Hardy, W. N.; Morrow, M.; Jochemson, R.; and Berlinsky, A.J. : Physica, Vol. 109 (1982), pp 1964-1977.

/3/ Vanier, J. and Audoin, C. : *The quantum Physics of Atomic Frequency Standards* (Adam Hilger, Bristol, 1989), pp. 478-534.

/4/ ®Vacuumschmelz GMBH, Germany.

/5/ Ramsey, N. F. : *Molecular Beams* (Oxford University Press, New York, 1963), p.377.

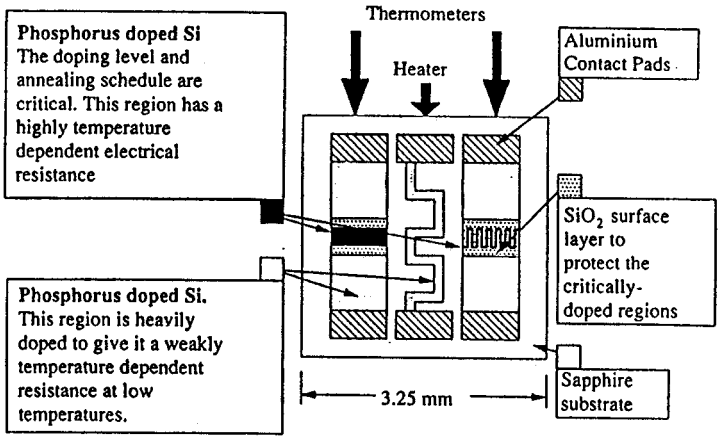


Fig.5. The microcalorimeter

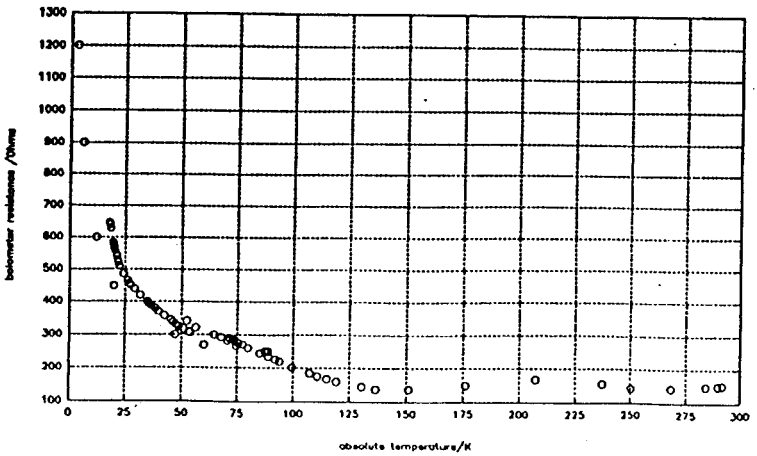


Fig.6. Resistance characteristic of the thermometric element of the microcolorimeter.

Preliminary Design and Breadboarding of a Compact Space Qualified Hydrogen Maser Based on a Sapphire Loaded Microwave Cavity

Laurent-Guy Bernier
Neuchâtel Observatory
Switzerland

1. Abstract

A spaceborne compact atomic hydrogen maser is being developed by Neuchâtel Observatory for the EXTRAS joint ESA-RSA H-maser experiment foreseen to be flown on the Russian Meteor 3M satellite at the end of 1997.

A compact sapphire loaded microwave cavity was designed at Neuchâtel Observatory to be used in the space H-maser. The ON-100M sapphire loaded cavity has an internal volume of 5 liters as compared to the 20 liters of a conventional full size cavity. The storage volume for hydrogen is 1.7 liter, i.e. not much smaller than a full-size design. The first ON-100 M sapphire resonator was evaluated in an aluminium test cavity.

The loss in atomic quality factor and in frequency stability in a compact design based on the ON-100 M compact cavity is small as compared to a full size design and the gain in size and mass is remarkable. The design goals of the compact space qualified instrument are 50 kg total mass and 60 W total electrical power. The structural and thermal design philosophy of the ON-SHM space hydrogen maser is reported.

IMVP, the Russian Institute of Metrology for Time and Space, was the first to design and to operate a sapphire loaded miniature microwave cavity into a compact hydrogen maser. In collaboration with IMVP, Neuchâtel Observatory has evaluated one of the Russian "SAPPHIR" compact cavities into an EFOS full size maser produced by ON. An operating quality factor of 1.5×10^9 was demonstrated for a hydrogen storage volume which is only 0.65 liters.

2. Introduction

This paper reports three related activities that were performed at Neuchâtel Observatory in a spaceborne hydrogen maser development project. Firstly the evaluation of the Russian "SAPPHIR" sapphire loaded microwave cavity in a full size EFOS hydrogen maser manufactured by ON is reported. Secondly the design by ON of the ON-100M sapphire loaded microwave cavity optimized for the largest storage volume for hydrogen is decribed as well as the microwave evaluation of the first prototype. Thirdly the general design of the ON-SHM spaceborne design now under development is summarized.

3. The Joint ESA-RSA "H-maser in Space Mission"

The opportunity for an in-orbit experiment using two hydrogen maser frequency standards and the associated microwave and optical tracking equipment on the Russian METEOR 3 M Earth observation satellite has been identified by ESA and RSA /1/.

Twenty-five METEOR 2 and five METEOR 3 Earth observation satellites have been launched so far. The METEOR 3 M spacecraft will be the first of the 3 M series and is

planned to be launched at the end of 1997. The baseline orbit is sun-synchronous at an altitude of 1000 km. 200 kg of payload are still available on the 3000 kg spacecraft for an opportunity flight. The nominal operational lifetime of the spacecraft is three years /2/.

A combination of Russian and West-European equipment will be flown on the EXTRAS joint ESA-RSA mission. Spaceborne instruments include two hydrogen masers, one Russian designed by IMVP, one West-European designed by ON, the two-way microwave ranging instrument PRARE /3/ which have been flown successfully very recently on a METEOR 3 S/C, TROLL an improved laser ranging equipment similar to LASSO /4/, and a GLONASS/GPS receiver.

The goal of the experiment is very precise in-orbit time/frequency generation, as well as very precise time transfer, synchronization and dissemination with by products in the areas of orbit determination, point positioning and atmospheric sounding by GPS/GLONASS occultation techniques. The foreseen future applications of the precise in-orbit time & frequency technologies to be demonstrated in the experiment include very precise navigation, global geodesy and geotectonic, atmosphere physics and radio wave propagation, fundamental physics and relativistic effects, time synchronization and dissemination.

4. Optimization of the Physics Parameters

The fundamental white phase noise and white frequency noise processes produced by the physics of the hydrogen maser determine the following Allan deviation limit:

$$\sigma_y(\tau) = \sqrt{\frac{3\, f_c\, kT}{2\pi\, \nu_0^2\, P_0}\left(\frac{\beta}{1+\beta}+\frac{F}{4}\right)\frac{1}{\tau^2}+\frac{kT}{2\, P_0\, Q^2}\left(\frac{\beta}{1+\beta}\right)\frac{1}{\tau}} \qquad (1)$$

assuming there is an isolator between the cavity coupling loop and the receiver. f_c is the bandwidth of the measurement, P_0 the maser output power, β the coupling factor of the cavity, ν_0 the frequency of the atomic signal, F the noise figure of the receiver and Q the quality factor of the atomic line. The signal power P_0 is determined by the available hydrogen flux I [atoms s^{-1}] and by the threshold of oscillation. The atomic quality factor Q is determined mainly by the spin exchange relaxation rate, dependent upon the hydrogen flux, by the "wall" relaxation rate, dependent upon the geometry of the storage bulb and upon the quality of the Teflon coating which covers the inside wall of the storage bulb, and by the escape rate γ_b [s^{-1}] of the storage bulb, determined by the geometry of the storage bulb collimator.

For a given flux I and a given storage bulb geometry there is an optimum escape rate γ_b that maximizes the frequency stability of the hydrogen maser. The optimization criteria and procedures are described in details in /5/. Note that the optimum escape rate does not depend upon the relaxation characteristics of the Teflon wall coating but on the hydrogen flux and on the geometry only. It is also shown in /5/ that for a given flux the optimum escape rate is higher in a compact H-maser with reduced storage volume for hydrogen than in a full size maser.

5. Evaluation of IMVP "SAPPHIR" Compact Cavity

5.1. Introduction

Within the scope of a preliminary space H-maser development phase performed for ESTEC at ON in preparation to the joint experiment, a Russian miniature sapphire loaded

microwave cavity, on loan from IMVP-VNIIFTRI, was evaluated in a full-size EFOS hydrogen maser built by ON. IMVP, the "Institute of Metrology for Time and Space" is a section of NPO-VNIIFTRI, the "Russian National Scientific and Research Institute for Physical-Technical and Radio-Technical Measurements". The tested cavity is of the same type currently used in the compact "SAPPHIR" hydrogen maser produced by IMVP-VNIIFTRI /6/. In fact this IMVP realization is the first application of a sapphire loaded microwave cavity to an active H-maser design.

The IMVP cavity design and its evaluation at ON is fully reported in /7/. The main results are reproduced below in section 5.3.

5.2. Theoretical Verification of the Microwave Design

The IMVP theoretical design analysis takes into account metallic losses only /8/. Dielectric losses are assumed negligible. The design procedure starts with a given inner radius r_1 of the sapphire tube and a given length h. Then for a variable outer radius r_2 of the sapphire cylinder, the radius r_3 of the circular metallic cavity is adjusted for having the TE_{011} mode resonating at the 1420.405 MHz hydrogen resonant frequency. The thickness of the sapphire cylinder is $e = r_2 - r_1$. The larger r_2, the thicker the cylinder, the larger the dielectric loading and the smaller the surrounding metallic cavity. The smaller the metallic cavity and the smaller the quality factor as determined by the metallic losses. In fact the quality factor depends upon the ratio between the stored energy, proportional to the inner volume, and the metallic losses, proportional to the inner surface. The more miniature the cavity the worse the volume to surface ratio. The "SAPPHIR" design is about the smallest possible cavity with a reasonable storage volume for hydrogen and a quality factor sufficient for active maser operation.

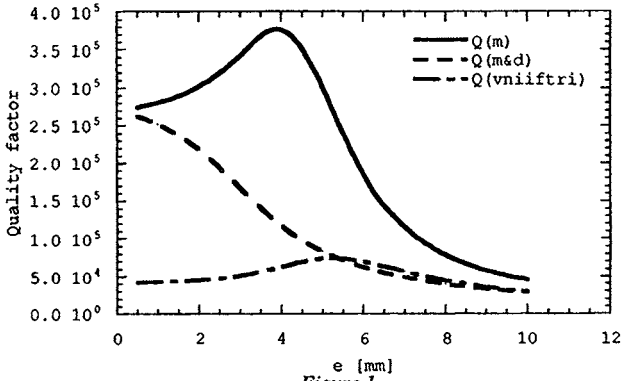


Figure 1
Verification of IMVP Computation for r1 = 40 mm

The IMVP analysis for $r_1 = 40$ mm was verified using ON own analysis method /9/. On figure 1 the quality factor Q(m) takes into account our calculation of the metallic losses only. Q(m&d) takes into account our calculation of both the metallic and dielectric losses. Q(vniiftri) on the other hand is the IMVP calculation that takes into account only metallic losses. Note that in both ON and IMVP calculations an optimum quality factor exists when the dielectric losses are neglected. When both the dielectric and the metallic losses

are taken into account, as does ON calculation, the quality factor decreases monotonously as the thickness of the dielectric tube is increased. For e = 7 mm, which corresponds to the "SAPPHIR" cavity, both ON and IMVP calculations yield the same quality factor. The difference is that IMVP calculation adjusts the surface conductivity parameter in order to justify the experimental quality factor of the "SAPPHIR" cavity through metallic losses only. ON calculation, on the other hand, uses an independent estimation of the surface conductivity of silver and adjusts the dielectric losses of sapphire in order to match the experimental "SAPPHIR" quality factor. By the way this is how the assumed design parameters of table 9 were determined. The problem here is that precise published values for the dielectric losses of sapphire at room temperature are not available.

It was pointed out by Dr. Gaigerov of IMVP that ON computation of the quality factor for e = 0 is optimistic with respect to the analytical calculation of the quality factor in an empty circular cavity. ON calculation based on the 3 layer dielectric formalism yield Q = 263'000 while the analytic formula yield Q = 115'000. Note that e = 0 is an extreme case of a circular cavity with a height of 140 mm and a diameter of 388 mm. In the case of the TE_{021} cavity of /11/, on the other hand, the computed quality factor is 68'000 while the analytical value is 65'000 for the same assumed surface conductivity. Therefore it seems that the discrepancy between the computed and the analytical quality factor is a function of the ratio between the radius and the height. In the case of an empty circular cavity with a height to diameter ratio close to one, the computed value using the 3 layer formalism is very close to the analytical value. The reason for the discrepancy is still to be investigated.

In any case our conclusion after this verification was that according to our own estimations the dielectric losses of sapphire could not be neglected and that it was important to check experimentally on a prototype the exact contribution of the metallic and dielectric losses.

5.3. Full Evaluation in an EFOS Maser at Neuchâtel Observatory

As mentioned above the evaluation at ON of the "SAPPHIR" compact microwave cavity & sapphire storage bulb assembly is fully reported in /7/. The atomic line quality factor was measured for different values of the hydrogen flux and the spin-exchange broadening effect was removed by extrapolation to zero flux. Neglecting magnetic relaxation, the extrapolated quality factor Q(I=0) contains only two contributions: the "wall" relaxation rate γ_{2w} associated with the atomic collisions on the Teflon coated storage bulb and the storage bulb escape rate γ_b. The latter can be calculated from the geometry of the storage bulb and of the collimator. The measured physics parameters of the storage bulb are reported in table 1. For comparison purposes the physics parameters of the EFOS storage bulb for a Teflon coating having the same relaxation probability per collision p_{2w} as the Russian Teflon used in the "SAPPHIR" storage bulb was computed and is reported in table 2. These values are typical of recent EFOS storage bulbs. This indicates that the IMVP Teflon coating has relaxation properties very close to those of the Teflon coating used in recent EFOS masers manufactured by ON.

Q(I=0)	2.67×10^9
γ_b [s^{-1}]	0.99
γ_{2w} [s^{-1}]	0.683
p_{2w}	1.6×10^{-5}

Table 1
Measured Storage Bulb Parameters of IMVP "SAPPHIR" Cavity

Figure 2 shows the atomic quality factor versus the maser output power. The circles indicate the experimental values while the solid line represents the theoretical model. The model is the same as in /5/ and the model parameters are set to the values of tables 3 and 4.

Q(I=0)	4.13×10^9
γ_b [s^{-1}]	0.74
γ_{2w} [s^{-1}]	0.34
p_{2w}	1.6×10^{-5}

Table 2
Theoretical Storage Bulb Parameters
Assuming IMVP Teflon Coating Properties
and Standard EFOS Storage Bulb Geometry

The correspondence between experimental values and the model is very good. Note that the filling factor of a dielectric loaded cavity is theoretically larger than the assumed $\eta' = 0.45$. However the fitting of the model is not improved if the value of η' is increased.

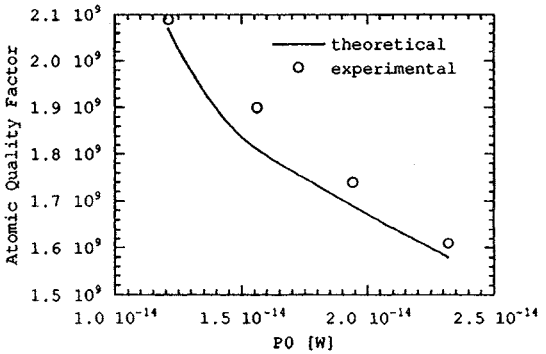


Figure 2
IMVP "SAPPHIR" Compact Design
Atomic Quality Factor vs Atomic Signal Power
Experimental vs Theoretical Model

For a -105 dBm maser signal power the measured operating quality factor of the "SAPPHIR" cavity was 1.5×10^9. This is reported in table 3 as well as the other measured operating parameters. The filling factor η' and the noise figure F of table 4 are assumed theoretical values. Using the parameters of table 3 and 4 the fundamental limit to frequency stability defined by equation (1) was computed and is shown in table 5. Note that the theoretical limit does not include the flicker noise floor and other environmental effects.

V_b = 0.65 liter
$\beta = 0.1$
Q_c = 47'000
$Q = 1.5 \times 10^9$ for P_0 = 105 dBm
$\gamma_b = 0.988\ s^{-1}$
$\gamma_{2w} = 0.683\ s^{-1}$

Table 3
Experimental Physics Parameters
Measured with "SAPPHIR" Compact Cavity

$\eta' = 0.45$
$F = 2$
f_c = 1 Hz

Table 4
Theoretical Parameters Assumed for the Computation
of the "SAPPHIR" Frequency Stability Limits

tau [s]	Allan dev.
1	1.5×10^{-13}
10	2.2×10^{-14}
100	5.5×10^{-15}
1000	1.7×10^{-15}
10'000	5.3×10^{-16}

Table 5
Theoretical Frequency Stability Limits
of IMVP "SAPPHIR" Cavity & Bulb Assembly

6. Theoretical Design of ON-100M Microwave Cavity

6.1. Introduction

ON theoretical analysis is based on a general theory that takes into account both metallic and dielectric losses for an arbitrary number of concentric layers of dielectric material in a circular geometry /9/. Table 6 defines the geometrical parameters of the circular cavity loaded by a single cylinder of sapphire. Our design procedure was the following. We fixed the size of the circular metallic cavity by fixing r_3 and h. h is 160 mm in all cases and r_3 is 80, 90 or 100 mm. Then we varied r_1, which determines the storage volume for hydrogen, and we computed the thickness $e = r_2 - r_1$ of the sapphire cylinder which produces the correct frequency of the TE_{011} mode and the associated quality factor. The result of the optimization process is the cavity that yields the largest storage volume for hydrogen for a given volume of the metallic cavity and with the minimum acceptable quality factor.

r_1 (D_1)	inner radius of sapphire cylinder
r_2 (D_2)	outer radius of sapphire cylinder
r_3 (D_3)	radius (diameter) of metallic circular cavity
h	height of sapphire cylinder identical to height of circular metallic cavity

Table 6
Geometry of Circular, Partially Filled, Dielectric Loaded Microwave Cavity

ON Designs	ON-80	ON-90	ON-100
h [mm]	160	160	160
r_1 [mm]	39	51	63
r_2 [mm]	46.8	57.3	68.4
r_3 [mm]	80	90	100
e [mm]	7.8	6.3	5.4
storage volume [cm^3]	760	1'310	2'000
cavity volume [cm^3]	3'220	4'070	5'030
Q	46'466	45'571	45'042
assumptions : dielectric losses of sapphire $\tan(\delta) = 1.5 \times 10^{-5}$ surface conductivity of silver $\sigma = 6.29 \times 10^7\ \Omega^{-1}\ m^{-1}$			

Table 7
Characteristics of ON Selected Design Options

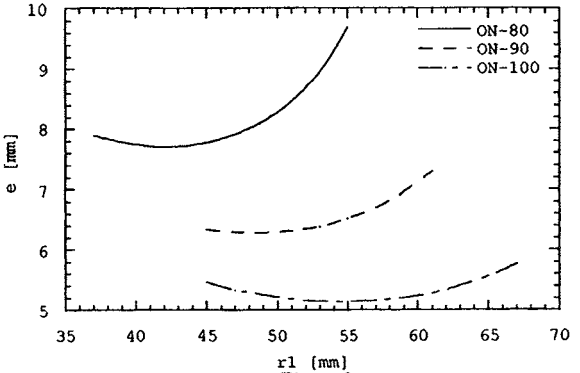


Figure 2
e vs r_1 for ON Design Options

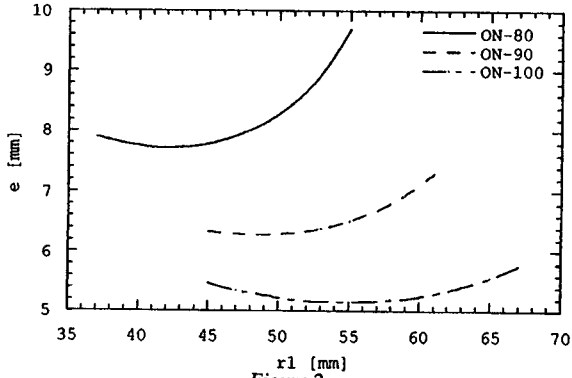


Figure 2
e vs r_1 for ON Design Options

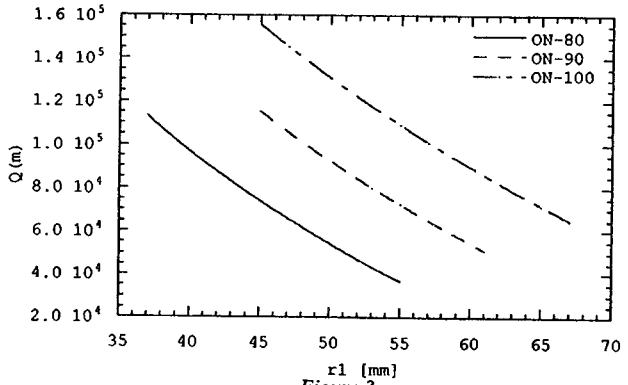


Figure 3
Metallic Losses vs r_1 of ON Design Options

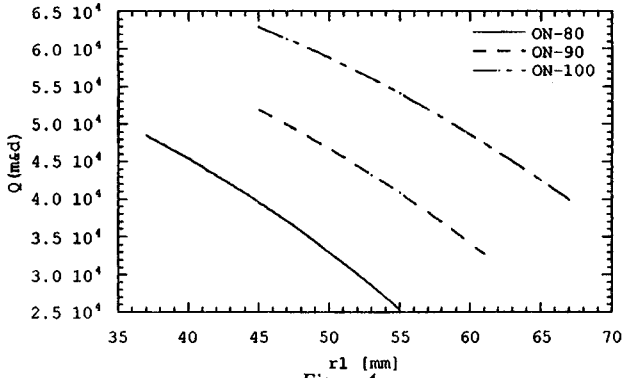


Figure 4
Metallic and Dielectric Losses vs r_1 of ON Design Options

6.3. The ON-100 M Sapphire Loaded Cavity Design

The ON-100 M sapphire loaded cavity was designed at ON as a development for ESTEC. The parameters of the ON-100 M design are summarized in table 8. The symmetry is truly cylindrical since the sapphire cylinder extends up to the metallic covers of the circular metallic cavity. For design purposes the cavity is assumed to be silver plated titanium like the IMVP cavity. The dimensions of the sapphire resonator and of the metallic cavity were computed theoretically for the resonance of the TE_{011} mode at the hydrogen frequency of 1420.405 MHz using the assumed design parameters of table 10.

The dielectric constants of sapphire are anisotropic, i.e. depend on the direction of the electric field, and depend also on both temperature and frequency. The dielectric constants found in ceramic handbooks and review articles usually come from reference /10/. Precise values of the dielectric parameters of sapphire as a function of the frequency and temperature is not available.

D_1 [mm]	125
D_2 [mm]	137
D_3 [mm]	200
h [mm]	140
V [cm^3] storage volume for hydrogen	1'700
ν_0 [MHz]resonant frequency of TE_{011} mode	1420.405
unloaded quality factor Q of TE_{011} mode	47'000

Table 8
ON-100M Sapphire Loaded Cavity Theoretical Design

The assumed dielectric constant of table 10 was taken from /10/ as the best estimation available. The assumed loss tangent was estimated from the published experimental quality factor of the IMVP "SAPPHIR" cavity /6/ and by removing the metallic losses contribution using the effective conductivity of silver and our own theoretical model.

In fact the metallic losses cannot be computed precisely using the DC value of the surface conductivity of the metal because at microwave frequencies the roughness of a machined metallic surface is of the same order of magnitude as the skin depth. The effective value of $4 \times 10^7\ \Omega^{-1}\ m^{-1}$ indicated in table 10 is to be compared to the DC conductivity of silver which is $6.29 \times 10^7\ \Omega^{-1}\ m^{-1}$. The effective value was deduced from the experimental quality factor measured in a silver plated TE_{021} cavity used in a Rubidium maser /11/.

The theoretical unloaded quality factor indicated in table 8 is based on the assumed design parameters of table 10. On the other hand the geometrical dimensions of table 8 will yield the correct TE_{011} resonant frequency only if the assumed dielectric constant of sapphire is correct.

7. Evaluation of The ON-100 M Test Microwave Cavity

Once the theoretical design of the ON-100 M compact cavity was ready, the first prototype ON-100 M sapphire resonator was ordered. The edges of the sapphire cylinder are optically polished. The diameters are fine ground. The cylinder is oriented along the optical axis in order to minimize dielectric losses. The tolerances on the length and radiuses are $\pm$ 25 µm. The sapphire crystal is not entirely of optical grade. In fact it is not possible to grow a crystal this large with optical grade quality throughout. Meanwhile an aluminium test cavity was manufactured at ON to the ON-100 M dimensions in order to verify the microwave design.

The theoretical sensitivities of the TE_{011} mode to the diameters and to the length of the cavity are given in table 9.

dv_0/dD_1 [MHz/mm]	-35
dv_0/dD_2 [MHz/mm]	-33
dv_0/dD_3 [MHz/mm]	-4.9
dv_0/dh [MHz/mm]	-2.3

Table 9
Theoretical Sensitivities to Diameters and Length of TE_{011} mode in the ON-100 M Cavity

The thermal coefficient of the ON-100 M compact cavity was computed theoretically assuming a titanium metallic cavity. The theoretical coefficient of the TE_{011} mode in a titanium ON-100 M cavity is -69 kHz/K°. The main contribution to the thermal coefficient is not the thermal expansion of the materials but the variation of the dielectric constant of sapphire with temperature.

The experimental measurements made on the aluminium test cavity yielded an unloaded quality factor of 68'000 at 2.115 GHz for the TE_{011} mode in the empty cavity, i.e. without the sapphire resonator. This result was used to deduce the effective surface

conductivity of aluminium reported in table 10. An unloaded quality factor of 33'000 was measured at 1416 MHz for the TE_{011} mode with the sapphire resonator in place. This result was used to deduce the effective dielectric losses of sapphire reported in table 10. The actual length of the test cavity is not 140.0 mm but 139.6 mm because of a groove in the bottom plate used to center the sapphire cylinder. The resonant frequency of the TE_{011} mode at room temperature is 1416 MHz instead of the theoretical 1419 MHz. The difference in frequency is smaller than the uncertainty due to the manufacturing tolerances of the sapphire cylinder. Therefore the dielectric constant $\varepsilon = 9.36$ assumed for sapphire when the electric field is normal to the optical axis is correct.

Table 10 compares the parameters assumed in the theoretical design with the same parameters adjusted against the experimental verification of the design. The main conclusion is that the dielectric losses in sapphire are a little higher than expected and in any case not negligible as compared to the metallic losses.

A theoretical calculation based on the adjusted parameters of table 10 yields an unloaded quality factor of 69'000 for the ON-100 M sapphire resonator in a silver plated cavity assuming the dielectric losses of sapphire are negligible. The quality factor drops to 39'000 if the dielectric losses are taken into account.

	assumed for design	adjusted after experiment
σ [Ω^{-1} m^{-1}] aluminium		3.2×10^7
σ [Ω^{-1} m^{-1}] silver	4×10^7	
tan(δ) sapphire E field normal	1.5×10^{-5}	1.7×10^{-5}
ε sapphire E field normal	9.36	9.36

Table 10
Assumed Design Parameters versus Experimental Verification

8. Outline of ON-SHM Space Hydrogen Maser Design

8.1. Introduction

Our special design approach for the compact ON-SHM physics package includes the following features:

- Sapphire-loaded microwave cavity
  As mentioned above the ON-100 M design reduces mass and size of the microwave cavity considerably. Note that also the mass of the magnetic shields is greatly reduced from 38 kg in the EFOS maser to about 4.2 kg in the SHM.

- Bulk getter pumped vacuum system
  Avoids heavy ion pumps as used in the ground masers. Only a pair of light-weight miniature ion pumps (2 l/s) are used to pump non getterable gases.

- Solid-state hydrogen storage
  This solution avoids pressure tanks, a potential fracture critical item, and saves volume and mass.

- Three stage thermal control of the microwave cavity
  Good temperature control is required because of the relatively high temperature

coefficient of a sapphire loaded cavity. The three stage thermal control is specified to regulate the temperature of the microwave cavity with a residual Allan deviation of $\sigma_T(10'000\ s) < 0.008\ °C$.

- Automatic pulse interrogation cavity tuning system
  Assures good long-term stability without degrading the spectral purity.
- Advanced teflonization technique for hydrogen storage bulb. Reduces ageing and assures high atomic resonance quality factor.

8.2. Maser Physics Design

I [atoms/s]	1×10^{12}
V_b [liter]	1.7
γ_b [s^{-1}] optimum	0.77
γ_{2w} [s^{-1}]	0.62
Q (operating)	2.2×10^9
P_0 [dBm]	-104
Qc (unloaded)	34'000
β	0.25
F	2

Table 11
ON-100 M Design Goal Physics Parameters

For the design goal parameters of table 11 based on the ON-100 M geometry the fundamental limit of the Allan deviation in a 1 Hz bandwidth yields:

$$\sigma_y(\tau) = \sqrt{1.98 \times 10^{-26} \tau^{-2} + 2.42 \times 10^{-27} \tau^{-1}} \qquad (2)$$

The fundamental limit is reported in table 12 and is not very far from the limits of a full-size design.

τ [s]	$\sigma_y(t)$
1	1.5×10^{-13}
10	2.1×10^{-14}
100	5.1×10^{-15}
1000	1.6×10^{-15}
10'000	4.5×10^{-16}

Table 12
Fundamental Frequency Stability Limitation
from ON-100 M Design Goal Physics Parameters

8.3. Mechanical Design

Figure 6 shows a top assembly drawing of the ON-SHM physics package. The main mechanical features are the following. The microwave & magnetic shield assembly is mounted on a structural titanium tube that provides a high rigidity and at the same time a

relatively low thermal conductivity to the base plate. The sapphire resonator will be sealed to the titanium covers using an aluminium diffusion bonding technique that provides a vacuum seal and at the same time a high strength mechanical link to the microwave cavity.

8.4. Thermal Design

The thermal design philosophy is to limit as much as possible radiative thermal exchanges with outer space by use of thermal vacuum inside the physics package and MLI insulation blankets around the instrument. The thermal link by conduction through the structural tube between the microwave cavity & magnetic shields assembly and the base plate is controlled by a three stage active thermal control system. The critical electronics that requires thermal control is mounted on a special electronics ring along the structural tube.

8.5. Electronics Design

Figure 5 shows the block diagram of the maser electronics. The main features will be the following. There are two quartz oscillators, one slave at 100 MHz and the master at 10 MHz controlled by the main PLL. The thermal control of the microwave cavity is specified to ± 0.008 C° in order to maintain frequency fluctuations due to the second-order Doppler effect at a negligible level. The cavity pulling effect will be maintained to a very low level by the use of an automatic cavity tuning system.

9. Conclusion

The activities reported in this paper were performed for ESTEC in the course of a feasibility study. This phase has come to an end in December 1993. Further preliminary development and breadboarding activities have been performed since and will be reported soon. The main phase of development of the ON-SHM spaceborne hydrogen maser is expected to start in the second half of 1994.

10. References

/1/ Proc. of the *First Working Meeting on "Hydrogen Maser Clocks in Space for Solid-Earth Research and Time-Transfer Applications, A Joint West-European and Russian Experiment"*, ESTEC, Noordwijk, the Netherlands, June 29th and 30th 1993.

/2/ *METEOR 3 M Spacecraft Accomodation Handbook*, document ref. HM-SP-001-KT, issue 1, Kayser & Threde gmbh.

/3/ Schäfer W., "The Application of the PRARE Principle for Two-Way Time Transfer to Space-Based Atomic Clocks", *Proc. 6th European Frequency and Time Forum*, ESTEC Noordwijk, March 17-19, 1992, ESA-SP 340, pp.41-47.

/4/ C. Veillet, "Time Transfer between USA and France through LASSO", *Proc. 7th European Frequency and Time Forum*,, March 16-18 1993, Neuchâtel, Switzerland, pp. 637-640.

/5/ Bernier L.G., "Optimization of the Escape Rate for Best Frequency Stability in the Atomic Hydrogen Maser", *Proc. 7th European Frequency and Time Forum*, Neuchâtel, March 16-18, 1993, pp. 373-377.

/6/ Gaygerov B.A., Rusin F.S., Sysoev V.P., "Portable Atomic Clock on the Basis of an Active Hydrogen Maser "Sapphir" ", *Proc. 5th European Frequency and Time Forum*, Besançon, March 1991, pp. 293-294.

/7/ Busca G., Bernier L.G., Silvestrin P., Feltham S., Gaygerov B.A., Tatarenkov V.M., " Hydrogen Maser Clocks in Space for Solid-Earth Research and Time-Transfer Applications - Experiment Overview and Evaluation of Russian Miniature Sapphire Loaded Cavity ", *Proc. PTTI '93*, Marina del Rey, Nov. 29 - Dec. 2, 1993.

/8/ Rusin F.S., Gaygerov B.A., Ponomarenko S.F., "Calculation of a Cylindrical Cavity Partially Filled with a Dielectric", Measurement Techniques, vol. 28, no. 5, 1985, pp. 396-398, Transl. from Izmeritel'naya Tekhnika no.5, May 1985, pp. 15-16.

/9/ Sphicopoulos T., Bernier L.G., Gardiol F., "Theoretical Basis for the Design of the Radially Stratified Dielectric Loaded Cavities Used in Miniaturized Atomic Frequency Standards", *IEE Proceedings*, vol. 131, part H, no. 2, April 1984, pp. 94-98.

/10/ von Hippel, *MIT Technical Report N°126* (1953, 1959).

/11/ Têtu M., *Le Maser au Rubidium 87 et sa Stabilité à Court-Terme*, Thèse de Doctorat (PhD Thesis), Département de Physique, Université Laval, Québec, Canada, Novembre 1972.

/12/ Dick, G.J., Strayer D.M., "Measurements and Analysis of Cryogenic Sapphire Dielectric Resonators and DRO's, " *Proc. 41st Annual Frequency Control Symposium*, Philadelphia, May 1987, pp. 487-491.

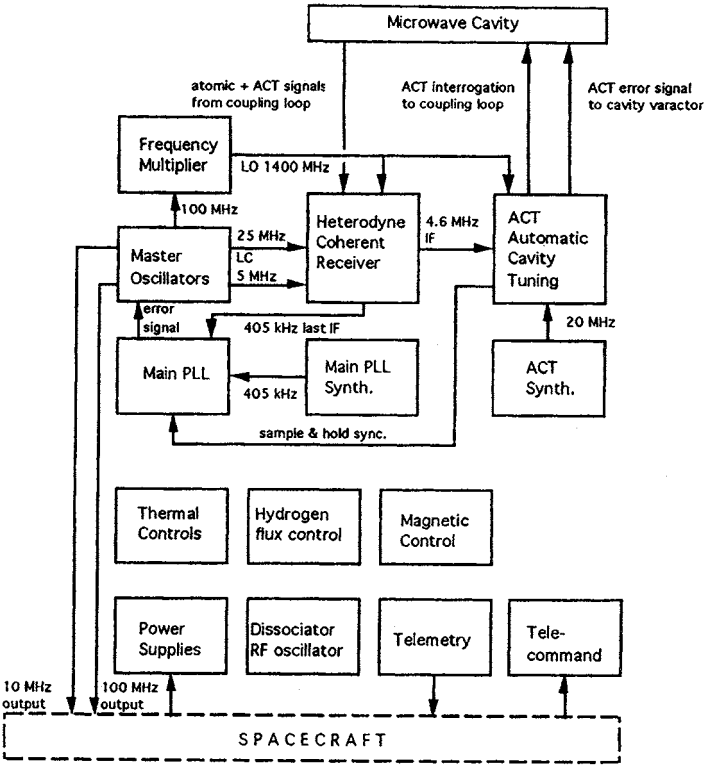


Figure 5
Block Diagram of Electronics Package

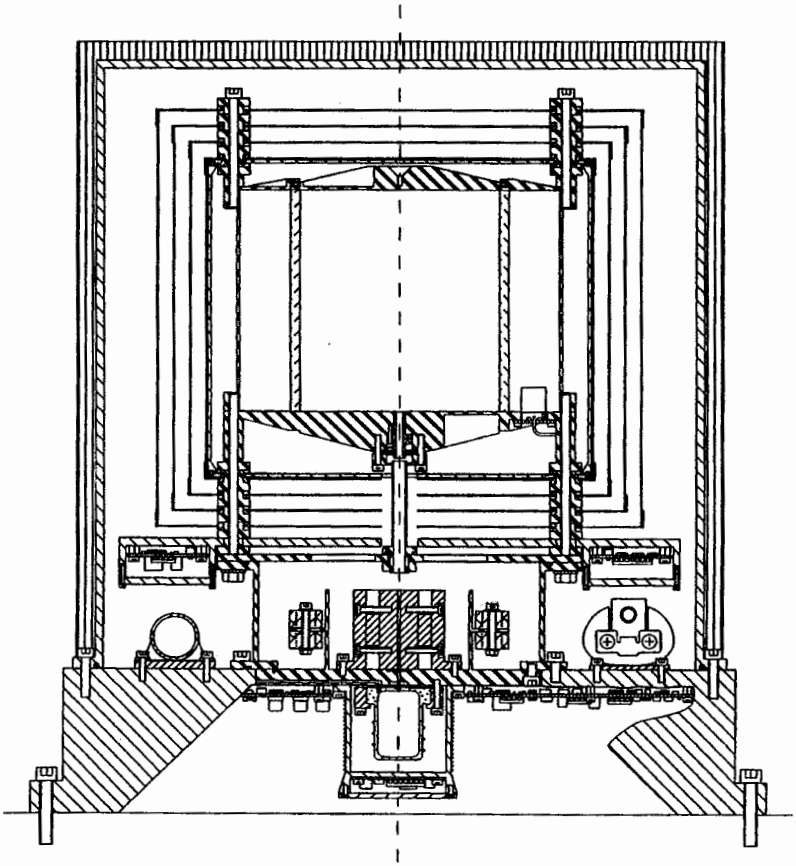


Figure 6
Top Assembly Drawing
ON-SHM

STUDY

OF TROPOSPHERIC CORRECTION FOR GPS COMMON-VIEW TIME TRANSFER BETWEEN THE BIPM AND THE OCA

W. Lewandowski
Bureau International des Poids et Mesures
Pavillon de Breteuil
92312 Sèvres Cedex, France

P. Grüdler, F. Baumont
Observatoire de la Côte d'Azur
Av. Copernic, 06130 Grasse, France

ABSTRACT

In GPS common-view time transfer, simple global models are used for the estimation of the tropospheric delay. At present, different models are in employment and they can differ by a few nanoseconds. Comparisons of these models with a semi-empirical model based on weather measurements show differences of several nanoseconds. This paper reports on a study of tropospheric correction during GPS common-view time transfer over a baseline of about 700 km between the BIPM and the OCA. It is shown that the use of the same simple model of the troposphere within a region of the similar climate does not affect time transfer by more than a few hundreds of picoseconds.

1. INTRODUCTION

During the last few years significant progress has been made in improving GPS common-view time transfer by the introduction of ultra-accurate ground-antenna coordinates, precise satellite ephemerides and dual-frequency ionospheric measurements [1]. Estimated precision and accuracy of some GPS common-view time links reaches a few nanoseconds [2]. Further progress is still possible towards sub-nanosecond region.

Among the improvements required to reach this goal is increased accuracy in the estimation of the tropospheric delay. At radio frequencies, the troposphere is a non-dispersive medium and its effect on pseudo-ranges cannot be estimated from

dual frequency measurements as is done for the ionosphere. Instead, models are used for the estimation of the tropospheric delay [3, 4, 5]. It has been assumed until recently that for the needs of GPS time transfer at the level of 1 ns to 2 ns and for observations performed at elevation angles above 30°, a simple global model is sufficient. However, carrying out common-view time transfer over long distances (9000 km), elevations of 20° are sometimes unavoidable. Also, different types of receivers use different tropospheric models which can differ by a few nanoseconds for low elevation angles [1, 6]. On this issue progress is possible by implementing recently established standards for receiver software including a common model for estimating signal delays due to tropospheric refraction [7].

To obtain an accuracy of a nanosecond or better in worldwide GPS common-view time transfer, more sophisticated models of the troposphere which incorporate local meteorological measurements are needed. Comparison of the models currently used by GPS time receivers with a semi-empirical model based on weather measurements show differences of several nanoseconds [8, 9]. This discrepancy increases for observations performed in hot and humid regions of the world.

This paper briefly describes and compares tropospheric models, and reports on a study of tropospheric correction during GPS common-view time transfer over a baseline of about 700 km between the Bureau International des Poids et Mesures (BIPM) and the Observatoire de la Côte d'Azur (OCA). It is shown that the use of the same simple model of the troposphere within a region of similar climate does not affect time transfer by more than a few hundreds of picoseconds.

2. TROPOSPHERIC DELAY

With regard to the transmission of light waves and microwaves, the atmosphere may conveniently be considered as two distinct strata: the troposphere and the ionosphere. The ionosphere, the upper of the two layers, extends from approximately 50 km to 500 km. The troposphere is the region below this, extending from ground level to the base of the ionosphere. Most of the common climatic variations therefore occur in the troposphere. The treatment of the two strata differs fundamentally due to the nature of the effects they have.

At radio frequencies, unlike optical frequencies, the troposphere is a non-dispersive medium. Thus, the tropospheric delay cannot be estimated from two-frequency measurements as can the ionospheric delay. Instead, estimation of the delay relies on the use of one of a number of models.

The tropospheric excess delay Dt is given by

$$D_t = 10^{-6}/c \int N(s)ds$$

where N is the atmospheric refractivity given by $(n-1)10^6$ with n the index of refraction of air and c the velocity of light in vacuum [3]. The integral is evaluated along the signal path. According to Smith and Weintraub [10], the atmospheric refractivity is given by

$$N = 77,6 \times P/T + 3,73 \times 10^5 \times e/T^2$$

where P is the total atmospheric pressure, e is the partial water vapour pressure, both in hPa, and T is the absolute temperature in kelvins. This is not the only formula available [11,12] although it has been shown to be accurate to 0.1% [12].

N is commonly expressed as the sum of two effects:

$77,6 \times P/T$ - the "dry" component,

and $3,73 \times 10^5 \times e/T^2$ - the "wet" component.

The table below gives typical tropospheric delays at radio frequencies.

Elevation Angle (°)	Dry Component (ns)	Wet Component (ns)
90	7,7	0,7
30	15,3	1,3
10	43,4	4,0
5	86,7	7,7

In general, the dry component accounts for about 90% of the total effect, but it has been shown [13] that the integral of the dry component through the troposphere, in the direction of the local zenith, can be accurately modelled from surface measurements of pressure alone. The contribution of the wet component, although small compared to that of the dry component, is more difficult to model accurately, since measurements at the antenna site are generally not representative of conditions along the line of sight. Only one instrument, a microwave radiometer, is capable of providing measurements along the observational line of sight [14, 15]. Such measurements reduce error due to the wet component of the tropospheric delay to around 30 ps. But this is an expensive technique. The usual approach for estimating the tropospheric effect is therefore to apply one of a number of models.

3. MODELS OF THE TROPOSPHERE

A number of tropospheric models exist, mainly because of the difficulty in modelling the water vapour content. Usually the integrals are evaluated in the zenith direction. The zenith corrections are then 'mapped' down to lower elevation angles using mapping functions. Models are whether semi-empirical based on surface measurements of the local temperature, atmospheric pressure and relative humidity or empirical based on general reference atmosphere only requiring the station height and the elevation angle to the satellite.

Of the semi-empirical models, some of the best known have been developed by Hopfield [13] and Saastamoinen [16] and are widely used within the geodetic community. In this paper we use as reference a model developed by the Jet Propulsion Laboratory (JPL) for its deep space missions [17,18]. Evaluated against balloon measurements, this model was found to be able to predict the zenith tropospheric delays with an accuracy at the subnanosecond level if the daily surface measurement (pressure P, temperature T and relative humidity H) was made near local noon.

The tropospheric corrections D_t currently used by the timing community are computed according to general empirical models which take into account only the station height h, referenced to mean sea level, and the elevation El of the satellite. They neglect the contribution due to the wet component which may be as much as 15% of the total correction. Consequently, the resulting errors of these simple models may exceed 3 ns in a one-way range delay at 20° elevation angle.

The three models usually implemented are:

* the NBS model, used in NBS type receivers [19], with

$$D_t = 7{,}255x(1+X+X^2)/\sin El$$

where $X = -1.4x10^{-4}h$, h is expressed in m and D_t in ns (the NBS model is divergent for a nul elevation angle).

* the STI model, used in receivers manufactured by Stanford Telecommunication Inc. [20], with

$$D_t = 8{,}081xEXP(Y)/(\sin El+0{,}026)$$

where $Y = -1{,}335x10^{-4}h$, h is expressed in m and D_t in ns.

* the STANAG model proposed in the STANAG document [21], used in the receivers manufactured by SERCEL (France), with

$$D_t = 1/(\sin El + 0{,}00143/(\tan El + 0{,}0455))x\ R(h)$$

for $El<90°$

where h is expressed in m, D_t in ns and R is written in several equations depending on the station altitude, and is not reported here.

The STANAG model was recommended by the Group on GPS Time Transfer Standards [7] as the standard model for international GPS common-view time transfer.

We have made some comparisons of the models described. First, we compared two GPS time receivers of the NBS and the SERCEL type, located at the BIPM, for a period of 17 days. Both receivers were connected to the same clock and were programmed with 32 tracks of an international GPS common-view schedule. We computed daily averages [22] of the differences between two receivers for two cases:
- the tropospheric models employed were the original ones, the NBS model for the NBS receiver and the STANAG model for the SERCEL receiver,
- the STANAG model for the SERCEL receiver was replaced by the NBS model, so two receivers used the same tropospheric model.

The results are reported in Figure 1. A constant bias of about 1 ns between two computations can be observed. This demonstration stresses the need for common tropospheric model.

Next, we computed the tropospheric delays given by the three usual models for an altitude h equal to 125 m and for elevation angles increasing from 10° to 90° by step of 1°. We then compared these values to the tropospheric corrections given by the JPL semi-empirical model for the same elevation angles and for two different situations:

* summer afternoon in a temperate region
  (P = 1000 hPa, T = 25 °C, H = 40 %) in Figure 2
* tropical conditions
  (P = 1000 hPa, T = 35 °C, H = 100 %) in Figure 3.

In the first case (Figure 2), the four models agree within 1,5 ns for elevation angles greater than 40°. At lower elevation angles, the STANAG model is in good agreement with the JPL model and the NBS and STI models show discrepancies of several nanoseconds. In the second case (Figure 3), the three empirical models are consistent within 1 ns at high elevation angles, but give smaller values than the JPL model. This is because they do not take into account the wet component, which is not negligible here. For low elevation angles, the discrepancy between the different estimations of the tropospheric delay is always of several nanoseconds and can even reach 13 ns.

This demonstration illustrates that the use of a simple empirical model for GPS time transfer between areas of the globe of very different climate can lead to noticeable biases in the results.

4. THE COMMON-VIEW EXPERIMENT

The comparisons of the tropospheric models described above are restricted to one site. One could expect that during a common-view time transfer an inaccuracy in the estimation of the tropospheric delay will be at least partially removed if the same model is used on both sites. However between different climatic zones this reduction can be somewhat limited. The ideal experiment to explore this issue would consist of the examination of the tropospheric delays for common-view time transfer between a site located in a cold and dry area and a site located in a hot and humid area. As no suitable meteorological data is available for time laboratories located in such areas we report here the results for two time laboratories, the BIPM and the OCA, separated by a distance of only 700 km. The climates of these two sites differ slightly, however, as the BIPM is located in northern France in the Paris area at a height of about 100 m, and the OCA is located in a mountainous region of southern France at a height of about 1300 m. We had only a few weeks of meteorological data from the OCA. Among these data we chose the two days, 22 and 23 April 1993, on which the meteorological conditions at these sites showed the greatest difference (Figures 4, 5 and 6).

The GPS time receivers of the same NBS type were operated on both sites, thus using the same model of the troposphere. The two receivers were connected to caesium clocks and programmed with the BIPM international schedule number 20. For the two days in question, 64 common views were observed. Figure 7 reports satellite elevation angles as observed at the BIPM (elevation angles at the OCA are only slightly different).

In the experiment, each track at both sites was computed using the NBS simple empirical model and the JPL semi-empirical model based on surface weather measurements. The data were then compared. The results are given in Figures 8 to 11. Differences between the two models ranging from 0.4 ns to 1.1 ns can be observed. The discrepancy between the two models is larger at the BIPM, which can be explained by the more humid climate.

Next, the common views were computed between the two sites using the NBS model and the JPL models. The differences between the two computations for each common view do not exceed a few hundreds of picoseconds and are reported in Figure 12. These results illustrate that the use of the same simple empirical model of the troposphere for GPS common-view time transfer within a region of similar climate is sufficient for 1 nanosecond accuracy.

5. CONCLUSIONS

1. The use of a standardized tropospheric model in GPS time receivers is essential for accurate time comparisons.

2. For GPS time links within a region of similar climate, the use of a simplified standard tropospheric model is sufficient for 1 nanosecond accuracy.

3. For intercontinental GPS time links, the use of a more sophisticated model based on surface measurements should be considered and further studied.

Acknowledgements

The experimental part of this work was done at the BIPM thanks to the loan from the USNO (Washington, DC, USA) of a commercial caesium clock. The staff of the Time Section of the BIPM is grateful to the USNO for its generosity.

References

[1] W. Lewandowski and C. Thomas, "GPS Time Transfer," Proc. IEEE, vol. 79, pp. 991-1000, July 1991.

[2] W. Lewandowski, G. Petit and C. Thomas, "Precision and Accuracy of GPS Time Transfer," IEEE Trans. on Instr. and Measur., vol. 42, no. 2, April 1993.

[3] A. H. Dodson and J. C. Hill, "The effects of atmospheric refraction on GPS measurements," in Proc. Seminar on the GPS, Nottingham, U. K., 12-14 April 1989.

[4] G. M. R. Winkler, "Path delay, its variations, and some implications for the field use of precise frequency standards," Proc. IEEE, vol. 60, pp. 522-529, 1972.

[5] G. M. Resch, "Atmospheric limitations to clock synchronization at microwave frequencies," Proc. 16th Annual PTTI Meeting, pp. 249-259, 1984.

[6] D. Kirchner, H. Bessler and S. Fassl, "Experience with two collocated C/A code GPS receivers of different type," in Proc. 3rd European Time and Freq. Forum, pp. 94-103, March 1989.

[7] The Group on GPS Time Transfer Standards, "Technical Directives for Standardization of GPS Time Receiver Software," Rapport BIPM-93/6,1993.

[8] W. Lewandowski, G. Petit and C. Thomas, "GPS Standardization for the needs of Time Transfer," in Proc. 6th European Time and Freq. Forum, pp. 243-248, 1992.

[9] D. Kirchner and C. Lentz, "Tropospheric Corrections to GPS Measurements Using Locally Measured Meteorological Parameters Compared with General Tropospheric Corrections," in Proc. 25th PTTI, 1993 (in press).

[10] S. K. Smith, and S. Weintraub, "The constants in the equation for atmospheric refractive index at radio frequencies," Proc. IRE, 41, pp. 1035-1037, 1956.

[11] International Association of Geodesy, Resolution 1 of the 13th general Assembly, Bull. Geod., no 70, 390, 1963.

[12] G. D. Thayer, "An improved equation for the radio refractive index of air," Radio Science, Vol. 9, pp. 803-807, 1974.

[13] H. S. Hopfield, "Tropospheric effect on electromagnetically measured range: Prediction from surface weather data," Radio Science, 6, pp. 357-367, 1971.

[14] G. Elgered and G. L. Lundqvist, "A three year study of radio wave propagation delays caused by tropospheric water vapor," IEEE Trans. Antennas Propag., 32, 390, 1984.

[15] R. H. Ware, C. Rocken, C. C. Goad and J. B. Snider, "Experimental verification of improved GPS measured baseline repeatability using water vapor radiometer corrections," IEEE Trans. Antennas Propag., 23, 467, 1985.

[16] J. Saastmoinen, "Contributions to the theory of atmospheric refraction, II," Bull. Geod., no 107, 13, 1973.

[17] G. A. Madrid, C. C. Chao, et al, "Tracking System Analytic Calibration Activities for Mariner Mars 1971 Mission," JPL Technical Report 32-1587, March 1, 1974.

[18] H. F. Fliegel (The Aerospace Corporation, El Segundo, CA, USA), personal communication, 1992.

[19] M. A. Weiss (NIST, Boulder, Co, USA), personal communication, 1989.

[20] D. Kirchner (TUG, Graz, Austria), personal communication, 1991.
[21] NATO Standardization Agreement (STANAG) 4294, Arinc Research Corporation, 2551 Riva Road, Annapolis, MD, 21401, USA, Publication 3659-01-01-4296, 1 August 1990.
[22] W. Lewandowski, "Determination of the differential time corrections between GPS time receivers located at the Observatoire de Paris, the Observatoire de la Cote d'Azure and the Technical University of Graz, Rapport BIPM-91/6, 1991.

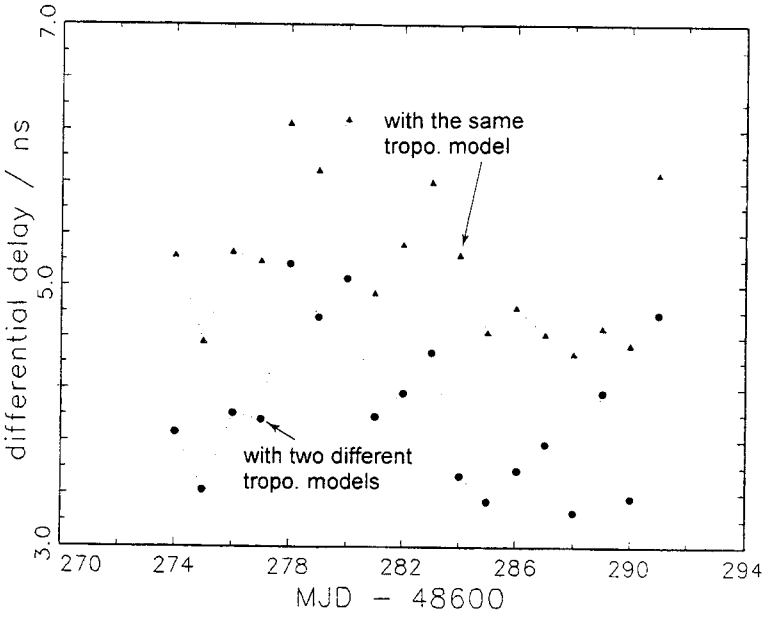


FIGURE 1. Daily means of the [BIPM Cs Clock - GPS time] by the NBS receiver minus [BIPM Cs Clock - GPS time] by the SERCEL receiver with the use of original tropospheric corrections and after unifying these corrections.

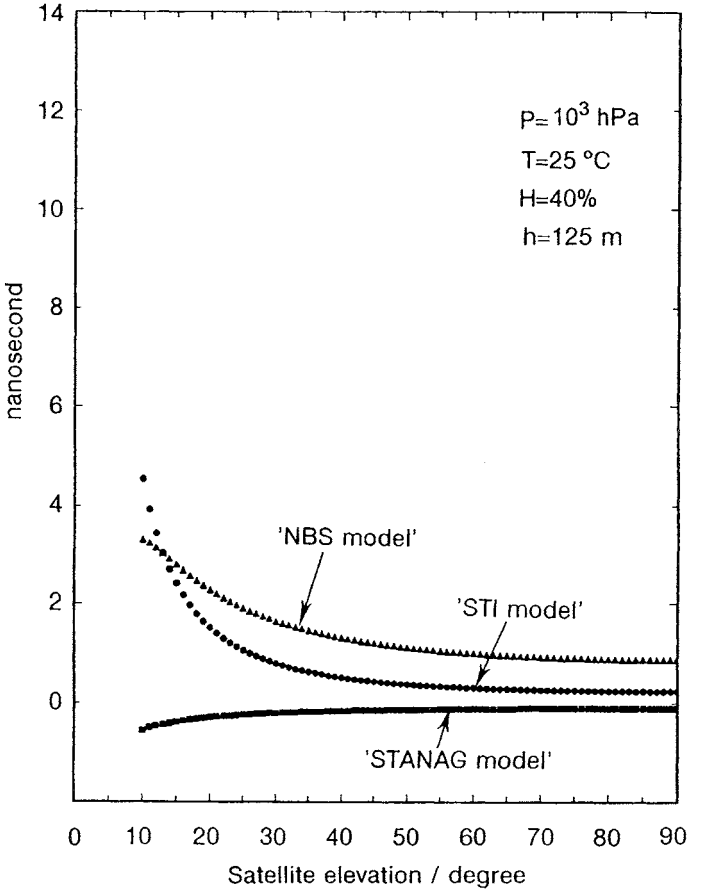


FIGURE 2. Comparison of tropospheric delays obtained from the JPL semi-empirical model based on weather measurements and from the usual models (NBS, STI & STANAG models) implemented in GPS time receivers, for all satellite elevations. Atmospheric conditions of a summer afternoon in temperate climate.

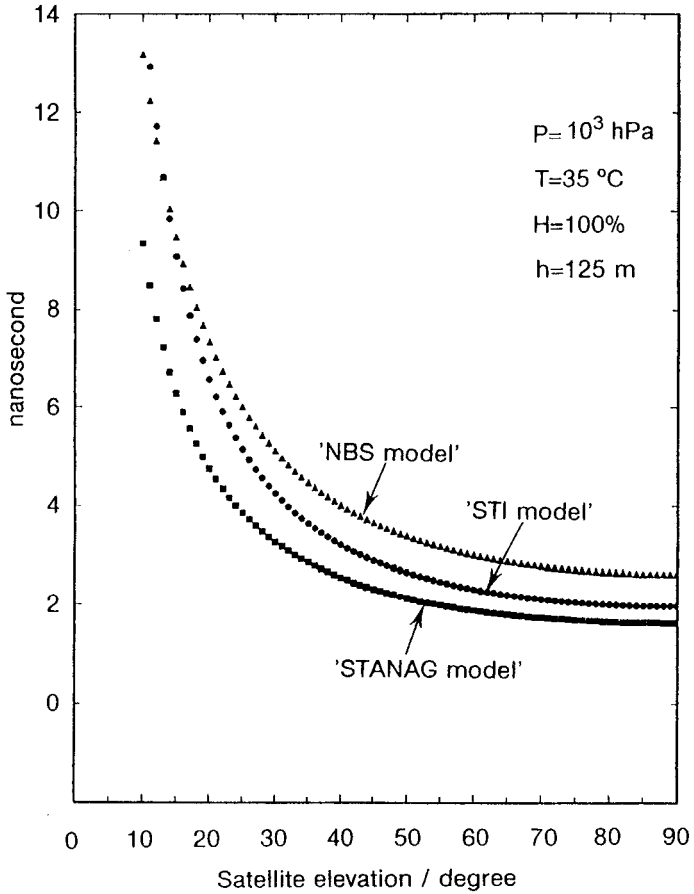


FIGURE 3. Comparison of tropospheric delays obtained from the JPL semi-empirical model based on weather measurementsp and from the usual models (NBS, STI & STANAG models) implemented in GPS time receivers, for all satellite elevations. Atmospheric conditions of a tropical climate.

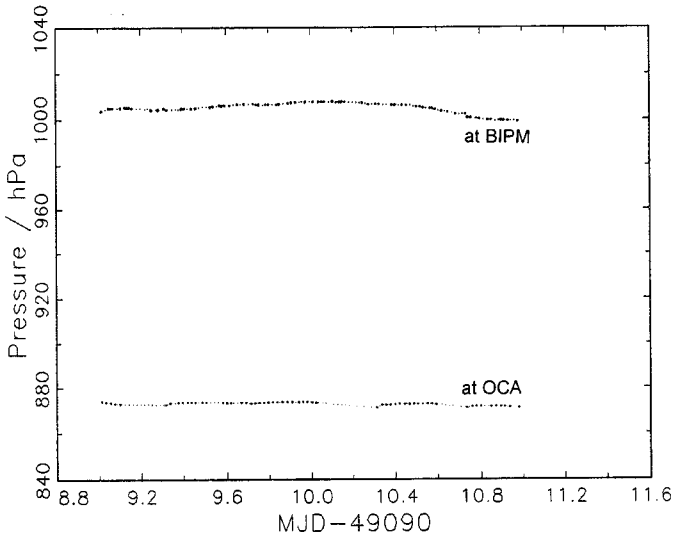


FIGURE 4. Atmospheric pressure on 22-23 April 1993 at the BIPM and the OCA.

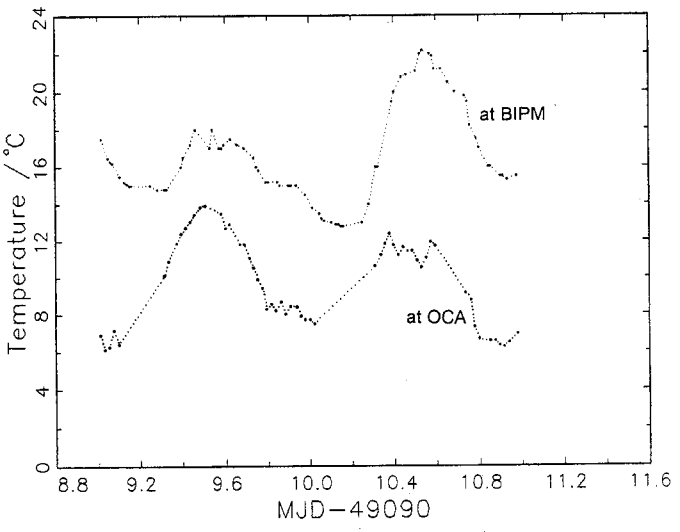


FIGURE 5. Temperature on 22-23 April 1993 at the BIPM and the OCA.

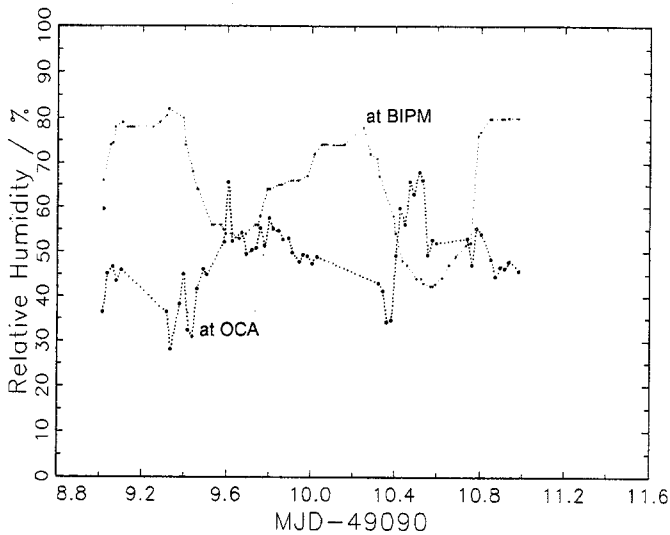


FIGURE 6. Relative humidity on 22-23 April 1993 at the BIPM and the OCA.

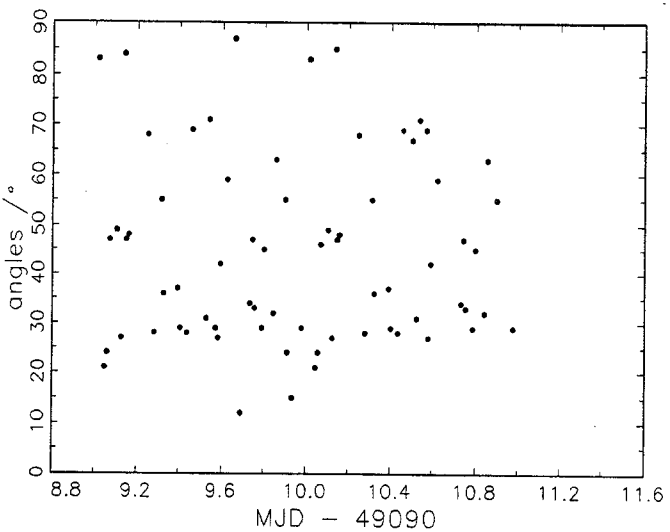


FIGURE 7. Elevation angles of each track on 22-23 April 1993 at the BIPM.

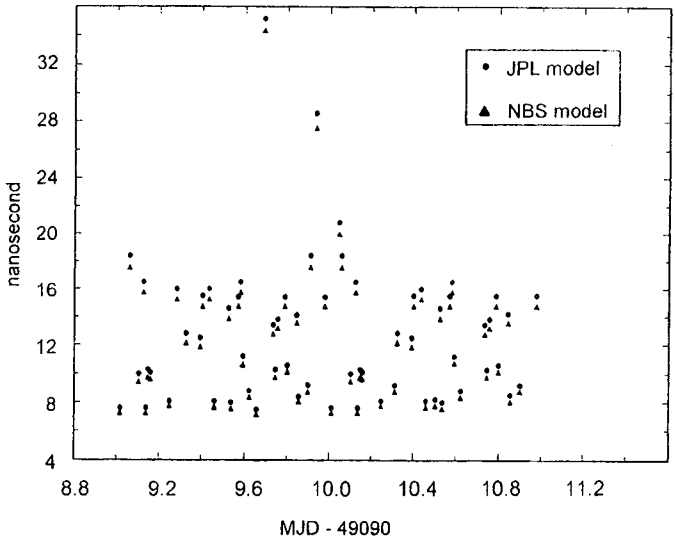


FIGURE 8. Tropospheric delays according to the JPL and the NBS models for each track on 22-23 April 1993 at the BIPM.

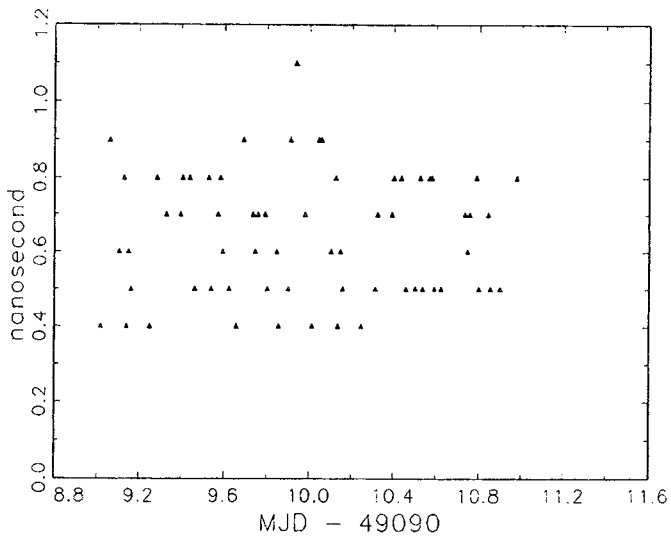


FIGURE 9. Differences between the JPL and the NBS tropospheric models for each track on 22-23 April 1993 at the BIPM.

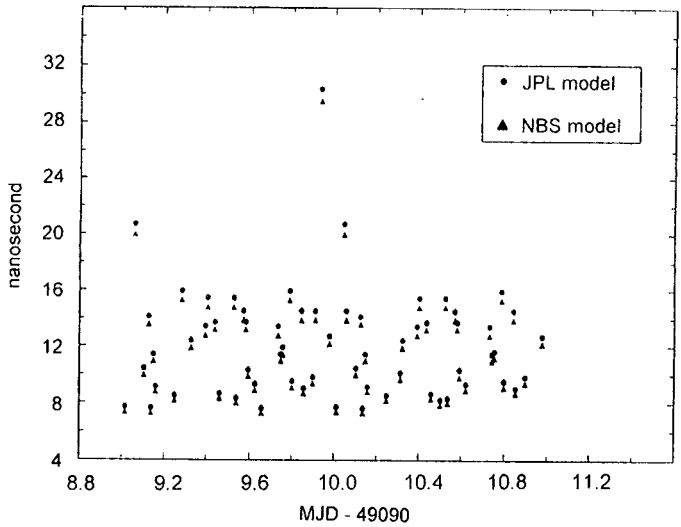


FIGURE 10. Tropospheric delays according to the JPL and the NBS models for each track on 22-23 April 1993 at the OCA.

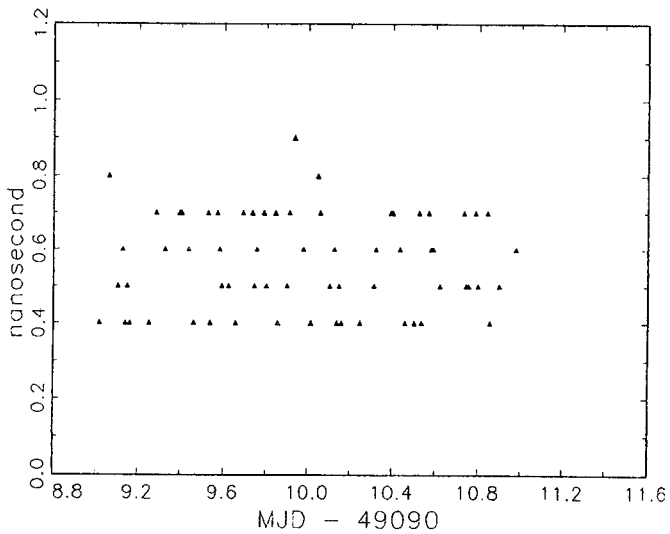


FIGURE 11. Differences between the JPL and the NBS tropospheric models for each track on 22-23 April 1993 at the OCA.

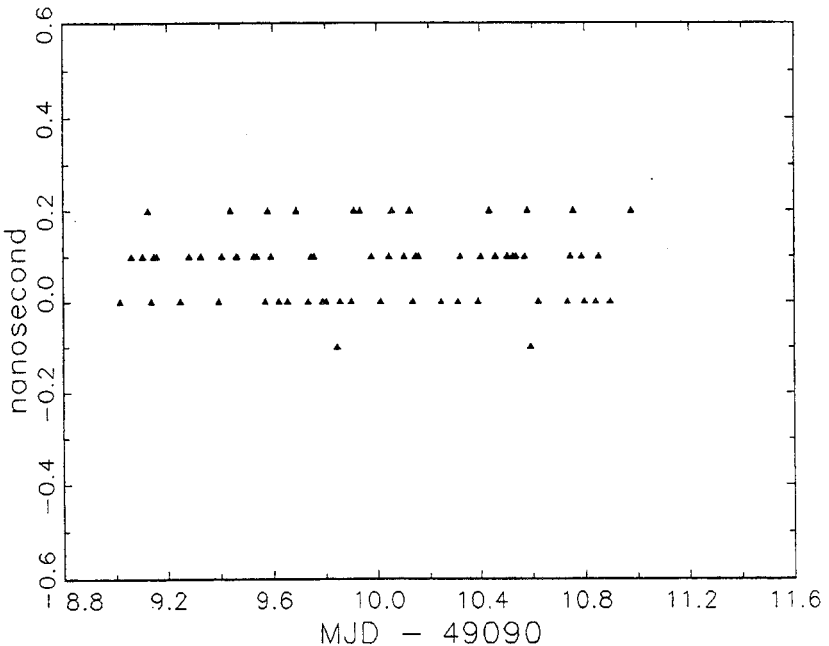


FIGURE 12. [BIPM Cs clock - OCA Cs clock] as obtained by GPS common-view with the NBS tropospheric model minus [BIPM Cs clock - OCA Cs clock] as obtained by GPS common-view with the JPL tropospheric model for each track on 22-23 April 1993.

TEMPERATURE VARIATION IN TEST CLOCK CABINETS WITH DIFFERENT FITTINGS.

Jari Mannermaa, Kalevi Kalliomäki
Dept. of Electrical Eng., University of Oulu
SF - 90570 Oulu
Finland

Abstract

Four test racks with different device configurations were studied. The racks were either open, totally closed, closed with natural or forced ventilation. The test racks each had two dissimilar sets of shelves for three devices.
The temperature distribution inside and outside the test racks was recorded. Air and surface temperatures were measured. The test racks were placed individually in a large blind basement room. The configuration of the heat supplies was varied.
The different test racks could be classified by the variation (SD) of the inside temperature. A point inside the racks representing "ambient temperature" of any device could also be determined. Time constants of the rack configurations were calculated.
When heat releasing equipment were housed inside cabinets, the best order was from bottom to top.

Introduction

Temperature is the most important environmental factor affecting the performance of precision devices. It is difficult to create a known and homogeneous temperature distribution in confined spaces like instrument cabinets. Electrical devices heat their surroundings creating varying conditions around them. If there are a few separate clocks in a cabinet, then the distribution is a complicated one. It depends essentially on the location of the clocks in the rack housing. In addition, the ventilation plan is important because the air should circulate in the cabinet.
Heat can be removed in three different ways from a sample into its surroundings, namely by conduction, convection and radiation. In this case, convection, especially a forced one, dominates and thus, other means can be neglected. An important quantity is the thickness of the boundary layer.
Four different cabinets were studied. The location of the devices was varied. Temperature distribution was studied in each case.

Methods

The object of the measurements was to determine the temperature distribution inside the test rack (30cm x 30cm x 90cm) which was on a table in the middle of a large, blind laboratory room. A computerized 16-channel measurement system was built to record the temperature distribution inside and outside the test rack. Air and surface temperatures were measured. The air flow was also monitored. The ambient temperature was known and stable. The configuration of the heat supplies (devices) was varied.
The temperature sensor configurations of the cabinets, the forms of shelves used and their location in the racks are presented in Figs. 1 a), b), c) and d).

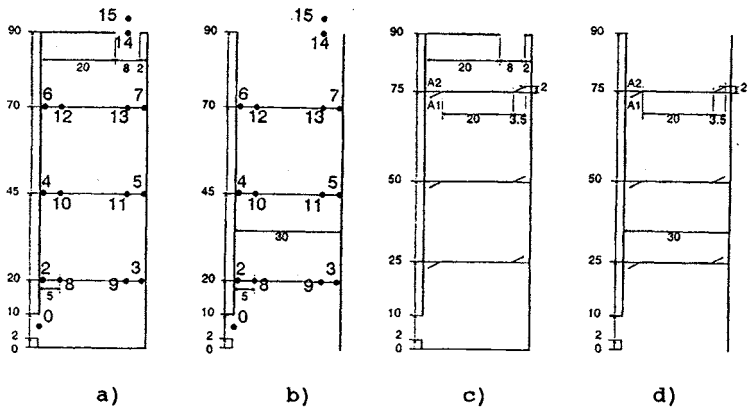


Fig. 1. a) & b) Temperature measuring points.
b) & d) The shapes (A1 and A2) of the shelves and their location.

Twelve sensors (2...13) were in the plane parallel to the front and rear panel and at the midpoint between them. Six sensors (2...7) were on the surfaces and six (8...13) 5 cm from the surfaces of the left and right panels, respectively. Sensor 1 was far away from the rack but also in the laboratory.
The test rack had two openings for ventilation. The input hole situated in the mid of left panel, 6 cm from the bottom and the output one in the top panel, 15 cm from the front and 6 cm from the right panel. The air flow was activated either by natural heat of the devices or it was speeded up by a fan to mix the air and homogenize the temperature distribution. The sensors 0 and 14 were at the midpoints of the input and output holes, respectively.

The studied cases were as follows:

Closed cabinet	Open shelves,	CC(OS)
Closed cabinet	Closed shelves,	CC(CS)
Closed cabinet	Open shelves & Fan	CC(OSF)
Open cabinet	Open shelves,	OC(OS)
Open cabinet	Closed shelves,	OC(CS)

Moreover, all these cases included the following types of heat source configurations: B (Bottom), M (Middle), T (Top), BM, BT, MT and BMT. The combination of letters gives the list of the active devices (heating) in each case. In addition, the letters R (right), L (left) and W (wall or panel) are also used to identify the location.
The heat conduction along the panels was studied. Then the sensors were located on the vertical line from top to bottom in the mid point between the front and rear panels on the surface of the right panel. The heat source was in the same line, 8 cm from top outside the rack. two sensor at this point measured the temperature inside and out side the cabinet. The other sensors lay at equal intervals of 12 cm downwards from the source on both the inside and outside surfaces.
The temperature distribution was studied in each case by measuring the temperature before and after the activation of the heating devices. The average values and standard deviations were calculated. The results were classified in different ways to determine essential facts e.g. temperature gradients in horizontal and vertical directions, suitable reference point for temperature, the relative velocities of the air flow in each case, etc.

Results

Typical measuring data is presented in Fig 2..

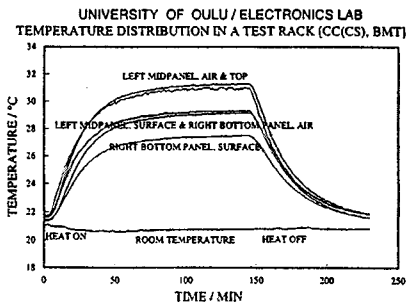


Fig 2. Typical temperature curves from the measurements.

The averages of the maximum and minimum temperature values (^{o}C), the standard deviation (SD) of the maximum values and the ratio SD/(Tmax-Tmin) are presented in the Table 1.

Table 1. Main characteristics of measured values.

Case	Comb	MAXave	SD	MINave	SD/dT %
CC(OS)	B	21.51	0.65	18.90	25
CC(OS)	B	22.53	0.69	19.76	25
CC(OS)	M	22.10	2.29	19.54	89
CC(OS)	T	20.55	0.83	19.19	61
CC(OS)	BM	22.32	0.63	19.16	20
CC(OS)	TM	22.36	2.39	18.86	68
CC(OS)	BMT	25.04	1.83	19.72	34
CC(CS)	B	23.34	1.57	19.69	43
CC(CS)	M	23.62	1.21	20.43	38
CC(CS)	T	23.16	1.78	20.62	70
CC(CS)	BM	26.94	1.10	20.53	17
CC(CS)	BT	26.05	1.09	20.52	20
CC(CS)	MT	26.24	2.43	20.52	42
CC(CS)	BMT	29.96	1.70	21.47	20
CC(OSF)	B	24.16	0.59	22.15	29
CC(OSF)	T	21.72	0.50	21.73	-5000
CC(OSF)	BM	23.22	0.31	21.65	20
CC(OSF)	BT	22.79	0.49	21.63	42
CC(OSF)	MT	22.44	0.65	21.37	61
CC(OSF)	BMT	23.54	0.71	21.33	32
OC(CS)	B	21.90	0.34	21.32	59
OC(CS)	M	22.04	0.56	21.36	82
OC(CS)	T	21.81	0.61	21.26	111
OC(CS)	BM	23.21	0.54	21.84	39
OC(CS)	BT	23.31	0.33	22.25	31
OC(CS)	MT	23.72	0.73	22.44	57
OC(CS)	BMT	24.10	0.57	22.15	29
OC(OS)	B	21.31	0.27	20.12	23
OC(OS)	M	20.87	0.40	19.3	25
OC(OS)	T	20.23	0.39	19.68	71
OC(OS)	BM	22.73	0.48	21.14	30
OC(OS)	BT	22.04	0.26	20.97	24
OC(OS)	MT	21.53	0.44	21.37	275
OC(OS)	BMT	23.32	0.49	21.54	28

Table 2. Absolute and relative variation (SD) of temperature with different rack configurations.

Case	SD (°C)	SD/dT %
OC(OS)	0.39	33
OC(CS)	0.53	58
CC(OPF)	0.54	37
CC(OS)	1.33	46
CC(CS)	1.55	36

Table 3. Absolute and relative variation (SD) of temperature with different heater element configurations.

Comb	SD (°C)	SD/dT %
BT	0.36	33
B	0.69	34
BM	0.69	24
T	0.82	78
BMT	1.06	29
M	1.12	59
MT	1.33	57

To determine the optimal measuring point for rack temperature control, the typical deviation of each channel from the average value in each of 34 cases was calculated:

Table 4. Typical deviation in degrees Celsius

Raw data (°C)

TLW	0.16	TLA	0.35	TRA	0.56	TRW	0.12
MLW	-0.19	MLA	0.58	MRA	0.27	MRW	-0.14
BLW	-0.89	BLA	-0.41	BRA	-0.68	BRW	-0.86

Weigted values (°C)

TLW	0.22	TLA	0.41	TRA	0.69	TRW	0.26
MLW	0.08	MLA	0.45	MRA	0.49	MRW	0.07
BLW	-0.81	BLA	-0.26	BRA	-0.83	BRW	-0.69

From Table 4 it can be seen that the deviation from the mean is smallest at the midlevel. In addition, wall values are more close to average temperature than the air values.

The standard deviations of the temperature gradients in

the horizontal and vertical directions are presented in the next table.

Table 5. The standard deviations of the horizontal and vertical gradients in the test rack. Unit is $^{o}C/m$.

Horizontal gradients			Vertical gradients			
			Left		Right	
Wall	Air		Wall	Air	Air	Wall
T 1.0	3.4					
		MT	4.1	6.0	5.7	3.4
M 1.0	8.3					
		BM	3.1	8.2	5.5	3.3
B 0.9	3.6					

Table 6. Absolute time constants (TCave) and relative variations (Ratio = TCsd/TCave) of TC:s in different cases.

Case	TCave /min	TCsd /min	Ratio
CC(OS)	37	10	0.26
CC(CS)	36	9	0.25
OC(CS)	12	3	0.44
CC(OSF)	11	3	0.19
OC(OS)	9	2	0.53

Table 7. Absolute time constants (TCave) and relative variations (Ratio = TCsd/TCave) of TC:s with different heater element configurations.

Heat	TCave /min	TCsd /min	Ratio
T	28	21	0.45
M	23	12	0.29
BM	23	15	0.21
MT	20	13	0.39
BMT	17	11	0.15
BT	17	12	0.26
B	16	10	0.58

Results of the heat conduction studies are presented in Figs 3 a) and b).

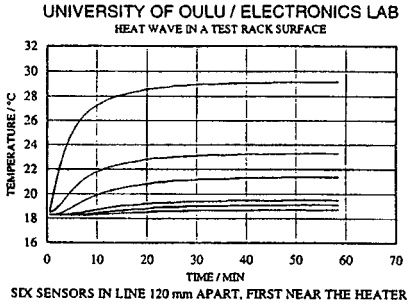


Fig 3. a)
Temperature curves of the heat wave studies.

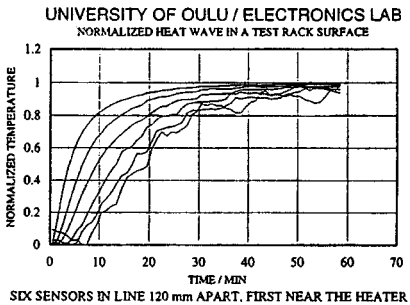


Fig 3. b)
Normalized temperature curves of the heat wave studies.

Conclusions

Rack configurations and heater element combinations used in the studies and some basic results are presented in Table 1. The case CC(OS) with B was repeated and shows the repeatability of the results.
The ranking of the tested cabinets and different heating element combinations on the basis of temperature stability are presented in Tables 2 and 3 as functions of the standard deviations (SD). The best ones the were OC -cases and CC with the fans both combinations of B (bottom) types. This means that the selection between open and closed racks and open and closed shelves depends

mainly on the stability of the room temperature. The results helped to determine the optimal rack configuration and instrument housing, at known and stable temperatures. When heat releasing equipment are housed inside cabinets, the position of the devices is important. These results indicated that, clearly, the best order was from bottom to top.
The optimum measurement point where the variation of temperature is expected to be smallest ("ambient temperature") was near the midrack. By regression fitting this point was located about 40 % from the bottom.
Even when the test devices were equal, the temperature distribution in the closed (CC(OS), CC(CS) and CC(OSF)) cabinet was asymmetric, both horizontally and vertically, despite the effective air mixing and symmetric placement of the heating units (Table 4). Moreover, the maximum standard deviation of the temperature gradient was about 8 ^{o}C/m in the closed rack. An intresting observation is that gradients are clearly smaller on the surfaces than in the air and the situation is better in the horizontal direction compared with the vertical one.
The time constant of the temperature changes varied strongly in the different cases. The OC and CC with fan cases, both the B -heater element combinations, responsed quickest to disturbances. A surprising observation was that heat was conducted downwards towards the bottom of the rack from the top. Time constants of the rack configurations were calculated. It was determined that changes in the power of the heating supplies had an effect of half-an-hour on the surroundings. The typical time constant was about 25 min (SD = 14 min).
The heat wave in the panel had the propagation speed of about 7 mm/s and the spatial constant of 100 mm. The values can be characterized as quick and non-spread if they are compared with the temporial changes in frequency standards and scales of normally used racks.
Especially studied with these measurements were the influence of temperature disturbance due to energy sources of the devices used and the malfunction of the room temperature stabilizing systems.These studies were yielded up some unexpected results which are not able to explain by known thermodynamic principles. Nowadays, the only way to approach these problems is to use chaos theory.

TIME TRANSFER OF CS-CLOCKS VIA A TELEPHONE LINE

K. Kalliomäki, T. Mansten

Technical Research Centre of Finland/Automation
Measurement Technology P.O. Box 1304
FIN-02044 VTT, FINLAND

Abstract

We have several commercial Cs-clocks at 20 km distance from KMNL (National Time and Frequency Laboratory at VTT) in Finland. The national time is traced from a GPS- and LORAN-C (Sylt) receivers. The link to three Cs-clocks is under construction using a telephone line.
The delay of transmission lines depends on frequency and temperature. Calcutions and measuring results are given. Especially standard twisted copper cables are problematic. A compromise between attenuation and delay variation is necessary. A novel delay measurement and compensation method, which does not need extra line is presented.

Introduction

Finland has no official and independent timebase of its own. KMNL has secondary time/frequency standards and time comparison receivers only. The Finnish Post Office has several Cs-clocks at 20 km distance from KMNL. Those clocks are primarily intended for digital telecommunicat ions network syncronization purposis. Time transfer from Cs-clocks to KMNL via TV-method, GPS-comparisons or via telephone line has been studied. The first two methods have been rejected do to reliability and serviceability, the access to Cs-clocks is subject to licence.
Therefore time transfer via telephone line has been selected as a primary method. It will be later supported by GPS comparison method.

Phase delay and attenuation in a telephone line

Starting from telegraph equation one can calculate phase and attenuation factor to a lossy transmission line. Line parameters mentioned in table 1 are picked from literature. The effect of temperature is assumed to be resistance variation in copper conductors, lengthening and other possible factors are ignored.
Carrier frequency is assumed to be between 1 and 100 kHz. Because only phase (zero crossing) is measured in practice, the carrier frequency cannot be high. 100 kHz means that higher than 10 us phase jumps are ambiguos.

	Frequency (kHz)			
	1	10	100	
Telephone cable				
Attenuation:	1.0	2.5	3.3	dB/km
Temp. coeff.	1.7	6	11	mdB/km/C
Phase factor:	18	7	5.3	us/km
Temp. coeff.	29	7	0.23	ns/km/C
Balanced airline				
Attenuation:	0.11	0.13	0.13	dB/km
Temp. coeff.	0.3	0.4	0.4	mdB/km/C
Phase factor:	4.1	3.6	3.6	us/km
Temp. coeff.	2600	48	(0.48)	ps/km/C

Table 1: Theoretical attenuation and phase factor of common telephone lines. Telephone cable (100 ohm, 0.7 mH, 40 nF per km), balanced airline (20 ohm, 2.5 mH, 5.1 nF per km)

The effect of frequency on phase factor and temperature coefficient is dramatic in a basic telephone line. The higher the frequency the better the values. The problems are rapidly increasing attenuation and telephone line standards, which do not allow for the use of high frequencies.

As a reference similar calculations are applied to low loss, balanced transmission line. Attenuation is much lower and phase factor approaches the free space value. What is best is the very low temperature coefficient of phase factor (resistance effect only). Kihara 1) has measured temperature coefficient of 80 ps/km,C for optical fiber ja Walls 2) 30 ps/km,C for a high quality coaxial cable.

Measuring results from a test line

We have used a standard telephone line (about 20 km) to syncronize clocks of Finnish Broadcasting Company. Both time signal (beeps) and TV-syncronizing frequencies are locked since 1975. Therefore is quite easy to monitor daily and annual delay variations. Telephone cables are buried into ground which decreases the effect of ambient temperature by an unknown factor, this factor is taken to be 1, however.

	Frequency (kHz)		
Telephone cable	1	10	
Attenuation:	1	3	dB/km
Temp. coeff.			mdB/km/C
Phase factor:	17.5		us/km
Temp. coeff.	30		ns/km/C

Table 1: Measured attenuation and phase factor of a 20 km long telephone line.

The measured attenuation and phase factor values (table 2) seem to be in fairly good agreement with theoretical ones (table 1). Therefore one can believe that the values above 10 kHz are quite true except the super low temperature coefficient for balanced line at 100 kHz.

Delay measurement in a single telephone line

Conventionally transmission delay time measurement requires two lines (round trip method). In our case it is possible to send identical signals in both directions in turn and measure phase differencies at both ends against stable atomic reference.

One can easily show, that by adding those two phase differencies together results twice the line delay and substraction results twice the desired time difference. The only problem is, that the far end (Cs-clocks) of the line has to be capable of sending the phase difference back to KMNL.

The electronics needed is descibed in figs. 1 and 2. We selected 10 kHz signal as a compromise between phase drift and attenuation.

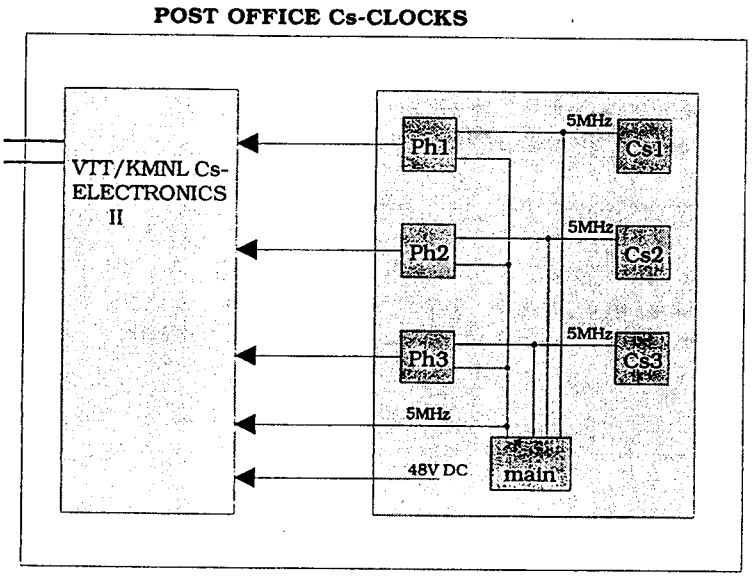


Fig 1. Electronics at the far end.

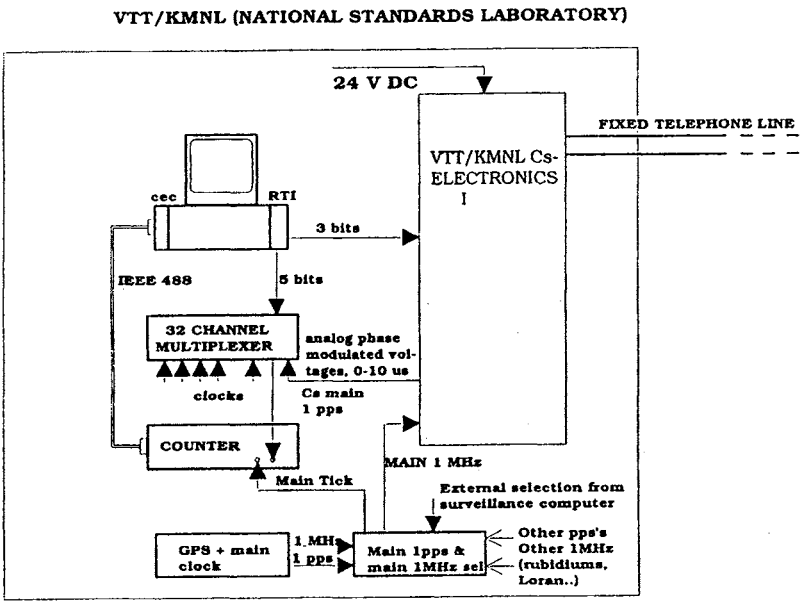


Fig. 2 Electronics at KMNL

References:

1) M. Kihara, A. Imaoka: Timing and Time Signal Distribution in Digital Communication Networks, Proc. EFTF 92, Noordwijk 1992.

2) F. Walls, L. Nelson, G. Waldez: Designing for Frequency and Time Metrology at the 10exp-18 Level, Proc. EFTF 92, Noordwijk 1992.

COMPARISON BETWEEN DIFFERENT DUAL-FREQUENCY GPS RECEIVERS MEASURING IONOSPHERIC DELAY FOR TIME AND FREQUENCY APPLICATIONS

C. Thomas, G. Petit and P. Moussay
Bureau International des Poids et Mesures
Pavillon de Breteuil, F-92312 Sèvres Cedex

Summary

The Global Positioning System is an outstanding tool for time transfer purposes. One important limitation to accuracy is the estimation of the signal delay in the crossing of the ionosphere. Since several years specific dual-frequency GPS receivers have been developed for measuring it. The Time Section of the BIPM regularly operates four different units of these devices, one decodes the P-code and three are codeless systems. This paper investigates a comparison study involving these four receivers together with another unit operating at the Paris Observatory. It appears that all units are sensitive to multipath effects and there is also some systematic variations of the measured delay with outside temperature. Differential internal delays are estimated between pairs of units. Application of the three-corner hat technique allows measurement of the intrinsic precision of ionospheric measurements over the usual 780 s track length. This precision is of order 2 ns.

Introduction

The Global Positioning System is now widely used for daily time comparisons between laboratories distributed worldwide. The observation of satellites by the common view method /1/ makes it possible to cancel, by difference, the error of the on-board clock. This procedure is of utmost interest in the presence of Selective Availability (SA), /2/.

The major errors affecting clock comparisons by the common view method come from satellite position, receiving antenna position, ionospheric and tropospheric delays, receiver differential delays and radio-signal multipaths. GPS receiver transportation and adapted means of reception greatly reduce some of them. Short distance comparisons (up to 1000 km) are weakly affected by satellite position error and by ionospheric effects. In addition, differential corrections have been determined for the antenna coordinates /3/, allowing time transfer with uncertainties of order 2 ns within a continent /4/.

However, the accuracy of long distance time links is degraded to several tens of nanoseconds. This is due to the use of broadcast satellite ephemerides, which are given with an uncertainty up to 10 m, and of estimated ionospheric delays derived from the ionospheric compensation model which provides a relative precision of

50%. Previous work has shown /5/ that this performance is improved at the level of 3 ns to 4 ns when rough data is corrected with post-processed precise ephemerides, produced for instance by the International Geodetic Service (IGS), and with measured ionospheric delays taken at both sites. The availability of measured ionospheric delays along the line of sight of GPS satellites is thus primordial for international clock comparisons.

Basically the ionosphere is a dispersive medium, so signal delay in crossing it can be measured by dual-frequency methods, in particular through the observation of GPS signals which are emitted on two carrier frequencies: L_1 (1575,42 Mhz) and L_2 (1227,6 Mhz). Several years ago, two different dual-frequency GPS receivers, specifically designed for ionospheric measurements and not able to demodulate the P-code, were developed. These are the GPS receiver GTR2, designed by Michito Imae /6/ and the NIST Ionospheric Measurement System (NIMS) /7/. Prototypes of these devices are now in operation in timing centres, as stand-alone units or associated with traditional single-frequency GPS receivers. Data from some of them are used in the computation of the International Atomic Time TAI.

While these codeless techniques were being developed, a new generation of GPS time receivers arose. These are derived from geodesic GPS receivers and operate in a multi-channel dual-frequency receiving mode. In such receivers demodulation of the P-code improves the precision of raw measurements by a factor 10 when compared to usual C/A-code GPS time receivers. An example of such a receiver is the Allen Osborne Associates model TTR-4P, one unit of which is at present under test at the BIPM.

In this paper we report a comparison of devices which measure ionospheric delays along the line of sight of GPS satellites. The aim is to identify systematic errors which can bias the results, arising from the receivers themselves or from the receiving conditions. Another important issue is the differential calibration of internal delays between pairs of receivers, necessary for improving the accuracy of clock comparisons.

To carry out this study, several receivers of different design were used to observe signals crossing the same part of ionosphere: data from five measuring systems is collected at the BIPM and at the OP, distant by 10 km, for several weeks in January and February 1994. Section 1 briefly describes the different receivers used. Sections 2 and 3 describe two different experiments: a comparison of 15 mn ionospheric data from two NIMS units in operation at the BIPM and at the OP, and a comparison of 780 s data, corresponding to usual common-view tracks, between four dual-frequency receivers of different design, in operation at the BIPM.

A comparison between two receivers, REC1 and REC2, can be made by taking the difference of ionospheric measurements obtained from the observation of the same satellite at the same time. In what follows, the sequence of points which is obtained is denoted [REC1 - REC2]. A statistical analysis of the noise affecting

these differences makes it possible to identify systematic effects and measure the noise intrinsic to each receiver.

1 Receiver descriptions

1.1 Codeless receivers: GTR2, NIMS, NITSUKI, NBSC

Several years ago, two different dual-frequency GPS receivers, specifically designed for ionospheric measurements, were developed. These do not demodulate the P-code:

* One is the dual-frequency GPS receiver, denoted GTR2, developed by Michito Imae, a Japanese engineer from the CRL (Tokyo, Japan), during his stay at the BIPM in 1987 and 1988 /6/. In this, measurements are based on the cross-correlation of the received P-codes carried by L_1 and L_2. It uses a directional antenna which points sequentially at a number of satellites according to a schedule specified in advance. Results are given every four minutes.
* The second is the NIST Ionospheric Measurement System (NIMS), developed by Marc Weiss and Dick Davis at the NIST (Boulder, Colorado, USA) in 1989 /7/. The measurement is based on the phase comparisons of the P-codes carried by L_1 and L_2 with a coherent signal generated locally. It observes all satellites in view, thanks to its omni-directional antenna, and results are given every 15 minutes.

The NIMS and GTR2 units are not used to measure time differences between a local clock and GPS time. No 1pps input is necessary and rough coordinates for the ground station are sufficient. But satellite almanacs obtained from conventional GPS time receivers must be provided regularly for these measuring systems.

In principle GTR2 directional antenna should minimize the spurious effects of multipath propagation of the GPS signals. For what concerns the NIMS omni-directional antenna, it is strongly recommended to protect it from multipaths by the installation of a choke-ring plane just below the antenna.

Both of these devices have also been coupled with traditional single-frequency GPS time receivers to provide conventional 780 s data following the International GPS Common-View Schedule and directly corrected for measured ionospheric delays:

* The GTR2 technique is now used in the commercial Nihon Tsushinki GPS time receiver, model 7633 RealTime Tecmeter, usually designed under the name NITSUKI (contraction from NIhon and TSUshinKI). Basic ionospheric measurements are taken every second to correct short-term timing data. A linear fit is then applied over 780 s and transferred to the middle-point of the common-view track.

* NIMS units can be connected to conventional NBS type GPS time receivers, their coupling in tandem being designated as NBSC ('C' for 'corrected'). Basic 15 mn ionospheric measurements are averaged or interpolated to provide an estimate at the middle-point of the common-view track.

The first GTR2 prototype is still in operation at the BIPM. It operates with a specific schedule optimized to cover the European part of the International GPS Common-View Schedule. Its 4 mn results are averaged or interpolated in order to obtain estimations of ionospheric delays valid for the usual 780 s common-view tracks.

A NIMS unit is also in operation at the BIPM, operated in tandem with a NBS time receiver. The 15 mn ionospheric data for all observable satellites is available, as well as corrected time data for the ensemble NBSC. Another NIMS unit is in operation at the OP as a stand-alone device and 15 mn ionospheric data is available. These two NIMS units are on loan from the NIST.

We also have at our disposal ionospheric data, corresponding to the usual 780 s common-view tracks, from a NITSUKI receiver, which was generously offered to the BIPM by Nihon Tsushinki.

In section 2 we compare the 15 mn ionospheric data from the two NIMS units, and in section 3 the 780 s ionospheric data from the GTR2, NBSC and NITSUKI units together with results obtained with the Allen-Osborne Associates TTR-4P unit in operation at the BIPM.

1.2 A decoding receiver: AOA TTR-4P

The Allen-Osborne-Associates TTR-4P is a GPS receiver adapted to high accuracy time measurements. It is derived from the largely digital Turbo-Rogue geodesic receiver. It decodes the P-code, its pseudo-range measurements are then 10 times more precise than those from conventional C/A-code receivers. It receives both of the GPS frequencies, so that it delivers timing data directly corrected for measured ionospheric delays. The calibration between the L_1 and L_2 signals inside the receiver is given by the manufacturer. In case of Anti-Spoofing (AS), it uses a special cross-correlation mode, so that results are still corrected for measured ionospheric delays.

The operation of a TTR-4P unit at the BIPM began in November 1992. To our knowledge, this is the sole unit operating in a civil timing laboratory. Since December 1993, it has operated with a 20 s update rate, although theoretically it could operate with a 1 s update rate. Short-term data are treated in real-time through a processing as close as possible to that in the Technical Directives for Standardization of GPS Time Receiver Software /8/. This makes it possible to recover timing data and to evaluate ionospheric measurements, both referred to the mid-date of usual 780 s common-view tracks.

The AOA TTR-4P has 8 physical channels which makes it possible to observe up to 8 satellites simultaneously. This performance is not used for the work described in this report but can serve other purposes /9/.

2 Comparison between two NIMS units

The two NIMS units installed at the OP and the BIPM operated continuously and simultaneously between 24 January and 1 February 1994. More than 2000 common points, corresponding to 15 min observations of all visible satellites, were recorded. This means that an average of 3 satellites were seen every quarter of an hour simultaneously by the two units.

Figure 1 shows the ionospheric delays measured by the NIMS(OP) and the NIMS(BIPM) for three different satellites observed from rise to set. Values measured for successive days show characteristic variations with the satellite elevation (see Figs 1.a and 1.c). Delays recorded by the NIMS(BIPM) are systematically larger than those from OP. However the shift [NIMS(BIPM) - NIMS(OP)] varies throughout the day, so that it cannot be completely explained by a lack of differential calibration between both devices. In addition the measurements from each unit are fully coherent from one day to another. This day-to-day repetition of the pattern is characteristic of multipath propagation /10/ which does not seem to be fully avoided even when using choke-ring planes on both sites. The shifts between values measured by the two NIMS thus present, in addition to a constant calibration delay, systematic variations largely explained by the effects of multipaths. The same phenomenon has already been observed between two NIMS units operating in collocation at NIST /11/.

Although the ionospheric measurements at our disposal correspond to an average of 3 satellites seen simultaneously, the 2000 differences between the two receivers can, for the purposes of statistical analysis, be treated as sequential. Two consecutive points are thus distant by an average time τ_0 of about 330 s. The sequence [NIMS(BIPM) - NIMS(OP)] is shown in Fig. 2.a and the corresponding Allan deviations, $\sigma_y(\tau)$, computed for averaging times τ between τ_0 and 1 day, are given in Fig. 2.b. Short-term oscillations corresponding to the preceding systematic variations may be observed: these increase the instability for averaging times of order 4 10^4 s and thus correspond to oscillations of period close to 1 day.

Globally the variation of $\sigma_y(\tau)$ with τ indicates the presence of phase noise. The level of white phase noise can be deduced from the value of $\sigma_y(\tau_0)$ and is estimated to 2,58 ns.

Phase noise can be greatly smoothed out by taking averages over 10^4 s (32 consecutive intervals of length τ_0). This makes it possible to obtain 62 different estimations of the differential delay. These different delay estimations present an average of 2,64 ns and a RMS of 0,88 ns. This RMS value is of the same order of magnitude that the level of residual white phase noise over 10^4 s, 0,5 ns

(2,58/ $\sqrt{32}$). This indicates that, beyond short-term instabilities due to multipaths, the sequence of 62 delay estimations presents long-term phase noise, which can be reduced to a level of 0,1 ns (0,88/ $\sqrt{62}$) by averaging.

For the period 24 January - 01 February 1994, the differential shift [NIMS(BIPM) - NIMS(OP)] is 2,6 ns and is characterized by an uncertainty of 0,1 ns (1σ).

3 Comparison of BIPM dual-frequency GPS receivers

In this section ionospheric delay estimations over 780 s intervals and corresponding to the middle-points of the tracks scheduled in the International GPS Common-view Schedule No 22 are used. The AOA TTR-4P being theoretically the best receiver, it serves as a reference for comparison between pairs of receivers. The AOA TTR-4P has activated its cross-correlation mode since 31 January 1994, the first date of continuous implementation of AS.

3.1 Comparison between pairs of receivers

3.1.1 *[TTR4P - NBSC]*

For the period 19 January-14 February 1994, a total of 698 common points between the AOA TTR-4P and the NBSC units is available. Figure 3.a, which represents the sequence of differences of ionospheric measurements [TTR-4P - NBSC], shows a sudden increase of the level of noise affecting the values and a shift of several nanoseconds of the differential delay between both receivers. It appears on the 12th day of comparison, 31 January 1994, and corresponds to the date of implementation of AS and of the cross-correlation mode of the TTR-4P unit. In addition, a monotonous variation of the differential delay takes place during the last days of comparison.

Allan deviations, corresponding to both of these periods, are shown in Fig. 3.b. The noise affecting the sequence of differences of ionospheric data is phase noise for the period 19 January-30 January 1994, the level of white phase noise being estimated to the value 1,99 ns. For the second period, 31 January-14 February 1994, Fig. 3.b reveals the presence of systematic effects superimposed to phase noise, the level of white phase noise amounting to 3,18 ns.

These systematic effects are partly due to multipath propagation and also to sensitivity with temperature of at least one of the receiver: in Fig. 4 the differential delay [NBSC - TTR-4P] follows the slow decrease of temperature which took place in the Paris area between 26 January and 14 February 1994. Surprisingly the rapid variation of the temperature experienced between 18 and 24 January was not felt: it is possible that the TTR-4P becomes much more sensitive to temperature when it operates in its cross-correlation mode.

For the period 19 January-30 January 1994, the differential delay [TTR-4P - NBSC] can be estimated by averaging individual shifts. It leads to the value -10,2 ns, characterized by an uncertainty of 0,1 ns (1σ).

For the period 31 January-14 February 1994, a procedure similar to that described in section 2 can be followed. The four last days corresponding to a large decrease of temperature are eliminated. The obtained differential delay is -7,3 ns characterized by an uncertainty of 0,3 ns (1σ).

3.1.2 *[TTR4P - NITSUKI]*

The NITSUKI unit did not operate till 31 January, so only 388 common points are available between the AOA TTR-4P and NITSUKI units. The corresponding sequence of differences of ionospheric data [TTR-4P - NITSUKI] is shown in Fig. 5.a: a long-term trend of the differential delay appears to be correlated with the decrease of temperature which occurred during the first two weeks of February (see Fig. 4). This phenomenon is similar to that observed on [TTR-4P - NBSC]. Earlier, timing data indicated a clear sensitivity to temperature in our NITSUKI unit. It is thus probable that the TTR-4P and the NITSUKI receivers both contribute to the variation with temperature which is now observed on ionospheric measurements. Values of daily delays [NBSC - NITSUKI] are also reported in Fig. 4, these seem to be less affected by temperature changes.

The values of the Allan deviation in Fig. 5.b show an increased instability for τ of order one day. This is probably partly due to changes of outside temperature. Following the procedure described earlier, the differential delay [TTR-4P - NITSUKI] can be estimated to -3,5 ns with an uncertainty of 0,1 ns (1σ).

3.1.3 *[TTR4P - GTR2]*

A similar study was made for the TTR-4P and GTR2 receivers, during the longer period 2 January to 14 February 1994. This includes 1300 common points and leads to Figs 6.a and 6.b. Again a sudden change of the differential delay [TTR-4P - GTR2] may be observed at the date of implementation of AS. The observed noise, which comes mainly from the GTR2 unit, keeps a nearly constant level.

The Allan deviation values of Fig. 6.b indicate the presence of phase noise for averaging times as long as 2 days. The level of white phase noise is estimated to the value 2,98 ns. The differential delay [TTR-4P - GTR2] can be computed by simple average over the month of January 1994. It amounts to -7,6 ns and is characterized by an uncertainty of 0,1 ns (1σ).

3.2 Study of stability

Suppose a triplet of receivers (REC1, REC2, REC3) operates continuously on the same period. They are compared by pairs, [REC1 - REC2], [REC1 - REC3], [REC2 -

REC3], which makes it possible to characterize the stability of each receiver relative to the other two for averaging times from τ_0 to $n\tau_0$ (n integer), τ_0 and n being identical for each pair. If the three receivers are independent, the three-corner hat technique allows calculation of the absolute stability of each receiver for averaging times from τ_0 to $n\tau_0$.

Different and independent systems able to measure ionospheric delays during the usual 780 s common-view tracks of the International Schedule operated continuously and simultaneously on a common site during the period 31 January-14 February 1994. This provides an opportunity to apply the three-corner hat technique to the results of comparisons between pairs of these receivers and so to characterize the stability in time and the intrinsic noise of a single measurement for each receiver.

3.2.1 *Stability in time*

The four different receivers have a common period of continuous operation from 31 January to 14 February 1994. They form four triplets:

* (TTR-4P, NBSC, GTR2),
* (TTR-4P, NBSC, NITSUKI),
* (GTR2, NBSC, NITSUKI),
* (TTR-4P, GTR2, NITSUKI).

Allan deviations are computed for each pair of each triplet. The number of common points is different for each pair, so that the averaged interval between two common points depends upon the pair which is considered. It varies from 2612 s to 7243 s. It is necessary to work with a common basic interval, chosen here to be τ_0 = 5000 s. We have thus interpolated the Allan deviation values in order to estimate the relative stability of each receiver for five different averaging times: τ_0, $2\tau_0$, $4\tau_0$, $8\tau_0$ and $16\tau_0$.

The three-corner hat technique is applied on each triplets for the five different averaging times. This results in the curves reported in Fig. 8. The artificial choice of τ_0 = 5000 s does not make it possible to obtain absolute values for individual stabilities, because basic measurements are not taken over intervals of 5000 s but over intervals of 780 s. During a 5000 s period one particular receiver has produced, on average, one unique 780 s measurement, while others have recorded two or three 780 s measurements. The Y-axis unit in Fig. 7. is thus not specified and only stability variations in time can be compared.

It may be observed that the NBSC, NITSUKI, and GTR2 units present random noise, white phase noise and flicker phase noise, for all averaging times under study, as can be expected from a measuring instrument. On the opposite the TTR-4P receiver is affected by systematic effects, probably due to its sensitivity to outside temperature.

3.2.2 *Intrinsic noise of measurements over 780 s track lengths*

It is also possible to evaluate the intrinsic noise affecting the ionospheric measurements performed by each receiver over the basic interval of actual measurement, 780 s. For this study, we do not consider values taken by GTR2 because they are basically averaged over 240 s and not 780 s.

When the three-corner hat technique is applied to the triplet (TTR-4P, NBSC, NITSUKI), for the period 31 January-14 February 1994, and for τ_0 equal to 780 s, one obtains the absolute Allan deviation values and the corresponding levels of white noise affecting basic measurements taken by each device:

TTR-4P	$\sigma_y(\tau_0) = 4{,}6\ 10^{-12}$ and $\sigma_x = 2{,}1$ ns
NBSC	$\sigma_y(\tau_0) = 5{,}1\ 10^{-12}$ and $\sigma_x = 2{,}3$ ns
NITSUKI	$\sigma_y(\tau_0) = 3{,}8\ 10^{-12}$ and $\sigma_x = 1{,}7$ ns

These values are close to each other, indicating that none of the studied receivers is dramatically less noisy than the others.

Conclusions

A comparison of five collocated GPS receivers, each capable of measuring ionospheric delay along the line of sight of GPS satellites, has been carried out.

This shows that all units are more or less perturbed by multipath propagation of the signals, although antenna protections are developed. Some receivers also show sensitivity to the outside temperature.

Differential calibrations of internal delays have been performed for different pairs of receivers. It was necessary to choose periods when the outside temperature is stable. The obtained differential delays are of order several nanoseconds and can be determined with uncertainties ranging from 0,1 to 0,3 ns (1σ). A shift of the internal delay of the Allen-Osborne Associates TTR-4P unit appeared suddenly when it activated its specific cross-correlation mode on the first date of definitive implementation of Anti-Spoofing (31 January 1994). This shift amounts to 3 ns.

The application of the three-corner hat technique shows that three dual-frequency GPS receivers, of different design and operating independently, present a comparable level, about 2 ns, of intrinsic noise for ionospheric measurements over the conventional 780 s track length.

Acknowledgements

The Time Section of the BIPM is grateful to timing laboratories, the USNO (Washington D.C., USA) and the NIST (Boulder, Colorado, USA), and to private

companies, Hewlett-Packard (Santa Clara, California, USA) and Nihon Tsushinki (Tokyo, Japan), for the loan of equipment without which the reported studies would not have been carried out.

The authors also thank Dr Pierre Uhrich from the LPTF (OP, Paris, France), Dr Michito Imae from CRL (Kashima Space Research Center, Japan), and Dr Marc A. Weiss from NIST (Boulder, Colorado, USA) for scientific support on the subject of ionospheric measurements.

References

/1/ Allan D.W. and Weiss M.A., Accurate Time and Frequency Transfer during Common-View of a GPS satellite, *Proc. 34th Annual Symposium on Frequency Control*, 1980, 334-346.

/2/ Allan D.W., Granveaud M., Klepczynski W.J. and Lewandowski W., GPS Time Transfer with Implementation of Selective Availability, *Proc. 22th PTTI*, 1990, 145-156.

/3/ Lewandowski W., A worldwide unification of GPS antenna coordinates for high accuracy time transfer, *Proc. Journées des systèmes de référence spatio-temporels*, Observatoire de Paris, 1992, 142-147.

/4/. Lewandowski W. and Thomas C., GPS Time Transfer, *IEEE Special Issue on Time*, 1991, **79**, 991-1000.

/5/ Lewandowski W., Petit G. and Thomas C., Precision and Accuracy of GPS Time Transfer, *IEEE Transactions on IM*, **42**, 2, 1993, 474-479.

/6/ Imae M., Lewandowski W., Thomas C. and Miki C., A dual frequency GPS receiver measuring ionospheric effects without code demodulation and its application to time comparisons, *Proc. 20th PTTI*, 1988, 77-85.

/7/ Davis D., Weiss M.A. and Vidmar M., A codeless ionospheric calibrator for time transfer applications, *Proc. 2nd International Technical Meeting ION GPS-89*, 1989.

/8/ The Group on GPS Time Transfer Standards, Technical Directives for Standardization of GPS Time Receiver Software, *Rapport BIPM-93/6*, 1993.

/9/ The use of the AOA TTR-4P GPS receiver in operation at the BIPM for real-time restitution of GPS time, *Proc. 25th PTTI*, 1993, in press.

/10/ Weiss M.A. and Thomas C., A study of data from two GPS ionospheric calibrators, *Annual Report of the BIPM Time Section*, **2**, 1990, part D.

/11/ Weiss M.A., personal communication, February 1994.

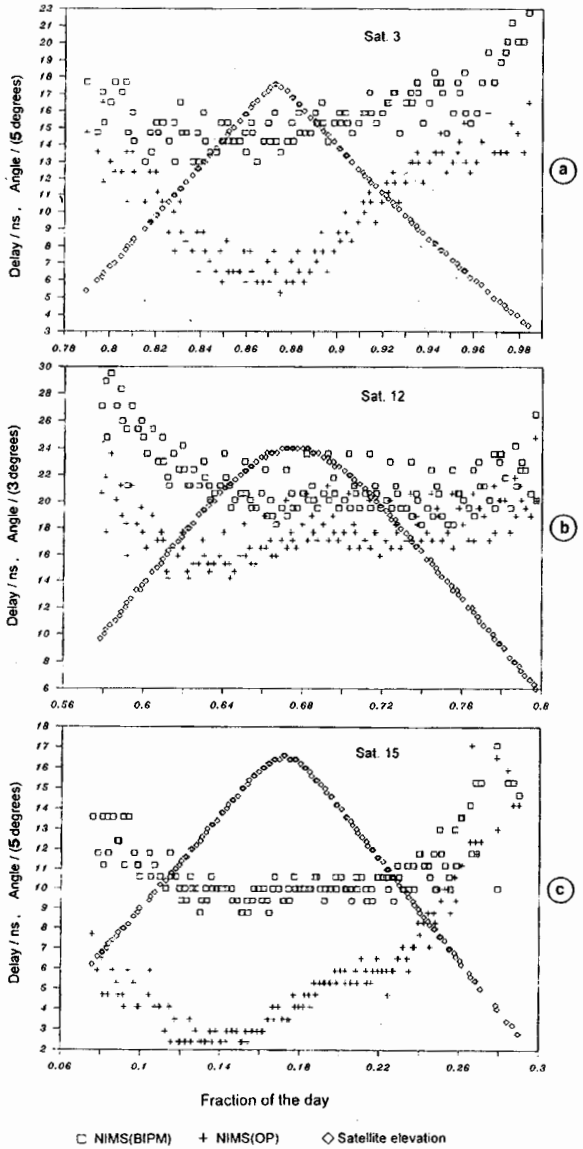


Fig. 1. Ionospheric delays measured by the NIMS units in operation at the OP and the BIPM for three satellites, together with the satellite elevations. Measurements, obtained for 7 successive days, are reported to the same date taking into account the sidereal repetition of the GPS constellation.

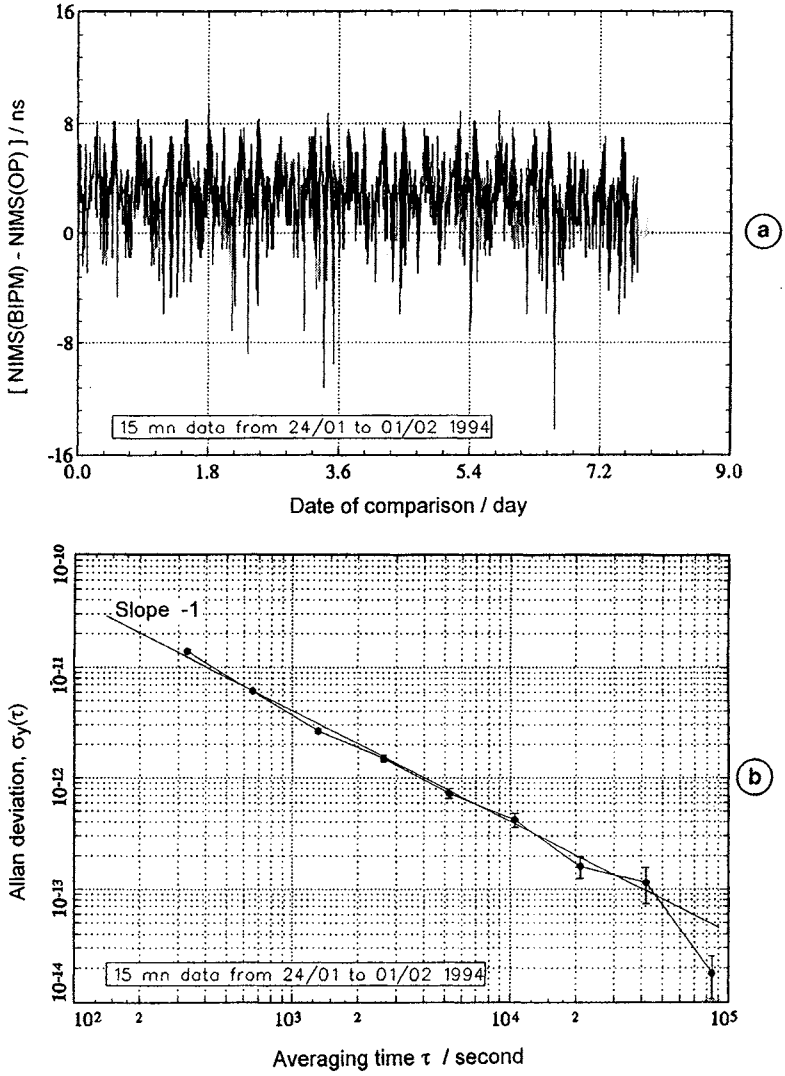


Fig. 2. Difference between 15 mn ionospheric measurements taken by two NIMS units located at the BIPM and at the OP for all satellites in view:

2.a Sequence of differences [NIMS(BIPM) - NIMS(OP)].

2.b Corresponding Allan deviation.

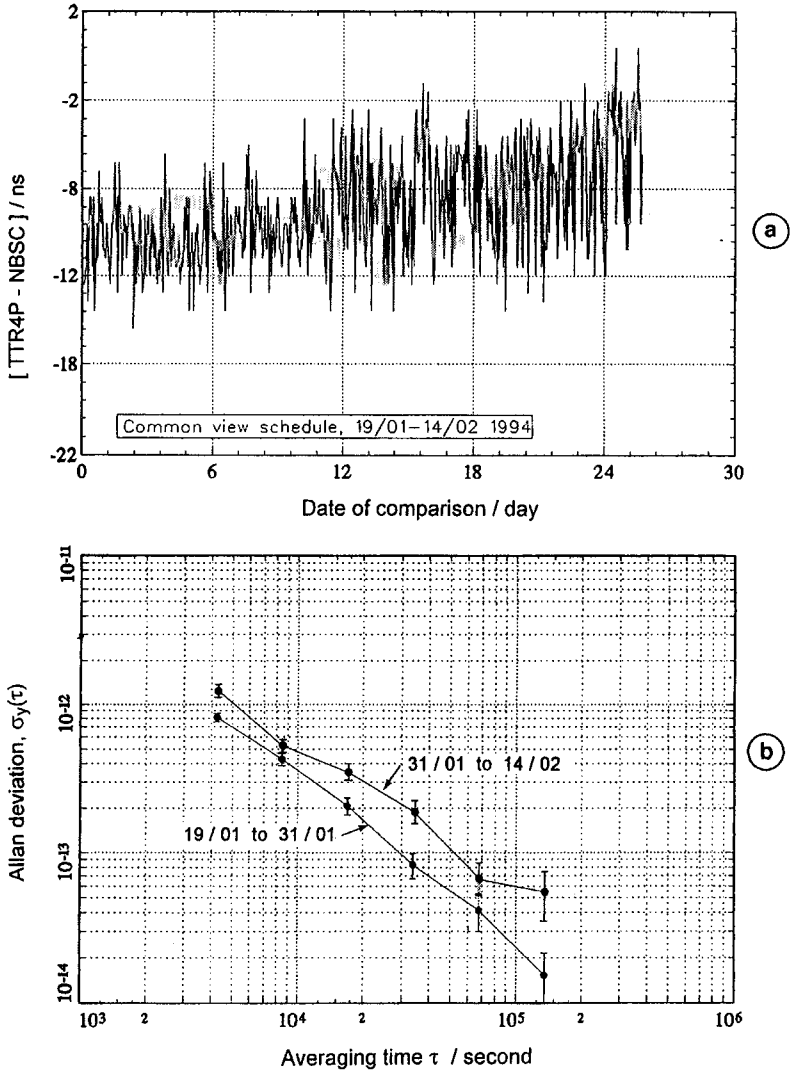


Fig. 3. Difference between 780 s ionospheric measurements used for correcting common-view timing data taken by the TTR-4P and NBSC units at the BIPM:

3.a Sequence of differences [TTR-4P - NBSC].

3.b Corresponding Allan deviation.

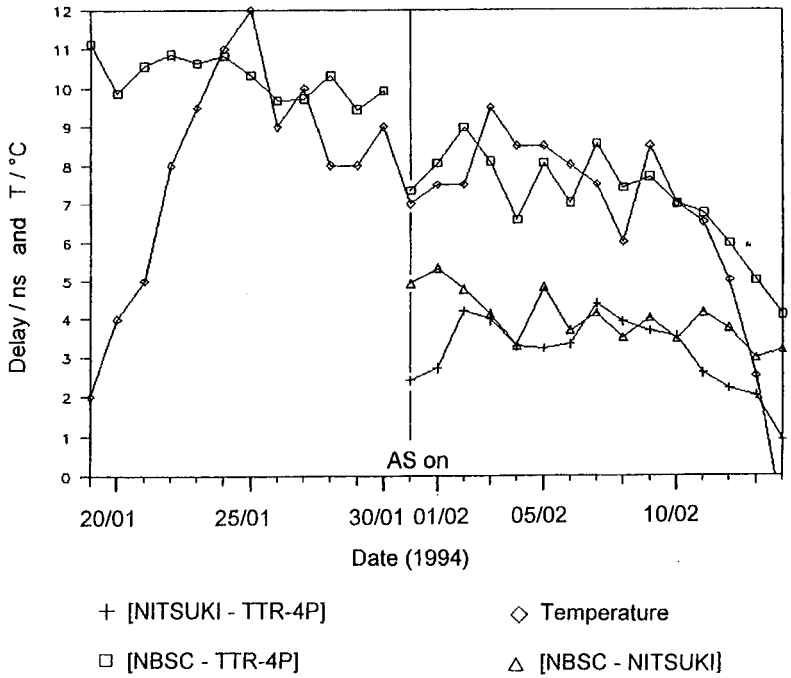


Fig. 4. Daily averages of differential delays between 3 pairs of systems measuring the ionosphere from the same site. The outside temperature is added in the graph.

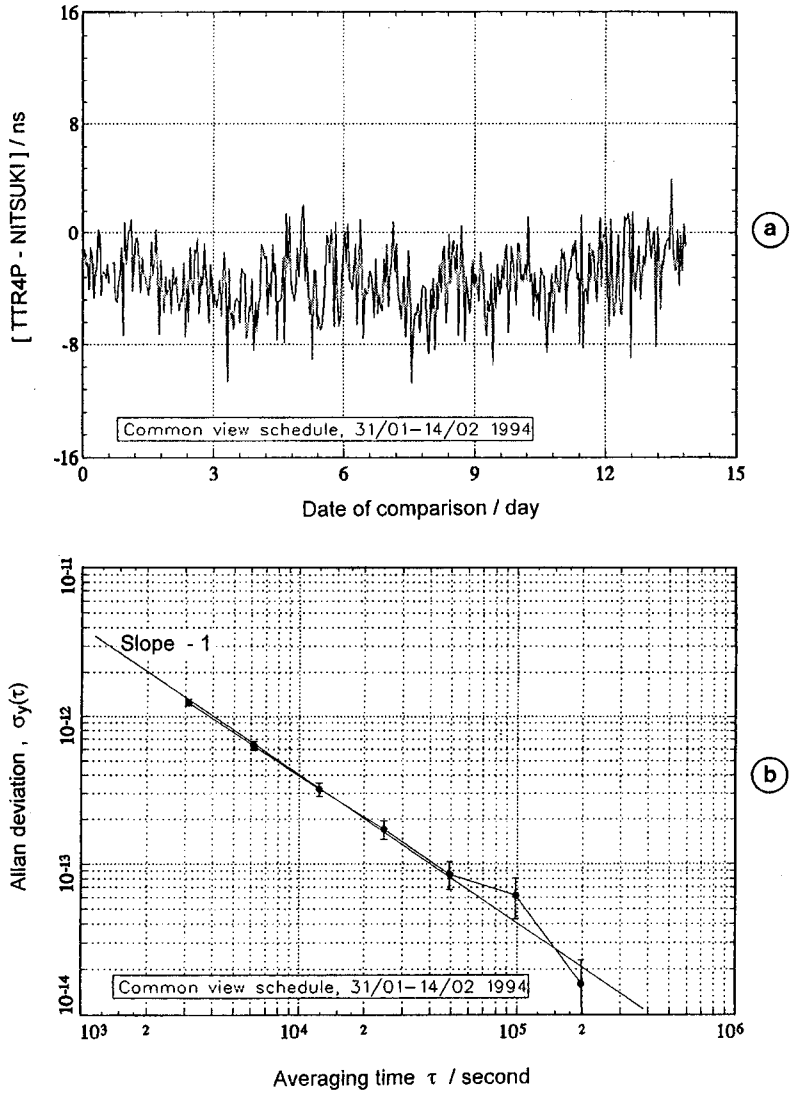


Fig. 5. Difference between 780 s ionospheric measurements used for correcting common-view timing data taken by the TTR-4P and NITSUKI units at the BIPM:

5.a Sequence of differences [TTR-4P - NITSUKI].

5.b Corresponding Allan deviation.

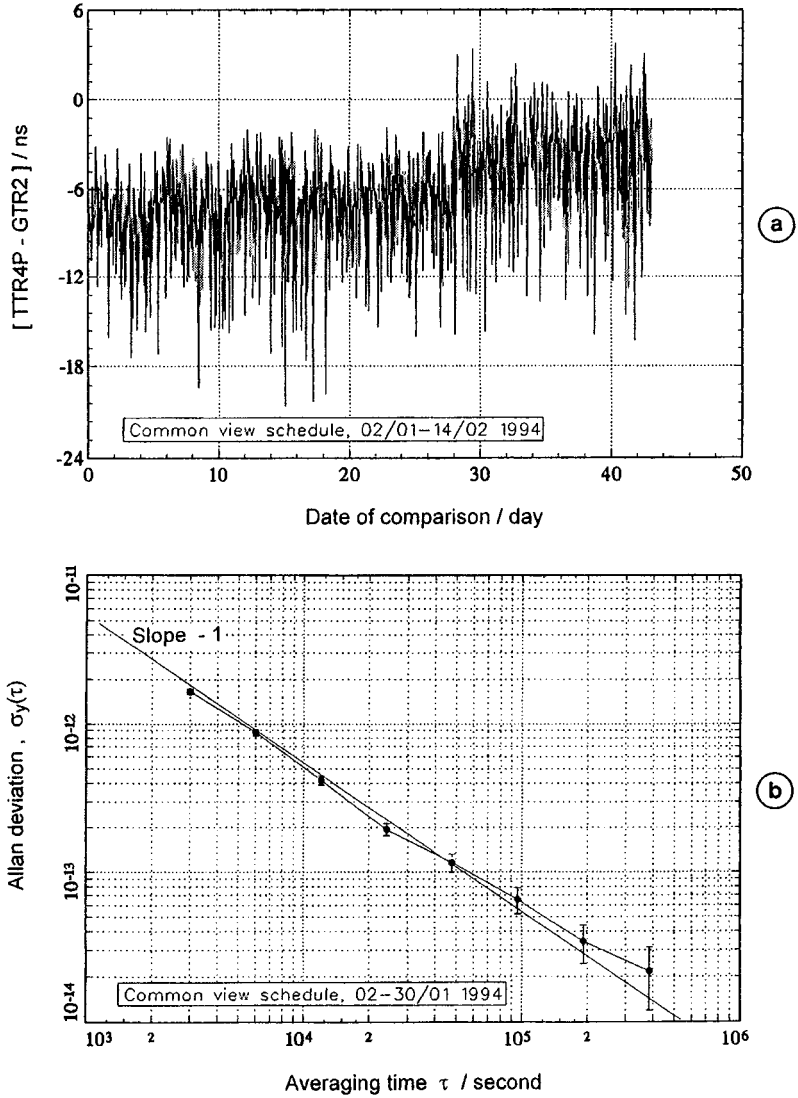


Fig. 6. Difference between 780 s ionospheric measurements used for correcting common-view timing data taken by the TTR-4P and GTR2 units at the BIPM:

6.a Sequence of differences [TTR-4P - GTR2].

6.b Corresponding Allan deviation.

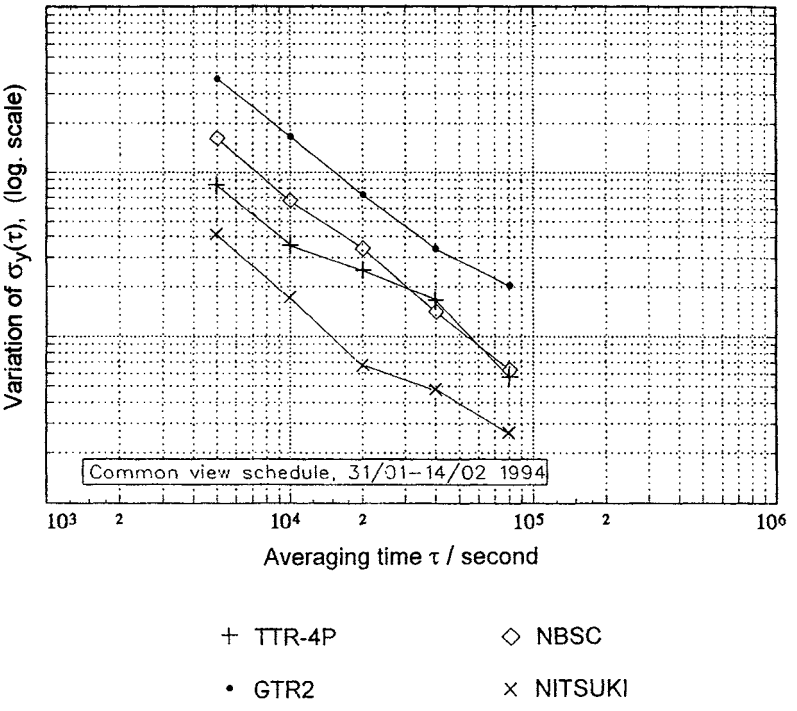


Fig. 7. Characterization of noise type of four ionospheric measurement systems.

Time and frequency traceability sources in the Italian Calibration Service

F. Cordara ,V. Pettiti (*), P. De Giorgi (**)

(*) Istituto Elettrotecnico Nazionale Galileo Ferraris
Torino, Italy.

(**) ESAT s.n.c.
Torino, Italy.

Summary

The Italian Calibration Service (SIT) was established in 1977 to perform secondary laboratories accreditation in the field of electrical, mechanical and thermal quantities and ionizing radiations.

In this frame, the IEN time and frequency laboratory has already accredited 13 calibration centres for frequency and performs the calibration of their reference oscillators to ensure the traceability to the national time and frequency standard.

The measurement systems at the user site are mostly based on three different synchronization techniques: passive television method, GPS satellites or standard time signals transmitted by the Italian Broadcasting Company (RAI) .

Two devices that use the GPS signals in the one-way mode or the RAI signals to correct the frequency of quartz or rubidium oscillators have been developed by ESAT and tested at the IEN as regards to their accuracy and to their short term behaviour.

The results obtained in an operative enviroment of these devices are presented.

1 - Introduction

The national calibration system (SNT) in Italy, was established on 11 August 1991 with the law n. 273 under the sponsorship of the Ministers of University and Research and of Industry and Trade, among others. According to this Act, the system consists of three primary Institutes and of calibration centres and it is under the responsability of these two Ministers. The main task of SNT is to ensure the traceability of the measurement results to the national standards maintained by the primary institutes: IEN for electrical quantities, IMGC for mechanical and thermal quantities and ENEA for ionizing radiations. The primary institutes main duties consist in studying and realizing the primary standards, in comparing these standards internationally and in disseminating them within the SNT context. The primary standards maintained in the three institutes become "national standards" by a decree of the Minister of Industry and Trade. The calibration centres agreed by the primary institutes and taking part to the SNT are secondary level calibration laboratories, selected according to their technical competence and organization, and equipped with secondary standards periodically compared to the national standards.

In practice, the calibration centres are accredited according to the Italian Calibration Service (SIT) rules that are based on the European standards on testing, certification and accreditation activities EN 45000. More than 60 calibration centres have

been accredited by SIT since its beginning in 1977, and among these 13 have been accredited by the IEN for calibrations in the field of frequency.

The IEN time and frequency laboratory, that generates the national time scale based on five cesium clocks, is responsible for the laboratories accreditation in this field. The traceability to the national time standard of the secondary standards maintained in these laboratories, can be obtained by means of three different synchronization systems. In the next paragraphs, the results obtained in this activity together with the characterization of a new device using the GPS signals to control a frequency standard are presented.

2 - Traceability by passive television method

The distribution in the Italian territory of the centres accredited for frequency is shown in Fig. 1, meanwhile Fig. 2 represents the classes of uncertainty recognized to the

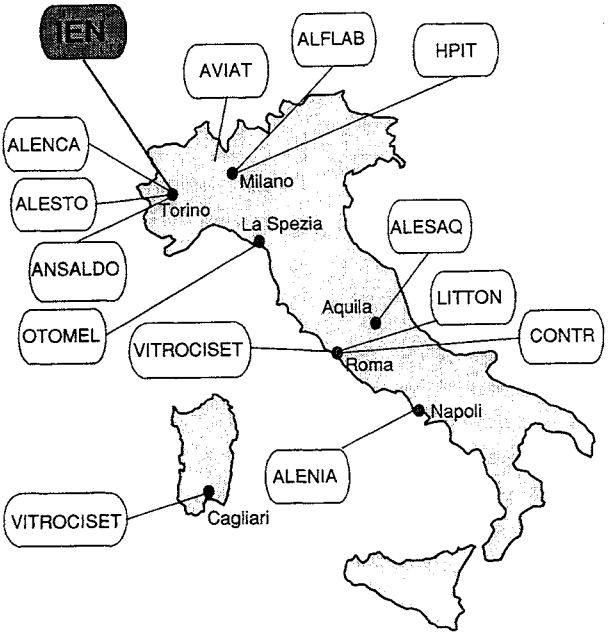


Fig. 1 - Calibration centres accredited for frequency in Italy.

centres.

The secondary standards of these centres are mostly referred to the IEN national time standard by means of daily time comparisons on the television synchronizing pulses extracted from RAI UNO transmissions. This method has been in use since the end of the sixties /1/, and allows to determine the relative frequency departure of distant oscillators at the $1 \cdot 10^{-13}$ level for observation times of one month.

The measurement system at the users side, performs automatically 60 time interval readings between its reference clock and the selected TV pulses and the data are processed monthly at IEN to supply the SIT centres with calibration certificates reporting the relative frequency departure and the frequency drift of their reference oscillators. The typical standard deviation of the daily differences over the 6 synchronization results exchanged is of 10 ns, in case of common-view, and 60 ns for the farthest centres.

Some example of the results obtained in 1993 with this method are shown in Figures 3, 4 and 5 relatively to two rubidium oscillators and one cesium standard. The mean value of the frequency departure is computed over at least 15 data points, after a filtering of

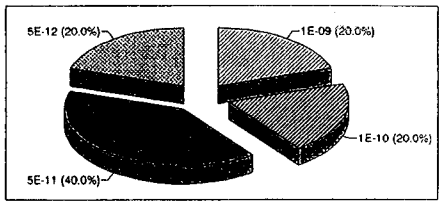


Fig. 2 - Distribution of the uncertainty levels.

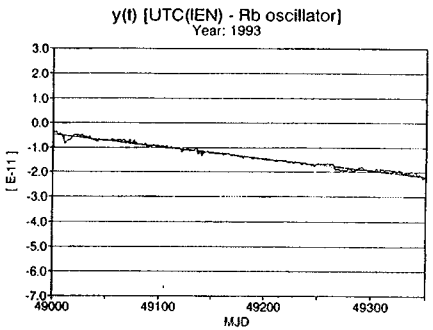


Fig. 3 - Accredited laboratory standard traceable via TV (same transmitter).

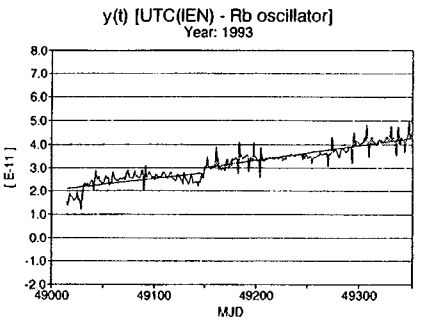


Fig. 4 - Accredited laboratory standard traceable via TV (long distance).

the synchronization results, and its uncertainty can range from $5 \cdot 10^{-12}$ to $1 \cdot 10^{-13}$.

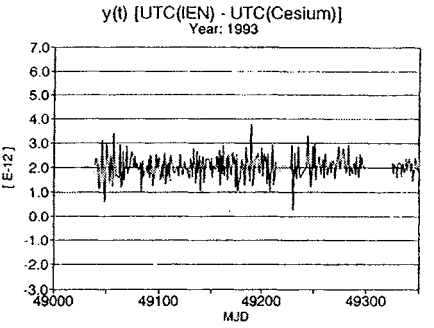


Fig. 5 - Accretdited laboratory standard traceable via TV (long distance).

3 - Traceability by means of coded time signals

Another possibility of traceability for local reference oscillators in Italy is offered by the coded time signals generated by IEN and broadcasted by the RAI FM and AM transmitters15-20 times per day in each program. These signals are used both for automatic synchronization of remote clocks, allowing precision better than 1 ms, and for disciplining the frequency of an oscillator with thermostatic control at the $1 \cdot 10^{-9}$ level /2/. Some of these devices, such as RAD 100 model, developed by ESAT and certified by IEN, are operating in the calibration laboratories of industries and in scientific institutions.

The synchronization results and the frequency corrections applied to the voltage controlled oscillator inside the instrument, are printed every day to check its performances, stored and sent monthly by disk or E-mail to IEN where they are compared with the data measured at the same reception times by a similar device supplied with UTC(IEN) frequency reference.

The RAD 100 clock is synchronized to UTC(IEN) as soon as a coded time signal is recognized; from that moment, the time interval measurements between the decoded time signals and the 1PPS derived from the local oscillator are accumulated according to the time constant T selected, and finally a linear fit of the synchronization data is computed and a statistical filter applied to eliminate the outliers. This process is iterative and extended over the last T days (where T can vary from 1 to 30), and the average time error of the clock and the mean frequency departure of the oscillator are computed every day, and consequently, a frequency and time correction are applied. The time constant, in the case of quartz oscillators with frequency drifts of the order of $1 \cdot 10^{-10}$/d, is usually selected in the range from 3 to 7 days.

In Fig. 6 have been reported the results of 100 days comparisons between UTC(IEN) and the coded time signals received from RAI-FM transmissions at 92,1 MHz in Torino, that represent the long term behaviour of the synchronization delay of the link

Torino-Roma-Torino (about 1500 Km). It can be seen that the maximum synchronization error is within 2 ms for the individual data points. The reason for the relevant variation of the transmission delay is explainable by the different locations and equipment of the RAI studios in Roma involved in the signals broadcasting during the day.

As a sample, the frequency behaviour of a quartz oscillator controlled by these time signals, used as reference in a calibration centre located at 20 kilometers from IEN and accredited with a $1 \cdot 10^{-9}$ uncertainty, is reported in Fig. 7. The graph has been obtained by subtracting the synchronization results stored in the RAD 100 from the same kind of data collected at IEN and referred to its time scale. Afterwards the mean frequency departure for T=1 day and T=5 days have been computed. It can be seen that the frequency drift is compensated and that the residual frequency departure is kept within $\pm 5 \cdot 10^{-10}$.

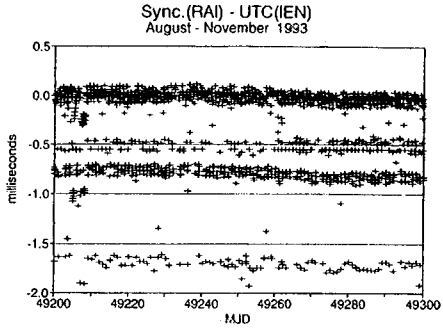


Fig. 6 - IEN coded time signals reception delays in Torino.

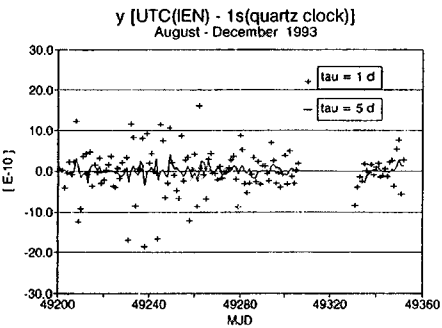


Fig. 7 - Accredited laboratory standard disciplined by IEN time signals.

4 - Oscillators controlled by GPS signals

In addition to the systems mentioned above, some calibration laboratories have recently chosen, as a frequency reference, a high quality oscillator controlled with the time information transmitted by the GPS satellites. These systems, using GPS in a one-way mode, allow a traceability to the UTC time scale of the U. S. Naval Observatory at a level variable from 10^{-10} to 10^{-12}, depending on the selected oscillator, the instrument features, and the integration times used /3/. The GPS system is used by the national

laboratories to relate to the international atomic time scale, using the common view technique and following a daily measurement schedule supplied by BIPM /4/. It is therefore possible to relate a GPS disciplined oscillator to the national time standard, via the GPS synchronization results, as requested in the calibration systems.

The GPS 100 Time and Frequency Reference is one of such devices that has been designed by ESAT, taking into account the problem of the traceability to a national standard. The block diagram of ESAT GPS 100 is reported in Fig. 8. It is a microprocessor controlled device made of a commercial GPS receiver card (Magellan OM-5000), a voltage controlled oscillator, a digital divider, a digital phase stepper and a time interval counter with a 20 ns resolution. The GPS signals received are used to

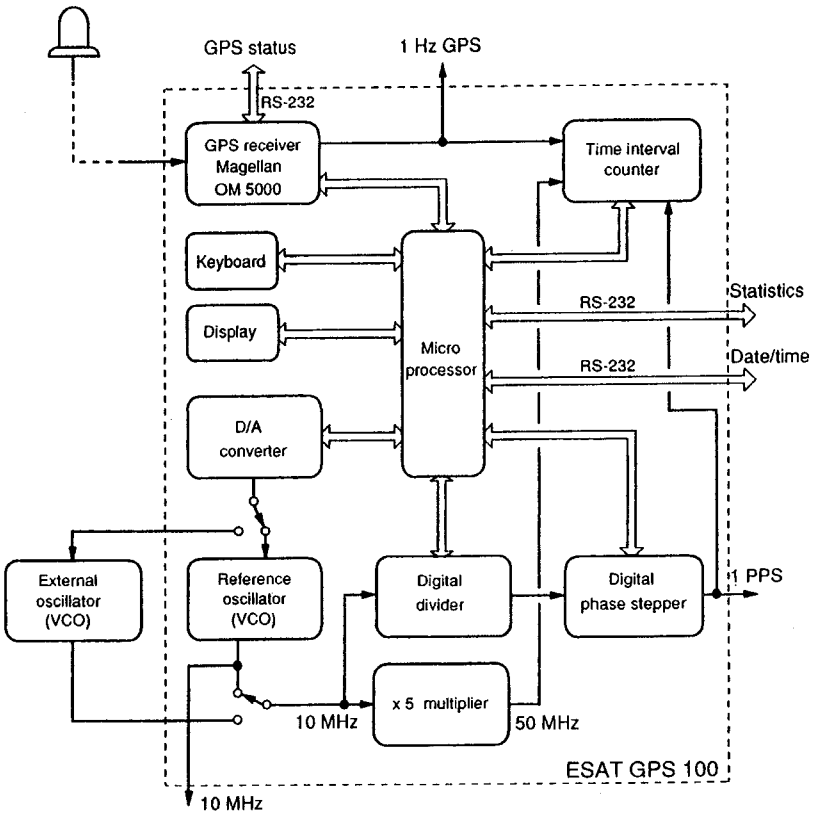


Fig. 8 - ESAT GPS 100 block diagram.

compensate for the frequency departure and drift of the local oscillator or, if the main interest is a reference of time, to maintain the 1PPS output signal derived from the local oscillator in close agreement with the GPS time and improve its short term stability. The internal reference can be either an ovenized quartz oscillator or a small rubidium. If a reference oscillator is already available in the laboratory, its frequency can be supplied to the GPS 100 that can provide externally a control voltage, if applicable to the standard used. The frequency and time corrections are computed, by linear regression, from the time interval measurements, performed each second, between 1Hz GPS and the local 1PPS after having applied a statistical filtering. The periodicity of the computation cycle, and the consequent updating of the corrections applied, can be selected in the range from 1 minute to several days according to the oscillator specifications and the expected uncertainty and short term instability. The resolution of the time correction of the 1PPS output, performed by the digital phase stepper, is 1 ns.

The GPS card can receive near simultaneously five satellites, and the best four are used to compute the antenna coordinates and to generate a 1Hz GPS output. When the coordinates are well known, they can be entered in the receiver that can be switched to its "Time Mode" operative condition. In this last case, the GPS time reference is obtained from an average of the four best satellites in view. Three RS-232 interfaces, allow the ESAT GPS 100 to: a) communicate the GPS receiver status and receive e.g. initialization parameters, b) send the results of the statistical evaluation performed by the microprocessor unit, c) supply both UTC and local date and time information updated each second and information on the satellites tracked.

To investigate on the capabilities of the ESAT GPS 100, a characterization of the device with an internal BVA 8601/02 quartz oscillator and with an external HP 5065 Rubidium oscillator has been performed at IEN.

The 1Hz(GPS) and 1PPS output signals of ESAT GPS 100 have been continuously compared with UTC(IEN) using an external counter. With the same counter, the frequency of the controlled oscillator versus UTC(IEN) has been also measured. For the results obtained in 20 days of time comparisons between UTC(IEN) and 1Hz(GPS) with a repetition rate of 1 s and 1 minute, it has been computed the standard deviation and the Mod $\sigma_y(\tau)$. Typical values of the standard deviation observed over 1 minute ranged from 30 ns up to 250 ns, due to the Selective Availability on the GPS signals. From Fig. 9, where the Mod $\sigma_y(\tau)$ is reported, it can be seen that for T>200 s up to $5 \cdot 10^4$ s, the data fit quite well with a $\tau^{-3/2}$ slope, corresponding to white phase noise. This means that, increasing the size of the averaged data, an increased accuracy of the frequency control can be obtained.

The test with the BVA oscillator, which has a frequency drift of less than $1 \cdot 10^{-11}$/d, has been performed choosing on the ESAT GPS 100 the "frequency correction" mode and using a time constant of 3 hours. The results obtained as regards to the GPS 100 output frequency and 1PPS signal are reported in Fig. 10 and 11 respectively. The relative frequency departure of the BVA oscillator, averaged over 1 hour, has been found within $\pm 1 \cdot 10^{-11}$ meanwhile the time reference output varies of ±200 ns around its mean value. Since for the period considered the difference between UTC(IEN) and GPS time measured with Allen Osborne Receiver TTR-5 was 0,16 μs in the average, the residual error is the calibration value of the GPS 100 .

For the second test, performed with an HP 5065 rubidium oscillator with a frequency drift of $7 \cdot 10^{-13}$/month, the "time correction" mode has been selected in the GPS 100. In Fig. 12 are shown the results of 15 days of time comparisons versus UTC(IEN) of the 1Hz(GPS 100) output, 1 minute apart. The time constant chosen, was of 6 hours. The time error varies within ±60 ns around its mean value. As far as

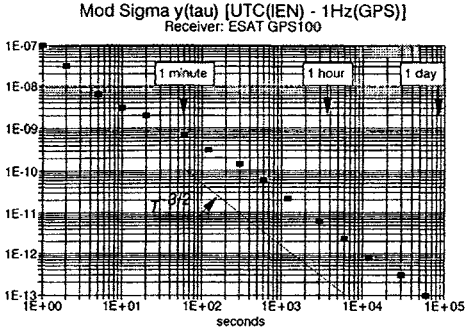


Fig. 9 - Modified Allan deviation of GPS signals.

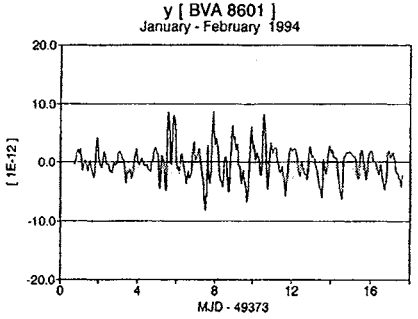


Fig. 10 - Frequency departure of the disciplined BVA oscillator vs. UTC(IEN).

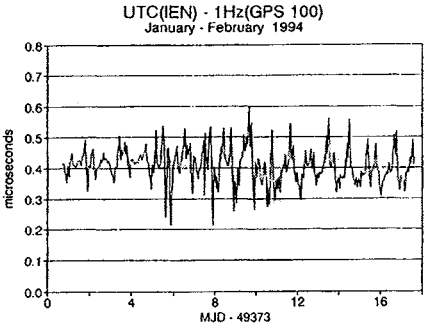


Fig. 11 - Time error of the GPS 100 with BVA oscillator.

the absolute value of the difference, it is applicable what reported above.

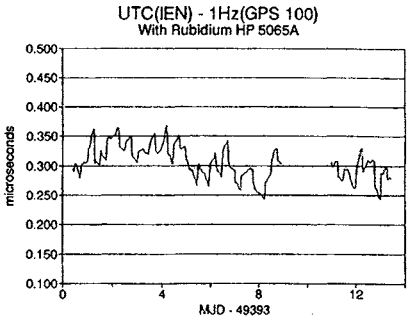


Fig. 12 - Time error of the GPS 100 with rubidium oscillator.

5 - Conclusions

The results obtained with different synchronization techniques, applied in the control of the secondary standards maintained in the accredited laboratories of the Italian Calibration Service (SIT), have been reported. Traceability levels to the national time standard from 10^{-9} to 10^{-12} can be obtained using the IEN coded time signals, the passive television method or the GPS signals. A new device, developed by ESAT - Torino, providing time and frequency traceability by means of the GPS system, has been characterized at IEN showing that a long-term accuracy of $\pm 1 \cdot 10^{-11}$ for a disciplined BVA quartz oscillator is obtainable. A reference of time in agreement with UTC(IEN) within ±60 ns, using an external Rubidium oscillator, has been achieved with the same device.

References

/1/ CCIR Report 1016-1, Reports of the CCIR, 1990 - Annex to vol. VII Standard frequency and time signals.

/2/ F. Cordara, V. Pettiti, P. De Giorgi: Remote oscillators frequency control by means of coded time signals. Proc. of the 1st European Time and Frequency Forum, March 1987.

/3/ C. Thomas: Real time restitution of GPS time. Proc. of the 7th European Frequency and Time Forum, March 1993.

/4/ W. Lewandowski, C. Thomas: GPS time transfer. Proc. of the IEEE, Vol. 79, n. 7, July 1991.

PHASE SENSITIVE DETECTION OF LIGHT REFLECTED FROM A FABRY-PEROT INTERFEROMETER

E.Bava, M. Bonaiti and F. Massari
Dipartimento di Elettronica e Informazione - Politecnico di Milano
Piazza Leonardo da Vinci 32, Milano, Italy

Summary. Phase sensitive detection techniques are currently employed in laser frequency stabilization. Excellent short term stability has been obtained using a Fabry-Perot interferometer as a frequency discriminator (Pound-Drever method). In this paper an analysis is presented of the Fabry -Perot reflection response to a phase-modulated light by considering the general problem of the lock-in detection at the pth harmonics of the rf modulating frequency. Suitable FM conditions for servo locking purposes are obtained and investigated, and the values of modulation index which maximize the sensitivity for the first, third and fifth harmonics are found. Finally, the effects of the residual amplitude modulation introduced by the electro-optic frequency modulator, the presence of laser amplitude noise and the dependence of the achievable closed-loop frequency fluctuation spectrum on the modulation index and detector noise are considered.

1. Introduction.

Optical heterodyne of frequency-modulated laser signals /1,2/ are a useful tool both in high-resolution spectroscopy and in laser frequency stabilization. These techniques have been successfully implemented for many sources affected by different noise spectra : dye /3,4/, solid-state /5,6/, semiconductor /7/ lasers, etc., as the only requirement for a full exploitation of the method is the use of a modulation frequency much higher than the spectral feature of the molecular or cavity resonance.

Interesting results in high-resolution laser spectroscopy of I_2 have been presented in /8/, whereas recent achievements on solid-state laser stabilization with reference to a high-finesse Fabry-Perot, concerned also with gravity wave detection, have been reported in /9/; moreover stabilization problems are discussed with a comprehensive analysis in /10/.

Whereas the properties of the frequency response of a Fabry-Perot transmission cavity to frequency modulated light have been discussed in /11/, in this work the properties of the Fabry-Perot in the reflection mode are considered in general, limiting the specific analysis to the case of a modulation frequency much higher than the cavity line width; moreover an attempt is made to

examine the effects of the laser residual amplitude modulation and noise and of the detector noise.

2. Frequency response of the Fabry-Perot discriminator after lock-in detection.

In analyzing the response of a Fabry-Perot cavity operating in the reflection mode and excited by a phase-modulated light, the same method used in /11/ can be followed, just substituting in the relevant expressions the reflection to the transmission coefficients. Suppose the incident laser radiation, with amplitude A and angular frequency ω_l be modulated at ω_m with a modulation index m, then the total field can be expressed as

$$E_i(t) = \mathrm{Re}\Big\{ A\big[1+a(t)\big]\exp\big[j\big(\omega_l t + m\sin\omega_m t + \phi(t)\big)\big]\Big\}$$
$$= \mathrm{Re}\Big\{ A\big[1+a(t)\big]\sum_{n=-\infty}^{\infty} J_n(m)\exp\big[j\big((\omega_l + n\omega_m)t + \phi(t)\big)\big]\Big\} \qquad (1)$$

where terms $a(t)$ and $\phi(t)$ account for amplitude and phase fluctuations. In the following of this section, just concerned with the system response to a frequency difference $\Delta\omega = \omega_l - \omega_0$, where ω_0 is the cavity resonance, such fluctuations shall not be considered and only in section 4 suitable expressions shall be introduced to study particular problems of noise.

The total reflected field $E_0(t)$ is the inverse Fourier transform of the product of the cavity reflection coefficient $\Gamma(j\omega)$ times the Fourier transform $E_i(\omega)$ of the field expression inside the curled brackets in (1), namely

$$E_0(t) = \mathrm{Re}\Big\{ \int_{-\infty}^{\infty} E_i(\omega)\Gamma(j\omega)\exp(j\omega t)\,df \Big\} = \mathrm{Re}\{z(t)\} \qquad (2)$$

This field incident in a photodetector produces a photocurrent proportional to the field intensity

$$I = \frac{\eta e}{h\nu_l} c\varepsilon_0 \frac{zz^*}{2} S \qquad (3)$$

where S is the detector area, $\nu_l = \omega_l/2\pi$, η is the detector quantum efficiency and e, c, ε_0 and h are the electron charge, the velocity of light, the vacuum permittivity and Planck's constant respectively. The expression of zz^* is

$$\frac{zz^*}{A^2} = \sum_{n=0}^{\infty}\sum_{k=0}^{\infty} J_n(m)J_k(m)\Gamma_n\Gamma_k^* e^{j(n-k)\omega_m t} + \sum_{n=0}^{\infty}\sum_{k=1}^{\infty}(-)^k J_n(m)J_k(m)e^{j(n+k)\omega_m t}\Gamma_n\Gamma_{-k}^*$$
$$+\sum_{n=1}^{\infty}\sum_{k=0}^{\infty}(-)^n J_n(m)J_k(m)e^{-j(n+k)\omega_m t}\Gamma_{-n}\Gamma_k^* + \sum_{n=1}^{\infty}\sum_{k=1}^{\infty}(-)^{n+k} J_n(m)J_k(m)e^{-j(n-k)\omega_m t}\Gamma_{-n}\Gamma_{-k}^* \qquad (4)$$

where

$$\Gamma_k = \Gamma\big[j(\Delta\omega + k\omega_m)\big] \qquad (5)$$

and it appears similar to (7) of ref /11/ provided $L(x+kx_c))$ is substituted with $\Gamma\left[j(\Delta\omega+k\omega_m)\right]$. According to expression (4), the contribution $Y_p(t)=\left(zz^*\right)_p/2A^2$ at the frequency $p\omega_m$, $p\geq 1$, is

$$Y_p(t)=\mathrm{Re}\sum_{n=0}^{\infty}J_n(m)J_{n+p}(m)\left\{\left[\Gamma_n^*\Gamma_{n+p}+(-)^p\Gamma_{-n}\Gamma_{-(n+p)}^*\right]\exp(jp\omega_m t)\right\}$$

$$+\mathrm{Re}\sum_{n=1}^{p-1}{}'(-)^{(n+p)}J_n(m)J_{p-n}(m)\Gamma_{-(p-n)}^*\Gamma_n\exp(+jp\omega_m t) \qquad (6)$$

where $\sum'$ holds for $p\geq 2$, and the continuous term

$$Y_0=\frac{1}{2}\left\{\sum_{n=-\infty}^{+\infty}J_n^2(m)\Gamma_n\Gamma_n^*\right\} \qquad (7)$$

To obtain the signal after lock-in detection, the coefficients of $sin\omega_m t$ and $\cos\omega_m t$, the in-phase and quadrature components Y_{ps} and Y_{pc} respectively, are to be separated in (6). By introducing

$$\begin{aligned}
h_n(t)&=h_n'+jh_n''=J_n(m)J_{n+p}(m)\Gamma_n\Gamma_{n+p}^*\\
h_{-n}(t)&=h_{-n}'+jh_{-n}''=J_n(m)J_{n+p}(m)\Gamma_{-n}\Gamma_{-(n+p)}^*\\
k_n(t)&=k_n'+jk_n''=J_n(m)J_{p-n}(m)\Gamma_n^*\Gamma_{-(p-n)}
\end{aligned} \qquad (8)$$

the two components become

$$\begin{aligned}
Y_{ps}&=\sum_{n=0}^{\infty}h_n''+(-)^{p+1}\sum_{n=0}^{\infty}h_{-n}''+\sum_{n=1}^{p-1}{}'(-)^{p+n}k_n''\\
Y_{pc}&=\sum_{n=0}^{\infty}h_n'+(-)^{p}\sum_{n=0}^{\infty}h_{-n}'+\sum_{n=1}^{p-1}{}'(-)^{p+n}k_n'
\end{aligned} \qquad (9)$$

The equations (9) yield the relevant information on the frequency-modulated laser radiation reflected by a Fabry-Perot cavity and synchronously detected. From (6) and the relationship $\Gamma\left[j(\omega)\right]=\Gamma^*\left[j(-\omega)\right]$ it turns out that $Y_p(t)$ is an odd or even function of $\Delta\omega$ according to p is odd or even, therefore for p odd the two detected components are suitable for frequency discrimination around $\Delta\omega=0$.

To go on with this analysis we must specify the model of the Fabry-Perot cavity. For sake of simplicity a lossless symmetric structure is assumed and consequently the reflection coefficient can be expressed as

$$\Gamma(j\omega)=\sqrt{R}\,\frac{1-\exp(-j2\omega\tau)}{1-R\exp(-j2\omega\tau)} \qquad (10)$$

where 2τ is the round-trip time and $R=r^2$ is the mirror reflectivity. Then the in-phase ($sin\omega_m t$)and quadrature ($\cos\omega_m t$) current components , I_{ps} and I_{pc} respectively, are

$$I_{ps} = \frac{\eta e}{h\nu} c\varepsilon_0 Y_{ps} A^2 S = \frac{\eta e}{h\nu} c\varepsilon_0 R(1-R^2) A^2 S$$

$$\{ \sum_{n=0}^{\infty} J_n(m) J_{n+p}(m) \left\{ \frac{\sin[2(\omega_l + n\omega_m)\tau] - \sin[2(\omega_l + (n+p)\omega_m)\tau] + \sin(2p\omega_m\tau)}{D_n^p(\omega_l, \omega_m)} \right.$$

$$\left. + (-)^p \frac{-\sin[2(\omega_l - n\omega_m)\tau] + \sin[2(\omega_l - (n+p)\omega_m)\tau] + \sin(2p\omega_m\tau)}{D_n^p(\omega_l, -\omega_m)} \right\}$$

$$- \sum_{n=1}^{p-1} (-)^{n+p} J_n(m) J_{p-n}(m) \frac{\sin[2(\omega_l + n\omega_m)\tau] - \sin[2(\omega_l - (p-n)\omega_m)\tau] - \sin(2p\omega_m\tau)}{D_n^{-p}(\omega_l, \omega_m)} \Bigg\} \quad (11)$$

$$I_{pc} = \frac{\eta e}{h\nu} c\varepsilon_0 Y_{pc} A^2 S = \frac{\eta e}{h\nu} c\varepsilon_0 R A^2 S \Bigg\{ \sum_{n=0}^{\infty} J_n(m) J_{n+p}(m) *$$

$$\left\{ \frac{(1+R)^2 N_n^p(\omega_l, \omega_m) + (1+R^2)\cos(2p\omega_m\tau) + 2R\cos[2(2\omega_l + (2n+p)\omega_m)\tau]}{D_n^p(\omega_l, \omega_m)} \right.$$

$$\left. + (-)^p \frac{(1+R)^2 N_n^p(\omega_l, -\omega_m) + (1+R^2)\cos(2p\omega_m\tau) + 2R\cos[2(2\omega_l - (2n+p)\omega_m)\tau]}{D_n^p(\omega_l, -\omega_m)} \right\}$$

$$+ \sum_{n=1}^{p-1}{}' (-)^{n+p} J_n(m) J_{p-n}(m) *$$

$$\frac{(1+R)^2 N_n^{-p}(\omega_l, \omega_m) + (1+R^2)\cos(2p\omega_m\tau) + 2R\cos[2(2\omega_l - (p-2n)\omega_m)\tau]}{D_n^{-p}(\omega_l, \omega_m)} \Bigg\} \quad (12)$$

where

$$N_n^p(\omega_l, \omega_m) = 1 - \cos[2(\omega_l + n\omega_m)\tau] - \cos[2(\omega_l + (n+p)\omega_m)\tau] \quad (13)$$

and

$$D_n^p(\omega_l, \omega_m) = \{1 + R^2 - 2R\cos[2(\omega_l + n\omega_m)\tau]\}\{1 + R^2 - 2R\cos[2(\omega_l + (n+p)\omega_m)\tau]\} \quad (14)$$

In (11), (12), (13) and (14) ω_l can be replaced with $\Delta\omega$ as $2\omega_0\tau = 2\pi$. A careful analysis confirms that the two components given by the above expressions are even or odd functions of $\Delta\omega$ according to p is even or odd. The same expressions give the well known patterns of Fig. 1, which specifically refer to p=1, $2f_m/\ \Delta\nu_c = 50$ and $m = 1.08$ (see also /3, 10/).

3. Some properties of the detected signal.

Up to the last expressions, the mathematical development has been rather general. However it has been recognized that a good choice for the ratio between modulation frequency and cavity line width is $p f_m/\Delta\nu_c \rangle\rangle 1$, as appears also from the patterns of Fig. 2, which specifically refers to Y_{1s}. Under this assumption, as the output (11) is more suitable to realize a frequency discriminator around $\Delta\omega = 0$ than the output (12), because of the much higher sensitivity, more

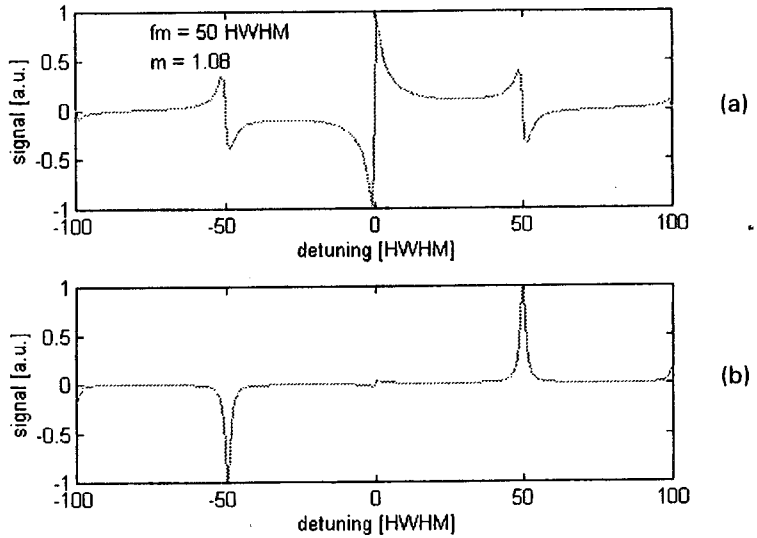


Fig. 1 - In phase (a) and in quadrature (b) components of the demodulated signal.

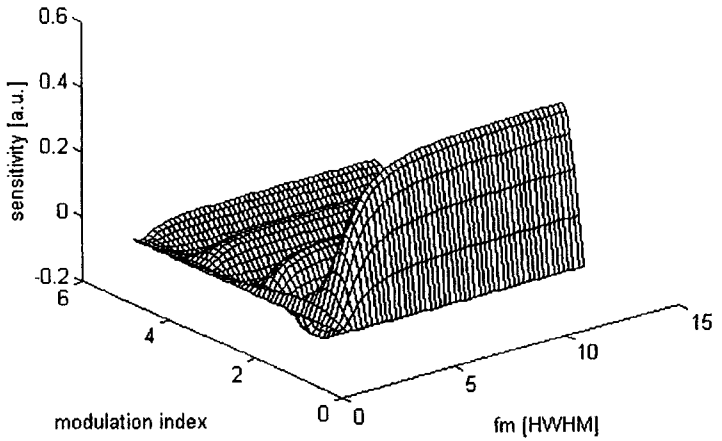


Fig. 2 Sensitivity of the in-phase component at p=1

attention shall be paid in the following to the in-phase component. Indeed if $(d\Gamma/d\omega)_{n\omega_m} \approx 0$, from (4) a simpler expression is found

$$Y_p = 2J_0(m)J_p(m)\Gamma'(\Delta\omega)\left[\Gamma''(p\omega_m)\cos p\omega_m t + \Gamma'(p\omega_m)\sin p\omega_m t\right] \qquad (15)$$

where $\Gamma = \Gamma' + j\Gamma''$. The expression (15) turns out a quite satisfactory approximation for $\Delta\omega \langle\langle \omega_m$. Both components Y_{ps} and Y_{pc} show the same dependence on m and $\Delta\omega$ and their responses are proportional to (1-R) and $sin\, p\omega_m\tau$ respectively, if (10) is used for $\Gamma(j\omega)$. From the relationship $\Delta\nu_c = 2(1-R)/\left(4\pi\tau\sqrt{R}\right)$ it turns out that $\Gamma'(p\omega_m) \rangle\rangle \left|\Gamma''(p\omega_m)\right|$ and therefore $Y_{pc} \langle\langle Y_{ps}$. As (15) in nearly linear in $\Delta\omega$ for $\Delta\omega \langle \omega_c$, varying m the product $J_0(m)J_p(m)$ determines the condition of maximum sensitivity. The dependence of sensitivity on m, for p=1,3,5, are reported in Fig.3; the maximum values are obtained for m=1.08 , 3.95 , 6.84 .

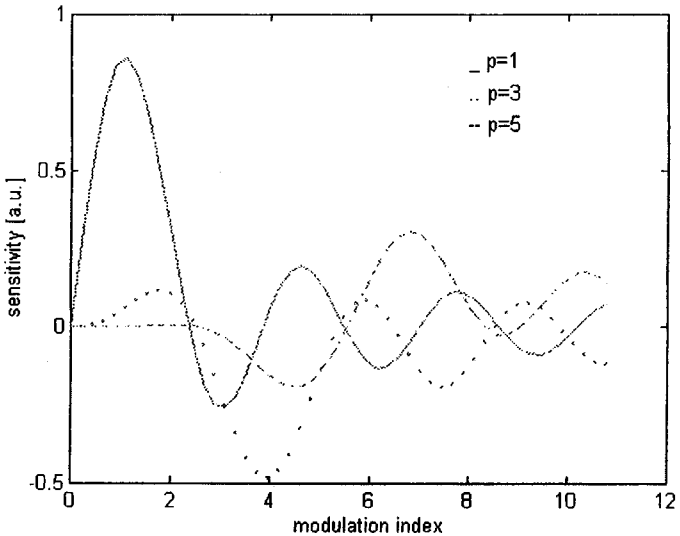


Fig. 3 Sensitivity of the 1st, 3rd and 5th harmonics of the in phase component. fm=25 HWHM

4. Unwanted amplitude modulation and noise and detector noise effects.

In this section a few problems concerned with the presence of noise and non ideal behavior of devices shall be considered, aiming at evaluating their effects on a laser stabilized by means of the Pound-Drever method.

4.1. Effects of unwanted amplitude modulation at ω_m.

If the device used to produce the frequency modulation of the laser output at ω_m introduces a small amplitude modulation as well, $\phi(t) = 0$, $a(t) = a_1 \sin(\omega_m t + \phi_1)$ and $\Delta\omega = 0$ are to be substituted in the expression (1). The normalized current is

$$Y_p = \mathrm{Re}\left\{ \sum_{n=-\infty}^{+\infty} \Gamma(n\omega_m)\Gamma^*\left[(n-p)\omega_m\right] L_n L^*_{n-p} \exp(j\omega_m t) \right\} \tag{16}$$

where

$$L_n = J_n(m) + \frac{a_1}{2j}\left[J_{n-1}(m)\exp(j\varphi_1) - J_{n+1}(m)\exp(-j\varphi_1) \right] \tag{17}$$

By approximating $\Gamma(n\omega_m) = |\Gamma(\omega_m)|$ for n≠0, it follows that Y_{ps} and Y_{pc} are proportional to $\cos\varphi_1$ and $\sin\varphi_1$, respectively. For Instance in the case p=1 the most significant terms are

$$Y_{1s} \cong a_1 |\Gamma(\omega_m)| \cos\varphi_1 \left\{ \Gamma(0)\left[J_0(J_0 - J_2) + 2J_1^2 \right] + |\Gamma(\omega_m)|\left[J_1(J_1 - 2J_3) + J_2(2J_2 - J_0) \right] \right\} \tag{18a}$$

$$Y_{1c} \cong a_1 |\Gamma(\omega_m)| \sin\varphi_1 \left\{ \Gamma(0) J_0(J_0 + J_2) + |\Gamma(\omega_m)|\left[J_1(J_1 + 2J_3) + J_2(2J_2 + J_0) \right] \right\} \tag{18b}$$

where $\Gamma(0) = \Gamma(\Delta\omega = 0)$ and the common argument of the Bessel functions is m. The output dependence on ϕ_1 suggests that there may be a suitable phase shift in the demodulator which makes this offset minimum. Apart from the different effects of ϕ_1, the two offsets are of the same order of magnitude.

4.2 Effect of laser amplitude noise.

A simple evaluation of the consequences of the presence of amplitude noise is obtained by assuming in (1) $a(t) = a\sin(qt + \alpha)$, $\phi(t) = b\,sinqt$ and $\Delta\omega = 0$. Accordingly one obtains

$$\begin{aligned} \frac{z}{A} &= J_0(m)\left\{ \sum_{k=-\infty}^{\infty} \exp\left[j(\omega_0 + kq)t \right] \Gamma\left[j(\omega_0 + kq) \right] M_k \right\} + \\ &\left\{ \sum_{n,k}{}' \exp\left[j(\omega_0 + n\omega_m + kq)t \right] \Gamma\left[j(n\omega_m + kq) \right] J_n(m) M_k \right\} \end{aligned} \tag{19}$$

where

$$M_k = J_k(b) + j\frac{a}{2}\left(e^{-j\alpha} J_{k+1}(b) - e^{j\alpha} J_{k-1}(b) \right) \tag{20}$$

A rather cumbersome development shows that both the demodulated outputs have biases rigorously proportional to $a\sin\alpha$. Suitable approximatios of the offsets are

$$\overline{Y}_{ps} \cong 4a\sin\alpha J_0(m) J_p(m) \Gamma'(p\omega_m) \sum_{k\geq 1} \Gamma''(kq) J_k(b)\left[J_{k+1}(b) + J_{k-1}(b) \right] \tag{21}$$

$$\overline{Y}_{pc} \cong 4a\sin\alpha\left\{\sum_{k\geq 1} J_k(b)\left[J_{k+1}(b)+J_{k-1}(b)\right]\right\}\left\{\begin{array}{l} J_0(m)J_p(m)\Gamma''(p\omega_m)\Gamma''(kq) \\ +\sum_{n\geq -(p-1)/2, n\neq 0} J_n(m)J_{n+p}(m)\left|\Gamma(n\omega_m)\right|^2 \end{array}\right\} \qquad (22)$$

If either $a = 0$ or $b = 0$ there is no bias at the two outputs, in the opposite case even if a control loop decreases frequency noise, an offset is produced due to the presence of amplitude noise. The former contribution in (22) is very low if compared with that in (21), because $\Gamma'(p\omega_m) \rangle\rangle \left|\Gamma''(p\omega_m)\right|$, however the most important term in (22) is the latter.

4.3 Dependence of the signal to noise ratio at the detector input on the modulation index.

When the loop gain of the servo system is high the laser frequency fluctuations are strongly reduced and the main contributions to noise are the detector shot noise and the electronic noise transferred to the detector input. Following /9/, the closed-loop spectral density of the laser frequency fluctuations becomes

$$S_{\Delta\nu,cl} = S_A \Big/ \left|D_A^2\right| \qquad (23)$$

where $S_A \left[A^2/Hz\right]$ is the spectral density of the current noise at the detector input and $D_A \left[A/Hz\right]$ is the discriminator sensitivity, which at the previous approximation level and for $p = 1$ can be expressed as a function of the power P_i incident to the cavity in the following way

$$D_A = \frac{8J_0(m)J_1(m)}{\Delta\nu_c}\frac{e\eta}{h\nu_l}P_i \cdot \Gamma'(\omega_m) \qquad (24)$$

The detector noise spectrum, sum of the shot term and of the electronic noise N transferred to the input, is

$$S_A = 2e\frac{e\eta}{h\nu_l}P_i\left[J_0^2(m)\Gamma^2(0)+2\sum_{k=1}^{\infty}J_k^2(m)\left|\Gamma(k\omega_m)\right|^2\right]+N \qquad (25)$$

From the substitution of (24) and (25) into (23) and with the approximation $\left|\Gamma(k\omega_m)\right|^2 \cong \left|\Gamma^2(\omega_m)\right|$, the closed-loop laser frequency spectrum is

$$S_{\Delta\nu,cl} = \frac{h\nu_l(\Delta\nu_c)^2}{32\eta P_i\left[\Gamma'(\omega_m)\right]^2}\left[J_0^2(m)\Gamma^2(0)+\left(1-J_0^2(m)\right)\left|\Gamma(\omega_m)\right|^2+\frac{Nh\nu_l}{2e^2\eta P_i}\right]\Big/ J_0^2(m)J_1^2(m) \qquad (26)$$

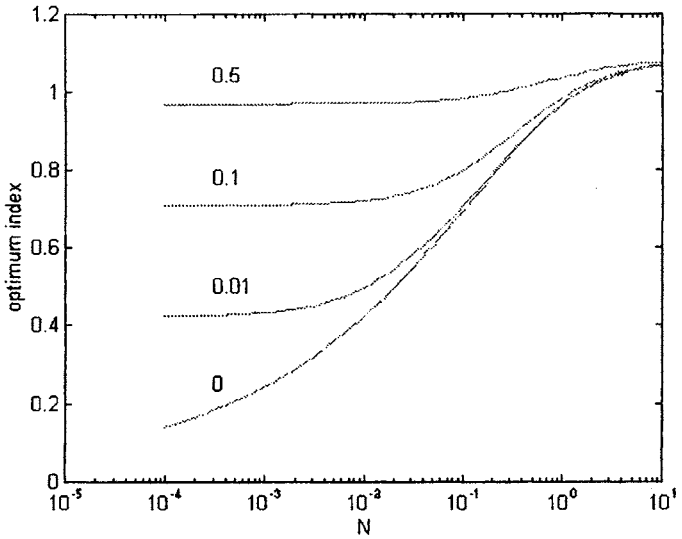


Fig. 4 Modulation index for optimum S/N for some selected values of $|\Gamma(0)|^2$.

To minimize the expression of $S_{\Delta\nu,cl}$ it is suitable to use as parameters $\Gamma^2(0)$ and $N' = Nh\nu_i/2e^2\eta P_i$, whereas $|\Gamma(\omega_m)|^2 \cong 1$, and the dependence of the optimal conditions on the parameter values is shown in Fig. 4. The diagrams show that to reduce $S_{\Delta\nu,cl}$ it is necessary to choose suitable m values which, for usual detectors and power levels are rather different from 1.08. For instance if $N = 2\cdot 10^{-24} A^2/Hz$, $\lambda = 1.06\mu m$, $P_i = 100\mu W$ and $\eta = .75$, $N' = 10^{-1}$ is obtained and from Fig: 4 it turns, out for $|\Gamma^2(0)| = 0.01$, $m \cong 0.65$. Moreover, if P_i increases and $\Gamma(0) \to 0$, the minimum frequency noise requires $m \to 0$, which is far from the condition of maximum sensitivity.

5. Conclusion.

The response of frequency-modulated light reflected from a Fabry-Perot cavity and detected in a phase-sensitive receiver has been studied , and properties have been examined when the modulation frequency is higher than the cavity line width. With reference to the problem of laser stabilization, expressions of biases at the discriminator output due to noise and to non-ideal behavior of the modulator have been given; moreover the working conditions to reduce the effect of detector noise on the frequency fluctuation spectrum of the locked laser have been obtained.

Acknowledgement

This work has been partially supported by the Italian Space Agency (ASI).

References

/1/ Bjorklund G.C., Levenson M.D., Lenth W., Ortiz C.: Frequency Modulation (FM) Spectroscopy, Theory of Lineshapes and Signal to noise Analysis. Applied Physics B, vol.32 (1983), pp.145/52.

/2/ Drever R.W.P., Hall J.L., Kovalsky F.V., Hough J., Ford J.M., Munley A.J.: Laser phase and frequency stabilization using an optical resonator, Appl. Phys. B., vol. 31 (1983), pp. 97/105.

/3/ Hall J.L., Baer T., Hollberg L., Robinson H.G.: Precision Spectroscopy and Laser Frequency Control Using FM Sideband Optical Heterodyne Techniques. Laser Spectroscopy V, Ed. McKellar,Oka, Stoicheff, Berlin, Heidelberg,New York: Springer Verlag 1981, pp. 15/24.

/4/ Houssin M., Jardino J., Gely B., Desaintfuscien M.: Design and performance of a few - kilohertz-line width dye laser stabilized by reflection in an optical resonator. Optics Letters, vol.13 (1988) no.10, pp. 823/25.

/5/ Day T., Nilsson A.C., Fejer M.M., Farinas A.D., Gustafson E.K., Nabors C.D., Byer R.L.:30-Hz -Linewidth, Diode Laser Pumped, Nd:GGG Nonplanar Ring Oscillators by Active Frequency Stabilization. Electronics Letters, vol 25 (1989) No. 13, pp. 81012.

/6/ Shoemaker D., Brillet A., Nary Man C., Cregut O., Kerr G.: Frequency-stabilized laser-diode-pumped Nd:YAG laser, Opt. Lett., vol. 14(1989), pp.609-611.

/7/ Ohtsu M., Murata M., Kourogi M.: FM noise reduction and subkilohertz linewidth of an AlGaAs laser by negative electrical feedback, IEEE J. Quantum Electron, vol. 26, (1990), pp. 231-234.

/8/ Arie A., Byer R.L.: Laser Heterodyne Spectroscopy of $^{127}I_2$ Hyperfine Structure Near S32 nm, J. Opt. Soc. Am. B, Vol. 10 (1993) n.11, pp. 1990/1997

/9/ Day T., Gustafson E.K., Byer R.L.: Sub-Hertz Relative Frequency Stabilization of Two-Diode Laser-Pumped Nd:YAG Lasers Locked to a Fabry-Perot Interferometer. IEEE Jour. of Quantum Electronics, vol. 38 (1992) No. 4, pp.1106/16.

/10/ Zhu M., Hall J.L.: Stabilization of optical phase/frequency of a laser system: application to a commercial dye laser with an external stabilizer. J. Opt. Soc. Am. B vol. 10 (1993) No. 5, pp. 802/16.

/11/ Hils D., Hall J.L.: Response of a Fabry-Perot cavity to phase modulated light. Rev. Sci. Instrum. vol. 58 (1987) No.8, pp. 1406/12.

THE VARIANCE OF PREDICTABILITY OF HYDROGEN MASERS AND OF PRIMARY CESIUM STANDARDS IN SUPPORT OF A REAL TIME PREDICTION OF UTC

by

David W. Allan
Allan's TIME
P.O. Box 66
Fountain Green, UT 84632

and

Marc A. Weiss
Time and Frequency Division
National Institute of Standards and Technology
325 Broadway
Boulder, CO 80303

ABSTRACT

The need for global synchronization has increased rapidly over the last few years. These needs have arisen from the telecommunications industry and from the monitor and control stations associated with navigation systems. This paper presents a possible solution for a real-time predicted estimate of UTC based on some of the best clocks in the world. It specifically addresses the variance of time predictability utilizing the excellent resources now available. These resources include the very stable hydrogen masers and the primary cesium-beam frequency standards located at several timing centers around the world.

One problem in providing a real-time estimate of UTC is the uncertainty in comparing the time and frequency of widely separated clocks. Based on real data, we will show the time predictability in a variance sense of the outstanding clocks available independent of this problem. Then we will discuss how an accuracy of about 10 ns for a real-time prediction of UTC can potentially both be achieved and be disseminated internationally.

I. OUTLINE AND PERSPECTIVE

The need for global synchronization has increased rapidly over the last few years, largely driven by the telecommunication and power industries and needs for synchronized monitor and control stations associated with navigation systems. These new timing needs are not yet being met. This paper will present possibilities for meeting these needs.

Contribution of the US Government, not subject to copyright.

The most desirable solution is for users to have ready access to a common real-time time scale. Universal Coordinated Time (UTC) is the international time scale generated by the Bureau International des Poids et Mesures (BIPM) in France. It is not a real-time scale, being provided about two months after the fact because of the care needed in its computation. To keep the best accuracy of the SI second as defined in the international system of units, UTC must be calculated after the fact. It is not possible to have a real-time UTC without compromising its reliability, stability, and rate accuracy. On the other hand, to be practically useful, a real-time estimate of UTC needs to be readily available. If UTC is needed in real time, it must be predicted forward in time [1,2].

All measurement units except time can be post-processed with effectiveness. The standard for time is different in that, by its very nature, it is most accurately realized when the prediction interval used to estimate the current time is minimized. For this reason, the BIPM is currently studying the calculation interval to determine whether it can be reduced to one month. Several timing centers provide real-time predictions of UTC (with the current two month delay in data), some keeping their real-time scale within approximately 100 ns of where UTC will be after it is computed.

This paper presents a method for obtaining a real-time prediction of UTC through combining some of the best clocks in the world using optimal estimation and prediction techniques. The clocks used include the very stable active hydrogen masers with cavity tuning and the primary cesium-beam frequency standards located at various timing centers around the world. We will show that it is possible to reduce the errors of prediction to the order of 10 ns, an uncertainty of the same order as the measurement noise. This uncertainty would be small enough to satisfy most current user needs.

In addition, the accessibility to this real-time prediction can be enhanced. We show how different national time scales, the UTC(k)'s, where "k" denotes a particular timing center, can be synchronized to about 10 ns. The value for users would be reduced concern about where they obtain a reference time. Such improvement would provide a reference one to two orders of magnitude better than is currently available. For example, within the USA the telecommunications industry has declared UTC as the official reference for time. However it cannot be used because it is not available in real time. Instead, UTC(NIST) and UTC(USNO MC), which track UTC to within about 100 ns, are used as secondary references. The synchronization needs of this industry are increasing. Hence, better coordination of the different sources of UTC is now needed. If the UTC(k)'s were commonly held within 10 ns of UTC, it could significantly assist international telecommunications.

II. THE VARIANCE OF TIME PREDICTION

This section describes the theory used to accomplish an improved real-time prediction of UTC [2,3,4]. For optimum time prediction there are two issues.

First, the systematics errors in a clock must be optimally estimated. These typically include the time offset, the frequency offset and the frequency drift. Second, the random variations need to be modeled. Optimum prediction algorithms can be constructed from these models. We then combine the optimum estimates of the systematic errors with the optimum estimation procedures for the modeled random processes. This parameterization of each clock can, in principle, be used to provide an independent estimate of a real-time UTC.

Once a prediction is obtained for each clock, the values are combined as a weighted set. If the clocks are independent, the optimum weighting factor for each clock is proportional to the reciprocal of the mean-squared prediction error for that clock. An optimum estimate of the systematic parameters will depend on the models used for the stochastic processes. Hence, the first step is to discuss these stochastic models.

Three noise processes provide parsimonious modeling for the random behavior of the clocks we will use. These are white-noise frequency modulation (FM), 1/f or flicker-noise FM and random-walk FM. The measurement noise also enters, and we will address this issue later.

Assuming an optimum algorithm is used, the optimum prediction error is given approximately by $\tau \cdot \sigma_y(\tau)$. Optimum means minimum squared error. This expression is exactly the optimum prediction error for white-noise FM and for random-walk FM. For flicker-noise FM the optimum prediction error is given by $\tau \cdot \sigma_y(\tau) / \sqrt{(\ln 2)}$.

The optimum estimate of clock frequency offset from a reference for white FM is the simple mean. The optimum time prediction value is the last time offset measurement. The optimum frequency drift estimate can be calculated from the linear regression to the residual frequency-offset measurements. New commercial cesium-beam frequency standards have negligible frequency drift and are dominated by white FM. In practice, the normalized mean frequency may be calculated by taking the last time offset measurement minus the first time offset measurement divided by the elapsed time [5].

The optimum algorithm for flicker noise FM requires a filter which converts the noise type to white noise FM, and then takes the mean. The filter is then inverted for prediction. This can be quite complicated. But there are two simple algorithms which are very close to optimum. The first is an exponential-filtered frequency estimate. This is particularly useful if the clock is well modeled by a composite of white noise FM and flicker noise FM [6]. The second is called a second-difference predictor. If τ is the prediction interval, it is simply formed by

$$x_p(t+\tau) = 2x(t) - x(t-\tau), \qquad (1)$$

where x(t) is the last known value of the clock relative to UTC. The value of τ

in our example would be two months with the current procedure for computing UTC two months after the fact. The BIPM is studying the impact of reducing this delay to one month.

The optimum algorithm for random-walk FM is the last value for the time error, the last slope for computing the frequency offset, and the mean second difference for computing the frequency drift. The difficulty is that, in practice, one model is never adequate. From Kalman-Busie theory [7], an exponential filter can be shown to be optimum for a composite of white FM and random-walk FM. If, for example, a clock is modeled by white noise FM for shorter integration times and by random-walk noise FM for longer integration times, then the optimum drift estimate is given by a three-point, second-difference estimate [8]. The three points are the first, the middle, and the last time readings.

Given a set of clocks which are independent of each other, the individual instabilities can be computed using an N-cornered hat approach [9,10]. We might also use a time scale algorithm to estimate the internal noise processes for each of the contributing clocks. [6] We will use this latter approach here, because it avoids negative variances.

We use a unique method to treat data from different locations taken at different times. Under the assumption of independence, we combine the prediction error variances to arrive at an estimate of the total prediction error for a real-time UTC. In this way we discover potential prediction accuracy, independent of time links. We will discuss the impact of this assumption later.

III. APPLICATION USING THE CURRENT BEST CLOCKS AVAILABLE

To study the performance of the best frequency standards available, we generated three experimental time scales, the first two using clocks colocated at particular sites, the third using time transfer to bridge two sites. At these sites we had access to clock data with outstanding performance. We used the TA2 time scale algorithm to generate these three independent time scales. TA2 is a retrospective smoothing algorithm, which runs a time scale both forward and backward in time. It combines the forward- and backward-calculated values and obtains a post-processed scale optimized for frequency stability [11,12]. This provided a reference for studying the instability of some of the most stable clocks in the world. The one-directional time scale algorithm, AT2 [13], is an extension of AT1, the algorithm used to generate the real-time unsteered time scale at NIST, with the addition of an estimate of confidence of the frequency estimate. This confidence estimate allows the forward and backward AT2 estimates of frequency to be combined into a smoothed estimate of frequency. A third pass in the forward direction uses this smoothed estimate of the frequency offset of each clock to estimate the time offsets. For estimating the predictability of clocks we are interested in their behavior only forward in time.

We obtained estimates of clock predictability from the TA2 algorithm only after compensating for the advantage gained from retrospective smoothing.

The clocks we used were active hydrogen masers (with cavity autotuning) and primary cesium-beam frequency standards. The first site was PTB (Physikalisch-Technische Bundesanstalt), which operates two laboratory primary cesium-beam frequency standards and two active hydrogen masers with cavity tuning. The time scale we generated from the PTB data we refer to as TA2PTB. The second site is the former Soviet Union (SU) timing center north of Moscow. We had the data from five active cavity-tuned hydrogen masers at this site. We call the time scale we generated from the data from these masers, TA2SU. The third scale combined two hydrogen-masers at the United States Naval Observatory (USNO) linked with one hydrogen-maser at NIST (National Institute of Standards and Technology) using the GPS common-view time transfer technique. We call the time scale generated from these measurement data, TA2M1.

The masers at PTB and at SU use a cavity tuning technique where the atomic line width of one maser is modulated and the cavity frequency adjusted under servo control until its oscillation frequency as measured against the other maser is independent of atomic line width. The reference maser and tuned maser are then reversed so that the cavity frequencies of both masers are autotuned. The masers at USNO and NIST use a different technique where the cavity frequency is modulated. A servo system then tunes the cavity frequency until the oscillation frequency is independent of small mistunings. This approach does not require an external frequency reference. Either approach is designed to remove the effects of cavity mistuning. Removing the effects of cavity mistuning is very important in achieving long-term frequency stability and time predictability. Cavity detuning is the major cause of frequency drift in conventional masers. [14,15,16] The cavity tuning usually reduces the frequency drift by more than an order of magnitude.

The variances of the clocks in this study were computed against their respective scales. The true variance of a clock is larger than the variance of that clock when measured against an ensemble containing that clock, since the ensemble is in part defined by that clock. An estimate of the true variance can be obtained by adjusting the measured variance for correlation with the scale. The relationship is

$$\sigma^2_{true} = \frac{\sigma^2_{measured}}{1-w}, \qquad (2)$$

where w is the normalized weight of the clock in the ensemble [17].

Figure 1a shows the estimated true variances for the clocks at PTB as seen by the TA2PTB scale. The data cover the period from MJD 48700 - 49050 (19 March 1992 - 4 March 1993). A short period of anomalous performance was

removed from the PTB H1 maser data (MJD's 48921 - 48953), to enable determination of the steady-state performance. Figure 1b shows the variances with estimated drifts removed. The drift values are tabulated in Table I below along with the confidence on each estimate [18]. The confidence intervals on the variance estimates are omitted from the plot to allow the variances to be seen clearly. We discuss the 10 d confidence below, since we use that to determine confidence on the 10 d prediction errors.

The TA2SU scale ran from MJD 48835 - 49000 (01 August 1992 - 13 January 1993). Figures 2a and 2b show the uncorrelated estimates of the true variances for the clocks at SU using the TA2SU time scale without and with frequency drift removed, respectively. As for the data from PTB, the drift values and confidences are tabulated in Table I.

The TA2M1 scale linked masers at NIST and USNO with once-per-day measurements from MJD 49050 - 49349 (4 March 1993 - 28 December 1993). The resultant variances for the masers as referenced by TA2M1 and adjusted for correlation are plotted in figure 3a, while the variances with drift removed from USNO maser N4 are plotted in figure 3b. The time transfer technique added measurement noise in the form of a time step in the comparison technique of the order of 1 ns. Such a shift in the measurement system would appear as white noise FM of the order of $1.2 \cdot 10^{-14} \cdot \tau^{-1/2}$. To help stabilize the TA2M1 scale three other masers at NIST were added for short periods each, as they were available. The result is that the TA2M1 algorithm assigned the time transfer noise to the USNO masers. The variances of the masers themselves are not visible until after τ=10 days. The measurement noise is averaged down for longer values of τ.

To estimate the error in prediction we ran the scales again, with data selected once every 10 days. We also added a "clock" to the TA2M1 scale which was an average of 33 commercial primary cesium-beam frequency standards. Figures 4 - 6 show the actual rms prediction errors for each clock versus its respective ensemble, adjusted to decorrelate clocks from their respective scales according to equation (2). We further adjusted the prediction errors to account for TA2 being a retrospective smoother. Since the prediction errors from TA2 are the root mean square of the forward and backward time prediction errors, we have multiplied the TA2 prediction errors by $\sqrt{2}$. We note in TA2M1 that for this run there is only one maser at NIST and effectively three clocks at USNO. For this reason, the algorithm confused the time transfer noise with the NIST maser. It appears that at τ = 10 days there is still some measurement noise contaminating the prediction error estimate of the NIST maser. Also note that the NIST H1 maser was out of the measurement system for some days for adjustments. When it came back in, the prediction error took a large jump until the scale learned the new frequency of the maser.

The values of the 10-day prediction errors after each clock reached steady-state were combined in a root-mean-square sense for the entries in Table II from the

three different ensemble time scales. We determine confidence on these values from the confidence on the $\sigma_y(\tau)$ values and the fact that the optimum prediction error is proportional to $\tau \cdot \sigma_y(\tau)$. The confidence interval for variance or prediction error is asymmetric. The 95% confidence fractional uncertainty at 10 d for the variances against each of the ensembles were: 0.83 - 1.2 for TA2PTB, 0.75 - 1.5 for TA2SU, and 0.8 - 1.4 for TA2M1. These values give the confidence intervals in Table II.

An optimum combination of these 10 d prediction errors -- excluding time transfer measurement noise -- would yield a 10 day prediction uncertainty of less than 1 ns. If we were limited by flicker-noise FM for prediction times longer than 10 days, then we would have a one-month prediction uncertainty of about 3 ns. If we were limited by random-walk FM, then the one-month prediction uncertainty would be about 5 ns.

IV. THE SIGNIFICANCE OF TIME TRANSFER TECHNIQUES

Since the high quality clocks used in the analyses are not colocated, transferring their frequency and time optimally is an important part of combining their readings. Obviously, the combined readings could give us the best prediction of UTC. This would also give us the advantage of independent clocks at separated sites. We could then better test for correlated effects that might exist between clocks at a given site [19]. Also, the more clocks we have, the better we can detect abnormal behavior in any one of the clocks. Such detection would minimize the effect on the composite prediction from the set of clocks we choose for minimizing the errors of prediction of UTC.

Currently, the best operational method for time transfer is obtained using GPS in the common-view mode. A receiver capable of participating in this kind of time transfer has been installed by the BIPM at VNIIFTRI (National Scientific and Research Institute for Physical-Technical and Radiotechnical Measurements) near Moscow. This is where the official SU time scale is generated. All of the sites used for the analysis in this paper participate in this method of time transfer. The measurement noise for this method is slightly dependant on location and the base-line distance between the two sites being compared. Typical numbers are 1 to 6 ns from day to day [10]. The measurement noise is typically white noise PM; hence, it will average as the square root of the inverse averaging time. If optimum filtering on the noise is utilized, it need not impact the prediction times for prediction intervals of one month. Sometimes the measurement noise is as dispersive as 1/f (flicker) PM, and even in this case it should average sufficiently so that its contribution will be significantly less than 10 ns at one month.

Having a high-confidence current estimate of UTC would provide several advantages. We have illustrated that less than 10 ns is probably achievable. Hence, each timing center could, in principle, steer their UTC(k) to this accuracy. If this were the case, industries like telecommunication could obtain

a synchronization reference from any key timing center -- a significant advantage for international telecommunication particularly for countries where multiple servers are employed.

If a high-confidence estimate of UTC were available in real time, then the disseminated time UTC(USNO) from GPS would track UTC corresponding to the confidence of the estimate. Hence, GPS could become a more accurate source of UTC, even as GPS is becoming more readily available. Recent work has shown that for timing, the SA modulation on GPS can be filtered to the order of 1 ns [20]. SA is the Selective Availability modulation which degrades the GPS broadcast signal as used by the civil sector. Since SA can be filtered to low levels, as UTC(USNO) steers more closely to UTC, GPS can provide a high accuracy, indirect access to UTC at a level of better than 10 ns [2].

Other time transfer techniques are being developed. The two-way satellite time and frequency transfer technique shows promise of 1 ns accuracy and subnanosecond day-to-day stability. It is now being developed among several time standards laboratories [21]. However, it is relatively expensive. There are possibilities for transmitting a time service from INMARSAT satellites which could potentially supply UTC world wide [22]. It could become the first UTC distribution method. The potential accuracy of INMARSAT timing could be 10 ns -- consistent with the confidence level for time prediction of UTC presented in this paper. Further, INMARSAT would operate in the receive-only mode, making it easier and less expensive to use than the two-way technique.

Another possibility is the use of the telecommunications system for precise time and frequency transfer. A number of telecommunication companies are studying two-way time transfer in optical fiber between nodes in their network [23]. Optical fibers have been shown to be an excellent media for time transfer [24]. Eventually, if the telecommunication industry has ready access to UTC, they in turn could become very accurate suppliers of UTC as well.

V. CONCLUSIONS, REVIEW AND PERSPECTIVE

In the last few years, clocks have been introduced with outstanding long-term stability. This long-term frequency stability translates to exceptional time predictability when optimum or near optimum prediction algorithms are used. These clocks include laboratory cesium-beam frequency standards and active hydrogen masers with cavity tuning capability. In addition, a particular commercial cesium-beam frequency standard now rivals the laboratory standards of recent years.

The analysis in this paper considered a subset of available clocks to demonstrate the thesis -- that UTC can be predicted a month in advance to an accuracy of less than 10 ns. This, of course, assumes that the rate of UTC will be stable to this same level. The subset we used were the two primary cesium-beam frequency standards (operating as clocks) at the PTB and hydrogen masers

located at NIST, USNO, PTB and VNIIFTRI. Figures 4-6 show the 10 d prediction errors for these clocks relative to their respective experimental time scales as the scales evolve into steady-state performance. The values after each clock reached steady-state were averaged for the entries in Table II.

There are now about 50 improved commercial cesium-beam frequency standards, like the ensemble of 33 at USNO, contributing as clocks to TAI and UTC. The specification of these can be characterized with a frequency stability, $\sigma_y(\tau) = 8 \cdot 10^{-12} \cdot \tau^{-1/2}$ for 10 s < τ < 10 d. In some cases this white noise FM is lower and persists for even longer values of τ. Using the optimum time prediction procedures outlined in this paper, the rms prediction error at 10 d for a single standard would be 7.4 ns. If all 50 of these standards were operating within their specification and were independent, optimally combining them would yield an rms dispersion of 7.4 ns/$\sqrt{50}$ = 1 ns. Of course in real manufacturing and laboratory conditions there may be problems causing correlations and improper performance of standards. The ensemble of 33 such cesium-beam frequency standards at USNO showed a predictability of 3 ns at 10 d, relative to the ensemble including them and 3 H-masers. The number of 3 ns is perhaps larger than it could be, since the clocks were combined as a simple average without regard to individual performance. Given a predictability of 3 ns for 33 standards, if we increase to 50 independent standards we find a predictability of ($\sqrt{33}/\sqrt{50}$)·3 = 2 ns at 10 d. For τ values longer than 10 d, we may conservatively assuming both a 10 d predictability of 2 ns and a model of flicker noise FM past 10 d. For this model the dispersion grows as time, and under the assumption of optimum prediction outlined herein, we would have 7.2 ns for a one-month prediction interval from these commercial cesium-beam frequency standards without the masers.

From the actual clocks analyzed herein, we have shown that a time scale can be created using three to five active hydrogen masers and/or laboratory cesium-beam frequency standards. Against such time scales the predictability of several masers taken together reached 1 ns for a 10 d prediction interval. We also saw that the noise type of many such systems was flicker noise FM at under $3 \cdot 10^{-15}$ out to one month or longer. Since the time dispersion due to flicker noise FM is proportional to τ, we can forecast a time dispersion at one month for a scale based on an ensemble of similar clocks. Shortly, there may be 25 such cavity-tuned masers in the world available for use in an ensemble approximately a factor of 5 greater than what we use in our study here to obtain 1 ns predictability at 10 d. If they are properly linked, an ensemble of independent standards with drift modeled perfectly should be $\sqrt{5}$ times better than what we have demonstrated here. Using a model of flicker noise FM from 10 d to 30 d, we find a one-month predictability of

$$\frac{3}{\sqrt{5 \cdot \ln(2)}} = 1.6\, ns.$$

If drift can be estimated to $5 \cdot 10^{-17}$/d as indicated in Table I, this could add a time dispersion of 2.2 ns for each maser. Conservatively, it is not unreasonable

to expect a one month rms prediction error of less than 3 ns for such an ensemble. For UTC to be that predictable it would need to have a flicker floor at about $1 \cdot 10^{-15}$. This is not currently available. If standards can be predicted below the 3 ns level at one month, it seems reasonable to expect that UTC will improve as it uses these standards, commensurate with their predictability.

While clocks are being improved, methods of time comparison need to be improved as well. It will take time to reach the goals demonstrated in this paper, but the prospects seem very good.

At the present time, practical consideration such as time steps and frequency steps in clocks, and uncertainty in estimating frequency drift limit predictability of UTC to significantly more than 1 ns. In addition, UTC is still only available with about a two-month delay. The noise in the two-month region might be as dispersive as random-walk FM. This would cause the time prediction error to grow as $\tau^{3/2}$. This might bring us up to about the 10 ns region at the present time using the best clocks available. This is still about an order of magnitude better than the current performance of most of the UTC(k)s.

ACKNOWLEDGEMENTS

The authors would particularly like to thank Nichol Koshelyaevshy for supplying the hydrogen-maser data from VNIIFTRI, Andreas Bauch for supplying the hydrogen-maser and primary cesium-beam clock data from the laboratory standards at the PTB, and Gernot Winkler for supplying the hydrogen-maser and commercial primary cesium-beam clock data from the USNO. Their data were essential to this paper. Several significant suggestions have been made by colleagues. In addition to the above mentioned contributors, these include Claudine Thomas of the BIPM and Donald Sullivan of NIST. The authors are also indebted to the readers for helpful comments. These include Prof. Neil Ashby of the University of Colorado as well as Fred Walls and Matt Young of NIST.

REFERENCES:

[1] Guinot, B., Thomas, C.," Establishment of International Atomic Time," *BIPM Annual Report*, 1988, Part D, pp. D1-D22.

[2] Allan, D. W., A. Lepek, "Trends in International Timing," *Proceedings of the Seventh European Time and Frequency Forum*, 16-18 March 1993, pp 221-230.

[3] *Characterization of Clocks and Oscillators,* Eds. D. B. Sullivan, D. W. Allan, D. A. Howe, and F. L. Walls, NIST Tech Note 1337, 1990.

[4] J. A. Barnes and D.W. Allan, "An Approach to the Prediction of Coordinated Universal Time," *Frequency,* pp. 3-8, November/December, 1967.

[5] D.W. Allan, "Time and Frequency (Time-Domain) Characterization, Estimation, and Prediction of Precision Clocks and Oscillators," in [3], pp. 121-128.

[6] M.A. Weiss, D.W. Allan and T.K. Peppler, "A Study of the NBS Time Scale Algorithm," *IEEE Transactions on Instrumentation and Measurement*, vol. 38, pp. 631-635, 1989.

[7] A. Gelb, ed., *Applied Optimal Estimation*, M.I.T. Press, Cambridge, MA, 1974.

[8] M.A. Weiss, D.W. Allan, and D.A. Howe, "Confidence on the Second Difference Estimation of Frequency Drift, A Study Based on Simulation," *Proceedings of the 1992 IEEE Frequency Control Symposium*, pp. 300-305, 1992.

[9] J. E. Gray, D. W. Allan, "A Method for Estimating the Frequency Stability of an Individual Oscillator," *Proc. 28th Symposium on Frequency Control*, May 1974, pp. 243- 246.

[10] M. A. Weiss, D. W. Allan, "An NBS Calibration Procedure for Providing Time and Frequency at a Remote Site by Weighting and Smoothing of GPS Common View Data," *IEEE Trans. on Instrum and Meas.*, vol. IM-36, no. 2, pp. 572-578, 1987.

[11] M. A. Weiss and T. Weissert, "Sifting Through Nine Years of NIST Clock Data WITH TA2," *Metrologia*, to be published, Spring 1994.

[12] M. A. Weiss and T. Weissert, TA2, A time scale algorithm for Post-Processing: AT1 plus frequency variance, *IEEE Trans. Instrum. Meas.*, 1991, **IM-40,** pp. 496-501.

[13] M. A. Weiss and T. Weissert, AT2, A new time scale algorithm: AT1 plus frequency variance, *Metrologia*, 1991, 28, 65-74.

[14] F. L. Walls, H. Hellwig, "A New Kind of Passively Operating Hydrogen Frequency Standard," *Proc. 30th Annual Symposium on Frequency Control*, pp. 473-480, Ft. Monmouth NJ, June 1976.

[15] F. L. Walls, D.A. Howe, "Timekeeping Potential Using Passive Hydrogen Masers," *J. Physique*, vol. 42, no.12, pp. C8-151 - C8-158, Dec 1981.

[16] F. L. Walls, "Frequency Standards Based on Atomic Hydrogen," *Proc. IEEE*, vol.74, no.1, pp. 142-146, Jan 1986.

[17] P. Tavella, J. Azoubib, C. Thomas "Study of the Clock-Ensemble Correlation in ALGOS Using Real Data," *Proc. 5th European Frequency and Time Forum,* Besançon, pp. 435-441, March 1991.

[18] M. A. Weiss, C. Hackman, "Confidence on the Three Point Estimator of Frequency Drift," *Proceedings of the 21st Annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting*, Nov. 1993, McLean, VA.

[19] P. Tavella and C. Thomas, "Report on Correlations in Frequency Changes Among Clocks Contributing to TAI," Report 91/4 of the Bureau International Des Poids et Mesures, July, 1991.

[20] D. W. Allan, W. Dewey, "Time-Domain Spectrum of GPS SA," *Proc. Institute of Navigation GPS Meeting*, Sept. 1994.

[21] D. Kirchner et al., "Two-way Satellite Time Transfer Beween USNO and TUG Using the Intelsat satellite at 307 Degrees East,"*Proc. 7th European Frequency and Time Forum,* Neuchâtel, pp. 135-140, March 1993.

[22] A. Brown, "Precise Time Dissemination Using the INMARSAT Geostationary Overlay," *Proc. Institute of Navigation GPS Meeting*, June 1993.

[23] A. Imaoka and M. Kihara, "Long Term Propagation Delay Characteristics of Telecommunication Lines," *IEEE Trans. Instrum. and Meas.*, vol. IM-41, no. 5, pp. 653-656, 1992.

[24] G. Lutes, A. Kirk, "Reference Frequency Transmission Over Optical Fiber," *Proceedings of the 18th Annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting*, Dec. 1986, Washington, D.C.

Table I

Standard	Lab	Estimated Linear Frequency Drift
CS1	PTB	$(-2.9 \pm 0.8)\cdot 10^{-16}$/d
H1	PTB	$(1.4 \pm 0.4)\cdot 10^{-16}$/d
H2	PTB	$(1.3 \pm 0.2)\cdot 10^{-16}$/d
H50	SU	$(-0.6 \pm 0.3)\cdot 10^{-16}$/d
H51	SU	$(0.7 \pm 0.3)\cdot 10^{-16}$/d
H53	SU	$(-0.2 \pm 0.3)\cdot 10^{-16}$/d
N4	USNO	$(1.0 \pm 0.2)\cdot 10^{-16}$/d

Table II

Measurement	RMS Uncorrelated Prediction Error at 10 d 95 % confidence limits:	lower	upper
For PTB, 21 points used from end of data, MJD's 48840 - 49050.			
CS1 - TA2PTB	8 ns	7	10
CS2 - TA2PTB	5 ns	4	6
H1 - TA2PTB	2 ns	2	2
H2 - TA2PTB	3 ns	2	4
For SU, 8 points used from end of data, MJD's 48915 - 48985			
H47 - TA2SU	2 ns	2	3
H49 - TA2SU	2 ns	2	3
H50 - TA2SU	2 ns	2	3
H51 - TA2SU	1 ns	1	2
H53 - TA2SU	2 ns	2	3
For TA2M1, 16 points used from end of data, MJD's 49171 - 49321			
N4 - TA2M1	2 ns	2	3
N5 - TA2M1	1 ns	1	1
33Cs - TA2M1	3 ns	2	4
NIST H1 - TA2M1	3 ns	2	4

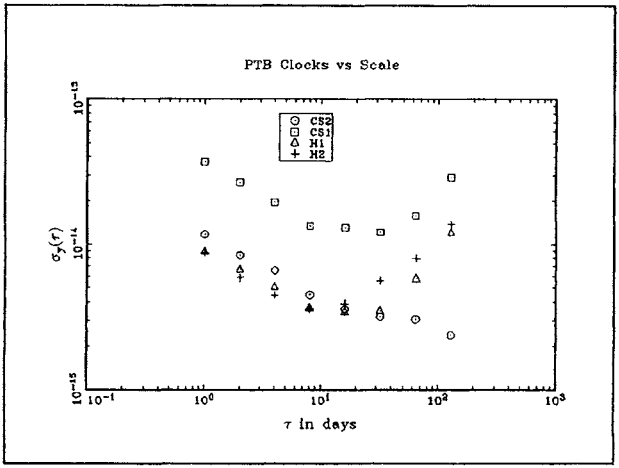


Figure 1a

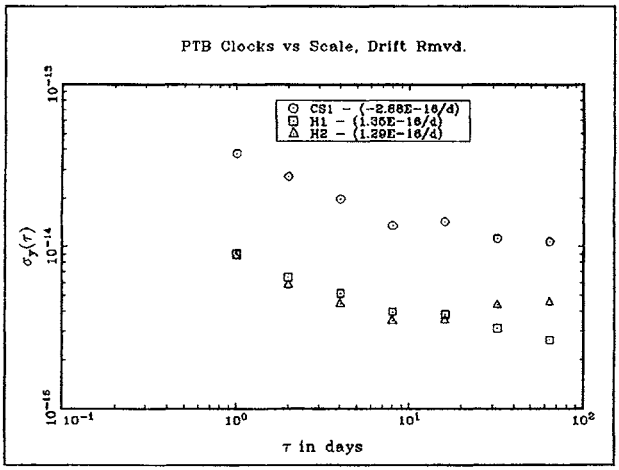


Figure 1b

Figures 1a and 1b use the data from the PTB primary cesium-beam frequency standards operating as clocks and two hydrogen masers. Figure 1a is a plot of the estimate of the unbiased instability, $\sigma_y(\tau)$, as inferred from measuring each clock against an ensemble -- then removing the bias resulting from membership of that clock as a member of the ensemble. Figure 1b is the same as figure 1a except a frequency drift has been subtracted from the data.

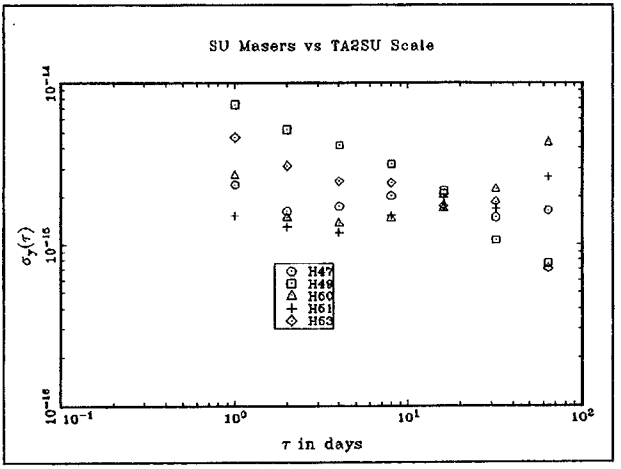


Figure 2a

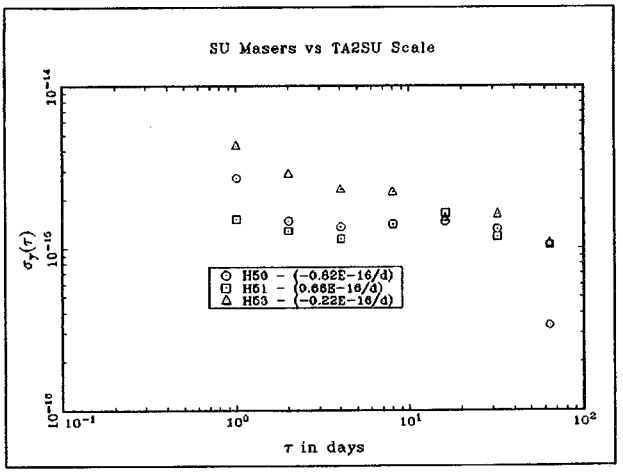


Figure 2b

Figures 2a and 2b use the data from clocks located at VNIIFTRI (the former Soviet Union reference time scale is still maintained for the different Unions). These are all hydrogen masers. Figure 2a is a plot of the estimate of the unbiased instability, $\sigma_y(\tau)$, as inferred from measuring each clock against an ensemble -- then removing the bias resulting from membership of that clock as a member of the ensemble. Figure 2b is the same as figure 2a except a frequency drift has been subtracted from three of the clocks which indicated some drift.

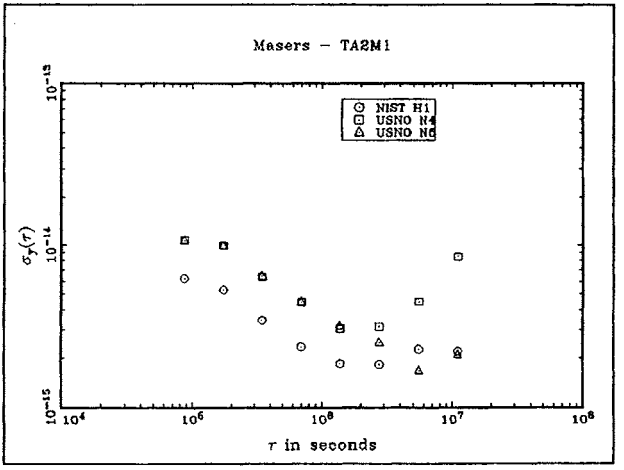


Figure 3a

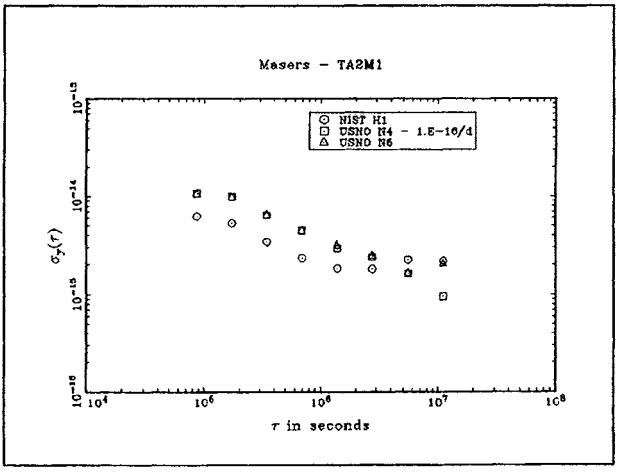


Figure 3b

Figures 3a and 3b use the data from clocks located at both USNO and NIST. The GPS common-view technique was used to compare the clocks at the two sites. Figure 3a is a plot of the estimate of the unbiased instability, $\sigma_y(\tau)$, as inferred from measuring each clock against an ensemble -- then removing the bias resulting from membership of that clock as a member of the ensemble. Figure 3b is the same as figure 3a except a frequency drift has been subtracted from the data for the N4 maser. For this data set the measurement noise contaminates the stability of the NIST maser.

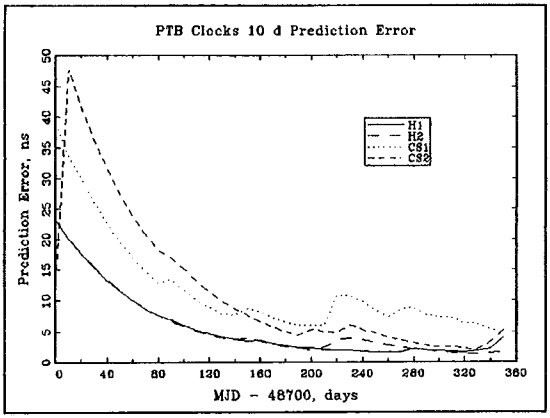


Figure 4

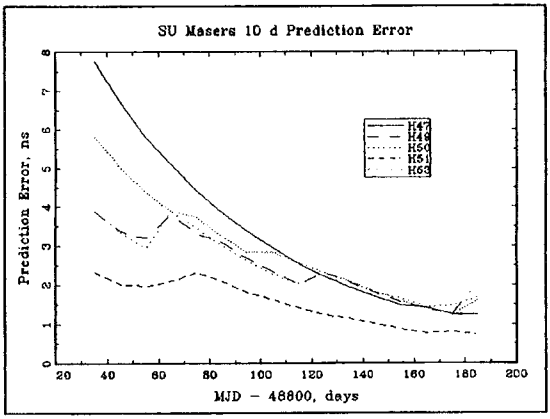


Figure 5

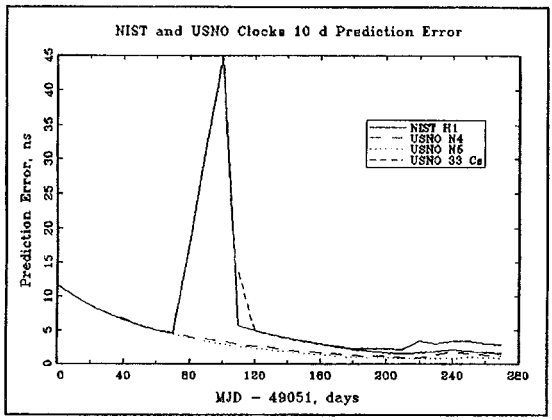


Figure 6

Figure 4 - 6 These three figures all estimate the 10 day rms prediction error as determined (internally) with respect to the TA2 time-scale algorithm. The figures correspond to the stability data shown in figures 1 - 3. Note, there is a turn on transient and occasional adverse perturbations in some of the clocks. Table I lists a root-mean-square of the asymptotic average for each of the clocks.