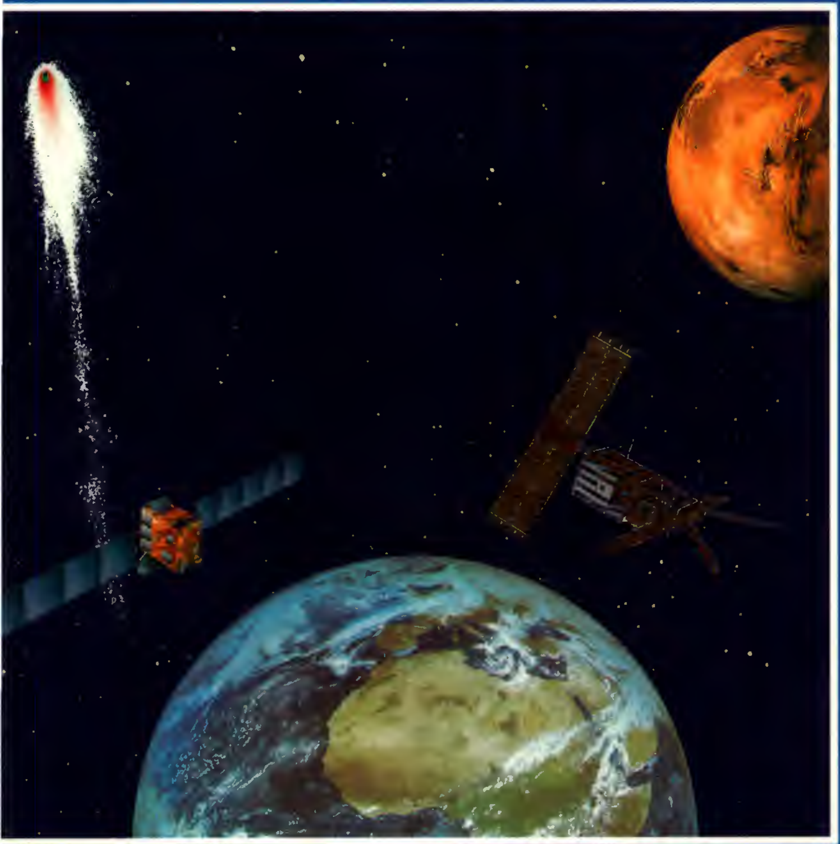
SP-423



The 32nd ESLAB Symposium on

Remote Sensing Methodology for Earth Observation and Planetary Exploration

ESTEC, Noordwijk, The Netherlands
15-18 September 1998

European Space Agency
Agence spatiale européenne

32nd ESLAB Symposium Organising Committee
Evert Attema, Clare Bingham, Gonnie Elfering, Roland Meynart, Cecilia Nilsson, Mike Rast, Doris Reinprecht, Gerhard Schwehm and Andrew Wilson

Publication	Proceedings of the 32nd ESLAB Symposium, 'Remote Sensing Methodology for Earth Observation and Planetary Observation', ESTEC, Noordwijk, The Netherlands (ESA SP-423, December 1998)
Edited by	E. Attema & G. Schwehm ESA Space Science Department & A. Wilson ESA Publications Division
Published and distributed by	ESA Publications Division ESTEC, Noordwijk, The Netherlands
Printed in	The Netherlands
Price	80 Dutch Guilders
ISBN	92-9092-692-9
Copyright	© 1998 European Space Agency

SP-423
December 1998

The 32nd ESLAB Symposium on

Remote Sensing Methodology for Earth Observation and Planetary Exploration

15-18 September 1998
ESTEC, Noordwijk, The Netherlands

European Space Agency
Agence spatiale européenne

Contents

Processing Technologies

Chairman: H. Svedhem

Welcome & Overview

INTRODUCTION

M.C.E. Huber

Head, Space Science Department, ESA/ESTEC, Noordwijk, The Netherlands

Dear Colleagues, Ladies and Gentlemen,

I would like to welcome you and to introduce to you ESA's Director of Science who will in a few minutes open this year's ESLAB Symposium, which deals with 'Remote Sensing Methodology for Earth Observation and Planetary Exploration'.

I hope that this 32nd ESLAB Symposium will provide a timely opportunity for participants from both the Earth Observation and Space Science communities to exchange experiences on how best to work in the new environment of faster, cheaper, smaller (but not necessarily better)! The Symposium will address the problems of future instrument development from a scientific and technological perspective, and will, I hope, make recommendations to the National Agencies and ESA to design strategies for an efficient instrument development programme that will meet the needs of remote sensing research in Earth Observation and Planetary Exploration.

Before I leave the podium I would like to explain two things about this Symposium: its name ESLAB and the function of the organising entity, the Space Science Department, within ESA.

In the first half of the 1960's, when European space research was organised through the predecessor of ESA, namely the European Space Research Organisation, ESRO, there was a laboratory called ESLAB, the European Space Research Laboratory. It was housed in a former hotel in Noordwijkerhout, not very far from here. Towards the end of the 1960's, ESLAB became the Space Science Department (SSD) and is now located here at ESTEC, the European Space Research and Technology Centre. SSD's old name survives in the annual ESLAB Symposia.

We are in the happy position of being able to choose an entirely different topic for each ESLAB Symposium. Sometimes we decide to look at a specific mission – be it during its scientific preparation, or, after flight, at its results. Another important aim is to look at the scientific technology for future missions – as we are indeed doing this year.

Now to the function of SSD within ESA. I have frequently noticed that the functions of this Department are not very well known outside ESA. I should therefore also explain these in a few words.

SSD is part of the Directorate of Scientific Programmes of the European Space Agency, and is often called the main interlocutor between ESA and the scientific community. Indeed, the Department's prime function is to provide Project Scientists; they ensure that ESA's missions fulfil their original scientific goals, and they work with the Principal Investigators (and their Experiment Teams) as well as with the Project Manager (who belongs to the other Department of our Directorate, the Scientific Projects Department) in the continuous process of optimising a mission, balancing scientific, technical and financial considerations. Another important activity in SSD is the science operations of observatory-type missions like, e.g. the Infrared Space Observatory, ISO, in the recent past or the X-ray Multi-mirror Mission, XMM, in the near future.

Space Science Department also provides a scientific environment in which Project Scientists (and, before they reach that stage, Study Scientists) can perform their own, original research. This, in our view, is mandatory, if Project Scientists are to stay in touch with the reality of the scientific process – the arduous step-by-step progress towards achieving new knowledge. To this end, SSD has not only an infrastructure for data analysis and computational modelling, we also maintain (here at ESTEC) the technical infrastructure for building flight hardware and, of course, implementing the associated software.

Ulysses, Geotail, WIND, SOHO, Bepposax, Polar, Equator and also Interball have experiments on-board that have been built (entirely or in part) in SSD –

Cluster II and Rosetta and, later, Planck will carry experiments with significant SSD contributions as well.

A recent addition to SSD is the Earth Sciences Division, which joined us when the exploratory (as opposed to the applications) part of Earth Observations was placed under the authority of ESA's Director of Science. Evert Attema of this Division is one of the main organisers of this year's ESLAB Symposium and he worked closely together with Gerhard Schwehm of the Solar System Division.

In closing, I wish you a fruitful and enjoyable Symposium and I dearly hope that in a few year's time, important new remote sensing developments can be reported with the remark: several ideas for the design of this instrument came out of the deliberations of the 32nd ESLAB Symposium.

I would now like to introduce to you, ESA's Director of Science, Prof. Roger Bonnet.

WELCOME ADDRESS

R.M. Bonnet

Director of Science, ESA

Welcome to the 32nd ESLAB Symposium and to ESTEC. It is a real pleasure to see scientists and engineers from institutes and industry coming together to exchange views on the important topic of 'Remote Sensing Methodology for Earth Observation and Planetary Exploration'. The Head of ESA's Space Science Department has already explained what ESLAB is and the history of the ESLAB Symposia and their general goals.

I can therefore concentrate on this year's topic, which deviates from the general themes of the past in that previously ESLAB Symposia addressed topics related to the scientific results of ESA's Science missions or addressed general scientific topics relevant to mission definition or to promoting synergy between different areas in space science.

This time ESA is trying something new and different.

You will discuss remote sensing methodology for Earth Observation and Planetary Science and this brings about two features that are important to me. Firstly, following the incorporation of the Earth Sciences Division into ESA's Space Science Department, this is the first opportunity in the Agency to demonstrate that the two communities do talk to each other and try to learn from each other. I am sure this will be of mutual benefit to both communities.

Dave Southwood will reflect on this in his keynote address – and he can approach it from two different points of view: as a prominent scientist active in solar system and planetary research and from his current position as being responsible for defining Earth Observation science strategies in my Directorate. I want to stress again that it is important that we look beyond our horizons to learn from others and to improve the way we do business.

Secondly – and this is at least equally important – one of the main goals of this symposium is to "discuss a strategy for pre-development of key instrumentation that will meet the requirements of the new, cost-effective approaches to mission implementation, and assure the competitiveness of European Research Institutions and Industry for participation in future international space enterprises".

I would like to expand briefly on this aspect as this has been occupying us for the past few years. In principle, there are many common features in remote sensing instrumentation for Earth Observation and Planetary Exploration, but the approach and the community appear to be worlds apart. Earth Observation missions at ESA usually carry large facility instruments, of course defined in consultation with the user community, but developed and funded by the Agency as part of an optional Programme. The data are calibrated, quality-controlled and provided to the science community by the Agency.

In general, for solar system and planetary research, we take the classical PI approach of experiments developed by one institute or a consortium of institutes and funded by institute resources and the national funding agencies. The instruments are much smaller in general even though they also look down at a planet. This provides great flexibility to the community to always propose a state-of-the-art instrument adapted to a specific mission and involves the Community much more directly with a 'hands on' approach.

We now need to cope with faster delivery schedules, requiring much more up-front effort and early development.

In the new ESA Earth Observation programme 'The Living Planet' which comprises science-driven missions, the Earth Explorer Missions, and operational application-driven missions, the Earth Watch Missions, the Agency has also to adopt new strategies: upper limits on cost and size are defined for scientific missions and for the operational applications the concept of co-funding by the private sector has been introduced.

The Study

Study

⇨ triggered by ESA Earthwatch initiative, but not limited to it

⇨ during study scope has been extended to a more generic assessment of “geospatial and environmental” information needs which can be met by EO

Objectives

⇨ identify EO information needs based on EU Policies (time horizon 2005-2010)

⇨ identify potential role/s of the Commission - in relation to other partners, i.e. ESA and industry in Member States

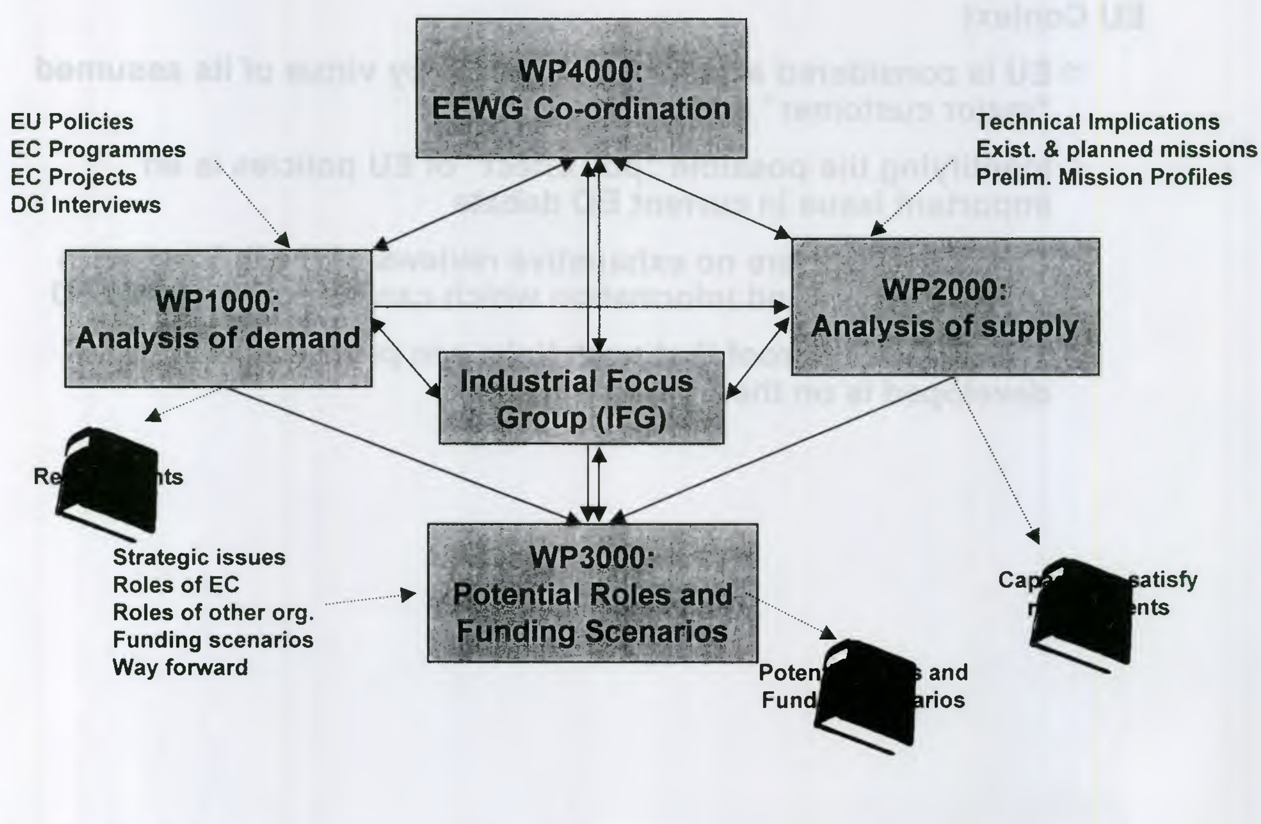
Implementation

⇨ joint initiative of DGs JRC and XII, led by CEO Unit

⇨ led by EC staff, with experts from industry and Member States

⇨ work schedule: Jan 98 to Jun 98

Overview of Study Activities



Key Deliverables

D1 - EU Geospatial and Environmental Information Needs: Policy Assessment

D2 - EU Geospatial and Environmental Information Needs: EO Contribution

D3 - EU Geospatial and Environmental Information Needs: Potential Involvement of the EC in EO Missions

D4 - EU Geospatial and Environmental Information Needs: Study Overview

---

MG1 - Earthwatch: Plan for the Participation of the EC

MG2 - EEWG Workplan

Industrial Focus Group (IFG)

Mandate & Tasks

- ⇨ role was advisory, the IFG will not undertake new work
- ⇨ "Reality check" of EEWG work
- ⇨ Early involvement of industry = expected future partner in Earthwatch missions

Composition

- ⇨ Members: Eurospace (6), EARSC (3)
- ⇨ Observers: DGIII, ESA, EUMETSAT, EEWG
- ⇨ Chairman: Roy Gibson

4 Key Meetings

- ⇨ Feb (preparatory), Mar, Apr, Jun 1998

Summary of Work

- A total of 16 policies/themes addressed
- Approx. 50 DG staff interviewed, in 25 Units within 15 DGs
- Thorough policy review performed and potential for geospatial and environmental information needs identified
- 5 Cases developed as starting point

- Parameters defined for 5 Cases (type of geo-physical information, resolution, coverage, repeat time, timeliness)
- Horizontal analysis of five cases performed

- Potential Roles of EC discussed, including 5 Implementation Scenarios

Policy Review

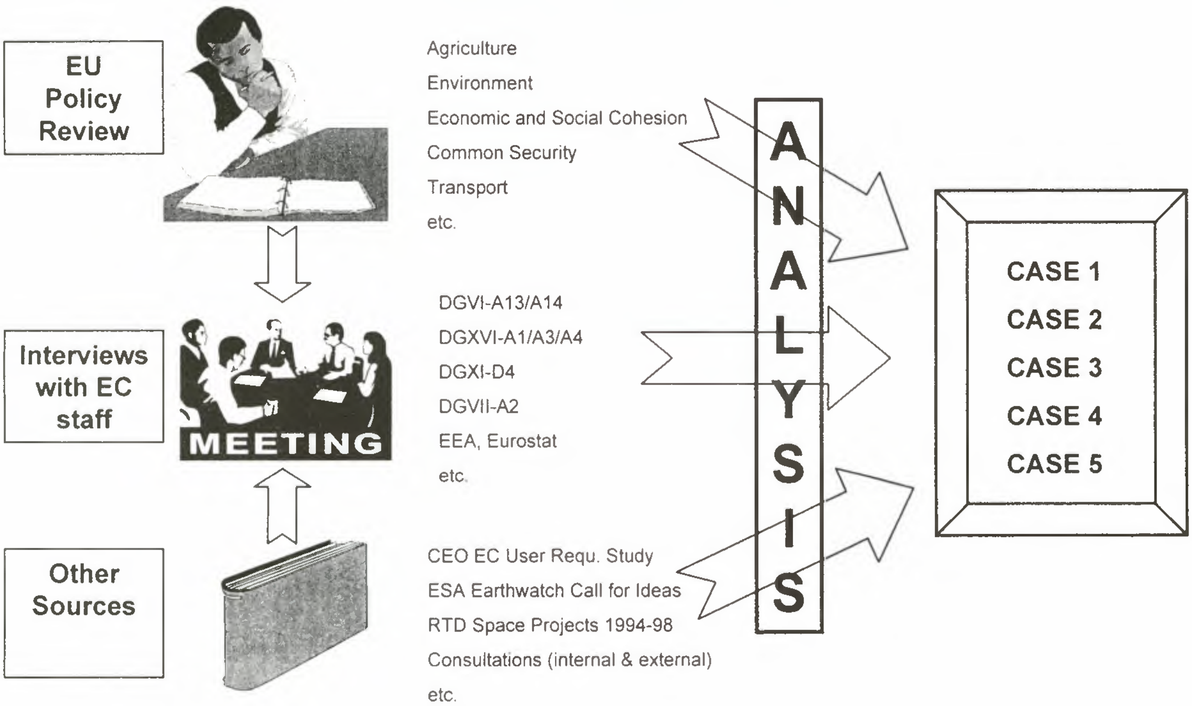
Policies or Themes addressed

⇨ Agriculture & Forestry
⇨ Regional Development incl. Economic & Social Cohesion
⇨ Environment
⇨ Development
⇨ Humanitarian Aid
⇨ Transport incl. TENs
⇨ Fisheries
⇨ Research and Technology
⇨ Tourism
⇨ Risks & Natural Hazards
⇨ Industrial Policy

⇨ Horizontal Activities: EEA
⇨ Horizontal Activities: Eurostat

⇨ Common Foreign & Security Policy
⇨ Enlargement of the EU

Development of CASES



The Five CASES

5 CASES developed - as a starting point

- ⇨ **Case 1: Regional development & cohesion prioritising spatial planning**
- ⇨ **Case 2: Land resources management prioritising agriculture**
- ⇨ **Case 3: Environmental conventions**
- ⇨ **Case 4: Environmental indicators**
- ⇨ **Case 5: Security including risks and hazards**

  NOTE: Needs under five Cases to some extent overlap, and Cases should not be regarded as mutually exclusive

Horizontal- strategic - issues

- ⇨ **Expansion of the EU**
- ⇨ **Need for harmonised (pan-European) and continuous (long-term) information**

  Need for independent data source

CASE 1 : Regional Development & Cohesion

Policy Basis

- ⇨ Single European Act of 1986: "Economic and Social Cohesion"
- ⇨ Regional Development and Cohesion - European Spatial Development Perspective (ESDP)
- ⇨ Transport - TENs

Key Issues

- ⇨ Single Market: social and economic integration
- ⇨ Preparation for expansion of EU to the East
- ⇨ Spatial information needed for strategic planning

Primary DG

- ⇨ DGXVI, (DGVII for transport)

CASE 2: Land Resources Management

Policy Basis

- ⇨ focus on Common Agricultural Policy (CAP): Art. 38-47 of Treaty
- ⇨ also other topics supported: forestry, etc.

Key Issues

- ⇨ CAP reform: focus on agro-environment and rural development
- ⇨ Crop production estimates and fraud control
- ⇨ Need for a global and independent agriculture monitoring system (EU expansion, foreign crop production)

Primary DG

- ⇨ DGVI, (DGXVI)

CASE 3: Environmental Conventions

Policy Basis

- ⇨ EU commitments in international conventions, i.e. Biodiversity, Climate Change (post-Kyoto initiatives)
- ⇨ Environmental Policy

Key Issues

- ⇨ Definition, monitoring and verification of international treaties and agreements
- ⇨ Example: monitoring of "sinks" within Climate Change Convention (post-Kyoto, preparation for COP-4, Buenos Aires)

Primary DG

- ⇨ DG XI

CASE 4: Environmental Indicators

Policy Basis

- ⇨ Environmental Policy
- ⇨ Support to other policies (CAP, rural development, transport, energy, etc.)

Key Issues

- ⇨ Monitoring of environmental pressure indicators
- ⇨ Need for tools to measure progress towards long-term sustainability
- ⇨ Support EEA and Eurostat initiatives

Primary DG

- ⇨ DGXI, (VI, VII, XVI, etc.)

CASE 5: Security

Policy Basis

- ⇨ Common Foreign and Security policy - Maastricht Treaty (third pillar), updated by Amsterdam Treaty
- ⇨ Risks & Hazards - various policy statements

Key Issues

- ⇨ Assuring safety of European citizens - monitoring of nuclear installations, civil protection, monitoring of European border-lines, etc.
- ⇨ Preparing for EU expansion to the East

Primary DG

DG IA, DGXI (for risks & hazards), others

Conclusions

The output from the study is

- ⇨ Initial objective of the study - to produce an "assessment"
- ⇨ the study set into motion a "process" whereby role of information services is being revisited with particular reference to EO
- ⇨ this process must be continued in order to reinforce the base on which future EO missions can be built
- ⇨ the study has shown that there is a need to further develop the interface role between demand and supply

Long-term vision

- ⇨ to progress towards the establishment of an operational and sustainable EO capability for Europe to serve its strategic and policy needs

Passive Microwave Technologies

MICROWAVE LIMB SOUNDERS

K. F. Künzi

Institute of Environmental Physics, University of Bremen
P.O. Box 33 04 40, D-28334, Bremen, Germany
Phone: +49 (0)421 218-3909
Fax: +49 (0)421 218-4555
e-mail: kunzi@physik.uni-bremen.de

ABSTRACT

With the increasing awareness for environmental issues in the seventies the need for accurate and reliable methods to investigate and monitor the atmosphere, ocean, and land became a more and more pressing issue. A particularly important question was, and still is, to better understand the processes leading to the destruction of the stratospheric ozone layer. In order to assess this danger a number of dedicated sensors have been proposed. A very successful technique is offered by spaceborne microwave radiometric limb sounding. Microwave limb sounders measure thermal emission from pressure broadened rotational lines of interesting molecules. Of particular importance are ozone and the key constituent in the anthropogenic ozone destruction cycle ClO. Microwave limb sounders have the advantage of being unaffected by aerosols (e.g. polar stratospheric clouds), and by measuring in emission the sensor can operate day and night.

Key words: remote sensing; atmospheric chemistry; limb sounders, mm- and submm waves.

1. INTRODUCTION

One of the key issues in manmade changes of our environment concerns changes in atmospheric composition resulting in changes of the delicate balance of atmospheric chemistry (e.g., "ozone hole"). Our limited knowledge of these important processes made it necessary to improve our inventory of measuring tools to learn more on the state and on the changes in this part of our environment.

Microwave radiometers have a number of advantages for monitoring the composition of the atmosphere. Since microwave sensors measure thermal emission they do not depend on solar illumination, and the rather large wavelength (> 1 mm for frequencies of < 300 GHz) involved, in comparison to infrared sensors, make these instruments virtually independent of aerosols, and also much less affected by clouds (Cirrus) than sensors operating at a much shorter wavelength.

Many key species of importance in atmospheric chemistry show strong and isolated resonances in the millimetre and sub-millimetre wavelength region making this band very attractive to simultaneously measure families of chemically active constituents related to catalytic ozone destruction cycles. On the other hand molecules not having rotational resonances (e.g. CO_2) can not be detected.

2. PRESENT STATUS OF MICROWAVE SENSORS FOR ATMOSPHERIC CHEMISTRY

In recent years microwave limb sounders have been used successfully to investigate the middle atmosphere, and in particular to perform observations in relation to anthropogenic ozone destruction. Two instruments have been deployed so far, the *MLS* (Microwave Limb Sounder) on *UARS* (Upper Atmosphere Research Satellite), and *MAS* (Microwave Atmospheric Sounder) on *ATLAS* (Atmospheric Laboratory for Applications and Science). The *MLS* launched in 1991 is still operational on a limited basis and *MAS* has flown on three *ATLAS/Shuttle* missions in 1992, 1993 and 1994. Both instruments have provided very interesting results and collected Data not available by any other Sensor.

A review of *MLS* can be found in Waters et al. (1996) and for *MAS* in Hartmann et al. (1996), furthermore a

Proc. 32nd ESLAB Symp., 'Remote Sensing Methodology for Earth Observation and Planetary Exploration', ESTEC, Noordwijk, The Netherlands, 15-18 September 1998 (ESA SP-423, December 1998)

large amount of papers making use of *MLS* or *MAS* data is available in the open literature. Listings can be found in the corresponding home pages (for details see the listing of home pages given at the end).

In this paper major emphasis is placed on new developments and in particular on new microwave limb sounders presently under construction.

3. TECHNIQUE OF OPERATION

An excellent review of microwave remote sensing techniques and their application in atmospheric research is given in the book edited by M.A. Janssen (1993).

The superheterodyne detection scheme used by all microwave radiometers provides excellent frequency resolution which permits to resolve pressure broadened emission lines. The limb sounding instrument will scan the atmospheric limb with a sufficiently narrow antenna beam (typically a few tenths of a degree). Figure 1 shows as an example the situation for a limb sounder on a polar orbiter scanning along track.

Earlier limb sounders such as *MLS* and *MAS* have been launched on platforms with a rather low inclination of roughly 60°. In order to cover also higher geographic latitudes the instruments observed perpendicular to the velocity vector. This results in a asymmetric global coverage of approximately 40° in one hemisphere and 80° in the opposite hemisphere, consequently for observing both polar regions, in the North and the South the platform has to be rotated in regular intervals. Furthermore the observing geometry used for *MLS* and *MAS* does not allow to measure true vertical profiles because the tangent points will change along the line of sight for different altitudes. Some new instruments to be discussed later do not use this geometry anymore but will observe along track (see also figure 1).

Vertical volume mixing ratio (VMR) profiles are retrieved from the measurements performed during one limb scan, for weak lines several limb scans need to be averaged resulting in a smearing in time and/or space. From radiative transfer calculations it can be shown that the vertical resolution can not be better than 1 or 2 kilometres, and based on this vertical resolution, the horizontal resolution along the line of sight is approximately 200 km. Since the typical groundspeed of a satellite at altitudes below 1000 km is of the order of 8 km/h and assuming the along track integration interval needs not to be better than the along of sight resolution, the integration time available for one scan is approximately 25 seconds. This time also imposes some requirements on acceptable receiver noise for retrieving single profiles of weak lines.

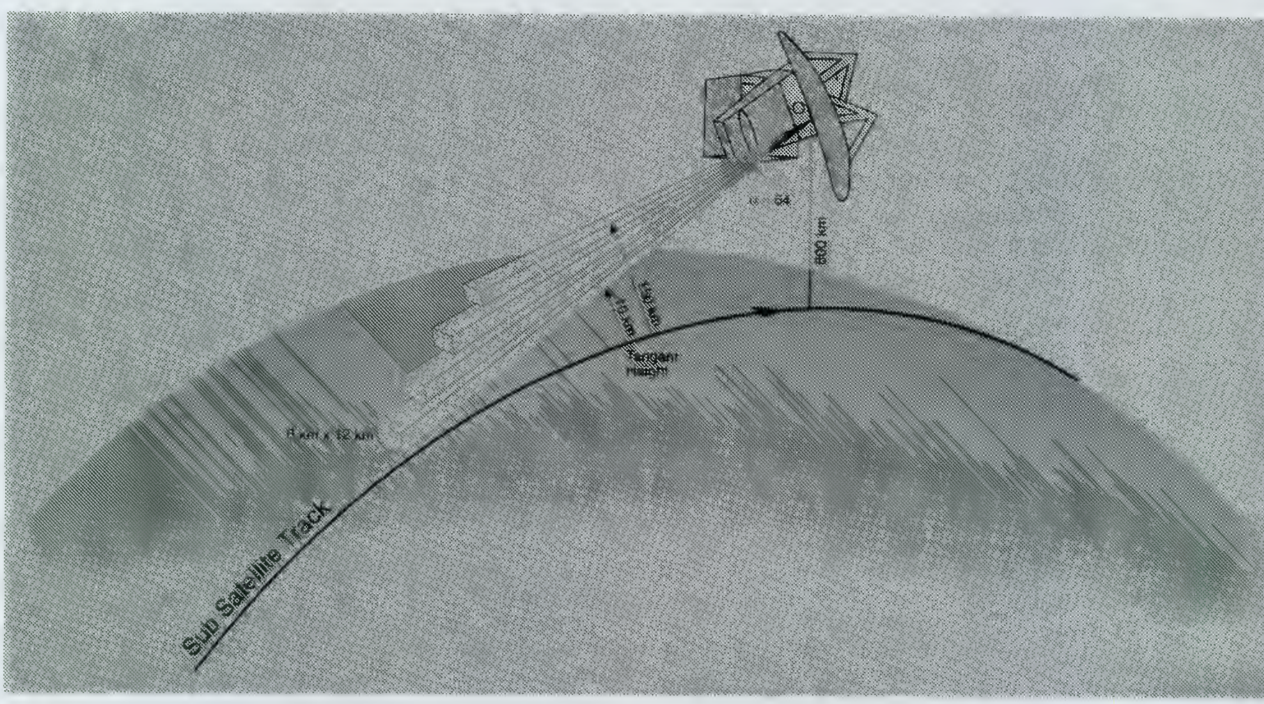


Figure 1. Viewing geometry for limb sounding. The observation is along track, earlier sensors on low inclination platforms (UARS and Shuttle) are observing in a cross track geometry.

Below 10 km of altitude the atmosphere becomes opaque and above 90 km the lines tend to be too weak for observation. Therefore profiles are obtained over an altitude region of 10 to 90 km with a vertical resolution of ≈2 km and a horizontal spatial resolution (pixel size) of up to 200•200 km^2 depending on the observing geometry. This rather coarse horizontal resolution limits the applications of a spaceborne limb sounder to the observation of a more synoptic nature, e.g. some small scale features of importance in stratospheric chemistry such as filaments at the polar vortex edge or lee wave phenomena can not be observed. It should be noted that such limitations are inherent to all limb sounders irrespective of wavelength or observing mode (emission or occultation).

For the wavelength range covered by present microwave instruments, and for temperatures found in the atmosphere, the emitted power is nearly proportional to the physical temperature (Rayleigh-Jeans approximation of Planck's Law), this results in a limited dynamic range required for the sensor, and allows very accurate and stable calibration using black bodies at different physical temperatures. Furthermore accurate knowledge of physical temperature is less important than for instruments operating near the peak of the Planck curve (10 μ). The physical temperature of the atmosphere can directly be obtained from measuring the emission of a uniformly mixed gas such as oxygen. The retrieval process also requires knowledge of atmospheric pressure at the observing altitude. This information is easily available from observing the emission of a uniformly mixed constituent. In a recent study performed for ESA/ESTEC it has been shown that temperature can also be retrieved accurately from emission lines of non uniformly mixed gases such as ozone (for details see ESA/ESTEC references at the end of the paper, Study on upper Troposphere/Lower Stratosphere Sounding").

For retrieving the VMR profiles from limb scans several techniques are available, however the technique most successfully used is the optimal estimation method (OEM) as described by Rodgers (1990). The OEM is a least square technique using a priori information for stabilising the retrieval. In the case of true Gaussian noise it can be shown that the retrieved profile is actually the best estimate. Again more details on retrieval and extensive sensitivity analysis can be found in the ESA/ESTEC studies listed in the reference section.

On the other hand a major disadvantage of microwave limb sounders are the rather large optics (antenna) required to obtain a sufficiently small vertical resolution at the tangent point. For spaceborne sensors such as MLS and MAS this results in antenna dimensions of 2 and 1 meter respectively. However with the increasing use of channels operating in the sub-millimetre range antenna dimensions can be smaller due to the shorter wavelength.

4. NEW DEVELOPMENTS

With the rapid improvements made in sensor technology at sub-millimetre wavelengths it is now feasible to extend the application of microwave radiometry to shorter wavelengths. Taking advantage of this development a new generation of microwave limb sounders is being prepared all extending the measuring capability into the submm wavelength range. The reason for this interest in these short wavelengths is of course the fact that many molecules of importance in photochemistry show emission lines only at these wavelengths (e.g. HCl and OH) or exhibit the largest features in this range (e.g. ClO). A list of important bands is given in some recent ESA/ESTEC studies conducted to prepare new instruments for future missions (for details see ESA/ESTEC references at the end of the paper). The new sensors under consideration at ESA/ESTEC are *MASTER* to measure in the upper troposphere and lower stratosphere, *SOPRANO* an instrument dedicated to middle atmosphere composition measurements, and *PIRAMHYD* operating in the terahertz range to measure the important minor constituent OH for which no suitable emission line at a lower frequency is available. Some details for these three instruments can be found in Table 1. As mentioned already for all three sensors scientific studies have been conducted or are in progress, and critical hardware elements have already been developed or are being developed on a breadboard level.

All microwave limb sounders presently under development, with some exceptions for bands below 300 GHz, cover a subset out of the bands listed in table 1. The first instrument of this new generation expected to be in orbit is the Swedish satellite *ODIN* (launch planned for 1999), developed in co-operation with Finland, France and Canada. Next the Japanese *JEM/SMILES* will become operational on the international space station in 2002, and scheduled for 2003 is the follow on instrument to MLS, the EOS-MLS. In the next paragraphs these three new microwave limb sounders will be briefly characterised.

ODIN: The *ODIN* will be the first sub-mm wave radiometer in space covering the frequency range 486.1-503.9 and 541.0-580.4 GHz, supplemented by one lower frequency channel 118-119 GHz. The sub-mm receiver is using an uncooled Schottky diode mixer, and the 118-119 GHz instrument has a low noise HEMT preamplifier. The *ODIN* concept combines two scientific applications in astronomy (star formation and early solar system), and aeronomy (measuring upper atmospheric composition) to get a better understanding of anthropogenic ozone destruction processes. Many important constituents in stratospheric chemistry can be observed with *ODIN* (e.g. O_3, N_2O, HNO_3, and ClO). However the very important species HCl in stratospheric chemistry is not covered.

JEM/SMILES (Superconducting Sub-millimetre Wave Limb Emission Sounder on Space Station, launch expected 2002)*:* The *JEM/SMILES* instrument will have two single side band sub-mm channels for the frequency band 624.2-628.6 GHz and 649.3-653.1 GHz for measuring key constituents in anthropogenic ozone destruction (e.g. O_3, ,ClO, HCl, HOCl, BrO, H_2O, HO_2, H_2O_2, N2O, NO, HNO3, SO_2, and $O^{18}O$ for temperature and pressure). At present the addition of a channel covering 318-326 GHz is being investigated extending the observing capabilities down into the upper troposphere (down to 6-8 km) for detecting O_3 and H_2O, and to obtain Cirrus cloud information. The major new technology to be employed for the first time in a space instrument are superconducting detector elements (SIS, super-conductor-insulator-superconductor) operating at 4 K for the higher frequency channels. The reduction in noise temperature of a SIS receiver over a conventional uncooled Schottky mixer is approximately a factor of 10. This improvement results in a reduction of measuring time (integration) of a factor 100. Therefore the *JEM/SMILES* instrument can measure single profiles of many constituents where a conventional, uncooled receiver can only provide averages over many days or zonal means. An additional advantage of SIS receivers are the very low power levels needed for the local oscillator (LO). Sufficient LO power is a major problem for uncooled Schottky mixers at shorter sub-mm wavelength.

SIS receivers operating also in the 600 to 700 GHz band have been used extensively on Aircraft during a number of large campaigns to study processes in the middle atmosphere. A description of the receiver can be found in Mees et al. (1995).

EOS-MLS: The *EOS-MLS* (launch December 2002) is the follow on instrument to the very successful *MLS*. This new sensor is very much improved, in addition to the lower frequency channels with LO-frequencies at 126.8000 GHz, 191.9000 GHz and 239.6600 GHz, additional double sideband sub-mm channels at 642.8700 GHz. and 2522.7816 GHz will be available, however without the low noise offered by SIS receivers. Therefore some weaker constituents will only be measurable with rather coarse spatial and/or temporal resolution. Main target molecules are: O_2, H_2O, N_2O, HNO_3, O_3, ^{18}OO, CO, ClO, HOCl, BrO, HCl, HO_2, and OH. Again the sensor will also cover the upper troposphere and provide Cirrus cloud information. A major step in new technology in the *EOS-MLS* project is the receiver at approximately 2.5 THz to measure OH, the only highly important constituent in atmosphere chemistry not yet accessible by spaceborne sensors. The technique favoured for this channel is a Schottky diode mixer pumped by a laser LO consisting of a methanol laser which in turn is pumped by a CO_2 laser.

This technique has been used successfully in an airborne system to measure water vapor and OH profiles in the stratosphere and mesosphere (see Titz et al., 1995).

5. CONCLUSIONS

The technology of microwave radiometry is now mature and well tested up to frequencies of 1.0 THz (Corresponding to wavelengths >0.3 mm). Complete receivers using uncooled detectors are available in space qualified form. For higher frequencies up to 3 THz as planned for *EOS-MLS* some difficult problems with limited sensor sensitivity and reliability of the local oscillator source have still to be solved. Also the newly developed superconducting detectors for *JEM/SMILES* need to demonstrate dependability in space, in particular the closed cycle refrigerator is a critical item. However these new instruments will eventually make extremely important contributions to be used in a number of different fields such as meteorology, atmospheric chemistry, and climate research.

However with all the new and exciting instruments to be flown in space it is equally important to pay attention to groundbased and airborne instruments because these sensors form the basis of long term monitoring programs, and will be indispensable for validating spaceborne sensors. Furthermore in order to study physical and/or chemical atmospheric processes (e.g. in campaigns), the flexibility offered by airborne sensors together with a higher spatial and temporal resolution makes these platforms a necessary element in all future and existing research programs.

It is important to keep in mind that only a well balanced set of spaceborne, airborne and groundbased sensors will provide the necessary information for improving our understanding of our environment, and to correctly assess existing or future dangers to our environment. Furthermore future spaceborne sensors will very much be based on experience collected by groundbased and in particular airborne instruments.

Table 1. Characteristics of microwave limbsounders under consideration at ESA/ESTEC with key species and parameters to be measured. Additional information can be found in study reports (see references).

MASTER: To measure composition of upper troposphere and lower stratosphere

Band	A	199-207 GHz	N_2O, H_2O^{18}
Band	B	296-306 GHz	O_3, N_2O, ^{18}OO (Temperature and Pressure)
Band	C	318-326 GHz	H_2O
Band	D	342-348 GHz	CO, HNO_3
Band	E	498-505 GHz	BrO, ClO, O_3

SOPRANO: To measure composition of the stratosphere and lower mesosphere

Band	A	497.5-504.75 GHz	BrO, O_3, ClO, CH_3Cl
Band	B1	624.5-626.6 GHz	HCl
Band	B2	628.2-628.7 GHz	HOCl
Band	C	635.5-637.5 GHz	CH_3Cl
Band	D	730.5-732.0 GHz	^{18}OO (Temperature and Pressure)
Band	E	851.3-852.8 GHz	NO
Band	F	952.0-955.0 GHz	NO, ^{18}OO (Temperature and Pressure)

PIRAMHYD: To measure OH

Band	A	2.5138-2.5148 THz	OH
Band	B	3.5507-3.5517 THz	OH

REFERENCES

Hartmann, G.H. et al.., Special Section : ATLAS, Geophysical Research Letters, Vol. 23, No. 17, pp. 2285-2411, 1996

Janssen, M.A., Editor , Atmospheric remote sensing by microwave radiometry, John Wiley & Sons, Inc, 1993.

Mees, J., S. Crewell, H. Nett, G. de Lange, H. van der Stadt, J.J. Kuipers, R.A. Panhuuyzen, ASUR - An Airborne SIS Receiver for Atmospheric Measurements of Trace Gases at 625 to 760 GHz, IEEE Transactions on Microwave Theory and Techniques, Vol. 43, No. 11, 1995

Rogers, C.V., Characterisation and error analysis of profiles retrieved from remote sounding measurements, Journal of Geophysical Research, Vol. 95, pp. 5587-5595, 1990

Titz, R., M. Birk, D. Hausamann, R. Nitsche, F. Schreier, J. Urban, H. Küllmann, H. P. Röser, Observation of stratospheric OH at 2.5 THz with an airborne heterodyne system, Infrared Physics and Technology, Vol. 36, pp. 883-891, 1995

Waters, J.W., Atmospheric measurements by the MLS experiments: results from UARS and plans for the future, Adv. Space Research, No. 10, pp 1363-1372, 1998-09-08

ESA/ESTEC STUDIES RELATED TO MICROWAVE LIMB SOUNDERS

Kraft, U., K. Künzi, H. Nett, P. Zimmermann, B. Albinson, Development of Limb-Sounder Critical Receiver Technologies for Remote Sensing of the Atmosphere, ESTEC/CONTRACT No. 8742/90/NL/PB (SC), 1992

Crewell, S., K. Künzi, J. Mees, Th. de Graauw, Spectroscopic Measurements in the Atmosphere at sub-mm Wavelengths, ESTEC/CONTRACT No 10655/93/NL/SF, 1994

Bühler, S., V. Eyring, H. Küllmann, The impact of Continuum Emission in the mm and Sub-mm Spectral Range, ESTEC/CONTRACT No. 10998/94/NL/CN, 1996

Bauer, A., J.M. Colmont, D. Priem, G. Wlodarczak, M. Birk, G. Wagner, S. Bühler, A. von Engeln, K. Künzi, A. Perrin, Study on a spectroscopic database for millimeter and submillimeter wavelengths, ESTEC/CONTRACT No. 11581/95/NL/CN, 1997

Goede, A., W. Bouwman, F. Helmich, M. Birk, D. Hausamann, F. Schreier, B. Schimpf, B. Carli, P. Ciarpallini, M. Carlotti, E. Venuti, B. Dinelli, B. Kerridge, J. Reburn, R. Siddans, K. Muirhead, PIRAMHYD retrieval study: OH profiling by far-Infrared limb sounding, ESTEC/CONTRACT No. 11719/95/NL/CN, 1997

In Progress are the following two final study reports (completion expected in 1998), "Study on upper Troposphere/Lower Stratosphere Sounding", ESTEC/CONTRACT No. 12053/97/NL/CN, and "Retrieval of Data from Sub-millimetre Limb Sounding", ESTEC/CONTRACT No. 119797/97/NL/CN

LISTING OF HOME PAGES RELATED TO MICROWAVE LIMB SOUNDERS

MLS : http://mls.jpl.nasa.gov

MAS: http://www-iup.physik.uni-bremen.de

ODIN: http://www-ssc.se/ssd/ssat/odin.html

JEM/SMILES: http://www.crl.go.jp/ck/ck321/smiles

EOS/MLS: http://eos-chem.gsfc.nasa.gov/mls.html

SOLAR SYSTEM RESEARCH WITH MICROWAVES

Paul Hartogh

Max-Planck-Institut für Aeronomie, Max-Planck-Str. 2, 37191 Katlenburg-Lindau, Germany, Phone +49-5556-979342
Telefax +49-5556-979240, e-mail: paul@linmpi.mpg.de

ABSTRACT

Microwave remote sensing offers unique features in the detection of physical and chemical parameters of surfaces and atmospheres of planets (including the Earth) , comets and asteroids. The high resolution of the heterodyne techniques allows the accurate analysis of the spectral shapes of the molecular emissions, from which the distribution of the molecules in the atmospheres as well as temperatures and wind velocities can be retrieved. The development of microwave technologies has made big progress in the last years. While mass, volume and power consumption could be reduced, the sensitivity and spectral coverage of receivers on the one hand and of spectrometer backends on the other hand could be improved considerably. However, there is still a lot of work to be done. We have to distinguish between the consequent implementation of existing technologies for space missions in the near future and basic research in instrumental development to extend the capabilities of the sensors near to the physical limits. A broad application of microwave instruments can be expected including highly integrated units for micro-satellites and ground-based extraterrestrial measurements, limb and nadir sounding instruments for small satellites and very complex large scale facilities.

Key words: microwave spectrometer, solar system

1. INTRODUCTION

There are a large number of scientific objectives concerning the solar system, which can be addressed with passive microwave instrumentation. An important objective of this ESLAB Symposium is to find out the interrelation between techniques being used for remote sensing of the Earth and other objects in the Solar System. Surely this interrelation is not only restricted to the technique. The similarities and differences in the chemistry, dynamics and circulations of different planetary atmospheres provide a context in which to understand the behaviour of Earth's atmosphere. The Mars atmosphere is so to say a second terrestrial atmosphere and climate system in that the general circulation of both planets are very similar. Much of the numerical modelling framework developed to study Earth's atmosphere can be applied in a relatively straightforward way to Mars including data assimilation techniques. Applications of these models and techniques to Mars can provide true test of climate parameterisations because Earth and Mars differ in certain interesting ways. Atmospheric data of the Mars atmosphere from ground up to 100 km altitude, which can only be provided by passive microwave sounding is required as model input to get unambiguous results.

The Venusian atmosphere is strictly different from Earth in structure, composition, solar energy disposition and planetary rotation state. There is little similarity in the global circulation and dynamics and a detailed explanation remains a major challenge.

The zonal flows in the atmospheres of the four giant planet-atmospheres are measured as large as hundreds of meters per second, but there is as yet no accepted theory about their origin, or even of their behaviour as a function of latitude. Even after Cassini the behaviour of the zonal flows with depth in the atmosphere-an all important dynamical quantity-will still be unavailable.

At the low temperatures prevailing in giant-planet atmospheres, hydrogen molecules separate into almost-distinct quantum species, with a slow conversion from one to the other and an associated latent heat. Investigations of the relative abundances will provide better information about the conversion process and its role in generating the large-scale flows and temperature fields seen in these bodies. At the same time measurements of flow patterns at deep levels in the atmosphere will tell us if these flows are superficial or extend deep into the interior, as is currently believed to be the case.

Water has been detected by ISO/SWS in the upper atmospheres of the four Giant Planets, however it could not be identified if its source is interplanetary dust or material originating from the ring systems. Depending on the source, different vertical and latitudinal distributions are expected. Only microwave sounders can provide this information.

Other important species may be observed in the Outer Planets. In particular the search of NH_3, PH_3 in Jupiter and Saturn and of CH_4 in the four Giant Planets would allow the still controversial stratospheric abundance of these species to be determined. On Titan, observing the

Proc. 32nd ESLAB Symp., 'Remote Sensing Methodology for Earth Observation and Planetary Exploration', ESTEC, Noordwijk, The Netherlands, 15-18 September 1998 (ESA SP-423, December 1998)

strongest lines of HCN near 1 THz would provide information on the unknown structure of Titan's mesosphere at 300-500 km.
The D/H-ratio in the atmospheres of the four Giant Planets provides information about the galactic evolution and the composition and origin of grains embedded in the promotive nebula.
Water dominates the composition of comets and their evaporation drives the dynamics of cometary outgassing. H_2O is also expected to play and important role in the thermal balance of the atmospheres of active comets. The D/H ratio in cometary water provides important information on the origin and evolution of comets. In three Oort-cloud comets (Halley, Hyakutake, Hale-Bopp) D/H was measured to be 3x10e-4, an order of magnitude higher than the cosmic abundance at the time of the formation of the solar system, as inferred from Sun and Jupiter. For comparison, water in Hot Cores near luminous protostars, which is thought to reflect the sublimation of interstellar ice grains, has a similar D/H value. This D/H ration is also large than the value of the Earth's oceans, which has implications on the volatile budgets of the inner planets. Only with microwave instruments the D/H ration can be determined in a sample of comets including Jupiter family comets and in that way can probe the conditions throughout the outer solar nebular where these comets were formed.
Furthermore the high sensitivity enables microwave sounders to look for water production in asteroids, to study the asteroid-comet interrelation by measuring water production in small Earth-orbit crossing asteroids or other unusual asteroids, to detect the extremely low amount of water in the Martian polar winter or on the Earth moon and other planetary moons, find surface sources of water and provide information about surface-atmosphere interactions, etc.
A quite large spectrum of different passive microwave sounding systems is required to address all the above mentioned scientific objectives. Before we go into detail, the basic principle of all these techniques should be recalled. Figure 1 shows a general sketch about a microwave system: A heterodyne receiver is used to down-convert the microwave radiation to an intermediate frequency range. Microwave is a rather general expression. Here it means the range from 20 GHz to at least 3 THz, i.e. it covers the centimetre-, millimetre- and submillimetre-waves. The IF-band may have a centre frequency of several GHz or even several 10 GHz, depending on the application. In the IF-band a real-time spectrometer is used in order to perform spectral analysis. The definition of real-time here is that the time domain information is converted into the frequency domain with a duty-cycle of 100%. This includes in general a data compression, because otherwise it would be almost impossible to handle the data rate, which typically would be spectrometer bandwidth in Hz times 1 byte per second. Maximum spectrometer bandwidths have been reported to be several GHz. The data compression is achieved by integrating (averaging) of spectra in real-time. This is required anyway, first because of the stochastic nature of the signals to be observed and second because of the fact that they are very weak compared with the receiver noise. The radiometer formula shows the relationship between rms-noise on the spectra Δt, receiver noise temperature T_{RX}, resolution bandwidth of the spectrometer B and integration time τ:

$$\Delta t = \frac{T_{RX}}{\sqrt{B * \tau}}$$

For a given resolution bandwidth the noise decreases

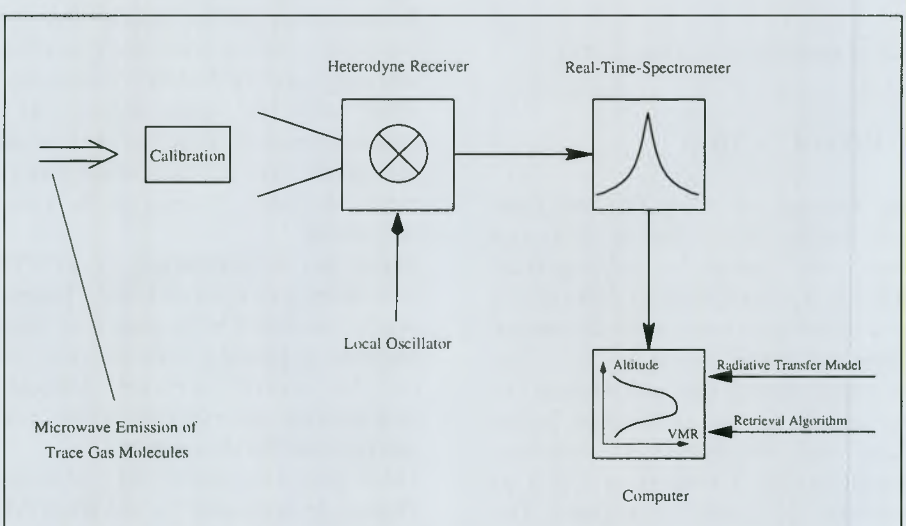


Figure 1: Passive microwave heterodyne system

linearly with the receiver noise temperature and reciprocally with the square root of the integration time, in other words the receiver should be as sensitive as possible and the whole system as stable as possible for long integration times. Except some systematic errors, the stochastic error on the measured spectra limits the degree of information gain in the retrieval process. Information can be retrieved from features of the molecular line shapes: pressure broadening, Doppler broadening, Doppler shift and excitation states of the lines. Examples will be given in the next chapter.

2. COMPARISON OF THREE SPACEBORNE MICROWAVE SPECTROMETERS

The comparison of three spaceborne microwave heterodyne spectrometers will illustrate the technical development during the last decade. We start with the Microwave Atmospheric Sounder (MAS, Figure 2), which has been flown during the Space Shuttle missions ATLAS1- ATLAS3 (Hartmann et al, 1996). MAS was a limb sounder detecting ClO, O_3, H_2O and O_2 (pressure and temperature) in the millimetre-wave range from 60 to 204 GHz. MAS has been developed mid of the eighties. It consists of an 1x1.3 m telescope, movable in one direction with a linear actuator, three whisker contacted Schottky receivers and a 240 channel filterbank with 200 KHz maximum resolution.

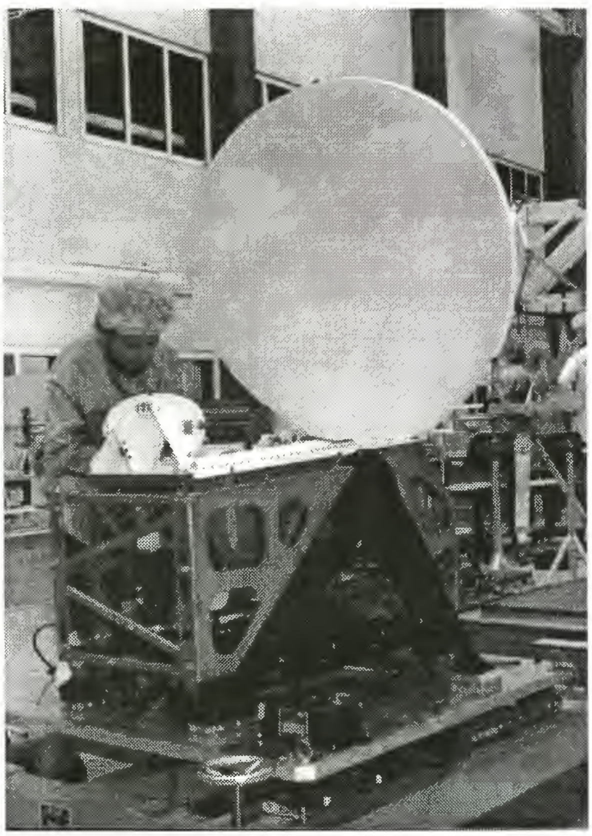


Figure 2: MAS during assembly

The MAS receiver temperature was in the range of 3000 to 5000 K SSB. The total mass of the experiment was about 200 kg, the backend mass about 65 kg. The total power consumption was 406 W (backend 120 W). The total dimensions were 128x134x173 cm^3. This relatively large and massive instrument could be flown on a carrier like the Space Shuttle, however there are in general much smaller resources available on spacecrafts being used for planetary or cometary missions.

The Microwave Instrument for the Rosetta Orbiter (MIRO, Figure 3) is a limb- and nadir sounder presently under development for the mission to the comet Wirtanen.

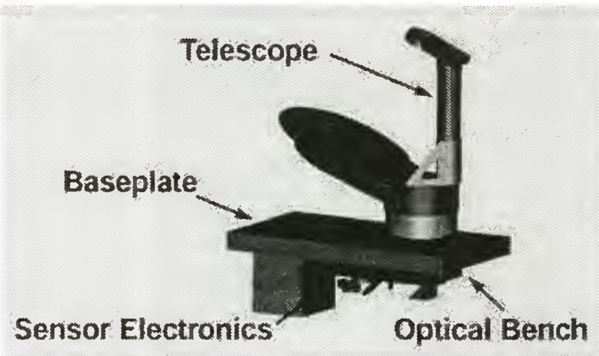


Figure 3: MIRO

MIRO will determine the abundances of the major volatile species, the fundamental isotopic ratios, the surface outgassing rate, temperature profiles in the inner coma, the nucleus subsurface temperature and kinetic velocity close to the nucleus surface and on its way to Comet Wirtanen the subsurface temperatures of asteroid targets and search for low gas levels in the asteroid environment (Gulkis, 1995, Beaudin et al. 1998).

MIRO is constructed to detect the following molecules: H_2O, $H_2{}^{17}O$, $H_2{}^{18}O$, CO, NH_3 and several lines of CH_3OH for the temperature sounding within the inner coma. MIRO uses a 30 cm telescope, two receivers with integrated planar mixer diodes in the 187-191 GHz – and the 547- 577 GHz range with a receiver temperatures between 2500 K and 5000 K DSB. A Chirp Transform Spectrometer with 4096 channels and a resolution of 43 KHz is used for the backend.

With its mass of only 18 kg (spectrometer 2.3 kg), power consumption of 16-63 (spectrometer 4-14W), depending on the observation mode and dimensions of 62x30x77 cm^3, MIRO profits a lot from recent technical developments. While during the MAS era submillimetre wave technology for space applications was not available, MIRO can take advantage of it. There are several advantages. In general the strength of molecular lines in the submillimetre wave range is considerably higher than in the millimetre wave range and therefore the number of observable lines increases dramatically. Furthermore taking into account that the spatial

Table 1: Development of Chirp Transform Spectrometers at MPAE from 1983 to 1996

PARAMETER	1983	1984	1985	1987	1991	1992	1996	UNIT
Input Centre Frequency	180	300	1350	300	300	1350	1350	MHz
Input Bandwidth	25	22	40	40	40	40	178	MHz
Spectral Resolution (-3 dB)	500	150	50	50	22	50	43	KHz
Passband Ripple	3	6	1	1	1	1	1	dB
Dynamic Range	18	27	26	15	30	30	29	dB
Frequency Linearity (<)	20	8	1	1	1	1	2	KHz
Allan Variance Minimum (>)	100	1100	900	100	1000	900	300	sec
Frequency Scale Stability	50	1	5	2	2	2	5	KHz
Power Consumption	420	520	530	430	50	30	14	W
Mass	22	20	20	25	10	10	2.3	kg

resolution of microwave sounders is diffraction limited, going to a submillimetre frequency allows to use smaller telescopes. MIRO observes the 557 GHz water line with almost the same spatial resolution as MAS the 183 GHz line with its much larger telescope.

The Chirp Transform Spectrometer is another big step forward. Compared with the MAS filterbank it has 4096 Vs 240 channels with only 12 % of the power consumption and 3.5 % of the mass. Originally it has been planned to use the CTS already for MAS, however the power consumption of this technique was much too high end of the eighties to make it applicable for space missions. (Table 1).

The third experiment called Microwave Investigation on Mars Express (MIME, Figure 3) has been proposed for the Mars Express mission (Hartogh, 1998a).

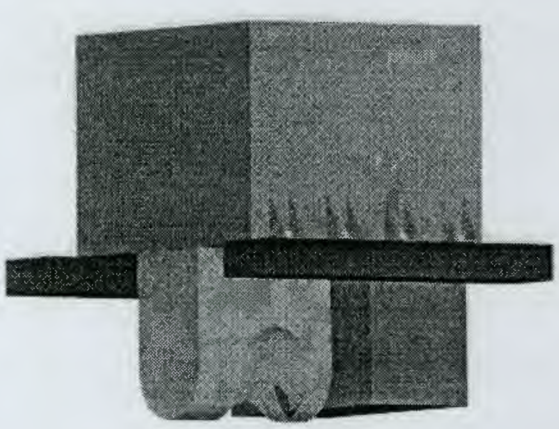


Figure 4: MIME

The purpose of MIME is to perform three dimensional mapping of the time variable atmospheric state parameters temperature and winds, three dimensional mapping of atmospheric water vapour and carbon monoxide in order to improve the phenomenological description of the present three dimensional general circulation and develop a quantitative description of the processes controlling the circulation. Furthermore it was proposed to perform global mapping of the subsurface brightness temperature in order to develop a quantitative understanding of the time-variable sources, sinks and atmospheric transport of water vapour.

MIME is a limb- and nadir sounder and detects the rotational transitions of H_2O, $H_2{}^{18}O$, CO and ^{13}CO with a 547-577 GHz receiver using integrated planar mixer diodes and 5000 K DSB noise temperature. The backend is a Chirp Transform Spectrometer similar to that used for MIRO.

The mass of MIME could be reduced compared with MIRO by using one receiver rather than two, a smaller telescope aperture of only 7 centimetres with integral shroud and mechanical/electrical improvements developed and space qualified at MPAE for the NASA´s Mars Polar Lander 1998 program (Robotic Arm Camera). This results in a mass of only 9.2 kg (spectrometer 2.3 kg) and a size of 33x24x25 cm^3.

The performance of MIME will be illustrated in the following figures. Figure 5 shows the coverage of MIME projected on MARS for nadir and limb observation mode. The main parameters determining

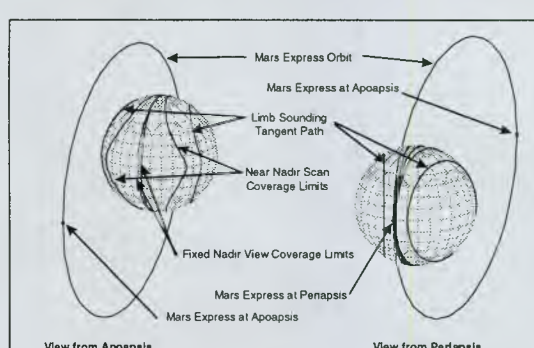


Figure 5: Coverage for nadir and limb observation modes projected on Mars.

the coverage are the sensitivity of the receiver, the orbit of the spacecraft and the requirements of the retrieval algorithm in terms of signal-to-noise-ratio.

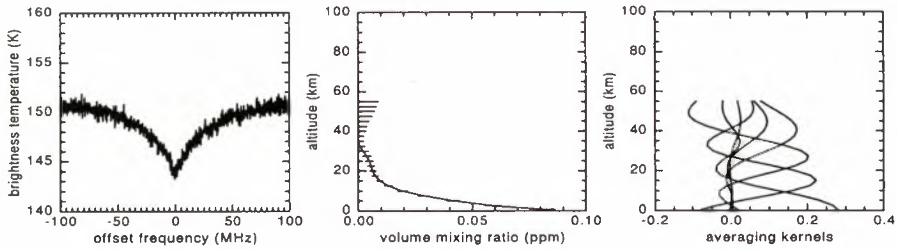


Figure 6: Nadir sounding of water in martian polar winter (<0.0001 pr μm)

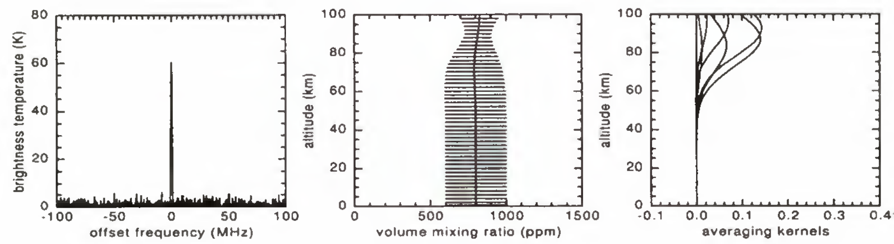


Figure 7: Limb sounding of CO

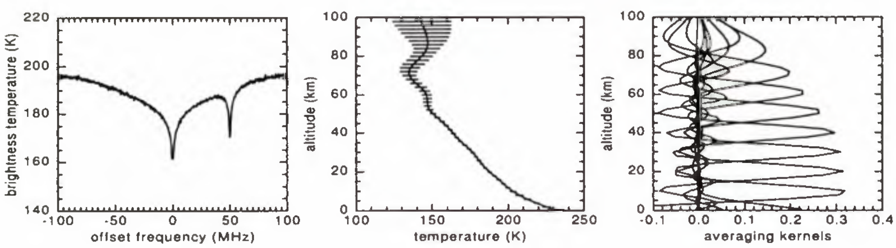


Figure 8: Nadir sounding of temperature

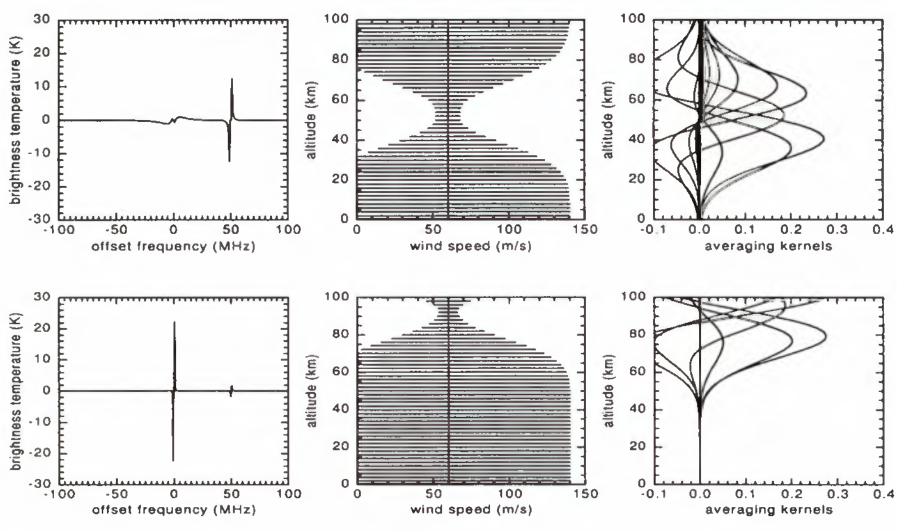


Figure 9: Limb sounding of wind for two altitudes

Figure 6 illustrates the extreme sensitivity of this instrument. MIME can detect water profiles in the Martian polar night with total water column amounts of < 0.0001 pr μm. The plot shows the modelled spectrum after 1 minute integration time, the retrieved water vapour profile and the averaging kernels, a measure for the altitude resolution of the retrieval.
Figure 7 shows a modelled limb spectrum of CO, with MIME pointing at 90 km altitude. MIME measures the ^{12}CO and ^{13}CO lines. The latter one does not appear in the plot, because its intensity is smaller than the noise. CO is assumed to be uniformly distributed in the Martian atmosphere with 650 ppm volume mixing ratio. It can therefore be used to retrieve the atmospheric temperature. However there is no measurement, which could confirm the assumption. Non-uniform distributions could not only have implications on the retrieval of temperature but on our understanding of the atmospheric chemistry of Mars.
Figure 8 shows a nadir measurement of the CO lines including the retrieved temperature, which has an accuracy of 1 –3 K over a wide altitude range.
Figure 9 shows a simulation of a MIME wind measurement in limb geometry. The spectra of two across track limb measurements are subtracted. In case of wind, the two lines are Doppler shifted and the subtraction results in a peak to which the radiative transfer model is fitted. Again the CO lines have been used for the simulation. For pointing at 50 km the ^{12}CO line is opaque and the wind information is retrieved from the ^{13}CO line. For pointing at 90 km the isotopic line is too weak and the information is contained by the ^{12}CO line. Winds can been measured in this way with an accuracy of about 10 m/s. The limit in this case is driven by the attitude knowledge and stability of the Mars-Express spacecraft.

3. FURTHER IMPROVEMENTS

In the above chapter it has been demonstrated that rather complex microwave sounders are now applicable for interplanetary space missions. Further improvements are required to offer a wider spectrum of applications. On the one hand the functional performance could be improved at lot. On the other hand any reduction in mass and power consumption will rise the chance for a microwave sounder to become payload of a spacecraft, in special on microsatellites. Other very interesting applications could be ground-based systems on planets, comets asteroids and moons, which could monitor the climate or the outgassing rate. Even ground-based limb sounders are an option (for instance on Olympus Mons on Mars). For these applications the mass has certainly to be reduced to 2-3 kg.
Improvement of functional performance means to reduce the noise temperature of the receivers, to extend the receiver bandwidth and to develop better spectrometers in terms of bandwidth and number of channels. The optimum here would be a spectrometer with large bandwidth and high spectral resolution at the same time.
Presently four types of receivers are applied in microwave spectrometers: HEMT-, Schottky-, SIS- and HEB receivers.
In the first case the microwave signal is amplified with a High Electron Mobility Transistor amplifier before the conversion process takes place. This has advantages compared with other techniques, because for instance sideband filters can be used after the amplification. In general quasi-optical techniques are applied in front of the receivers. These easily add unwanted reflections and cause residual baselines, which are very difficult to be removed. HEMT receivers are nowadays available for frequencies up to 110 GHz (Pospieszalski and Wollack, 1998). The sensitivity depends on the operational temperature of the HEMT amplifier. For 300 K operational temperature, noise temperatures of about 50 times the quantum limit (hv/k) can be achieved. Cooling down the HEMT amplifier to 20 K reduces to noise temperature to about 10 times hv/k.
Schottky receivers are still widely used, although there performance is not as good as the performance of SIS and HEB receivers. For frequencies above the range covered by HEMT receivers, there is still no alternative for this sensor in case there is no cooling is available. Schottky receivers are available for frequencies of up to 2500 GHz. For uncooled systems the noise temperature is more than 150 times the quantum limit below 700 GHz and about 200 times above 700 GHz. These numbers can be improved by a factor of 3 for receivers cooled to 20 K.
Superconductor Insulator Superconductor (SIS) receivers cover the frequency range from 100 to 1200 GHz (Carlstrom and Zmuidzinas, 1996, Karpov 1998). The upper frequency limit is given by the energy gap of the superconduction material, which is 700 GHz for Nb and 1200 GHz for NbTi. The down-conversion process can be achieved with near quantum-limited noise performance. Because the SIS device can carry currents up to those of the superconducting gap, there is no fundamental limit to the bandwidth of the IF. Practical SIS-receivers achieve sensitivities of 2-3 times the quantum limit below 400 GHz, 4-10 times between 400 and 700 GHz and 10-30 times the quantum limit between 700 and 1200 GHz. There is a permanent optimisation in progress and a design goal of SIS-receiver developers is to achieve about 3 times the quantum limit for the whole frequency range. The typical operational temperature of the SIS device is 4 K.
In a bolometric detector the resistance is changed by heating due to absorbed electromagnetic energy. Superconducting materials are most sensitive to this

effect and therefore well suited for bolometric detectors. Application of two signals with different frequency will modulate the resistance at an intermediate frequency. Useful IF-bandwidths require fast cooling and low heat capacity of the detector (Mc Grath et al. , 1997, Gol'tsman, 1998). In a Hot Electron Bolometer (HEB) this is done by diffusion of the hot electrons out of the detector contrary to the slow electron-phonon cooling mechanism in traditional bolometers. IF-Bandwidths of several GHz can be achieved with a HEB-mixer. The mixing properties of the HEB mixer does not depend on the individual photon energy, but on the total absorbed power. HEB mixers are therefore not limited by the energy gap of the superconductor. The intrinsic noise levels of HEB mixers are higher than in the SIS mixers, but with current technology HEB become preferred to SIS mixers at 1200 GHz. Noise temperature of 20 times the quantum limit have been achieved with conventional superconducting materials at 4 K, the goal is to reduce this by a factor of six.

Recently some work has been done on using high temperature superconducting HEB-mixers operating at 77 K. It is expected that noise temperatures of 50 times the quantum limit can be achieved. (Mc Grath, 1998).

It is planned to use the HTS-HEB device in space with a radiation cooler. The other devices have to be cooled either with liquid Helium or an active cooling mechanism such as Sterling, pulse tube or sorption coolers. The latter device seems to be most promising for future space applications, because of the absence of moving parts (no vibration) , the fact that the simple design allows loose manufacturing tolerances and simple electronics and that it is scalable over a wide range of power requirements simply by varying the size of the sorbent bed (Wade, 1998). Sorption coolers work by using metal hydride powders (chemisorption) or helium-charcoal (physisorption), that absorb the hydrogen or helium refrigerant through a reversible process. In the hydrogen sorption compressor the metal hydride is first heated to pressurise the hydrogen, and then cooled to room temperature to reduce the pressure. By sequentially heating and cooling the powder, the hydrogen is circulated through the refrigeration cycle. Very low temperatures are achieved by expanding the pressurised hydrogen at the cold tip of the refrigerator. This expansion actually freezes the hydrogen to produce a solid ice cube. The heat load generated by the device being cooled then sublimates the ice. This closed cycle operation is repeated over and over. Sorption coolers can be staged to provide cooling to temperatures as low as 200 mK. Application of sorption coolers for microwave spectrometers using SIS- of HEB receivers require an accelerated miniaturisation of this technique.

Presently five types of real-time spectrometers are used as backend of the microwave system. The classical filterbank is still in use for application where enough resources in terms of mass and power are available. The relatively low number of channels is still a limitation and therefore filterbanks make only sense of no high resolution is required. This is the case for limb sounders pointing at atmospheric pressure levels above a few hPa. For the Earth atmosphere we are speaking about the upper troposphere and the stratosphere. Filterbanks still seem to be the cheapest technique, however except the relatively small number of channels there are other disadvantages. The linearity and dynamic range of the used detectors is a weak point and that in principle each filter needs its own detector with its individual characteristics. This may add errors on the calibrated spectra, which may be negligible for strong atmospheric lines, however certainly cause problems for weak lines.

Acousto-Optical Spectrometers (AOS) are multichannel instruments and mainly used for broadband applications. Bandwidths of up to 2 GHz have been realised. There seems to be a performance optimum at 1 GHz bandwidth, because recent requirements for 4 GHz bandwidth (Heterodyne Instrument on FIRST, HIFI) have been described as a realisation of either four single 1 GHz AOS or by using an array-deflector (Schieder et al, 1996, Rosolen et al, 1998). AOS have in general 1000 independent channels. The first AOS operating in a science space mission will be launched end of this year with the Submillimetre Wave Astronomy Satellite (SWAS) . A competitive design will fly on the Odin satellite which is planned to be launched next year (Frisk, 1998).

A Digital Autocorrelator Spectrometer (DACS) with 8x100 MHz bandwidth and 768 channels will fly on Odin as well. (Emrich, 1997) .The principle of this type of spectrometer is to digitise the IF signal and then perform an autocorrelation. The Fourier transform of the correlated signal results in the wanted power spectrum. The difficulty is to sample the IF signal with a very high speed according to the Nyquist theorem and to perform the correlation in real-time. This only works, if the sampling is done with only 1 or 2 bits. In recent low power designs 3 steps are used. The disadvantage of this kind of sampling is that the spectral noise is increasing by about 20 % and requires an increased integration time. Furthermore the DC-component of the spectrum gets lost, which is of disadvantage for applications where the exact knowledge of the continuum is of advantage as in the case of sounding planetary atmospheres. A third problem, which has been reported is the so called platforming. Platforming means that if several correlators are operated in parallel to achieve a larger bandwidth, slight differences in the measured power density lead to steps in the spectrum over the total bandwidth. DACS based on the Odin design seem to be well suited for space application in terms of power consumption and mass, however practical measurements on telescopes with that design have not

Angle from Apoapsis, deg	Time from Apoapsis, hr	1) Along Track Spatial Resolution, km	2) Along Track Spatial Resolution, km	3) Along Track Spatial Resolution, km
0.0	0.00	111.7	39.5	3.2
22.5	1.70	143.0	50.6	4.1
45.0	2.81	253.5	89.4	7.2
67.5	3.41	476.7	168.5	13.9
90.0	3.74	828.0	292.7	23.5

Table 2: Along track spatial resolution of MIME for three receiver noise temperatures

been published yet. DACS with bandwidths of up to 5 GHz can be realised according to recent studies. However in present applications DACS are mainly used for high resolution applications. Typical channel numbers are 1000 – 2000.

Analog correlators have been recently described for large bandwidth applications. Bandwidths of up to 3250 MHz have been demonstrated (Harris et al, 1998). The number of channels is relatively small with about 100. Future developments are planned with 8 GHz bandwidth and 50-100 channels. Analog correlators are planned to be used on SOFIA and for broadband submillimetre wave arrays. At the present time nothing has been published about their space potential.

Superconducting (DiIorio et al, 1989) and High-Temperature Superconducting (HTS) Compressive Receivers (Lyons et al, 1996) with bandwidth of 6 GHz have been demonstrated. HTS spectrometers with about 10 GHz bandwidth and 2048 channels are under development. The problem of this instrument at the present time is that the sampling rate is equivalent to the bandwidth. Analog-to Digital Converters with 10 GHz bandwidth and clock rate are under development. The power consumption of this type of spectrometer is relatively high. It has been mainly used for military applications (cueing receivers), but could in principle be used for applications in microwave spectrometers.

Chirp Transform Spectrometers (CTS) and their development during the last years have been mentioned in the last chapter (Hartogh, 1997) A CTS performs an analog Fourier transform. The output of the analog part of the spectrometer is a time sequence, which has to be sampled and averaged in a so-called digital preprocessor, which is the main power consumer. In the MIRO design about 2 W are required for the analog and about 12 W for the digital part of the CTS. According to a study based on space qualified technology available in 1997 this could be reduced to 0.2 W for the analog part using a bipolar ASIC and 1.9 W for the digital part using a CMOS ASIC. (Hartogh, 1998b). The speed of this ASIC fits the requirement for a 200 – 250 MHz bandwidth. The mass of the CTS could be reduced to below 1 kg using this ASIC technology. (A 2x5x200 MHz CTS with 40000 channels based on this technology has been proposed for HIFI). Dispersive delay lines required to perform the analog processing can be designed for bandwidths up to 1.5 GHz. We estimate that a CTS with 1 GHz bandwidth, 10000 channels 0.5- 1 kg mass and a power consumption of < 2W can be realised within the next 5 years using state-of the art technology.

Even if it does not seem to be realistic from the today's point of view at least the question is very interesting what the performance of a MIME like sensor would look like using a cooled receiver and an improved CTS. We assume that mass of such an instrument would be not much higher, because the additional mass of the cooler could be partly compensated by the reduced mass of the CTS. The larger bandwidth of the spectrometer would enable to detect additional molecular transitions like ozone and hydroxyl radicals and would therefore make this instrument not only a dynamical, but also an chemical sensor. Table 2 shows the impact of reduced receiver noise temperature and increased spectrometer bandwidth on the along track spatial resolution. In the first case the receiver noise temperature is 5000 K DSB (MIME), in the second case it is assumed that the receiver is cooled to 100 K resulting in a receiver noise temperature of 2500 K DSB and in the third case it is assumed that the instrument uses an SIS mixer with a noise temperature of 84 K DSB (noise temperature of HIFI channel 1). The orbit is the same as in Figure 5.

4. SUMMARY

Solar System Research with spaceborne microwave sounders has become applicable during the last few years, because the functional performance of instruments could be increased, while mass and power consumption has been reduced at the same time. Further developments in the areas receiver-and cooling-as well as spectrometer technologies are required. The development goal on the receiver sector should be to make SIS and HEB technologies available for small instruments like MIRO and MIME. To achieve this the cooling technologies have to be improved. There is still a large development potential in the spectrometer sector as well. Development goal should be to increase the bandwidth and the number of channels and to decrease mass and power consumption.

REFERENCES

Beaudin, G., Gulkis, S., Frerking, M., Hartogh, P., Allen, M., Bockelee-Morvan, D., Crovisier, J., Despois, D., Encrenaz, P., Encrenaz, T., Germain, B., Hofstadter, M., Ip, W., Janssen, M., Lellouch, E., Mann, I., Muhleman, D., Rauer, H., Schloerb, F.P., Spilker, T., *A Microwave Radiometer/Spectrometer for the Rosetta Orbiter,* Proc. of 2nd ESA Workshop on Millimeter Wave Technology and Applications: antennas, circuits and systems, Proceeding Editors: J. Mallat, A. Räisänen, J.Tuovinen, 27-29 May 1998, MilliLab, Espoo, Finland, pp. 43-48.

Carlstrom, J.E., Zmuidzinas, J., *Millimeter and Submillimeter techniques,* Reviews of Radio Sciencs 1993-1995, ed. W.R. Stone, Oxford, The Oxford University Press, 1996.

DiIorio, M.S., Withers, R.S. and Anderson, A.C., *Wide-band superconductive chirp filters*, IEEE Trans. Microwave Theory Techn., vol. 37, pp. 706-710, 1989.

Emrich, A., *Autocorrelation Spectrometers for Space Borne (Sub)millimetre Spectroscopy,* The Far InfraRed and Submillimetre Universe, ESA Publication SP-401, ISBN 92-9092-287-7, Editor A. Wilson, pp 361-364, 1997.

Frisk, U.O., *The Odin Millimetre to Submillimetre Wave Satellite,* , Proc. of 2nd ESA Workshop on Millimeter Wave Technology and Applications: antennas, circuits and systems, Proceeding Editors: J. Mallat, A. Räisänen, J.Tuovinen, 27-29 May 1998, MilliLab, Espoo, Finland, pp. 16-21.

Gol'tsman, G.N. *Hot Electron Bolometric Mixers: New Terahertz Technology,* Proc. of 2nd ESA Workshop on Millimeter Wave Technology and Applications: antennas, circuits and systems, Proceeding Editors: J. Mallat, A. Räisänen, J.Tuovinen, 27-29 May 1998, MilliLab, Espoo, Finland, p. 233-238.

Gulkis, S, *Microwave Instrument for the Rosetta Orbiter,* Proposal in response to ESA Ro-EST-AO-000, Pasadena, 1995.

Harris, A.I., Isaak, K.G., Zmuidzinas, J, *WASP: A wideband spectrometer for heterodyne spectroscopy,* Proc SPIE Vol. 3357, pp. 384-394, 1998.

Hartmann, G.K., Bevilacqua, R.M., Schwartz, P.R., Kämpfer, N., Künzi, K.F., Aellig, C.P., Berg, A., Boogaerts, W., Connor, B.J., Croskey, C.L., Daehler, M., Dicken, H.D., Goldizen, D, Kriebel, D, Langen, J., Loidl, A., Olivero, JJ., Pauls, T.A., Puliafito, S.E., Richards, M.L., Rudin, C., Tsou, J.J., Waltman, W.B., Umlauft, G. and Zwick, R., *Measurement of O_3, H_2O and ClO in the Middle Atmosphere using the Millimeter-Wave Atmospheric Sounder (MAS),* Geophys.Res. Lett. **23**, no 17, Paper 96GLL01475, 2313-2316, 1996.

Hartogh, P., *Present and Future Chirp Transform Spectrometers for Microwave Remote Sensing,* Proc. SPIE, Vol3221-328-339, 1997.

Hartogh, P., *Microwave Investigation on Mars Express,* Proposal in Response to AO ESA/SCI(97)3, 1998a.

Hartogh, P., *A new Microwave Heterodyne Spectrometer Backend: The ASIC-CTS,* , Proc. of 2nd ESA Workshop on Millimeter Wave Technology and Applications: antennas, circuits and systems, Proceeding Editors: J. Mallat, A. Räisänen, J.Tuovinen, 27-29 May 1998, MilliLab, Espoo, Finland, pp., 333-338.

Karpov, A, *Receivers for the Ground Based mm Wave Radio Telescopes,* , Proc. of 2nd ESA Workshop on Millimeter Wave Technology and Applications: antennas, circuits and systems, Proceeding Editors: J. Mallat, A. Räisänen, J.Tuovinen, 27-29 May 1998, MilliLab, Espoo, Finland, pp. 227-232

Lyons, W.G., Arsenault, D.R., Anderson, A.C., Sollner, T.C.L.G., Murphy, P.G., Seaver, M.M., Boisvert, R.R., Slattery, R.L., Ralston, R.W., *High Temperature Superconductive Wideband Compressive Receivers,* IEEE Trans. Microwave Theory Tech. Vol. 44, No 7, pp. 1258-1277, 1996.

McGrath, W.R., Skalare, A., Karasik, B., Gaidis, M., Bumble, B., LeDuc, H.G., Burke, P.J., Schoelkopf, R.J., Prober, D.E., Verghese, S., McIntosh, K.A. and Brown, E.R., *Superconductive Hot Electron Heterodyne Detectors for Far-Infrared Remote-Sensing Applications,* Invited Presentation: Japan Society of Infrared Science and Technology, Tokyo, Japan, October 31, 1997.

Mc Grath, W.R., *personal communication*, 1998.

Pospieszalski, M.W. and Wollack, E.J., *Characteristics of Broadband InP HFET Millimeter-Wave Amplifiers and Their Applications in Radio Astronomy Receivers,* Proc. of 2nd ESA Workshop on Millimeter Wave Technology and Applications: antennas, circuits and systems, Proceeding Editors: J. Mallat, A. Räisänen, J.Tuovinen, 27-29 May 1998, MilliLab, Espoo, Finland, pp. 221-226.

Rosolen, C., Michet, D., Lecacheux, A. and Clerc, V., *Acousto-Optical Spectrometers for Millimeter and Submillimeter Limb Sounders,* Proc. of 2nd ESA

Workshop on Millimeter Wave Technology and Applications: antennas, circuits and systems, Proceeding Editors: J. Mallat, A. Räisänen, J.Tuovinen, 27-29 May 1998, MilliLab, Espoo, Finland, pp. 321-326.

Schieder, R., Horn, J., Siebertz, O.,. Klumb M, Frerick, F., Tolls, V., *Acousto-Optical Spectrometers in Space,* Proc. 30th ESLAB Symp., 'Submillimeter and Far-Infrared Space Instrumentation', 24-26 September, 1996, ESTEC, Noordwijk, The Netherlands, ESA SP-388 (December 1996).

Wade, L.A, *personal communication,* 1998, for more info contact lwade@squid.jpl.nasa.gov.

Optical Technologies

NEW CONCEPTS FOR HIGH RESOLUTION IMAGERS FOR EARTH OBSERVATION AND PLANETARY EXPLORATION

H. Michaelis, H.-G. Grothues, T. Behnke, H. Hirsch, E. Ress, I. Sebastian

DLR, Institute of Planetary Exploration
Rudower Chaussee 5, D-12489 Berlin, Germany

K. Schröter

Daimler Benz Aerospace
Jena-Optronik GmbH

1. Introduction

The field of terrestrial remote sensing and planetary exploration is an increasing demand in high and very high resolution imaging from space. In terrestrial remote sensing, high resolution imaging of the '1m- class' (ground resolution) is requested for commercial applications like hazard monitoring, town mapping, and planing or traffic monitoring. In planetary exploration, high resolution imaging is needed for searching of potential landing areas, high resolution surface mapping, or it is requested in asteroid fly-by and rendezvous missions.

An example of the technical requirements for a very high resolution imager is given in Tab.1.

Tab. 1: Technical requirements for a very high resolution imager

spatial resolution (GSD)	~1m @ 500 km ($\theta \leq 1''$), panchromatic
swath width	$\geq$ 10 km
radiometric resolution (SNR)	> 50 (100)
mass of sensor	< 20 kg (10kg)

The technical problem is that with increased spatial resolution the resources needed for the optics of the imaging system will rapidly increase which means very high cost for the spacecraft and the launch vehicle. Therefore, new concepts are needed for very high resolution imagers which can be accommodated on small spacecraft (less than about 250 kg).

2. The Technical Problem

The main technical problems for the class of high resolution imagers defined in Tab.1 are outlined by the following topics:

- minimum aperture
- focal length
- swath width
- signal-to-noise ratio (SNR)

<u>Minimum Aperture</u>

The minimum criteria for an aberration free optical system is that the diameter (entrance pupil) of the optics has to exceed the one given by the diffraction limit. The simplest approach is to apply the Rayleigh criterion where

$$D \geq 1.22\, \lambda\, s / GSD \qquad (1)$$

with s: ground distance of system (spacecraft),
GSD: ground sampling distance (unit: m / pixel).

In typical applications contrast has to be at least 15% for safe detection of ground features resulting in a somewhat higher aperture that is required, i.e.:

$$D = 1.35\, \lambda\, s / GSD \qquad (2)$$

The minimum aperture as a function of the ground resolution (GSD) according to (2) is shown in Fig.1. It is based on the assumption that:

s: 500 km
λ: 675 nm
$\Delta\lambda$: 180 nm
contrast: 15%

Proc. 32nd ESLAB Symp., 'Remote Sensing Methodology for Earth Observation and Planetary Exploration', ESTEC, Noordwijk, The Netherlands, 15-18 September 1998 (ESA SP-423, December 1998)

It is obvious that ground resolutions of 1 m / pixel require apertures of > 0.5 m. At 1.5 m / pixel ground resolution the diameter of the optics has to be at least 0.36 m.

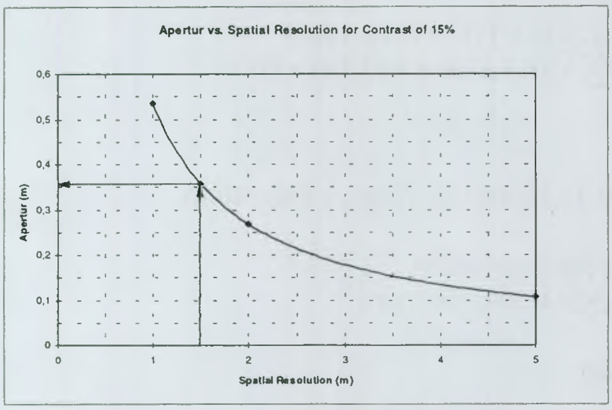


Fig.1: Minimum aperture as a function of resolution (GSD)

Focal Length

With the increasing resolution (GSD) the focal length (f) of the camera optics also has to increase according to:

$$f = P\,s / GSD \qquad (3)$$

where P: pixel size

This means for example that the focal length has to be 6 m at GSD of 1 m / pixel from an orbit of 500 km with a pixel size of 12 μm. Such a very large focal length can only be decreased using smaller pixels of the detector (e.g. in linear CCDs, see Fig. 2) which, on the other side, requires higher quality optics (improved MTF).

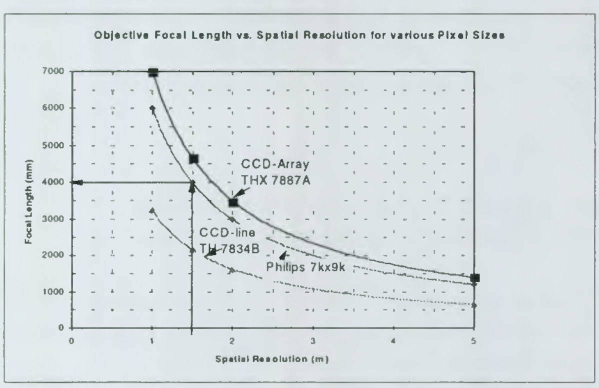


Fig. 2: Focal length as a function of ground resolution (GSD) for some CCDs

Swath Width

The swath width is a linear function of the number of pixels per detector line. In order to fulfil the specified requirement of at least 10 km swath width at a GSD of 'only' 1.5 m / pixel the detector needs to have at least 7000 pixels per line. This is the present limit for monolithic CCD area array detectors (cf. Fig. 3).

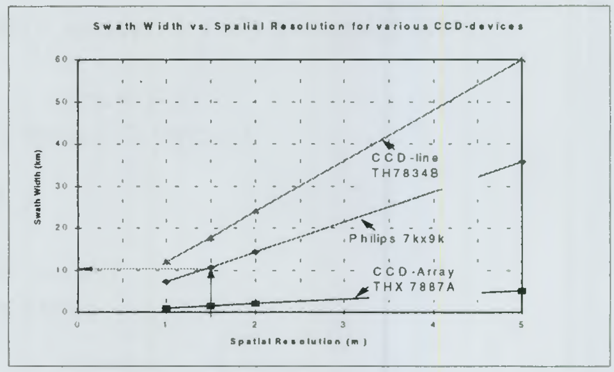


Fig.3: Swath width as a function of resolution (GSD) for some CCD detectors

Signal-to-Noise Ratio (SNR)

The present discussion about focal length and swath width gives the impression that linear CCD array are better suited for high resolution imagers than are area array detectors because of their smaller pixel sizes available, and the larger number of pixels per line. Another argument for linear detectors is their good quantum efficiency and their simple structure (lower cost).

Therefore, most of existing high resolution imagers for earth observation are typically operating with **CCD-lines in pushbroom mode**. This means that the scene on the ground is scanned by making use of the motion of the orbiting spacecraft relative to the ground. In order to prevent smearing over more than one pixel the integration time T_i for one line has to be less than the ground sampling distance (GSD) divided by the ground speed (v_g) of the spacecraft (which is about 7 km / s):

$$T_i < GSD / v_g \qquad (4)$$

That means that for a ground sampling distance of 5 m / pixel the maximum integration time which avoids smearing has to be less than 0.7 ms, (s: 500 km). This very short integration times require fast optical systems (low f-numbers) with large apertures to yield images with good radiometric resolution (SNR ≥ 50 at all illumination conditions), as demonstrated in Fig.4.

At higher resolutions the f-number and the aperture of the optics has to grow larger and larger. This in turn results in a steep increase of mass and cost of the systems.

Therefore, imaging systems with a resolution of 1 m / pixel or even better in an earth orbit have typically masses much in excess of 100 kg and cost more than 100 million US$. Moreover, such systems do not fit on small satellites, and will also not fit into the budget constraints of most scientific and commercial applications.

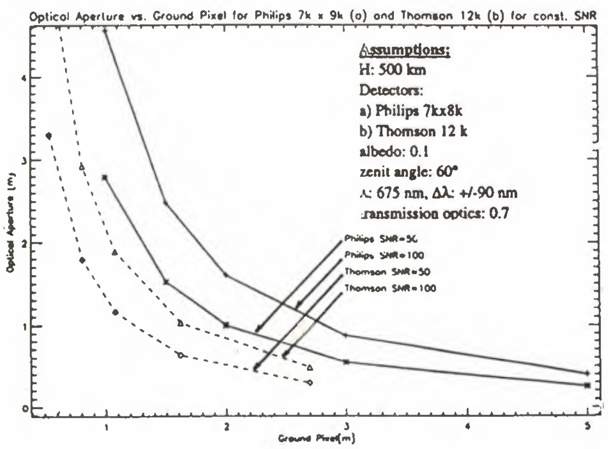


Fig.4: Aperture as a function of resolution (GSD) for some CCDs at a calculated SNR of 50

3. Conceptual Approach

Alternative concepts for very high resolution systems have been studied at DLR, which are suited to be installed on small spacecraft for earth observation and planetary exploration.

The concept proposed here for a space camera combining very high ground resolution with low mass and low volume is based mainly on following four steps:

3.1. Time Delayed Integration (TDI)

One of the basic ideas of the proposed concept is to decrease the diameter of the optics by increasing the integration times. This will be performed by use of area array CCDs instead of linear CCDs, and their operation in time delayed integration mode (TDI).

Therefore, the focal plane will consist of a single (monolithic) CCD or a mosaic of area array CCDs where the CCD lines are clocked synchronously with the projected image on the detector array. The advantages of area array CCDs over linear CCDs are:

- **integration times can be electronically adjusted** to the conditions of illumination simply by commanding the number of TDI steps (see Fig. 5)
- **no mechanical devices** (e.g. scanning mirrors) are required for mechanical drift compensation
- a **fixed geometry** is given by an area array CCD (image data are easier to be processed for accurate photogrammetry than scanning data of linear CCDs).

Therefore the f-number of the optics can be increased, and its diameter will be smaller than for applications with linear CCDs where TDI cannot be applied. This results in lower mass and cost of the imaging system. Furthermore, the depth of focus will increase owing to larger f-numbers (e.g. f / 8...f / 16) which will relax the requirements on the opto-mechanical and thermal subsystems.

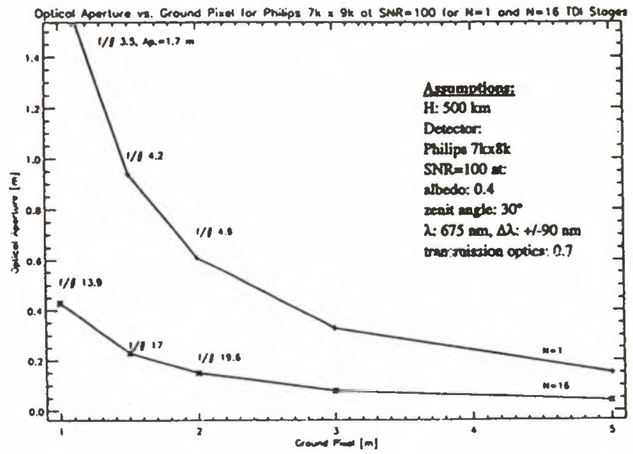


Fig. 5: Telescope aperture as a function of ground resolution (GSD) with a CCD operated in TDI mode. The number of consecutive TDI stages are N = 1 (no TDI) and N = 16

Compact Optical Design

Besides the telescope aperture of the camera optics which can be reduced by application of TDI, it is very important to have a short optical system in order to arrive at the most compact instrument which can easily be accommodated into a small spacecraft. Thus, a compact optical design is very important for very high resolution systems which folds the required long focal length into the shortest optical system that can be handled in terms of manufacturing the optical elements and the mechanical and thermal stability.

Therefore, a first optical design study was initiated by DLR which resulted in a very compact optical design with a length of only about 0.7 m at a focal length of 6 m which would be required for a resolution of 1 m / pixel at an orbit of 500 km with a 12 μm pixel- detector (cf. Figs. 6 to 8). This optical design is the basis for the Very High Resolution Camera (VHRC) shown below.

The design principally comprises as spheric primary and secondary mirrors with a field corrector for the large CCD detector (7k x 8k) that is foreseen for the focal plane.

Another example for a compact optical design which is based on simple spherical mirrors in an catadioptic-design is shown in Fig. 9 (SHRC / SMART-1). This system has a focal length of only 1 m and is well suited for low cost and light-weight imaging systems for planetary exploration.

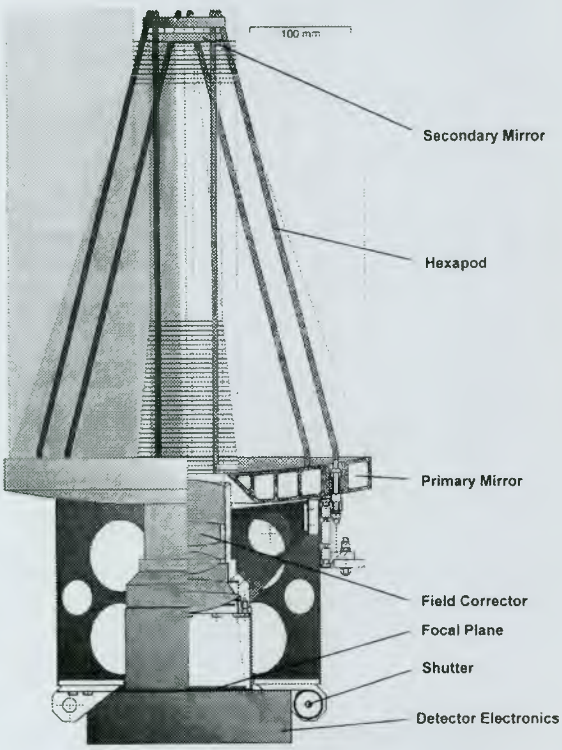


Fig. 6: Example of an opto-mechanics design for a high resolution space camera (VHRC) with f = 6 m

Optics	catadioptric system with aspheric reflector and field corrector, diffraction-limited
Aperture	0.35 m (TBC)
Focal length	6.00 m
f-number	17.1
Wavelength-range (chromatic correction)	0.4 - 0.8 μm
Focal plane diameter	130 mm (diagonal of Philips 7k x 8k CCD)
FOV (diameter)	degrees (10.8 km from 500 km orbit)
IFOV (12 μm pixels)	0.4 arcsec (1.0 m / pixel for 500 km orbit, TBC)

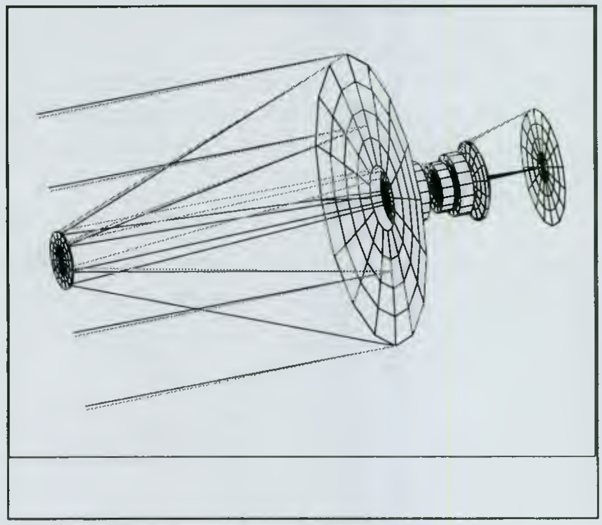


Fig. 7: Optical layout of the VHRC (for parameters see table below)

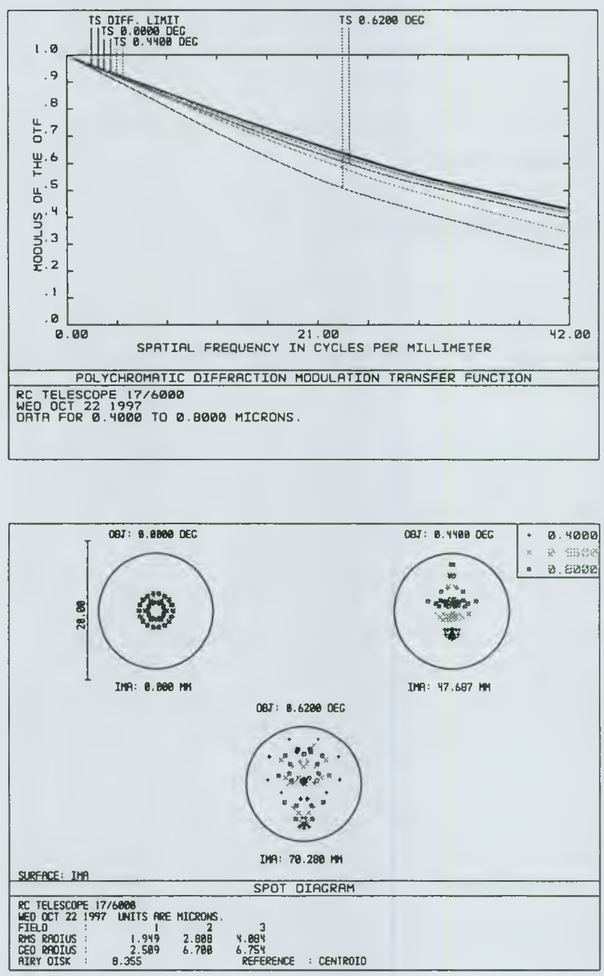


Fig. 8: MTF (top) and spot diagrams (bottom) for the VHRC optics

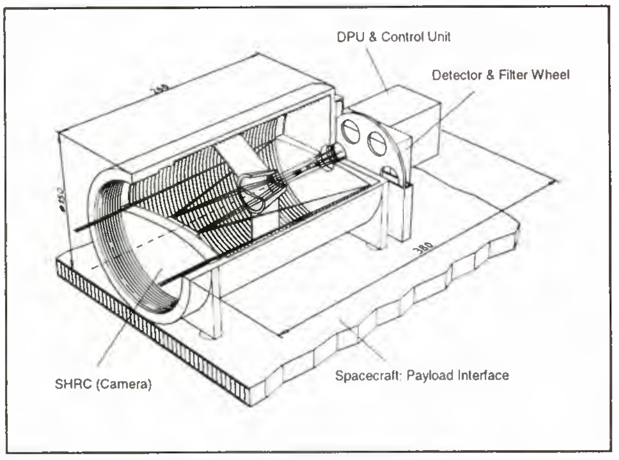


Fig. 9: Optical Layout of the SHRC / SMART-1 (GEMINI) optics

Materials for Optics and Structure

Another important step to reduce the mass of the imaging system is to use modern (compound) materials mainly for the optical components (mirror supports) and the mechanical structure of the camera. One example is the application of **C/SiC**, which is a ceramic composite material. This material is very well suited for light-weight imaging systems in space applications because of their low thermal expansion, high thermal conductivity and their high stiffness. A comparison between mechanical and thermal properties of C/SiC (and C/C-SiC) with conventional materials is shown in Tab. 2. To give an example, it is possible to manufacture primary mirrors with a diameter of 0.5 m and a mass of less than 5 kg.

Tab. 2: Comparison of mechanical and thermal properties of C/SiC and C/C-SiC with conventional materials (20°C)

Material	C/SiC (C/C-SiC)[1]	Zerodur	Invar-36[2]	Al
Density (g cm-3)	2.7 (1.9)	2.53	8.0	2.71
Youngs Modulus (GPa)	240 - 465[3]	90	140 - 150	70
CTE (10^{-6} K^{-1})	2.0 (1 - 2)[4] (4 - 6)[5]	-0.02	1.2 - 2.0	23
Thermal conductivity (W m^{-1} K^{-1})	135 (13)[3]	1.46	12.8	230
Optical Surface	yes	yes	no	(yes)

[1] values in brackets refer to C/C-SiC (carrier of FPA)
[2] 64 % Fe, 36 % Ni
[3] depending on structure of carbon fibers and processing
[4] parallel to carbon fibers
[5] perpendicular to carbon fibers

Another example is shown in Fig. 10 of an ultra-light weight scanning mirror in C/SiC with 520 x 802 x 104 mm^3 and a mass of less than 6 kg.

The proposed material can also be used for the support of the (CCD) detector in the focal plane. This is important because of the large dimensions of the focal planes needed to realize large swath widths. A standard material which has a similar coefficient of thermal expansion than the Si-detectors is Invar alloy which was used for testing the 7k x 8k CCD detector at DLR. However, an Invar-FPA has a much higher weight than a C/SiC focal plane which was alternatively fabricated by DLR and which is thermally very well suited to the CCD detector (cf. Fig.11).

Fig. 10: Ultra-light weight C/SiC scanning mirror (dimensions 520 x 802 x 104 mm^3) with a mass of less than 6 kg (manufactured by DASA-DSS / IABG)

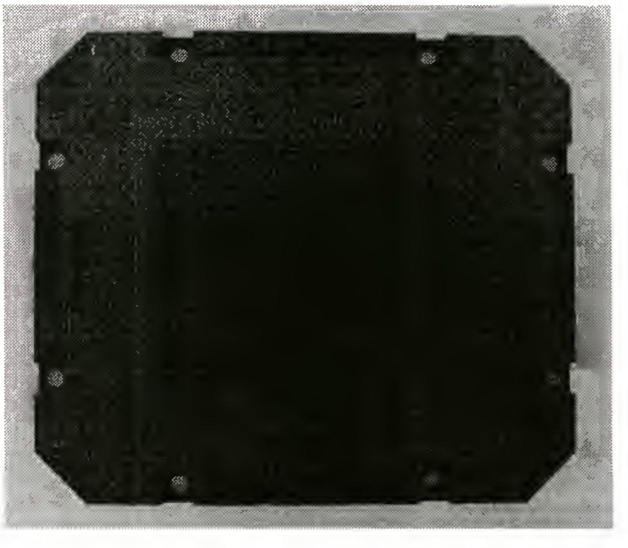


Fig. 11: C/SiC- FPA fabricated by DLR (dimensions 195 x 170 mm^2, thickness 12 mm, mass about 700 g as compared to approx. 3 kg of an equivalent Invar focal plate)

Highly Integrated Detector Electronics

The final, but very important step for the design of a compact, light-weight, and high resolution imaging system for space applications is the application of highly integrated electronics technology. This is a very important measure for the detector part in order to arrive at a high performance (e.g. low noise) at the high readout frequency of the detector that is required.

At DLR two concepts were tested successfully for the detector electronics. The first one is to implement parts of the electronics directly on the CCD-focal plane as shown in Fig. 12. This approach is very useful for the preamplifier and bias. However, at very high readout rates (> 2 MSPS) it becomes more and more important to implement the CCD clock-driver very close to the detector.

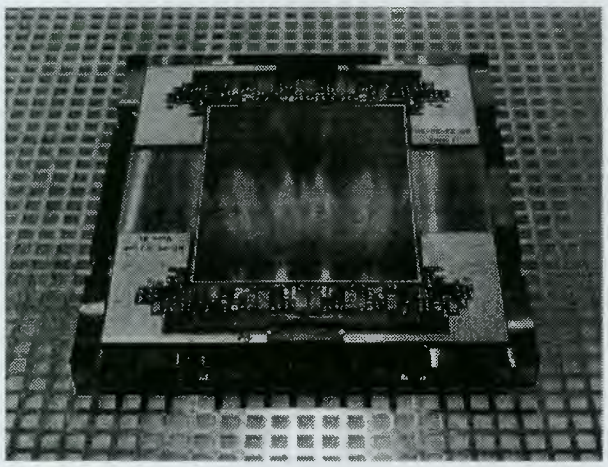


Fig. 12: Philips 7k x 9k CCD with highly integrated FPA-Electronics

Another but complementary concept is the design of highly integrated modules which can simply be expanded to a high performance digital camera as shown in Fig. 13. The highly integrated Modular Sensor Electronic System (MOSES) has been designed by DLR and shows excellent performance also at very high readout rates (10 MHz, 14 bit, cf. Tab. 3). It has the advantage of a large degree of modularity which means that it can simply be adapted to specific applications. Furthermore, MOSES can be fabricated in rather low cost technologies as it will be done for the Rosetta Lander Descent and Downlooking camera ROLIS-D.

Tab. 3: Parameters of the MOSES electronics (measured with CCD THX 7888A)

Parameter	Value
Pixel Readout Time	200 ns*
Noise Floor	45 e^- @ 295K
CCD Noise	43 e^- @ 295K
Electronic Noise	10 e
Responsivity	≅ 7 µV/ e^-
System Gain	15 e^-/DN
ADC Resolution	14 bit

*max. sample rate of the ADC: 10MHz

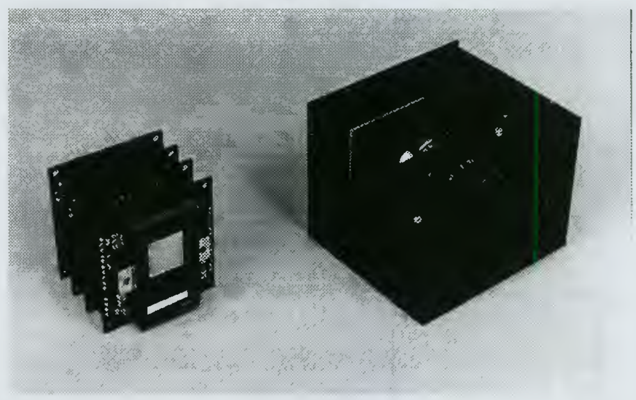


Fig. 13: MOSES (left) with Thomson THX 7888A 1024 x 1024 px Frame Transfer CCD (14 x 14 µm²) ; breadboard dimensions H x L x W = 26 x 48 x 52 mm³, (Mass: ~50 g)

Summary and Outlook

The presented concept shows that:

there is a need for very high resolution imaging in earth observation and planetary exploration;

there is an increasing need for compact, low-mass imaging systems which can be accommodated on small spacecraft;

there are concepts, technologies and new materials available which make low mass and very high resolution imagers feasible;

there are further studies and experiments urgently needed (in Europe) to verify the proposed concepts.

One first experimental step in a space experiment could be the ESA Technology Mission SMART-1 where DLR has successfully proposed a compact imaging system which will be based on the concepts proposed here.

IMAGING SPECTROMETERS FOR REMOTE SENSING FROM SPACE

D R Lobb, Sira Electro-optics Ltd
(Phone, 0044 (0) 181 467 2636, Fax: 0044 (0) 181 467 6515, e-mail: dan_lobb@siraeo.co.uk)

ABSTRACT

The requirements for hyperspectral imaging of Earth from space are briefly reviewed, concentrating on the spectral regions from visible to short-wave infrared. Basic design forms for hyperspectral imagers are compared, including conventional systems using refractive and diffractive dispersion, Fourier transform spectrometers, wavelength-graded filter methods and other devices. Typical performance limitations, including particularly constraints due to detector technology, are noted, and the relative merits of different optical design approaches are considered. The case is made for refractive and diffractive dispersion methods, taking examples from the ESA developments MERIS [1] and PRISM [2] (Medium Resolution Imaging Spectrometer and Process Research by an Imaging Space Mission).

1 IMAGING SPECTROMETERS AND THEIR USES IN SPACE

Imaging spectrometers are instruments that form extended images of remote scenes with significant spatial resolution – typically resolving some hundreds of elements across a scene width. They are distinguished from simple imagers, and from multi-spectral imagers, in providing detailed spectral data on each spatially-resolved scene pixel – a typical imaging spectrometer is capable of providing radiance or reflectance data in order of 100 resolved spectral bands. Imaging spectrometers are therefore sometimes called hyperspectral imagers.

The value of imaging spectrometers in remote sensing of Earth surface (and surfaces of other planets) is in the detailed physical data that can be derived from the spectra of the target surface. Imaging spectrometers are usually designed for the spectral region in which the natural radiance of the target is dominated by reflected solar radiation, including the visible and near infrared (VNIR) bands from about 400nm to 1000nm, and sometimes the short-wave infrared (SWIR) band between 1000nm and 2500nm. In these bands, the recorded spectra include absorption bands due surface chemistry (and also to the intervening atmosphere). The detailed spectral radiance data can be used for example to gauge the health of vegetation, distinguish some vegetation types, and identify minerals in rocks and soil. The imaging capabilities of the instruments allow physical parameters to be mapped over relatively large areas.

The basic advantages of remote sensing from space include frequent and reliable coverage of very large areas of Earth surface. Space-based multispectral imagers, with a few resolved spectral bands, have generated useful data for decades, with Landsat MSS missions starting in 1972 and continuing to date through many Landsat [3], SPOT [4] and other missions. Hyperspectral imagers have not yet been established in Earth orbit, although airborne systems, including AVIRIS [5], have been operated successfully for several years. However we expect imaging spectrometers to extend the success of remote sensing from space, particularly by providing physical data that can be derived only from fine spectral structure. Applications of space-based imaging spectrometers will for example include estimation of biological activity in ocean surfaces from chlorophyll fluorescence, using the MERIS instrument, due to be launched in 1999, and investigation of land surface processes using the PRISM instrument, now in development.

2 PERFORMANCE REQUIREMENTS AND PRACTICAL LIMITATIONS

To be useful, space-based imaging spectrometers must of course provide certain minimum performance in terms of spatial and spectral resolution, image size and spectral range, accuracy and signal-to-noise ratio. Together with the platform, ground stations (and possibly intersatellite communication links), the instrument must be capable of downloading a minimum number of images to ground over a given period, and in general provide repeat coverage of selected targets over a given mission life.

Ideally, of course, we would like to be able to achieve very high spatial resolution over very large scene areas, with large numbers of resolved spectral bands. However, performance must be limited, in different respects, by detector technology, electronics and data handling technology, and by economic scales of optics and structures. Some of the more important constraints are discussed in section 4 below, following a brief discussion of basic design in section 3.

A "typical" set of performance requirements, appropriate for spectral imaging of land targets from low Earth orbit, and achievable now or in the near future, is given below:

- Image area 25km x 25km
- Spatial resolution 25m

Proc. 32nd ESLAB Symp., 'Remote Sensing Methodology for Earth Observation and Planetary Exploration', ESTEC, Noordwijk, The Netherlands, 15-18 September 1998 (ESA SP-423, December 1998)

- Spatial registration ±2.5m
- Spectral range 400nm to 2500nm
- Spectral resolution 10nm
- Spectral registration ±0.5nm
- Signal-to-noise ratio 200 (typical for the band)
- Absolute accuracy ±2%

There are large ranges of actual and potential performance requirements for imaging spectrometers, driven by different mission objectives, so that this "typical" set may be regarded as merely a starting-point for discussion of the practical options. For example, the MERIS instrument, designed for measurement of ocean colour, does not require high spatial resolution and is not required to work in the SWIR region. These limitations allow high performance in other respects, including particularly: coverage of very large ocean areas, and very high signal to noise ratio (needed to resolve very subtle spectral features). **"Image area"**, listed above, is not appropriate within the MERIS specification, since it will image Earth continuously in the sun-lit part of each orbit. Limited image areas tend to be specified for instruments needing high spatial resolution (particularly for land targets) in order to limit problems in data handling.

Spectral resolution of 10nm is generally acceptable for measurement of the absorption features typical of land surfaces and most ocean colour effects, and will also resolve some useful atmospheric parameters, including the effects of aerosols. In principle, higher spectral resolution could be useful to resolve absorption features of trace gasses, such as ozone. However it is necessary in practice to trade spectral resolution against the demands of spatial and radiometric resolution, so that very high spectral resolution is not usually a sensible goal within the basic imaging-spectrometer concept: 10nm is a commonly-specified target.

In terms of radiometric resolution (**signal to noise ratio**) and spectral resolution, MERIS again provides an example in which there is a substantial departure from the "typical" specification. Because MERIS must resolve some very subtle spectral radiance effects, including chlorophyll fluorescence, the noise-equivalent albedo requirement is 0.05 rms. To achieve this very stringent requirement, it is necessary to sum signals from many detector elements, for each spatial and spectral resolution element. For this reason, MERIS has theoretical spectral resolution of 1.25nm, although it will probably not generally be used to resolve better than 5nm.

At present, targets of around 200 for signal to noise ratio, for land targets, and 2% for accuracy, appear reasonable. However, specifications will no doubt develop as experience is gained in application of the data that will be gathered by early-generation instruments.

Spatial registration refers to wavelength-dependent errors in the position of the sampled area on ground. Each sample is typically defined by a set of detector elements that receive the sample image in a set of different colours. In order to measure the spectral distribution of the sample accurately, it is important that the position of the sample shall be the same for all colours, to a small fraction of the sample size in the spatial domain. If not, the spectral distribution will generally be corrupted by the different distributions of neighbouring areas. Spatial registration errors can be partially corrected by re-sampling data. However, this is less satisfactory than use of well-registered data.

Spectral registration refers to variations, across the instrument field, of the resolved wavebands defined by the instrument. In conventional grating and prism spectrometers, these errors are produced by the image distortion called "smile". Spectral registration is inconvenient for data users, since it means that ideal algorithms for generation of higher-level data products should vary across the field. Again, partial corrections can be made by re-sampling of data, if more resolved wavebands are recorded, but this is much less satisfactory than use of well registered data.

3 BASIC LAYOUT AND PUSHBROOM SCAN

The most common basic design form for imaging spectrometers is illustrated in figure 3-1. The optical system is in two distinct parts: the telescope, which forms an image of the remote scene, and a spectrometer that provides spectral resolution of the scene image. The entrance slit of the spectrometer is located between these two parts, at the scene image formed by the telescope. The telescope, together with the slit, defines a line of spatially-resolved scene elements called the instantaneous field of view (IFOV). The spectrometer optics re-image this line of elements onto an area-array detector, so that each column of detector elements receives all the radiation from one spatially-resolved scene element. The number of spatially resolved elements is equal to the number of detector array columns. However, the spectrometer introduces spectral dispersion in the direction of the detector column. Thus, each row of detector elements receives the image of all elements in a narrow spectral band, and the set of signals from each column gives the spectral radiance distribution of one spatially-resolved sample (in each frame read out from the detector).

Signals from the detector array are read out during a frame period to provide detailed spectral data for each resolved point in the strip-shaped IFOV. For space-based systems the IFOV is usually oriented across the flight path direction (across-track), so that the satellite motion then moves the IFOV over the target surface (in the along-track direction). This is called a "push-

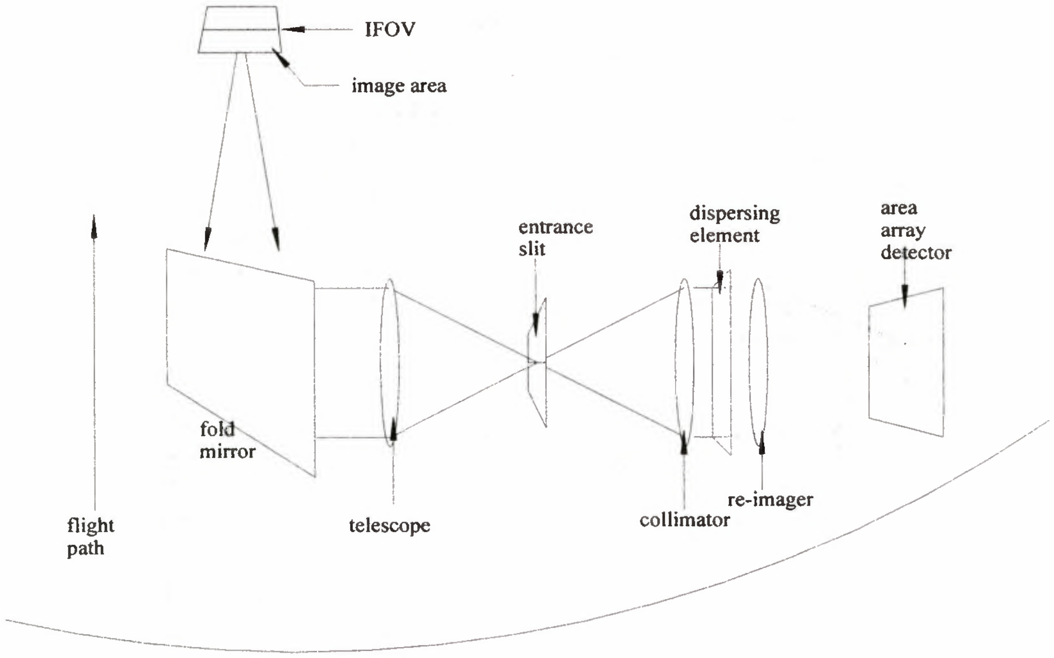


Figure 3-1 Conventional imaging spectrometer concept, with imaging concept

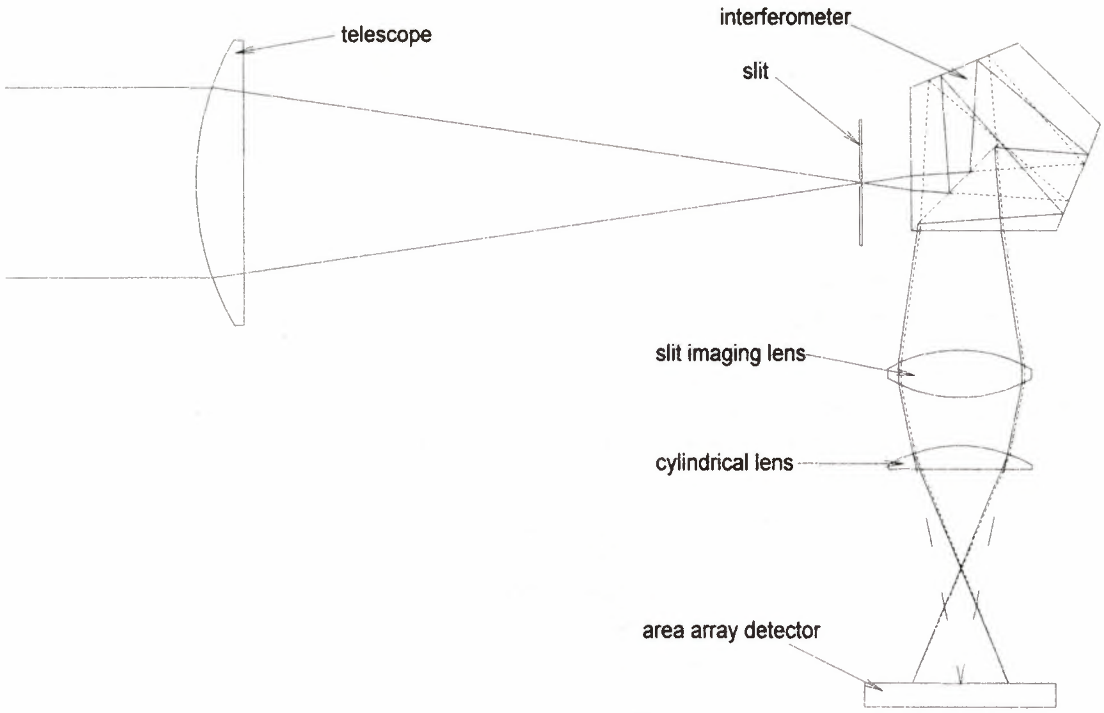


Figure 5-1 Spatially modulated Fourier transform spectrometer

broom" scan. Commonly, the system is designed so that the instantaneous field moves through its own width during the frame period; the distance moved in the detector integration period is called the ground sampling distance (GSD).

Thus in pushbroom imaging-spectrometer systems, the imaging and spectral resolution functions are divided as follows:

- Across-track image scan — detector rows
- Along-track image scan — satellite movement
- Spectral resolution — detector columns

Large numbers of detector array elements are in general required to perform the across-track spatial resolution and spectral resolution functions.

Before the advent of well-developed and large area-array detectors, across-track scan was performed, in whiskbroom imagers, by an opto-mechanical scan. The simplest whiskbroom system has a single detector element in the focal plane, assigned to each spectral waveband, producing a single-point IFOV (for each resolved spectral band) that is rapidly scanned across the total field of view by the mirror.

Whiskbroom systems have some significant advantages for some imaging radiometer specifications. However, they have not been considered very seriously for imaging spectrometers. It is generally desirable to avoid the need for a scanning mechanism, but the main reasons related to basic radiometry. A pushbroom system typically assigns several hundreds of detector elements to each spectral waveband, continuously, while a whiskbroom system assigns only one or a few. Pushbroom systems can therefore generate much larger total signal currents, for given optical aperture sizes, and offer much better signal-to-noise ratios.

4 DETECTORS AND SOME RADIOMETRIC CONSIDERATIONS

Area array detectors must be used for imaging spectrometers operating in the pushbroom scanning mode, as indicated in figure 3-1. In general, the size of the array in one dimension determines the number of spatially-resolved elements across the swath width, while the array size in the orthogonal direction determines the number of resolved spectral bands.

4.1 VNIR band arrays

For the VNIR band, there is generally no serious alternative to use of silicon area-array detectors – CCDs are normally used. Arrays exist with up to 4000 columns, but, for a variety of reasons, the arrays actually used for imaging spectrometers have been limited to about 1000 columns or less. As the size of the array is increased, it becomes more difficult to read out the array at the rates required for remote sensing, and of course the data generated becomes more difficult to handle. The ideal VNIR detector is usually a thinned, back-illuminated device, that has relatively high quantum efficiency (since the incident radiation does not pass through polysilicon electrode structures). The devices are preferably coated to optimise quantum efficiency at the extremes of the required spectral band – about 60% can be achieved at 400nm, and 10% at 1000nm.

Typically, CCDs can be used at the average temperature of the space instrument, without special cooling. Relatively large useful signal levels are required to overcome the effects of photon noise, and dark signal levels are generally low, because integration times are short. Again an exception, the MERIS detectors are moderately cooled to reduce dark signal, since MERIS has relatively long integration times (associated with low spatial resolution) and exceptionally stringent requirements on control of drift errors. In detailed radiometric analysis for space-based instruments, it is necessary to take into account the effects of ionising space-radiation on detector performance. For CCDs, the main effect is to increase dark signal levels and dark signal non-uniformity.

4.2 SWIR band arrays

For imaging spectrometers that cover both the VNIR and SWIR bands, it has generally been found necessary to use two separate area-array detectors. There is less history in development of arrays for the SWIR band, so that arrays are generally more difficult to obtain, more expensive, and more limiting on the overall instrument specification. At present, there is a relatively wide range of candidate materials to be considered, including: indium antimonide (InSb), cadmium mercury telluride (CMT) platinum silicide (PtSi) and indium gallium arsenide. PtSi can be made in large arrays, so that is generally worth some consideration at the start of any new development. However, it has low quantum efficiency, which tends to present serious problems for signal to noise ratio, particularly for the longer SWIR wavelengths where the solar spectral irradiance is low.

InSb area-arrays are available with up to 1000 columns, and a 1000-column CMT array (produced by butting smaller arrays) has been made as part of the HRIS development programme. (Another reason for limiting VNIR band detectors to about 1000 columns is that they then match current possibilities for the SWIR band, in instruments that include both bands.) InSb and CMT arrays provide fairly good quantum efficiencies over SWIR bands – typically over 50%. CMT arrays for the spectral range out to 2400 nm require cooling typically to around 150K. However, arrays with less stringent cooling requirements can be used for the SWIR band out to about 1700nm. InSb arrays need lower temperatures.

The need for cooling of SWIR detector arrays is likely to be a significant cost driver for imaging spectrometer instruments and platforms, so that there is a need to consider carefully where the long-wave limit should be set for any particular mission. Instruments working only to 1600 or 1700nm will probably be significantly less difficult than instruments covering the SWIR band beyond 2000nm, and of course instruments limited to the VNIR band will be much simpler. In this context, we should note that the main interest in wavelength above 1700nm appears to be in identification of minerals, using the atmosphere window between 2000nm and about 2400nm.

4.3 Signal levels

The signal charge generated on one detector element, in each frame, can be estimated using the relatively simple formula:

$$C = R.A.\Omega.T.S.\tau.d\lambda$$

Where R is the spectral radiance of the target surface, A is the detector element size, Ω is the solid angle imaged onto the detector, T is the optics transmission, S is the detector sensitivity (Amps/Watt), τ is the integration time, and $d\lambda$ is the spectral waveband imaged onto the detector element. For available large arrays, the detector element is typically around 0.02mm square. The optics transmission factor can often be large – for example between 0.5 and 1 - and the spectral waveband is around 10nm. Sensitivity levels are determined by detector quantum efficiencies, which can be high for most wavelengths in both the VNIR and SWIR bands. Integration times vary with the spatial resolution of the instrument (assuming a conventional pushbroom mode of operation). Instruments with the highest resolutions (that are at present reasonably realistic) have integration times of a few milliseconds, while MERIS, with a nominal resolution of 250m to 500m, has an integration period in tens of milliseconds.

Thus there are typical values for T, S, τ and $d\lambda$. Typical signal levels are then determined largely by the detector element area A, the f/number of the optics ($\Omega = \pi.(2.\text{f/number})^{-2}$ and by the radiance R of the target surface. When the target surface is the sun-lit Earth, it is generally fairly easy to design optics such that the charge density, C/A, on some detector elements, reaches the saturation level for the elements. For high resolution instruments, with the shortest integration times, saturation is typically reached, for some parts of the spectral range, when the optics aperture is about f/4. Two related conclusions are generally justified, for hyperspectral imaging of Earth from space:

(a) relative apertures of optics are generally limited to values (like f/4) that are fairly easy for optical design, and

(b) it is therefore sensible to design the optics such that maximum signal levels (typically from targets at about albedo 1 in direct sunlight) are close to detector saturation levels.

Of course these common conclusions will not be good for all imaging spectrometer developments. For example, an instrument designed only for the long-wavelength SWIR region, where Earth radiances are low, may need to be designed for higher relative aperture (requiring more difficult optics) if high signal-to-noise ratios are required. Higher apertures will be desirable for instruments designed specifically to work over scenes of lower radiance, possibly including outer planets. The conclusions will also be different, at least in detail, for airborne imaging spectrometers, for which integration times are much more variable (but usually longer, so that low relative apertures still tend to allow operation near to detector saturation).

4.4 Saturation levels and signal to noise ratios

The discussion above makes the point that signal levels, at least for most Earth-sensing imaging spectrometers, will tend to be limited by detector saturation levels. There is therefore of course a case for detector elements of relatively large size, for example >0.03mm square, to provide large saturation charge levels. However, there is generally a requirement to use arrays with fairly large numbers of elements (though with some consideration given to problems of signal handling and data handling), and arrays with large numbers of columns tend to have smaller elements. For a system providing 1000 spatially-resolved elements (using 1000 detector columns) detector elements will typically be about 20 μm square, and for such element sizes, element saturation charges are typically in the order of 500,000 electrons, per integration time.

Generally, Earth-orbiting instruments will be designed such that the maximum Earth radiance levels, derived from cloud and snow, give signals approaching saturation, so that most signal levels will be lower. For example vegetation reflectances in the visible range are typically 5% to 10%, rising to order of 50% in the near IR. However most signal of interest will be above 10^4 electrons, and will more typically be in the order 10^5 electrons.

Random temporal noise levels in hyperspectral imagers tend to be limited mainly by a combination of photon noise and detector read noise. The root-mean square (rms) photon noise for each element (in each frame) is calculated as the square root of the total signal charge, and is typically in the order 300 electrons. Read noise for CCDs is normally less than 100 electrons rms. SWIR detectors generally have somewhat larger read noise, but photon noise again tends to dominate.

High scene radiances within the instrument range will provide higher signal quality. But typical achievable signal to noise ratios, for high-resolution imaging spectrometers, tend to be in the region of 200, mainly because of the constraint imposed by detector saturation levels.

Of course this conclusion needs considerable qualification. In particular, larger detector elements will improve the achievable signal to noise ratios, in applications where it is reasonable to use relative low-number detector arrays or to develop special devices. It is also reasonable to consider summing signals from several detector elements and/or several successive frames to reduce the effects of noise. Using CCDs, summation of signals can be performed by charge handling on the chip before read-out, so that addition of signals is possible with read noise limited to that of one read-out per summed group. As noted above, the MERIS instrument is designed with excess optical resolution in both the spectral and spatial domains, so that signals from many elements are summed to produce the required spectrally and spatially resolved pixels. This allows signal to noise ratios of several hundred to be achieved for scene areas of moderate albedo, at low spatial resolution.

5 WAVELENGTH-DISCRIMINATION OPTIONS

5.1 Grating and prism dispersion

Classical spectrometers work by dispersion of the light from an entrance slit, as indicated schematically in figure 3-1. The dispersion devices may be either refracting prisms or diffraction gratings. The effect of these old-fashioned optical arrangements is simply to spread the radiation from each point in the slit to form a line spectrum on a detector column. Minimal data processing is required in these cases. The spectral radiance of each spatially-resolved point can be read simply as the column signal distribution, after calibration of the raw digitised data for offset and gain.

The main alternatives to prism and grating dispersion are use of two-beam interferometry, in Fourier transform spectrometer (FTS) configurations, and use of wavelength graded filters. Both these alternatives will be discussed briefly.

5.2 FTS

In FTS systems, the incoming radiation is split into two component beams that follow different paths before being recombined and imaged on the detector. The path difference for the two beams is modulated, usually by moving a mirror, so that interference between the two beams, at each wavelength, generates an intensity variation. The frequency of the intensity variation is inversely proportional to the wavelength, so that the source radiance at each wavelength can be deduced from the amplitude of the signal variation at the appropriate frequency. This means that the source spectral radiance distribution can be calculated as the Fourier transform of a detector signal distribution.

For space-based imaging spectrometers, the path-length modulation will probably not be produced by moving a mirror, since very accurate movements are very expensive in space developments. Also, if a linear detector array is used, with a movement, this implies that relatively little signal will be collected in total, with severe implications for possible radiometric performance. The likely form for a space-based FTS system uses a fixed interferometer with an area-array detector, as indicated schematically in figure 5-1. An instrument of this form is described by Otten et al [6]. Scene radiation is imaged on a slit and the slit is re-imaged onto rows of the detector array (as in conventional spectrometers). After the slit, the radiation passes through an interferometer, which splits and recombines the beam with a path length difference that varies as a function of the ray angles in the section orthogonal to the slit. A cylindrical lens is introduced into the recombined beam following the interferometer, which defocuses the slit image in the same section, spreading the image along detector columns. Each detector row receives the slit image, but each row is associated with a different path-length difference introduced by the interferometer. Each detector column therefore contains the Fourier transform of the spectral radiance of one spatially-resolved element.

FT systems have some distinct advantages in their many applications. They can give very fine spectral resolution. In comparison with scanning monochromators (using prism or grating dispersion) they provide a radiometric advantage, known as Felgett's advantage, due to the fact that they detect all wavelengths continuously during the measurement cycle. If movements are avoided, for example as described above, the optical system can be quite compact and robust.

However, the characteristic advantages of FTS do not appear to be well adapted to the most common requirements for space-based imaging spectrometers.

Very fine spectral resolution is not typically needed, and tends to be ruled out by radiometric considerations and data handling problems. There is no Felgett advantage in comparison with grating or prism spectrometers that also use area-array detectors. In fact there is a considerable radiometric disadvantage due to additional data processing required in FTS. The FTS raw signal associated with each spatially-resolved element and wavelength is distributed over all elements in the column. The transform that provides the radiance in the band is essentially a weighted sum of all the column signals. The noise associated with the result is therefore the typical element noise multiplied by the square root of the number of columns. This square root multiplier is the critical FTS disadvantage.

Using FTS, it is in principle possible to gather more signal in total, in comparison with a grating or prism spectrometer, by using more detector elements in the column. (In fact it is necessary to use more columns to achieve the same minimum spectral resolution.) However, the element signal is still limited by saturation, and the number of columns that can be read out will generally be limited by feasible detector size and/or by problems of signal and data handling. It will be difficult for FTS to compete radiometrically, except in relatively simple systems with low numbers of spatially and spectrally resolved elements.

5.3 Wavelength-graded filters

An effect very similar to that of a grating or prism spectrometer can be achieved, with much simpler optics, by using a wavelength-graded interference filter. In this case, as indicated schematically in figure 5-2, an area-array detector is placed at the focal plane of the telescope, with the filter immediately in front of it. The thin film layers of the filter are each wedge-shaped – an effect achieved typically by moving a mask across the substrate as each layer is deposited. The transmitted wavelength therefore varies across the filter area. Ideally, each line of constant wavelength is straight and parallel with a detector row.

It is necessary for the detector to image a two-dimensional area of the scene. It is assumed that the scene image moves over the detector, roughly or exactly parallel with detector columns. Each spatially-resolved scene element is therefore imaged in many successive frames, with the recorded wavelength changing.

The advantages of the wavelength-graded filter are fairly obvious. The optics can be very compact, and simplification of the optics reduces the likely overall system cost. Radiometric performance, in terms of signal-to-noise ratio, is likely to be similar to that of a prism dispersion spectrometer system with the same detector and the same spectral and spatial resolution.
One of the most important disadvantages of the concept, for space-based imaging spectrometers, relates to spatial registration, as discussed in section 2. Good spectral registration requires (a) that the scene image moves exactly parallel to detector columns, and (b) that the scene image moves at a speed of exactly one row in one frame period. These conditions can be achieved in principle when a space instrument with a small field is viewing the Earth at nadir. However, it is usually desirable for the instrument to be capable of working at a range of pointing angles to acquire targets that do not pass immediately below the satellite. The image of a target viewed at oblique angles does not move along parallel lines, or at a speed that is constant over the image, so that spatial registration is sacrificed. Instability of the platform may also present problems.

Other difficulties include:

- achievement of large spectral ranges, with suppression of filter side-bands,
- stray light generated by reflections between the detector and the filter and
- limited spectral resolution due to filter-detector separation and filter limitations.

However, for some applications, the wavelength-graded filter concept is likely to be found acceptable, and may be preferred for low cost and mass.

5.4 Chromotomographic systems

Various schemes have been suggested, for example by Mooney at al [7], in which a large area of image is used simultaneously, as in the wavelength-graded filter concept, with various kinds of dispersing or FT optics. For example, as indicated schematically in figure 5-3, an area of image is formed on a field stop by a telescope, and the whole image area is re-imaged onto an area array detector through a diffraction grating. The resulting image is the original scene image with wavelength-dependent shifts. The data can be processed by tomographic methods to yield the spectral distribution of each spatially-resolved element in the area.

In general, these methods seem to have a fairly severe radiometric disadvantage, since the tomographic data processing combines noise contributions from many detector elements into each spatially and spectrally resolved output datum. However, the methods may find some niche applications, particularly where it is necessary or convenient to take snapshot area images, rather than perform a pushbroom scan.

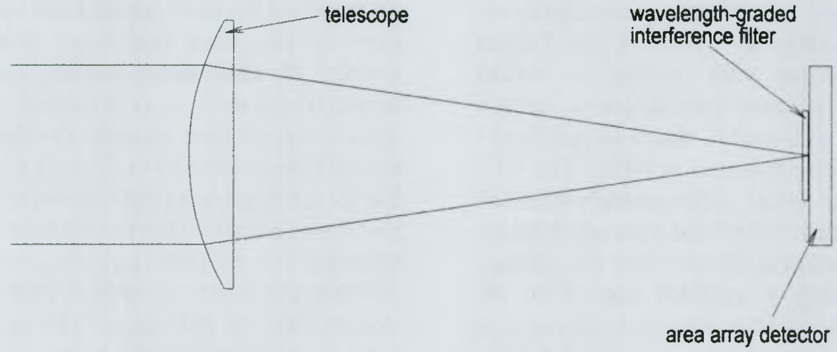


Figure 5-2 Wavelength-graded interference filter system

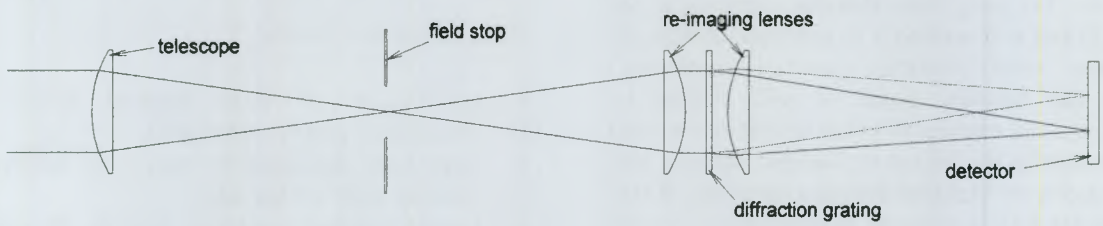


Figure 5-3 Chromotomographic spectrometer system

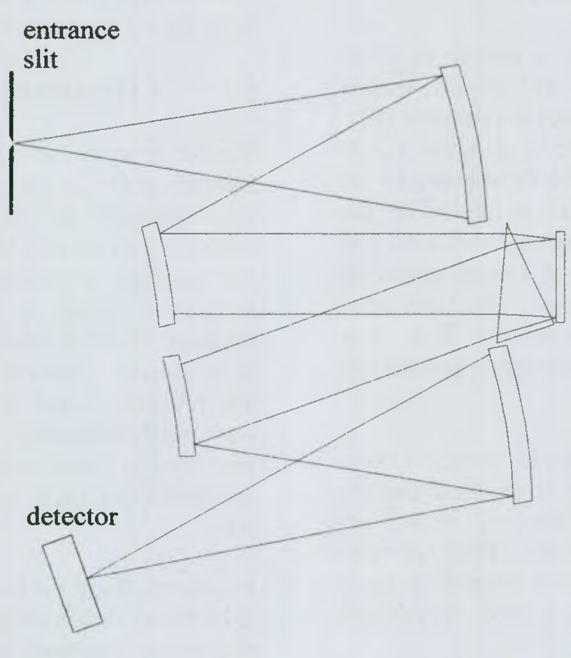


Figure 6-1 All-mirror collimator and re-imaging spectrometer systems - schematic

6 GRATING AND PRISM SYSTEM DESIGNS

Grating and prism spectrometers are probably the best options for space-based imaging spectrometers that are required to provide high performance, in terms of spectral, spatial and radiometric resolution and accuracy. The classical layout is shown schematically in figure 3-1. Following the telescope and entrance slit, the radiation is collimated by a first imaging system, then dispersed by a grating or prism, and then re-imaged onto area array detector by a second imaging system.

6.1 Relative merits of gratings and prisms

Gratings can produce much higher spectral resolution than prisms. However, this is not usually the most important consideration for hyperspectral imagers, since spectral resolution around 10nm is quite feasible in prism systems of acceptable scale. Gratings are less efficient than prisms (in transmission), particularly over large spectral bands, and they produce overlapping spectra in different orders. There is therefore a fairly strong tendency to prefer prisms to gratings for most imaging spectrometers.

As usual, there are significant exceptions. For the MERIS instrument, fairly high spectral resolution (1.25nm) was a useful means of increasing the numbers of detector elements from which signal charges could be integrated, to achieved the required very high signal to noise ratio. MERIS also has a very wide total field of view that can be achieved only by using several separate spectrometer modules (5 modules with 14° fields). A prism system would have been feasible in principle, and would have had some important advantages, but it would necessarily have been very large. The design team preferred a uniquely compact grating system, described briefly in 6.3.2.

Grating/prism combinations, called grisms, are used in spectrometer systems with relatively limited spatial-resolution requirements. High dispersion is provided by the grating, while the prism produces low dispersion in an orthogonal direction to separate the grating orders. Again, grisms do not generally appear to be appropriate for pushbroom-scanned imaging spectrometers, since the cross-dispersion domain is reserved for spatial resolution.

One objection to prisms is that they produce spectral resolution that varies with wavelength – high at short visible wavelengths, reaching a minimum at short SWIR wavelengths, and rising slightly at longer SWIR wavelength. Some designs have included compound prism arrangements, using two materials, to produce a more uniform spectral resolution across the spectral range. However, it is fairly easy to use the on-chip charge-handing capabilities of CCDs to sum signals from neighbouring sets of detector rows. The spectrometer optics can be designed simply, using a single prism material, to give spectral resolution typically 10nm at 1000nm. It will have spectral resolution "improving" to about 1.2nm at 400nm, but on-chip binning of signals from 8 rows can (if required) produce spectral resolution of 9.6nm at 400nm also. In this scenario, spectral resolution changes step-wise at selectable points in the VNIR band, but this is not obviously a significant disadvantage for data users. In fact, users are offered the possibility of finer control in selection of spectral bands in the visible range.

For space-based imaging spectrometers, the preferred prism material will usually be fused silica. The main reasons are that this material has good transmission over the whole spectral band from near ultraviolet through SWIR, and it is insensitive to the effects of space radiation. The material has low dispersion in the visible range, in comparison with optical glasses, but its dispersion is comparable with that of glasses in the near IR and SWIR regions, and more uniform over the whole VNIR and SWIR bands.

6.2 Design of telescopes, collimators and re-imagers

The optical design problems vary in detail for different missions, but are not usually very difficult: as noted in section 4.3, the f/numbers of the optics are moderate, so that relatively simple imaging arrangements can be used to resolve order of 1000 elements across the field. There are two special difficulties in general. First, the spectral ranges of imaging spectrometers are typically much wider than those of most imaging systems, so that there are few useable refracting materials and chromatic correction can be difficult. Secondly, it is necessary to control distortion of the spectrum image formed on the area-array detector to achieve the required spectral and spatial registration accuracies. In particular, it is necessary to correct the line-image curvature that is characteristically produced by both gratings and prisms, called "smile" by those involved in imaging spectrometers.

6.2.1 Refracting and reflecting options

The refracting options should not be discarded without some consideration, in any new development. Generally adequate chromatic correction can be achieved over the whole VNIR band and (separately) over the whole SWIR band, particularly by using fluorite for some elements, with conventional optical glasses. In general, some separate optics are required for VNIR and SWIR channels of a spectrometer that combines both bands, but this is not necessarily inconvenient, since separate detectors are in any case usually required. It is even possible, for low-aperture

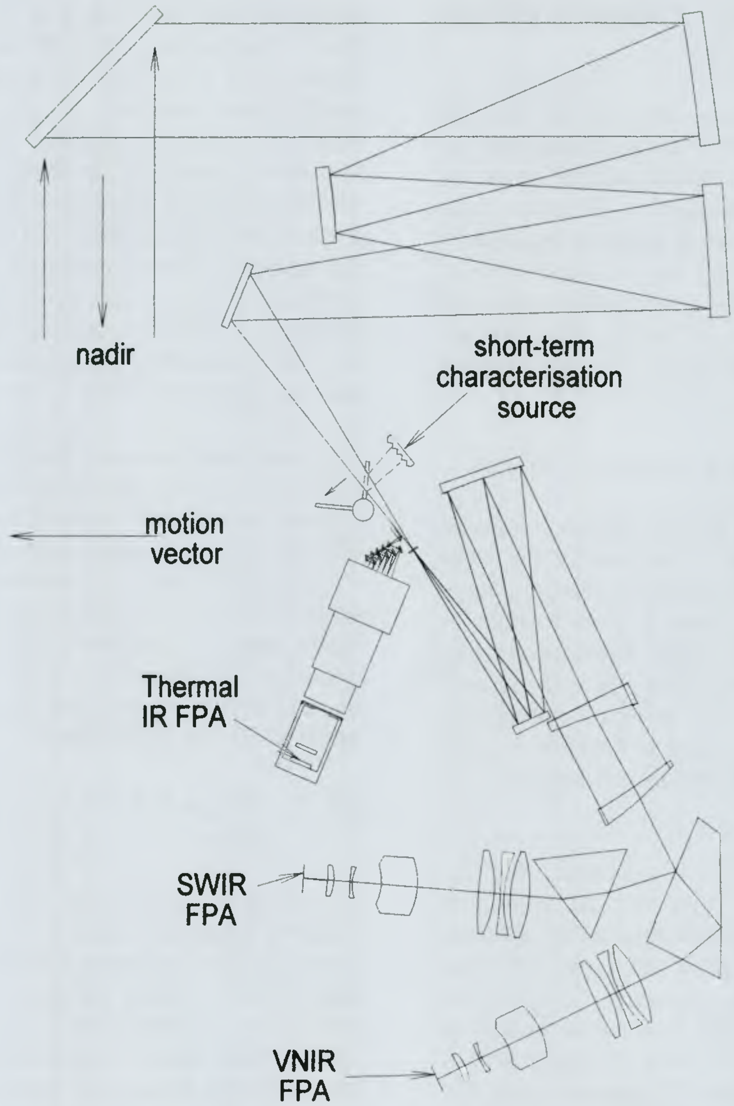


Figure 6-2 An early spectrometer concept for PRISM

Figure 6-3 Concentric grating spectrometer used for MERIS -schematic

telescopes (suitable for airborne spectrometers) to achieve adequate correction over the VNIR and SWIR bands together in a single refracting lens.

However, there is of course a tendency to use mirror imaging systems, in both telescopes and spectrometer sections, to avoid problems of chromatic correction over wide spectral ranges. For space-based systems, this decision tends to supported by other disadvantages of refracting systems: use of fluorite tends to make systems temperature sensitive, glasses must often be chosen from the very limited range of radiation-resistant types, and stray light problems are made more difficult by the inefficiency of wide-band anti-reflection coatings.

The conventional design approach using mirrors leads to some relative expensive optics, essentially because the beam is collimated at the grating or prism, to avoid aberrations produced by the dispersing element(s). Design of each imaging unit, the collimator and reimaging optics, requires at least two mirrors, as indicated schematically in figure 6-1. If only two mirrors are used, both must be aspheric, so that the whole spectrometer system tends to need at least 4 off-axis aspherics in total. This basic problem can be avoided by several strategies. At present, a very promising strategy is use of concentric system, described in section 6.2.3, which depart from the conventional collimator/disperser/re-imager configuration.

Other approaches include all-refracting (or mostly-refracting) optics, as discussed above. Catadioptric systems, in which refracting elements are introduced in mirror imagers essentially to avoid the need for aspherics, present a vast range of different design options, which have been investigated in detail in the ESA HRIS [11] programme, which may be considered a precursor to PRISM. Figure 6-2 shows an initial concept for the PRISM instrument, including an all-mirror telescope (a 3-mirror anastigmat), a catadioptric collimator and two separate refracting lenses for the VNIR and SWIR bands. The system also includes a set of thermal IR channels. There is a need for special care in design of catadioptrics to avoid serious stray light problems, but in general catadioptric designs can be expected to produce some acceptable compromises.

6.2.2 Smile correction tactics

Correction for "smile" is an added complication in design of spectrometers, but there are several possible tactics for control of the problem. It is important to remember that distortion in conventional lens and mirror-imaging systems produces curvature of images, so that one method to compensate smile is to make deliberate use of distortion in axially-symmetrical lens or mirror system, by using them in off-axis parts of their fields. Alternatively a single off-axis element can be assigned the main function of correcting smile – a decentred lens or a tilted spherical mirror close to the entrance slit can provide flexible smile control (and may also help in pupil imaging and/or field-curvature correction) without introducing high-order aberrations. In some cases it may be acceptable simply to curve the entrance slit, although this may mean that the IFOV is curved.

6.2.3 Concentric optics

Three optical design forms are drawn in figures 6-3, 6-4 and 6-5, to show some of the possible flexibility of optical design for spectrometers, and in particular to illustrate the concentric-optics design strategies [8,9]. The overall function of optics following the entrance slit is to image the slit onto the detector. This function can be performed by unit-magnification concentric systems, which require only simple spherical surfaces, and can therefore be relatively cheap. The problem is that concentric designs do not provide any collimated beam path, in which a conventional flat-surfaced grating or mirror can be placed without introducing large aberrations. However, it is possible to introduce gratings or prisms into concentric optics, with good control of aberrations, if the gratings or prisms are allowed to be curved. Concave spherical gratings and dispersing prisms with curved surfaces (Féry prisms [10]) have both been know for several decades for applications in simpler spectrometer systems.

Figure 6-3 shows the MERIS spectrometer design, which has a concave spherical reflecting diffraction grating, and a concentric refracting corrector. Figures 6-4 and 6-5 show alternative design forms produced for the PRISM instrument by two competing teams, both using a three-mirror (Offner relay) concentric imager in which curved refracting prisms are introduced. The essential optical logic is:

- spherical aberration and coma are initially corrected by near-constant values for the angles of incidence of all rays in each beam, at each surface,
- astigmatism is balanced between surfaces and
- Petzval sum is near zero.

In optimisation, particularly in the case of prism systems, the values of coma at individual surfaces are allowed to become substantial, but balanced, as correction for astigmatism and smile are optimised by changes in angles of the prism and mirror surfaces.

The concentric systems are not only relative cheap, in comparison with aspheric mirror systems, but also compact. They are capable of producing well-resolved images for the complete VNIR and SWIR range, since chromatic aberrations are very small (although the MERIS grating system is limited to the VNIR band).

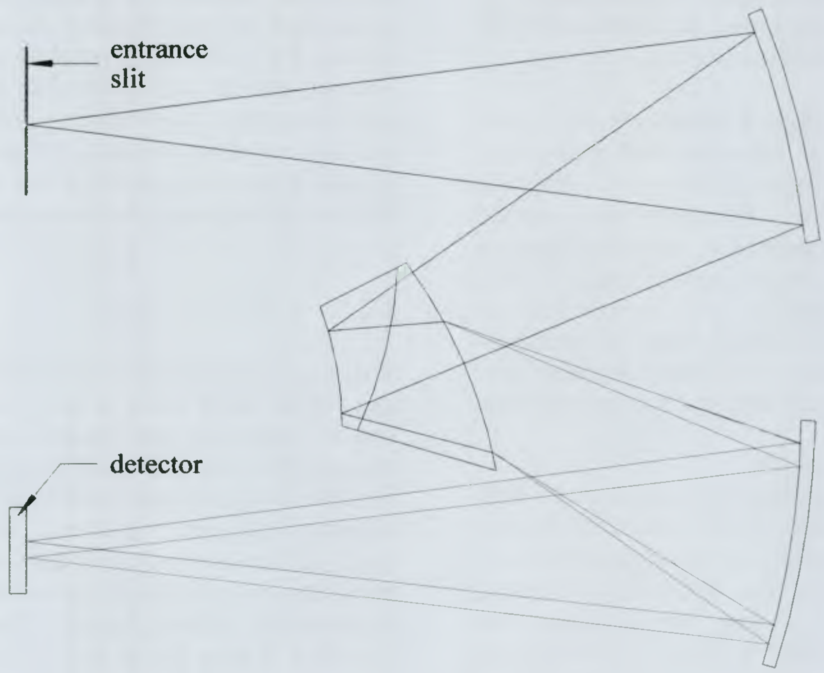


Figure 6-4 First curved-prism concept for PRISM spectrometer - schematic

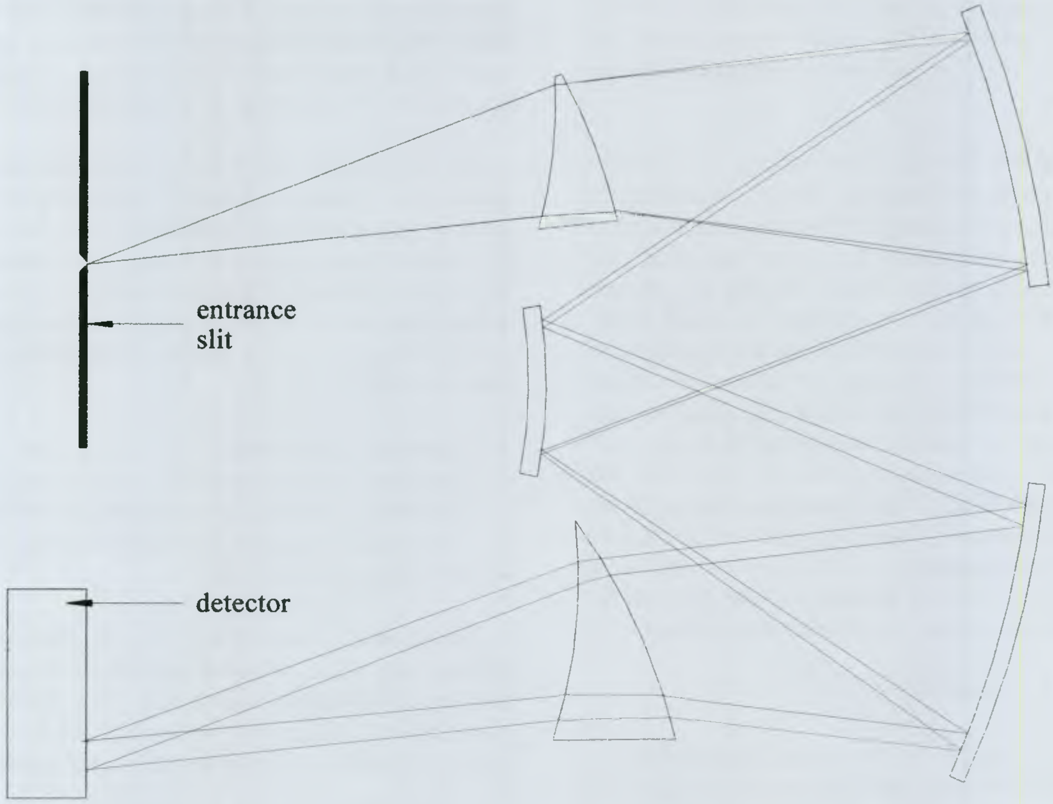


Figure 6-5 Second curved-prism concept for the PRISM spectrometer - schematic

The MERIS design gives near-perfect correction for smile. The curved prism systems correct smile with a little more difficult, good correction is feasible with only three spherical mirrors and two or three prisms (depending on the field size and f/number).

ACKNOWLEDGEMENTS

Most of the author's experience in design and analysis of imaging spectrometer systems derives from projects carried out under the direction of ESA, including MERIS (pre-phase A, and phases A phase B) PRISM pre-phase A and APEX phase A. Many concepts have been evolved and evaluated in collaboration with experts at ESTEC, Sira, Aerospatiale, DASA, MMS and Delft Sensor Systems. Sira also acknowledges the support of BNSC, NERC and ESA in the present development of the Sira CHRIS instrument, which will fly on the ESA PROBA platform.

REFERENCES

1 G Baudin, R Bessudo, M Cutter, D Lobb and J L Bézy, "Medium resolution imaging spectrometer (MERIS)", SPIE 1490, 1991.

2 U Del Bello and R Meynart, "Hyperspectral imager for the future land surface processes Earth Explorer mission", SPIE 2957, September 1996.

3 B L Markham, J L Baker, E Kaita and I Gorin, "Landast-7 Enhanced Thematic Mapper Plus: radiometric calibration and pre-launch performance", SPIE 3221, September 1997.

4 M Chevrel, M Courtois and G Weill, "The SPOT satellite remote sensing mission", Photogrammetric Engineering and Remote Sensing, Vol. 47, No. 8, August 1981.

5 S A Macenka and M P Crisp, "Airborne visible/infrared imaging spectrometer (AVIRIS) design and performance, SPIE 834, 1987.

6 L J Otten, J B Rafert and R G Sellar, "The design of an airborne Fourier transform visible hyperspectral imaging system for light aircraft environment remote sensing", SPIE 2480, April 1995.

7 J M Mooney, A K Brodzik, and M An, "Principal component analysis in limited-angle chromatography", SPIE 3118, July 1997

8 D R Lobb, "Theory of concentric designs for grating spectrometers", Applied Optics, Vol. 33, No. 13, May 1994.

9 D R Lobb, "Imaging spectrometers using concentric optics", SPIE 3118, July 1997.

10 C Féry, "A prism with curved faces, for spectrograph or spectroscope", Astrophysics J., Vol. 34, PP 79-87, 1911.

11 F Blechinger et al, "High resolution imaging spectrometer (HRIS): optics, focal plane and calibration, SPIE Vol. 1937, 1993.

THINNED CCD SENSORS FOR SPACE APPLICATIONS PRESENT STATUS AND FUTURE EXPECTATIONS

Gilles Ch. Boucharlat

Thomson-CSF Semiconducteurs Spécifiques
B.P. 123
Avenue de Rochepleine
38521 Saint Egrève, France
Tel. +33 476 58 31 46 / Fax +33 476 58 34 97 / gilles.boucharlat@tcs.thomson-csf.com

ABSTRACT

Solid State Image Sensors are nowadays widely established and commonly used in space instruments, as natural replacement parts of previous electron tubes or single detectors, as well for imagery as attitude sensing. Applications in cameras for earth observation, planet exploration, as well as star tracking, sun and earth sensing, X-ray telescopy, are nowadays flying and demonstrating quality, reliability and suitability of such components.
News trends in space program are in reducing drastically volume and weight of on-board instruments, and in designing micro-satellites, instead of wide space platforms.

As a response to these new considerations, optical imaging instrument designers have to reduce focal plane area, focal lengths and optical apertures, in order to improve the compactness of the whole system. They are then asking for reduced pixel size, increased detector responsivity while they are also looking for greater resolution and wider spectral range.

Pro and con's analysis of some potential technologies, including CMOS, are presented, and the capabilities of modern thinned backside illuminated CCD technology are considered as a natural way to solve the here-above dilemma.

1. INTRODUCTION

Permanent trend in miniaturizing object applies to space platform as well as to domestic appliance. Thanks to new generations of electronic components, home radio, TV set and walkman are smaller and smaller, minimum scale being now put by button size limitation according to human morphology more than limited by technical or physical considerations in the manufacturing process.

Space instruments take advantage of this scaling down tendency, as usually reducing component size is also a way to reduce mass and increase instrument compactness. Relying on this regular year to year reduction factor, and driven by the increasing need of information and communication, new concepts of micro-satellites are emerging for nevertheless ambitious space program.

2. HISTORY

At Thomson-CSF, CCD's are indebted to imaging Electron Tube Technology from the beginning in the early 70's, as they are the result of research in camera tube. At this time, CCD were developed in the clean room where Silicon targets for Vidicon tubes were being manufactured, taking full advantage of a perfect mastering of ultra-clean Silicon diode process.

Silicon CCD's have been competing with their mother technology, especially for space application, where their size and weight, their solidity and reliability were a major bounce ahead for their introduction in the design of space electro-optical instruments. According to the Fig. 1, showing a Vidicon tube and his CCD competitor, the technological gap is so clear that there is no more convincing demonstration to do : CCD is smaller, lighter, more reliable, less power consuming : the door is open to enter the space imaging era (Table 1).

	Vidicon tubes	Imaging CCD's
Overall dimension	Φ 1" x 5"	25.4 x 36 x 2 mm^3
Weight	65 g	7 g

Table 1.

3. SPACE ELECTRONIC IMAGING

Several ways were explored for imaging from space satellite. Beside the classical way using area array imagers in a snapshot mode, cleaver systems involve the natural movement of the satellite w.r.t. the earth for mechanically scanning the scene to be imaged.

3.1 Barrel rolling satellites

One way of scanning the earth surface is copied on the electron beam tube principle, in which the image focused on the tube target is electrically readout dot after dot, through an electron beam. In satellite

Proc. 32nd ESLAB Symp., 'Remote Sensing Methodology for Earth Observation and Planetary Exploration', ESTEC, Noordwijk, The Netherlands, 15-18 September 1998 (ESA SP-423, December 1998)

Fig.1 : A Vidicon electron tube and its son, a CCD image sensor

operating in the barrel rolling mode, MeteoSat for example, the image is analyzed with only one "light beam" readout by one detector : such satellites take advantage of the controlled spin of the satellite around its axis, to image the earth by a rotating "eye beam", focused on only one detector (Fig. 2) This system is very simple, as the inertial mass of the satellite gives a natural regulation of its rotating speed, and its move along its trace is regulated according to its altitude. The limitation of the imaging accuracy of such a system is tied to the distortion of the beam trace along the ideal line, due to the spherical aspect of the planet, the cylindrical path of the beam, and the poor sensitivity of this concept, related to the very short integration time of the detector. It is convenient for low resolution imagers. This way is currently used in the MeteoSat system, for

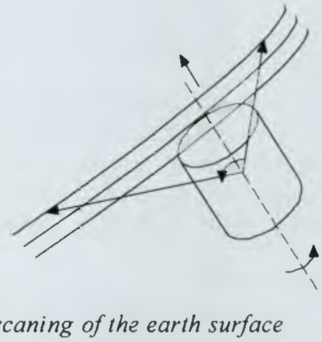


Fig 2 : scaning of the earth surface through a "barrel rolling" satellite

the first generation and the second generation satellite as well. The simple model of a sole detector is actually complemented in the detectors of the MeteoSat Second Generation (MS2G), by a multi detector approach, each of the detectors being devoted to a given spectral band, as described in Ref. 1.

3.2 Pushbroom imaging satellites

This system is based on the simultaneous analyze of one line of the scene to be imaged, through a line of corresponding detectors. The imaging resolution along field of view is then given by the number of detectors in the detector line. Again, the whole image analysis is based on the natural movement of the satellite along its trace, changing the line as far as the satellite *brooms* the image. A large number of identical detectors are needed to have a large swath in this mode, and these detectors are desired as similar as possible to each others, from an electro-optical and geometrical point of view. One of the most famous operational system is the SPOT one, in which the line of detectors is built around an optical beam spliter (known as DivOLi, i.e. Diviseur Optique de Ligne), merging the light path of four 1500-pixel linear CCD imaging arrays, giving 6.000 pixels in a line, this line looking at 60 km on the ground. Ground resolution is then 10 meters in this mode (SPOT 1 Panchromatic mode)

Such a 6.000-detector line uses one DivOLi and four CCD linear arrays, (see Fig. 3), occupying about $20 \times 20 \times 80\ mm^3$, and weighting around 125 g. A four band imager, as actually built for SPOT 1-3, uses four such lines - total weight : 500 g - and also a four-channel beam spliter, the volume of the whole optical bloc being roughly $130 \times 200 \times 185\ mm^3$.

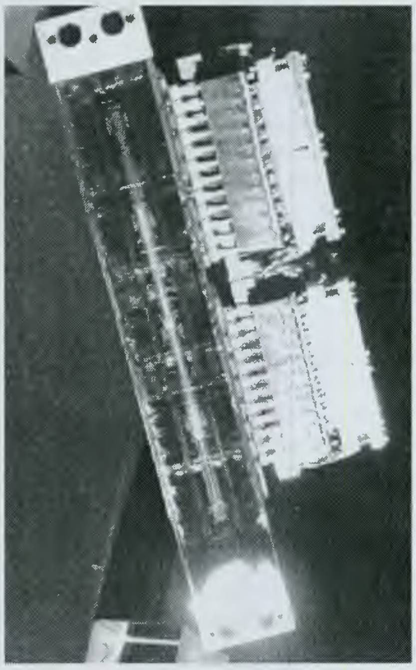


Fig 3 : Four 1500-pixel CCD linear arrays mounted on a glass DivOLi (courtesy of SODERN)

3.3 New imaging system capabilities

Looking at this large size optical bloc, and owing to an increasing mastering of the micro-electronic technologies, the integration of the four-1500 pixel arrays onto one Silicon chip is now possible, reducing drastically both weight and size of the linear detectors. One 6.000 pixel CCD die in a ceramic package is as small as $15 \times 100 \times 2\ mm^3$, and as light as 35 g. Constant progress in CCD technology allows now to build in a package as small as here above, a chip incorporating 12.000 pixels instead of 6.000 ! Doubling the resolution while dividing the functional mass by 4, that gives an idea of the permanent evolution of the space imaging instruments. This is the way driving the construction of the SPOT 5 satellite (Fig. 4).

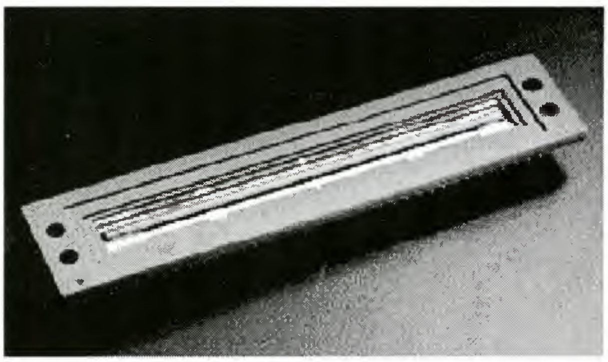


Fig. 4 : TH7834A,
monolithic 12,000 pixel linear array

The next effort is in integrating the four parallel lines onto one Silicon chip, leading to a quite perfect registration of the four-band imagery.

4. NEXT SYSTEMS

4.1 General view of the future

The up-to-date vision of the satellite world is driven by the concept "Smaller, faster, cheaper", now a common *leitmotiv* of the space industry, as the end user will be a commercial one, with his return on investment paid by his market, and no longer a government lab or a defense administration.

While keeping the overall performances of the instruments, several ways are available and concurrently used :

- Using COTS (Components Off The Shelves) for reducing development time (and related costs), and to take advantage of the most recent available technology (Ref. 2).
- Minimizing satellite size and weight, for reducing launch costs, by using less powered rockets, and by sharing launch costs among several users of one launch.
- Using currently available low power technologies, for limiting the requirement for the on-board power unit.
- Limiting component quality level, from ESA SCC space standards (or MIL-STD 883) down to commercial / industrial, sacrificing the reliability, overlaid by multiplying the satellite number.

4.2 Application to imaging systems

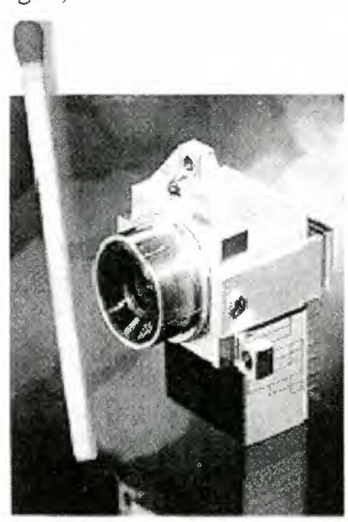
- Using COTS :

At the component level for imaging (i.e. detectors), we have to face an inadequate commercial offer for the space mission. Actually, the commercial components of this type are mainly devoted to two types of home appliances, driven by the increasing needs in the multimedia world : PC cameras, camcorders, fax machines, scanners and digital cameras. These systems are designed for terrestrial and in-house applications, far from the requirements of space missions.

- Detectors for TV-like imaging : these components suffer of poor resolution (576 lines, 720 pixels per line) and RGB color analyzer fixed by the TV standard (TV set CRT phosphor emissivity), fixed frame rate, low dynamic range, low fill factor, medium noise level.
- Detectors for line scanning are quickly progressing as the multimedia PC is more and more powerful, leading to medium resolution linear arrays (e.g. 6.000 pixels), but with limited color separation.

- Minimizing satellite size and weight :

An other way to reduce the cost is to lighten the satellite load. For imaging instruments, a major effect is obtained when reducing the camera size, closely related to the lens system diameter, and to the telescope focal length (Ref. 3). A specific assembly technique as MCM-V™ is a proposal for compacting cameras. As an example, the micro-camera developped by the tri-partner consortium -C.S.E.M, 3Dplus and TCS- shown on Fig. 5, based on a 1k x 1k TH7888 imaging



Courtesy of C.S.E.M.

Fig. 5 : Microcamera for planetary exploration

CCD. For a given mission (e.g. a given Ground Sampling Distance, GSD), this way requires to reduce the pixel size, to be compensated by an associated increase in the pixel responsivity.
Increasing the integration time is the way exploited in the Time Delay and Integration (TDI) approach, but this way suffers of a poorer sensitivity than a linear arrays of photodiodes. This tremendous dilemma will be developed in the next chapter 5.

- Using low power technologies :
Reducing the on-board power consumption will reduced the need for heavy batteries and/or large solar panel generators, then reducing the total mass of the satellite. In the imaging detector field, the emerging CMOS image sensors are one of the candidate for replacing CCD sensors in some future instruments (Ref. 4). This technology, while not dedicated to image sensing and then not monitored w.r.t. imaging parameters -e.g. dark current and quantum efficiency-, suffers of some un-reproducible and more or less medium performances, but has attractive new functionalities, as randomly addressable pixels or easy integration of analog-to-digital conversion.

- Limiting component quality level :
This way, related to the COTS view, is currently under actual investigations, and some results give not an enthusiastic confidence in this system reliability approach. A few satellites are already lost due to assumed in-flight failure of standard components.

5. THINNED BACKSIDE ILLUMINATED DEVICES

Reducing the optical instrument size and weight put a rise in the requirement of the image sensor responsivity. When photodiodes feature quantum efficiency as high as 80%, the remaining solution is to increase the integration time, and then to use the TDI concept. Unfortunately, frontside illuminated TDI sensitivity is less (about 1/3) than photodiode linear CCD' one, and the constraint increases on the number of TDI steps to be used to recover that loss. The backside illumination is an elegant way of maximizing definitively the sensitivity of an image sensor. This approach is almost as old as the photon light sensing devices had been imagined. Its major advantage is in using one side of a component for light incoming, and the opposite one for signal processing. This technique offers then a high level of integration, the fill factor is at its maximum value -100%-, and the whole area of the device is designed for increasing the sensitivity. For a Silicon detector as an example, this technique offers the maximum quantum efficiency the Silicon could ever offer. The Fig. 6 gives the theoretical quantum efficiency of a bare Silicon, 10 µm thick detector. This technique is very well suited to CCD detectors, for which the signal processing involves no electronic component, but only storage and transfer operations by

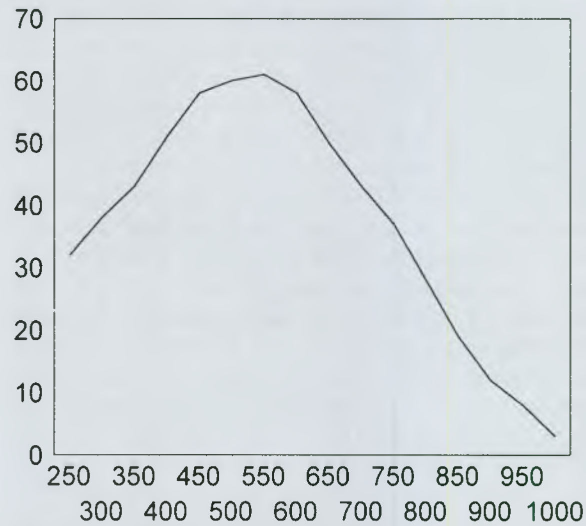


Fig. 6 : Theoretical Silicon quantum efficiency (%) vs wavelength (nm) for an effective thickness of 10 µm.

voltage gate control, these storage operations being not affected by light. On the contrary, a CMOS sensor could not accept light impinging on its basic device, the MOS transistor, the channel current of which being affected by the photogenerated current.

5.1 Thinned CCD development

At Thomson-CSF Semiconducteurs Spécifiques, thinned CCD sensors were in the natural development follow-on of imaging electronic tube, that have been using thinned Silicon 1" diameter membranes as the targets of the Vidicon tubes. This technique was used as it was for thinning CCD imaging sensors (Fig. 7) at the test device level, and for proving the capability of the technique. The reduced processed area, limited by the 1" diameter of the tooling permitted only devices as large as 15 mm diagonal -e.g. TH 7395M (Fig. 8)-, and then required to invest in extension of the 1" process to a 100 mm one, in order to manage large format CCD, as required by the market.

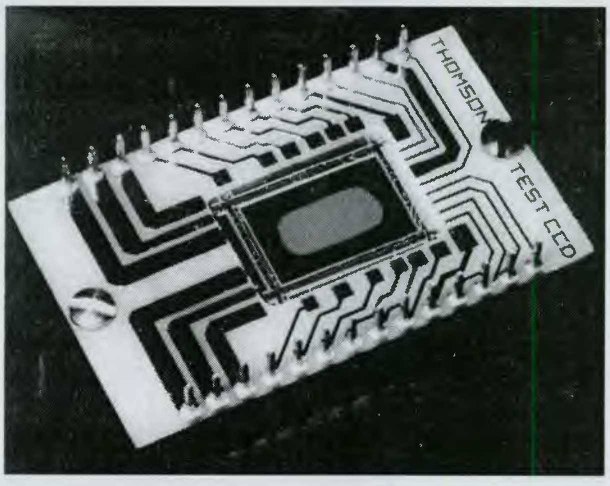


Fig. 7 : Test thinned device for development purposes

Associated to this full wafer thinning process, two assembly techniques were set up :

- for visible applications only : transparent glass support is used both to strengthen the thinned Silicon membrane and to protect the input back face of the CCD imager: The wavelength range bandwidth of this technique is limited by the materials used for the glass and the glass coupling.
- for wide wavelength range : a wafer scale flip chip strengthening technique, using lead-tin bumps to interconnect the front side pads to a thick Silicon supporting wafer. This technique allows to use the supporting wafer for its mechanical properties, and also for electrical connection or/and electronic processing, as the support wafer could be a processed one, including integrated functions as drivers and/or proximity electronics.

Both techniques are compatible with buttable devices, allowing to build large focal planes.

5.2 Anti-reflective coating development

As shown in Fig. 6, bare Silicon is not the best configuration for light detection, as an important part of the energy is reflected to the outside, at the air-Silicon interface, due to the high value of the Silicon refractive index. An improvement of the behaviour of the impinging light on the Silicon requires to deposit an accurately controlled thickness layer of medium refractive index material on the input face of the thinned Silicon. Specific anti-reflective coatings have been developed and are improving the performance of the bare device. The Fig. 9. shows such an improvement, with an anti-reflective coating dedicated to the visible band, 400-800 nm.

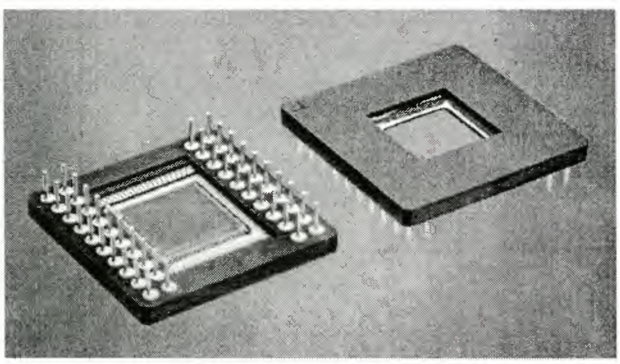


Fig. 8 : TH7395M, backside illuminated thinned 512 x 512, 19 x 19µm² pixel CCD area array,

5.3 The 150 mm TCS wafer fab

In 1996, December 31st, Thomson-CSF Semiconducteurs Spécifiques closed its multi-technology 100 mm wafer fab, and refurbished it for processing only 150 mm wafer CCD technologies. In the third quarter of 1997, occurred the first deliveries of

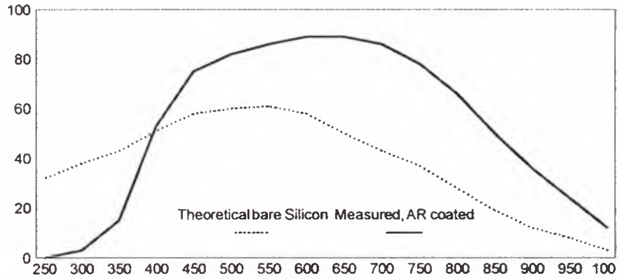


Fig. 9 : Measured quantum efficiency of TH 7395M with antireflective coating in the visible range 400-800 nm (solid line), with comparison of the bare silicon (doted line)

150 mm standard products, after a full validation of the processing steps.

The quality of this wafer fab is now quite proven, and the whole quality of the CCD processes in terms of controlled defects by cm², is at the top level in the scale of the wafer manufacturer standard.

5.4 TCS thinned CCD sensors

The wafer fab transformation into 150 mm having been mastered, it is now planned to validate the thinning process capabilities on 150 mm wafers, and the coating of anti-reflective layers as well. The next year will be devoted to this validation, and the production of flight model is expected within two years.

6. CONCLUSION

For economical reasons, new trends in electro-optic satellite architecture is to reduced drastically the size of the optical instruments. This constraint is reported at the same scale at the detector level, to tender a reduced pixel size. That puts the requirement on the responsivity of the imaging sensor to its maximum. The Time, Delay and Integration principle coupled with the advantages of the backside illuminated CCD is an elegant way to solve this dilemma. The CCD technology's from TCS are on the way to achieve such monolithic focal planes in the new high quality 150 mm wafer fab in the very near future.

7. ACKNOWLEDGEMENT

The improvement of the thinning process and the development of efficient anti-reflective coating layers were greatly helped thanks to the supporting interest and the partial funding from the French space agency C.N.E.S., and the product specifications were developed through fruitful discussion with numerous customers.

8. REFERENCES

1. Amore, O & al. 1998, MSG-SEVIRI Visible Near Infrared (VNIR) detectors development results, *The 32nd ESLAB Symposium Proceedings (SP-423), Remote sensing methodology for earth observation and planetary exploration.*

2. Benedetto, J 1998, COTS in space, *COTS journal, May-June 1998.*

3. Michaelis, H 1998, New concepts for high resolution imagers for earth observation and planetary exploration, *The 32nd ESLAB Symposium Proceedings (SP-423), Remote sensing methodology for earth observation and planetary exploration.*

4. Chautemps, J & Janvier, N 1998, Image sensors for star trackers, *The 32nd ESLAB Symposium Proceedings (SP-423), Remote sensing methodology for earth observation and planetary exploration.*

MSG-SEVIRI VISIBLE NEAR INFRARED (VNIR) DETECTOR DEVELOPMENT RESULTS

O. AMORE ; X. HUGON ; L. BRISSOT ; S. CORTIAL ; D. MOLLARET ; C. PAUTET

THOMSON TCS
BP 123
38521 SAINT EGREVE CEDEX
FRANCE
Tel : 33.4.76.58.30.00.
Fax : 33.4.76.58.31.05.
email : amore@tcs.thomson.fr

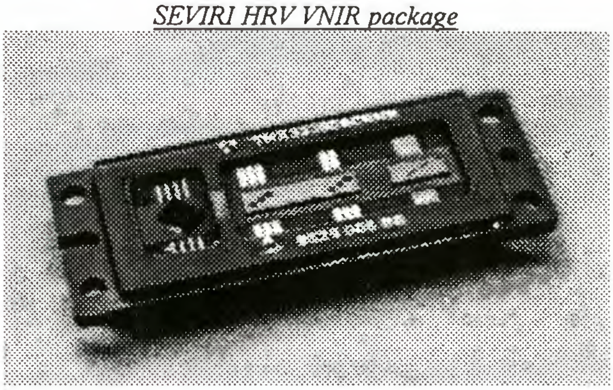
ABSTRACT

SEVIRI HRV VNIR detectors have been developed within the framework of Meteosat Second Generation (MSG) Project, which goal is the development of new multiple spectral band instruments for meteorological imagery. SEVIRI HRV VNIR photodetectors have been designed in order to cover four particular spectral bands defined in the required Visible-Near InfraRed (VNIR) spectral range (0.5 - 1.78 µm). As this wide spectral range can't be covered using a single semiconductor material, for the first time photodetectors made on different semiconductor materials for different wavelength ranges have been mixed in a single detector package. Thus, SEVIRI HRV VNIR detector results in a multichip electrooptical component including, in a custom designed package, dies of different technologies with different raw semiconductor materials. As high geometrical positioning accuracies were required, specific techniques have been developed and implemented. Absolute electrooptical measurements and accuracies requirements were also difficult to reach, leading to design specific test equipments.
Hereafter, HRV VNIR detector is described, pointing out main requirement difficulties. Technical solutions used by THOMSON TCS in order to overcome these difficulties are presented as well as final results which include electrooptical, mechanical and reliability aspects.

1. COMPONENT PRESENTATION

THX 33300 (HRV-VNIR) detector is a four channel spectral imaging component. It includes three visible channels and one near infrared channel. Each channel is fitted with one detector array. Visible channels comprise one high resolution wide spectral range channel (HRV) (0.5 µm up to 0.9 µm) and two low spatial resolution, high spectral resolution in the 0.56-0.71 µm (VIS 0.6) and 0.74-0.88 µm (VIS 0.8) range respectively. HRV detector array is formed by nine Silicon photodiodes, each one being a diamond area of 300 µm x 400 µm in diagonal, while each VIS channel includes three large 720 x 720 µm² Silicon photodiodes. The near infrared channel (NIR 1.6) is covering the 1.50-1.78 µm spectral band and uses three large 720 x 720 µm² InGaAs photodiodes. The eighteen photodiodes are reverse biased.

SEVIRI HRV VNIR package



All channels are implemented in a suitable housing including the window, the electrical output leads and the mechanical attachment and positioning interfaces. HRV VNIR detector package which operates at room temperature (20±2°C) is thus based on a 28 pins ceramic dual in line package. Light collection is obtained through a glass window on top of the photodiodes. It is translated into electrical signal by the photodiodes on the different materials while signal is output and available for each photodiode on the ceramic package pins. On MSG satellite, HRV-VNIR detector package is located within the SEVIRI specific focal plane assembly, and is interconnected to suitable preamplifiers. MSG SEVIRI satellite uses a spinner scanning operating mode which produces, at each revolution in the east-west direction, three contiguous lines for the VNIR channels and nine contiguous lines for the HRV channel. Therefore, it stands to reason that all photodiodes sensing area must be registered one with respect to the others with high precision while

Proc. 32nd ESLAB Symp., 'Remote Sensing Methodology for Earth Observation and Planetary Exploration', ESTEC, Noordwijk, The Netherlands, 15-18 September 1998 (ESA SP-423, December 1998)

focal plane position with respect to package mechanical references must be also accurate.

HRV-VNIR detector is using two kinds of semiconductor materials in order to cover the whole specified spectral range. HRV, VIS 0.6 and VIS 0.8 channels are manufactured on Silicon while NIR 1.6 uses InGaAs as active material. These materials have been chosen in order to be sensitive in the required spectral bands. Operating temperature together with dark current and noise requirements have been the driving performance for this choice. Therefore, as both technologies haven't reached the same technical achievement, development and manufacturing philosophies were quite different for each of them.
Silicon is a well known technology with large Front Ends and large size (6" at THOMSON TCS premise) multiwafer lots are manufactured at once. Most of the final component performances are due to design. In the contrary, InGaAs technology is using 2" wafers which are processed one by one. More, most of the main performances are present within the wafer to be processed, resulting in different performances with the same process.
In a same manner, performance requirements criticity lead to different kind of difficulties on the two materials. Absolute mean response, spectral template, cross talk and capacitance (which results in preamplifier noise increase) were the main difficult items pointed out for Silicon, while cut-off wavelength, noise, capacitance and cross talk were the most stringent ones for InGaAs.

Silicon requirements criticities

Items	Value	Units	Comments
Absolute mean response	≥ 24.6	$nA.m^2/W$	for HRV channel over 0.6-0.9µm
Spectral template	≤ ±5	%	for Silicon VNIR channels
Cross talk	≤ 0.2	%	for all Silicon channels
Capacitance	≤ 10	pF	for HRV channel

InGaAs requirement criticities

Items	Value	Units	Comments
Cut-off wavelength	≥ 1.717	µm	at 50% normalised spectral response
Noise	$\leq 10.10^{-14}$	$A/\sqrt{Hz}$ rms	over 2Hz-300kHz
Cross talk	≤ 0.325	%	
Capacitance	≤ 47	pF	

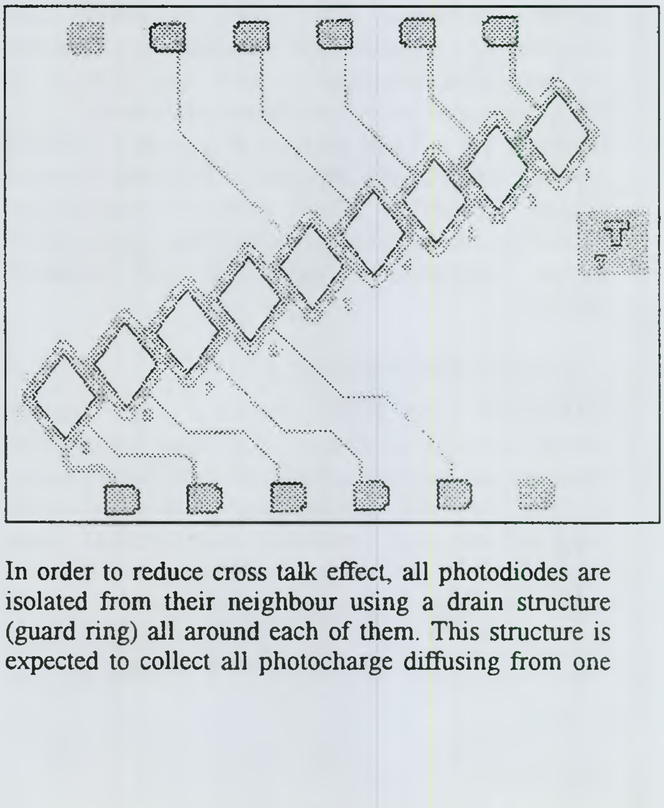
Thus, photodetector development has been driven in two different parallel manners.

2. DETECTOR DEVELOPMENT ACTIVITIES

2.1 Silicon arrays

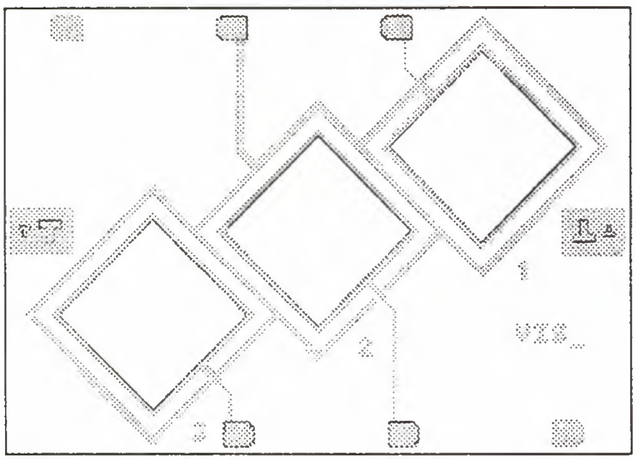
Silicon photodiodes, i.e. all visible channels, have been manufactured on P type substrate, which means common anode configuration. Photodiode response is driven, on red side of visible spectrum, by the active absorption depth, which must exceed several tens micrometer due to silicon low absorption coefficicent in this spectral region. However, increasing photodiode depletion zone leads to higher serial resistance which affects frequency response. In the lower side of visible spectrum (in the blue), photons are absorbed near silicon surface, therefore once technological design adjustments are completed (junction doping levels) antireflective coating becomes a major concern. Thus, a trade-off between serial resistance Rs and response has been fitted.
On TCS standard Charge Coupled Device (CCD) technology, thick silicon dioxide layer all over photosensitive area leads to interferometric effects. Spectral response becomes therefore highly uneven and up to ±25% overshooting could be observed. In order to avoid such effects, a specific technique called "DIOX" has been implemented which allows, at the end of process, getting a thin oxide layer acting as a single layer anti-reflecting filter.
In order to lower capacitance, care has been taken on interconnection pads and tracks design on the dies, especially on HRV channel on which maximum attention has been paid on capacitance evenness from diode to diode.

HRV channel



In order to reduce cross talk effect, all photodiodes are isolated from their neighbour using a drain structure (guard ring) all around each of them. This structure is expected to collect all photocharge diffusing from one photodiode to the other, or coming from the substrate. In addition, each photodiode optical area is limited by an opaque metallic mask for IFOV (Instantaneous Field Of View) accuracy improvement. Thus, whole non photosensitive surface is covered with a masking metallic layer preventing light from absorption in the substrate.

Silicon VNIR channel



2.2 InGaAs array

InGaAs photodiodes have been manufactured using a double heterostructure InP / InGaAs / InP grown by low pressure MOCVD on a N type InP substrate. The first InP layer acts as a buffer, to prevent substrate defects or impurities from disturbing the growth, the InGaAs layer is the actual active layer including the junction area and the top InP layer is used to improve junction leakage current.

InGaAs brings incomparable detectivity and quantum efficiency in the 0.8-1.7 μm spectral range. However, SEVIRI was asking for 1.78 μm extended response, which has imposed a specific InGaAs compositional. Thus, the main difficult item was relative to material growth since cut-off wavelength is controlled by InGaAs compositional. The goal was to find the epitaxial structure and it's growth process which leads to best results after photodiode process.

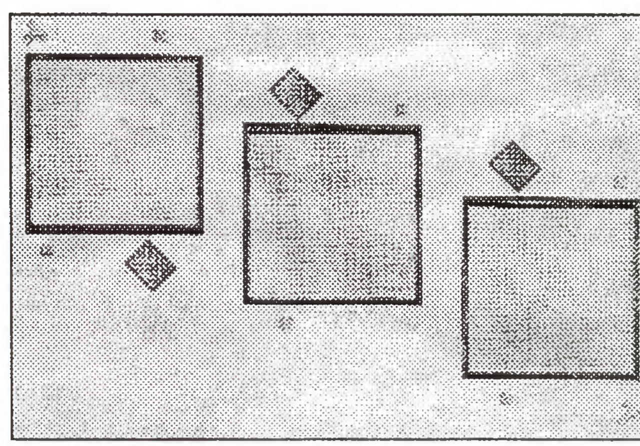
It must be pointed out that InGaAs is grown on an InP substrate. The lattice parameter matched InGaAs has a cut-off wavelength of 1665 nm at ambient temperature. Therefore, in order to reach more than 1720 nm, the InGaAs grown layer must be In richer which results in lattice parameter mismatching with the underlying substrate and the cap layer. Mismatch has dommageable effects on dark current (which increases exponentially with mismatch), thus on noise, due to increasing dislocation and defects densities. Furthermore and unfortunately in our case, the low capacitance value (regarding the diode area) imposes large reverse biasing which also increases dark current.

Consequently, after competitive trials, an epitaxial technique has been chosen in order to grow the InP / InGaAs heterostructure. This last is the very one used for qualified devices with direct epitaxy of lattice mismatched InGaAs without lattice adaptation buffer.

Photodiode process is issued from earlier SWIR qualified detector development. A classical Planar technology is used, in order to garantee high reliability and high stability of junction performances. P zone are obtained by Zinc diffusion through a SiO2 mask. This oxide layer constitutes the junction passivation. Further SiO2 layers are deposited for antireflection and electrical final insulation purposes.

Cross Talk is greatly reduced thanks to THOMSON TCS unique electrooptical pixel insulation technique which prevents outside incoming holes from being collected by the active junction. Thus, NIR 1.6 photodiodes are rounded with floating implanted drain, i.e. not biased, since active layer is very thin and consequently wide depletion is not needed in order to collect unwanted photocharges. In fact, the structure used is a high recombination rate zone. This structure is self sufficient and avoids optical mask needs, however, photodiode design includes an opaque optical mask in order to avoid response overshoot in photodiode junction perimeter area, leading to IFOV degradation.

NIR 1.6 channel

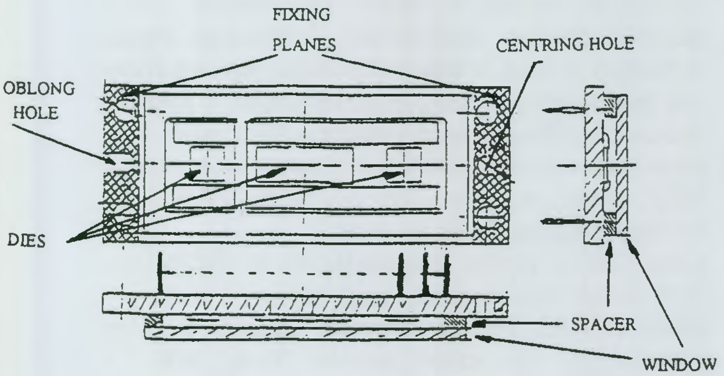


As noise in photodiode is mainly due to dark current (shot noise) a trade off between capacitance and dark current has been fitted leading to 5V reverse bias for NIR 1.6 channel.

2.3 Package and assembly

SEVIRI HRV-VNIR detector package is composed of three parts, a ceramic chip carrier covered with a cap in two parts, a window and a spacer. The high geometrical accuracy which is required (30μm flatness on package fixing planes, ±50μm for detector pattern position with respect to mechanical references) has not allowed the uses of standard dual in line ceramic package. Therefore, a specific design has been used.

HRV VNIR package details



Mechanical reference frame is defined with one centring hole and one aligning oblong hole for Y-Z axes references while two fixing planes acts as X direction reference. Package mechanical references, therefore, lay on top of ceramic on both small sides, where four holes are provided for fixing and two holes (one of which is the oblong one), precisely drill (H7), give mechanical references in the detector focal plane. Fixing and reference holes are separated in order to increase positioning accuracy. Fixing must take place in pressing ceramic fixing area top surface on the mounting system (instrument side) since only top surfaces are coplanar and flat (±20 μm).

So, due to high geometrical accuracy requirements, a specific design has been used in order to allow end process machining and polishing of the ceramic (both mechanical references and die mounting area) without injuring the electrical interconnection pads. A cofired ceramic process is used, where interconnection paths are buried within ceramic before firing. Bonding pads and pin brazing area are coated afterward. Pins are brazed at high temperature, using a classical "Pin Grid Array" package process. Fixing holes are realised before ceramic firing and therefore exhibit low mechanical accuracy. Mechanical reference holes, in the contrary are machined (grinded) after process completion. The top surface of the ceramic is grinded in order to obtain a high planarity level. Such machining is not possible in a cavity. Thus, the die report surface must be the upper plane of the ceramic and consequently, a spacer must be used between it and the window. This spacer, made of Kovar (FeNi), is sand blasted in order to minimise specular reflection and glued on ceramic. Conducting glue is used in addition to give electrical continuity with spacer interconnection pad.

Window is made of ZKN7 glass and glued on spacer. This material has been chosen regarding its good transmittance over the whole spectral range of concern. Window is antireflection (AR) coated on two faces. Two different and separate AR coating are implemented, one for visible channels and one for NIR 1.6 channel. A channel separation between NIR 1.6 and visible channels is implemented through spacer design.

HRV channel is settle within one single die, VIS 0.6 and VIS 0.8 channels are laid on the same common substrate (one die for both channels), NIR 1.6 channel of course uses a separate substrate. Interconnections take place within ceramic body (buried tracks). Redundancy is managed in doubling the critical signals pins. All substrate interconnections are doubled as well as guard rings on visible channels which are common to both HRV and visible. Two pins are dedicated to spacer electrical interconnections.

TCS uses its know how concerning high geometrical accuracy die report. HRV, VIS 0.6-VIS 0.8 and NIR 1.6 dies are assembled together in a single step to reach the best positioning accuracy. Tested photodiode dies are first positioned on an optically surfaced glass piece having metallic indexes showing the theoretical die positions. Visual superimposing of theses indexes to those which are laid on the dies is realised with micro mechanical positioners and through microscope observation. Glue is deposited onto ceramic baseplate which is brought under the dies and pressed. Then the package is removed from the glass piece for curing at higher temperature. This process leads to low geometrical distortion. Therefore, expected geometrical accuracy is even better than required.

3. ELECTROOPTICAL TEST BENCHES

Electrooptical measurements have been carried out on different measurement benches due to specificity of some measurement methods and high precision needed leading to some difficult measurements. Some of them are detailed hereafter.

3.1 IFOV test bench

Photodiode radiometric barycenter positioning knowledge with respect to package references was a major requirement. Radiometric barycenter have been determined in computing, for each photodiode, the barycenter of its IFOV mapping. As all photodiode mapping are done in a single batch, accurate relative position of photodiode barycenter are obtained. To link these radiometric informations to mechanical package references, geometrical positioning of photodiodes with respect to package references are measured using their optical mask. Finally, whole IFOV barycenter map is best fitted to visual geometrical center map.

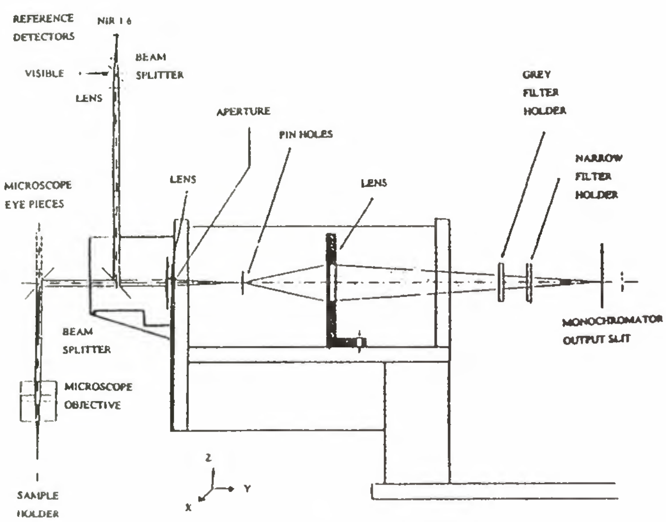
Due to optical technique used for assembly SEVIRI HRV VNIR devices, geometrical requirements are obtained from the tooling design (reference glass matrix, mechanical displacement systems, microscope objective ...). Thus, any die or photodiode positioning with respect to reference glass matrix is better than ±3 µm. Concerning the package, accuracy is worse (±30 µm) since focusing must be performed on reference hole edges located on black ceramic.

Thus, IFOV positions have been measured using an LBIC (Light Beam Induced Current) machine having an excitation beam of few microns in diameter. Whole detector (the four channels) have been scanned in one batch, in order to avoid repositioning relative errors.

3.2 Spectral response and Cross talk test bench

Electrooptical measurement bench is based on focused beam design.

Spectral response and cross talk test bench



A light beam is spectrally filtered with a monochromator. A secondary source is obtained with a grinded glass piece. A pin hole object is placed in order to reduce spot size at photodetectors level. One part of resulting beam is sent onto reference detectors (one for visible channels and one for NIR 1.6 channel), the other part is sent on measured detector, on which image of object pin hole is focused. In this way, reference detectors see at any time what is being seen by measured photodiode. Depending on photodiode size to be measured, spot size is set between 140 to 350 µm. Thus, all measurements are done with whole beam energy inside photodiode sensitive area. Spectral absolute response measurement are done by comparison with certified reference photodiodes. Absolute mean responses are computed from spectral absolute response integration over required spectral band. This test bench configuration has finally been optimised for best absolute response accuracy and spectral response reproducibility.

This bench is also used for cross talk measurement since large spectral wavelength beam can be obtained using narrow filters instead of monochromator wavelength selection. In this case, a smaller pin hole is used (35µm diameter) and focused on the central photodiode of a channel while the neighbour ones are in darkness. In this configuration, measurement accuracy is driven by light beam positioning within illuminated photodetector. Positioning error may result in parasitic lighting of neighbouring photodiodes and measurement errors.

3.3 Noise test bench

The last main measurement item concerns noise measurement. As no way was found to fulfil requirement in terms of bandwidth and background electronic noise, a substitute allowing to assess a noise maximum figure has been developed. Assuming that high frequency noise is dominated by shot noise while low frequency noise may be present and modifies results in the low frequency side, a CCD read out circuit has been designed which integrates photodiode dark current during about 60 µs leading to high cut off frequency of about 8kHz. Read out cycles are measured for 0.5 s in order to cover the low frequency requirement. Noise measurement on all sampled read out gives total measured noise B1 over a 2Hz-8kHz bandwidth. Noise in the 8kHz-300kHz bandwidth is computed from dark current measurement and is assumed to be shot noise dominated B2 (with $B2=\sqrt{2.q.Io.Df}$ where q is electron charge, Io dark current and Δf the 8kHz-300kHz measurement bandwidth). Total noise is then obtained using B1 and B2 results.

Using this method, NIR 1.6 channel noise is measured with acceptable accuracy while VIS 0.6 - VIS 0.8 channels noise reach the method limit, and HRV channel noise is lower than the measurement equipment noise. This last is, however, lower than SEVIRI specified requirement allowing to verify that HRV and visible channel noise are lower than specification, without knowing how much it actually is.

Test bench measurement accuracies

Test benches	Value	Comments
IFOV	±5µm	Radiometric barycenter measurement
Spectral absolute response	±4.5% ±6% ±2%	For Silicon channels For NIR 1.6 channel Spectral response reproducibility
Cross talk	0.06%	Absolute value
Noise (A/$\sqrt{Hz}$)	10.10^{-15} 6.10^{-16}	For NIR 1.6 channel For HRV channel

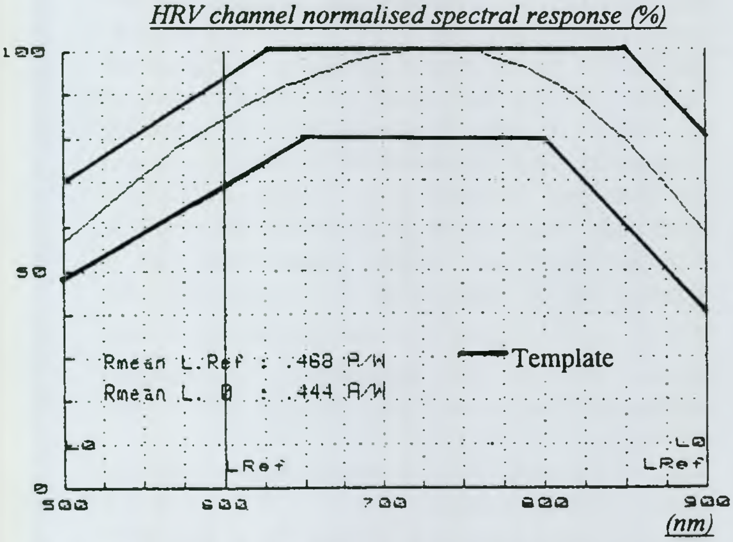
4. MAIN PERFORMANCE RESULTS

4.1 Mean response and Spectral response

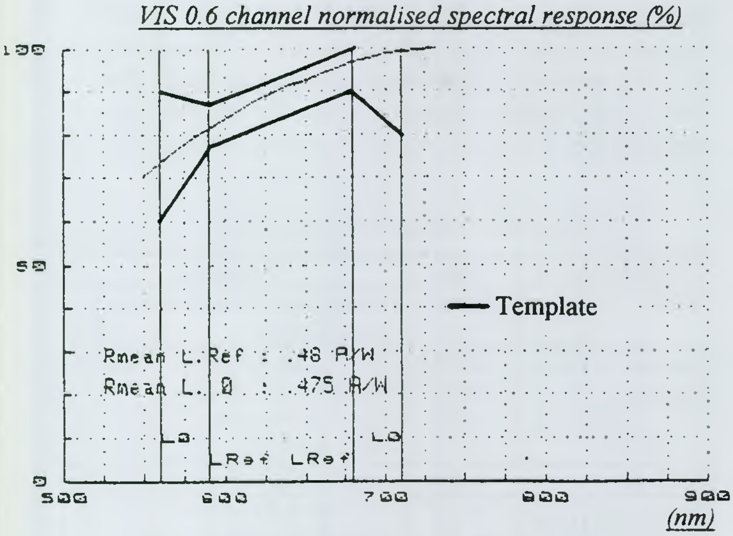
Spectral response stability measurement were quite difficult due to the highly stringent specification (±1%), therefore in most of measurements it has been run on larger temperature range (±5°C for Silicon channels) and extrapolated to the required range, ±2°C and ±0.5°C. These temperature variation ranges are very small with respect to detector behaviours with temperature, therefore all mechanism can be approximated to linear ones within this domain.

DIOX technology gives on all channels satisfying spectral response results, including HRV channel the most stringent one, which absolute mean response has been found about 10% higher than required one. Rmean stability has been demonstrated over the whole temperature range (±2°C) since less than 0.4% fluctuation has been observed.

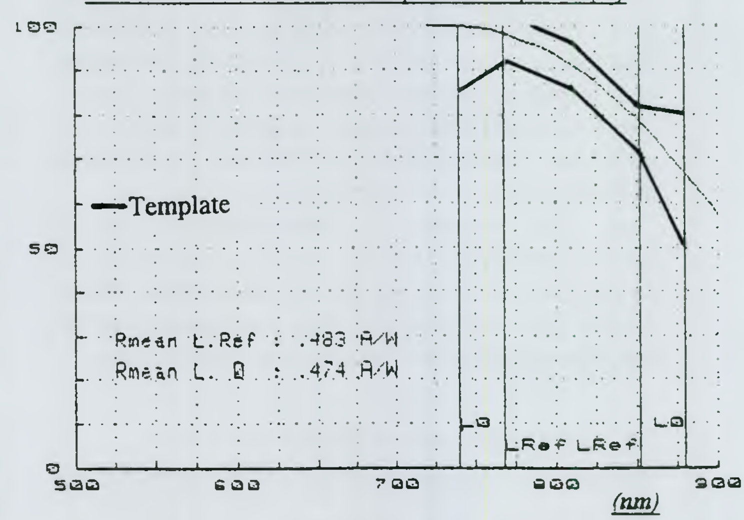
HRV channel normalised spectral response (%)



VIS 0.6 channel normalised spectral response (%)

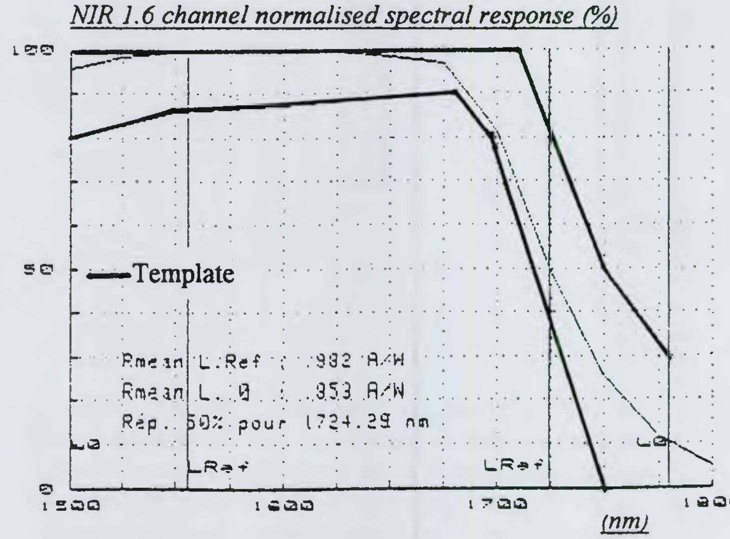


VIS 0.8 channel normalised spectral response (%)



NIR 1.6 channel measurements also give large mean response results. Mean response is affected by temperature since cut-off wavelength increases by 1 nm for each 1°C temperature increment. As spectral band is only 240 nm, such an effect leads to about 0.4% per °C temperature change, leading to a Rmean fluctuation of ±1% in the required temperature range (±2°C).

NIR 1.6 channel normalised spectral response (%)



Absolute mean response results (Rmean)

Channels	Results	Expected value
HRV		
Rmean (nA.m²/W)	27.5 to 29.0	≥ 24.6
Stability (%)	±0.4	±5
VIS 0.6		
Rmean (nA.m²/W)	247	≥ 182
Stability (%)	±0.4	±5
VIS 0.8		
Rmean (nA.m²/W)	247	≥ 197
Stability (%)	±0.4	±5
NIR 1.6		
Rmean (nA.m²/W)	506 to 523	≥ 467
Stability (%)	±1	±5

4.2 Noise

As said before, noise requirements lead to dark current specifications since at this low noise level, only shot noise is expected (low frequency noise if present will overstep the required levels). Low frequency noise is not observed in the specified frequency range in Silicon photodiodes, however lattice mismatched InGaAs often shows dark current frequency fluctuations.

For Silicon channels, measured dark current at typical bias (12V) in DIOX photodiode are very low and lead to by far lower shot noise than expected. Therefore, noise figure was easily within specification, the most difficult task has been t measuring it.

In the case of NIR 1.6 channel, this parameter was critical and part of a trade-off with cut-off wavelength. Requirement, when considering only shot noise, leads to 31 nA in dark current. However, measurements have shown that above 20 nA in dark current, photodiodes could exhibit low frequency noise which overcomes requirements.

Noise results

Channels	Results	Expected value
HRV		
Dark current (pA)	1	≤ 112 (187 nA/cm²)
Noise (A/$\sqrt{Hz}$ rms)	6.10^{-16}	$\leq 6.10^{-15}$
VIS 0.6 - VIS 0.8		
Dark current (pA)	3	≤ 1200 (230 nA/cm²)
Noise (A/$\sqrt{Hz}$ rms)	10.10^{-16}	$\leq 1.10^{-14}$
NIR 1.6		
Dark current (nA)	10	≤ 31 (6 μA/cm²)
Noise (A/$\sqrt{Hz}$ rms)	6.10^{-14}	$\leq 1.10^{-13}$

4.3 Cross talk

Measurement results give 0.15-0.18% and 0.26-0.28% ranges for respectively Silicon channels and NIR 1.6 channel.

4.4 Electrical characteristics

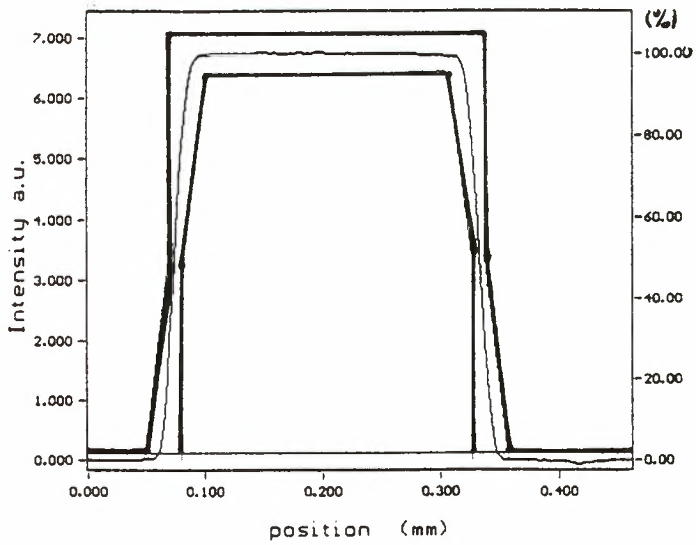
Silicon channel capacitances are less than 7pF for HRV channel and about 16pF for VIS 0.6 and VIS 0.8 channels. For NIR channel, although capacitance was a very stringent parameter regarding biasing and noise, measurement results give about 45 pF.

For each channel, serial and shunt resistances don't cause any difficulties since measurement results are by far into requirement specification. As example, at typical bias voltage, shunt resistance measured values exceed 500 MΩ for NIR 1.6 channel and 1000 GΩ for Silicon channels.

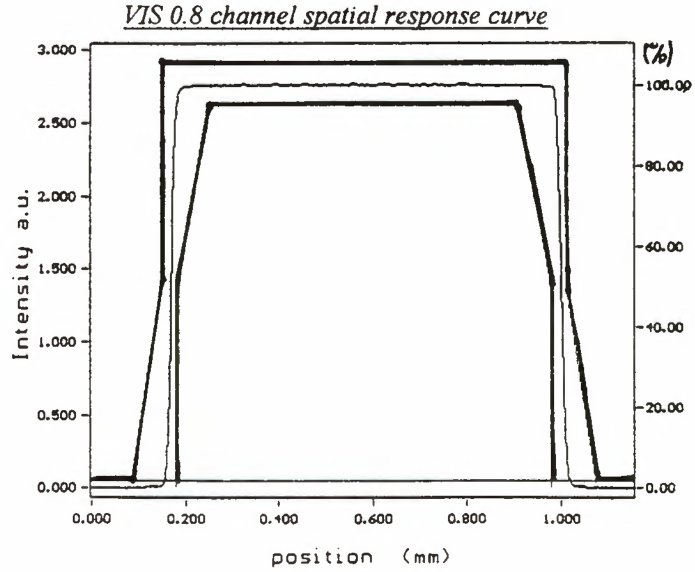
4.5 Geometrical measurements

Photodiode mechanical measurements have been verified on each device through IFOV measurement, defining the spatial response of each photodiode. Thus, the contour and the sensitivity of the whole theoretical optical area of each photodiode has been determined.

HRV channel spatial response curve



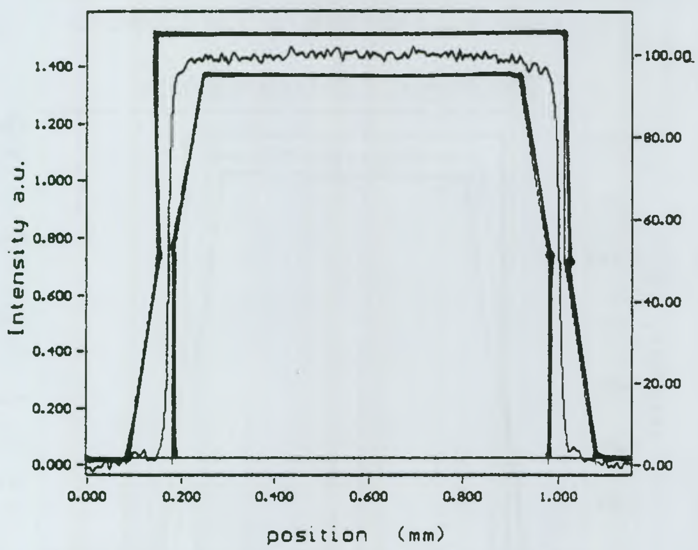
VIS 0.8 channel spatial response curve



From IFOV results, the actual radiometric center position (barycenter position) of each photodiode is computed and taken into account in order to determine the whole detector registration (position of each photodiode of each channel with respect to all others photodiodes within the package). It has been observed that barycenter position results correspond to photodiode theoretical geometrical position to within IFOV measurement accuracies (±5μm).

Finally, detector array position accuracies with respect to package mechanical references have been found less than 30µm, thanks to very specific THOMSON TCS tooling design and process assembly.

NIR 1.6 channel spatial response curve



4.6 Reliability aspects

Qualification tests taking into account classical environmental constraints as phases of transportation, handling, storage, launch, in-orbit, etc... have been performed. Results confirm that devices stay within requirement after all mechanical and thermal tests, as after 30 KRad Co60 irradiation cumulated dose, though dark current drift is very high (x100) for HRV-Visible channels, but an excellent spectral response stability. NIR 1.6 channel shows weak drifts insuring that InGaAs technology is quite insensitive to electrons radiation.

5. CONCLUSION

For the first time, photodetectors made on different semiconductor materials for different wavelength ranges have been mixed in a single detector package, providing a unconventional multichip electrooptical component dedicated to wide spectral analysis range for space application (0.5 - 1.78 µm). All specification requirements, either electrooptical or mechanical or environmental ones have been fulfilled.

Such multichip photodetector arrangement will provide, in future, drastic reduction in instrument complexity with corresponding cost and weigth saving. Even more complex assemblies could be done at detector manufacturing level for compact multispectral detection instrument. Finally, it must be noticed that with such highly integrated focal plane, major benefits could be obtained on optics' size with even more impact on weight, cost and size of instruments.

IMAGE SENSORS FOR STAR TRACKERS

J. Chautemps, N. Janvier

Thomson-CSF Semiconducteurs Spécifiques
B.P. 123
38521 Saint Egrève, France
Tel. +33 473 58 31 12 / Fax +33 476 58 34 97 / jacques. chautemps@tcs.thomson-csf.com

ABSTRACT

Optical attitude control of satellites can be performed by various concepts, depending mainly of the mission : Earth sensor, Sun sensor, and Star tracker. The purpose of this paper is to review evolution of star sensors requirements, discuss the possible architectures and technologies in near and longer term future.

1. INTRODUCTION

Star tracker is a crucial equipment of a satellite, as it controls its orientation in space and allows to stabilise it in a given direction.

After selecting a zone of the sky with a given field of view, the goal is to maintain dedicated star images at the same position in the successive images registered by a camera. Each position difference will be detected and attitude of the satellite corrected by micro-engines.

For over twenty years Charge Coupled Device technology is the preferred one for star sensors as it exhibits a high radiometric and electrooptical performance, and high reliability as well.

2. DISCUSSION AROUND AN EXISTING DEVICE

Thomson-CSF Semiconducteurs Spécifiques (TCS) TH7863 CCD area array is a well-known device, chosen as sensor for many star strackers in the past years. It is a 288x384 frame transfer CCD, operating in 4 phase mode. Pixel size is 23 x 23 μm^2.

2.1 TH7863D

Its main typical characteristics are summarised in Table 1.

PARAMETER	VALUE
Charge handling capability	450 ke-
Responsivity	8 V/μJ/cm^2
Photo Response Non Uniformity	1% (σ)
Dark current at 20°C	0.5 nA/cm^2
RMS noise in darkness	70 e-

Table 1 - TH7863D typical characteristics

Thanks to appropriate substrate technology, spectral sensitivity extends from 400 to 1100 nm. Good performance in near infrared spectrum (sensitivity, response uniformity and MTF) is important as faint objects are emitting a lot of photons in near infrared spectrum and as peak quantum efficiency is located at 700 nm. With no filtering, measured signal is 10 times higher than signal obtained with infrared filter in front of the CCD, with no noticeable decrease of response uniformity and limited decrease of resolution.

The large pixel size provides an even higher sensitivity and allows to obtain a good uniformity from pixel to pixel, by averaging microscopic non uniformity. These two parameters are of major importance when the field of view of the star tracker is reduced : there is a high probability that only faint stars are present : high sensitivity provides corresponding signal above background noise, while low Photo Response Non Uniformity (PRNU) allows to discriminate low amplitude useful signal from local spatial noise.

Four phase operation provides a high charge handling capability, thus achieving a high dynamic range. Even if neighbour stars exhibit a difference of several magnitudes in brightness, the CCD will be able to register an image free of blooming.

As regards radiation behaviour, pixel design is important : replacement of thick oxide insulation by planar P+ insulation allows to limit dark signal increase when the device is submitted to irradiation : after an irradiation dose of 5 krad, device with P+ insulation exhibits a dark signal multiplied by 3, while value is multiplied by 6 for a device with oxide insulation. Dark Signal Non Uniformity (DSNU) has the same variation rule : it means that detection of faint objects is improved in the same range.

Nevertheless, some performances of TH7863 can become limitation in some cases :

- resolution in terms of number of pixels becomes not sufficient when the field of view is increased.
- cooling of the device is necessary to decrease end of life dark signal, in order to keep good performance even after irradiation. Power consumption budget of the

Proc. 32nd ESLAB Symp., 'Remote Sensing Methodology for Earth Observation and Planetary Exploration', ESTEC, Noordwijk, The Netherlands, 15-18 September 1998 (ESA SP-423, December 1998)

satellite makes it sometimes impossible to completely cancel radiation generated dark signal.

2.2 TH7863M

In the recent years Multi Pinned Phase (MPP) operation has also be tested on TH7863 (Ref. 1). As the major contribution to dark current in buried-channel CCD is the interface generation states, applying negative voltages on sensor clocks with respect to substrate bring holes at the oxide-silicon interface : the CCD interface is thus in hole accumulation mode, thus

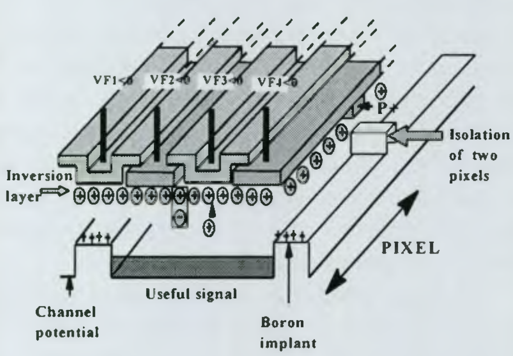


Fig 1 - MPP operation principle

recombining electrons generated at the interface by thermal and irradiation effects. Operation principle is shown in Fig 1.

Dark signal of TH7863M is drastically reduced to 20 pA/cm² at 20°C, and moreover it remains quite stable after 100 krad (Si) Co60 irradiation., assuming a margin of 7 volts on clock biases to compensate the threshold voltage drifts induced by irradiation. It should be mentioned that the main contribution to dark signal will be the transfer of charges : during that time, clocks are in non-inverted mode, thus creating non MPP dark current generation. Hopefully, this contribution acts only as an offset, as it is averaged by continuous transfer. Actual value is proportional to time during which clocks are positive compared to substrate bias, and P+ insulation contributes to limit this contribution.

3. A NEW DEVICE : TH7890

3.1 Application requirements

For coming years, the main sensor requirements for Star Trackers are summarised in Table 2.

As regards ionising radiation operating conditions, values are very different according to the application :

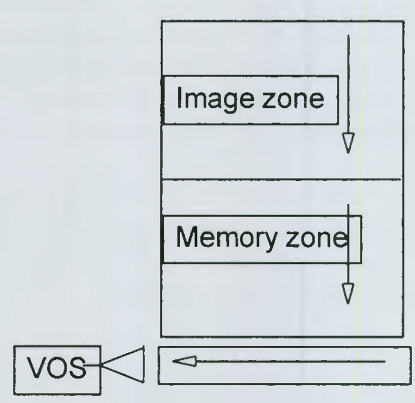
- for low orbit satellites, typical value is 3 to 10 krad (Si)
- for geostationary orbits, it becomes about 20 krad
- for satellite constellations, it can grow up to 100 krad (Si), due to irradiation anomalies. In that case, non ionising dose (displacement damage by neutrons) is also a concern. These effects were studied on TH7863 by G.Hopkinson and al. (Ref. 2).

FEATURE	VALUE
Number of pixels	512x512
Pixel size	15x15 to 20x20 µm²
Full well capacity	> 200 ke-
Peak quantum efficiency	> 40%
Dynamic range	> 11 bit
End of life dark signal	< 5 to 20% of full well
Output impedance	< 500 Ω

Table 2 - TH7890 main features

3.2 TH7890 presentation

The main features of this new device are provided in Table 3.



TH7890 MAIN FEATURES
Frame transfer organisation
Number of pixels : 512x512
Pixel size : 17x17µm²
100% fill factor (pixel aperture)
MPP operation
Four phase clocking
One output, up to 15 MHz data rate
Ceramic package, 24 pins
Integrated Thermo Electric Cooler (TEC) option

Table 3 - TH7890 main features

Pixel was chosen as the minimum size to achieve requested full well capacity : this performance is a trade -off between 4 phase operation and optimised MPP technology which increase this capacity, and P+ insulation which reduces it. The resulting image zone size remains compatible with 2/3 inch lenses. As TCS wafer manufacturing has shifted from 4 inch wafers to 150 mm wafers, deposited layers and etching are more uniform, which keeps the PRNU at a very low level even if pixel size is reduced. Smaller pixel makes the device less sensitive to irradiation.

Design rules, operation and technological process are very close to the ones used for TH7863M (MPP operation and P+ insulation). As a result, performance can be easily predicted and existing data as regards environmental tests, especially irradiation tests, remain valid.

3.3 TH7890 performance

It is summarised in Table 4.

PARAMETER	VALUE
Full well capacity	250 ke-
Peak quantum efficiency	45 %
Dynamic range	12 bit
Conversion factor	4 μV/e-
Dark current at 20°C, MPP mode	20 pA/cm^2
Dark current at 20°C, non MPP mode	500 pA/cm^2
Photo Response Non Uniformity	1% (σ)
Linearity error	< 1 %
Contrast Transfer Function at Fn	70% at 550 nm
Amplifier power consumption	120 mW
Output impedance	300 Ω

Table 4 - TH7890 typical performance

As regards behaviour after a 100 krad ionising dose, the expected performance changes at 20°C are listed in table 5.

PARAMETER	VALUE
Dark current in image & memory zones in MPP mode	50 pA/cm^2
Dark current in image & memory zones in MPP mode	5 nA/cm^2
Amplifier power consumption	250 mW

Table 5 - TH7890 behaviour after 100 krad irradiation.

Dark signal performance makes it possible to operate the device at an ambient temperature of 30°C without any additional cooling, in a low orbit satellite.

The device is now under development. It will be available in Q1 99.

In addition, TCS can offer the device in a specific package which includes a Thermo Electric Cooler (TEC). Only the CCD die is cooled, while the package itself is used as radiator. It saves a lot of room and power consumption as well. Environmental behaviour of this technology is now studied in depth by TCS in the frame of an Evaluation Test Program partially funded by ESA. Results will be available by end 1999.

4. APS / CMOS FOR NEXT GENERATION ?

4.1 Discussion

For some years, new image sensors based on CMOS technology are emerging. Designed as a « memory structure », these sensors (Fig 3) are no more based on the transfer principle as in CCD technology (Fig.2).

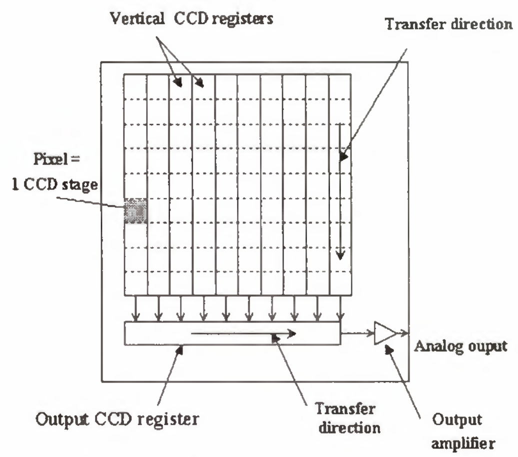


Fig.2 CCD organisation

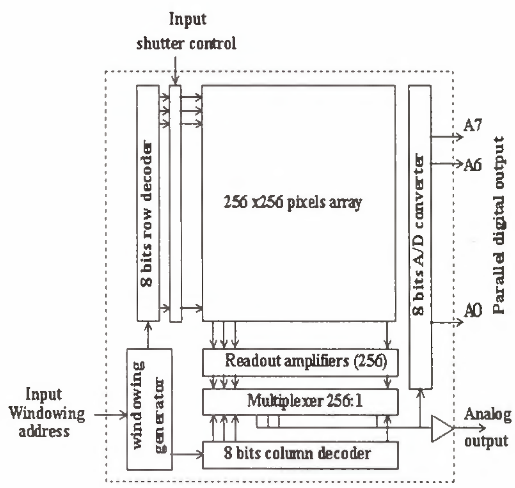


Fig. 3 CMOS organisation

New architecture and capability for electronic integration allow new functionalities of CMOS sensors :

- direct access to pixels
- simplified windowing capability

- on chip electronic integration (A/D converters, timing generation, ...). This last item can provide a significant decrease of the total cost of a system.

- power supply levels are 5V or 3.3V for CMOS operation, while levels up to 15 V are required for CCD operation.

Drawbacks of CMOS sensors are mainly linked to their electro-optical performance compared to CCD sensors :

- CCDs are manufactured in dedicated and optimised wafer processing (2,5 micron and 1,5 micron technologies are available in TCS), while CMOS process is mainly driven by memory manufacturing process.

- CCD exhibits a 100% pixel aperture while CMOS, with its associated transistors (fig 4), is limited to less than 50% aperture.

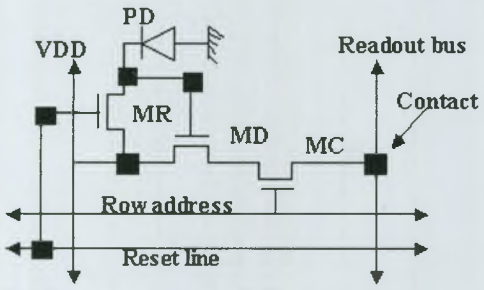


Fig 4 - Active pixel structure

The reduction in CMOS geometry will allow to increase the active area of a pixel compared to its dimensions, but others difficulty could appear, mainly in term of photodetection capability (epitaxial layer, junction depth, ...)

New structures -based for example on microlenses on top of the pixels- will increase sensitivity of CMOS pixels. But those are non standard technological levels.

4. 2 Performance comparison

There is typically a ratio of 10 between CCD and CMOS sensor on dark signal performance and a difference of 5 to 10 on the S/N ratio, depending on the chosen structure. Table 6 provide comparison results between different sensors technologies.

Technology	Dynamic range	Minimal sensitivity
0.7 microns	62 dB	3.76 lux
0.5 microns	65 dB	2.45 lux
0.35 microns	62 dB	2.09 lux
TCS CCD 1.5 microns Photomos	75 dB	0.88 lux

with :
pixel size : 14 microns
25 images/sec
F/2 aperture and 80% optical transmission

As far as the functionality is concerned, CMOS technologies seem to offer good opportunity to decrease the cost of star tracker systems via implementation of several functions on the same die. The evolution of the technologies, the effort that could be done on sensitivity and dark current, the reduction of lithography structure could permit to realise, in the future, star tracker imagers that could compete with some CCD devices.

That is why TCS, while continuing to develop the next generation of new CCD star tracker (in MPP mode), is working also on the evaluation of the CMOS technologies for opto applications.

Technology characterisations have started on 0.7 , 0.5 and 0.35 micron CMOS technology and, more over, components have been designed and manufactured on 0.35 microns technology for architecture evaluation.

5. CONCLUSION

TH7890 is a new CCD for star tracker applications which includes state-of-the-art technologies to better fit the requirements of the applications.

Thanks to continuous improvement of semiconductor technology, CMOS image sensors will pretty soon gain performance and become good candidates for next generation as they increase dramatically the integration level of the function and save a lot of power.

6. ACKNOWLEDGEMENTS

We would like to thank S.Caranhac, P.Cuniberti and J.L.Coutures for their contribution, and star tracker designers for their time and fruitful discussions we had together.

7. REFERENCES

1. S.Caranhac, J.L.Coutures, 1993, Comparative evolution of various CCD image sensors hardening techniques with ionizing radiation, *RADECS'93 proceedings, 2nd European Conference on Radiation and its Effects on Components and Systems*, 396-400.

2. G.Hopkinson. CCD Radiation Damage Study. *ESA contract* 9557/NL/LC (SC)

III-V SEMICONDUCTOR MICRO-MACHINED DEVICES FOR OPTICAL SENSING APPLICATIONS IN THE SWIR-MWIR SPECTRAL RANGE - TECHNOLOGICAL REVIEW

C. Pautet, S. Cortial, X. Hugon

THOMSON TCS
BP 123
38521 SAINT EGREVE CEDEX
FRANCE
Tel : 33.4.76.58.30.97.
Fax : 33.4.76.58.31.05
email : xavier.hugon@ tcs.thomson. fr

M. Garrigues, J. L. Leclercq, P. Viktorovitch

LEAME / ECL / UMR CNRS 5512
BP 163
69131 ECULLY CEDEX
FRANCE
Tel : 33. 4.72.18.60.43
Fax : 33. 4.78.43.35.93.
email : leame@ec-lyon.fr

ABSTRACT

Since a few years, innovative micro-machining technologies are developed on III-V compound semiconductors to gain benefit of their specific properties such as natural optoelectronic capacities. After a brief insight of the technology, we will focus on some specific devices like electrically tunable narrow band filters and electrically tunable selective photodetectors. A short review of some basic theoretical aspects of these devices will help to better understand their potential and limitations. Manufactured devices examples will be presented as well as their experimental performances and comparison with theory. The ability to build linear and 2D array of such devices will also be discussed with the resulting limitations.

1. INTRODUCTION

Micromachining technologies based on the silicon system have produced a large number of impressive realisations of optical and optoelectronic devices in the past 10 years. For a good overview see [1] and references therein.

However, the indirect band gap and/or operating spectrum of silicon imply the adjunction of external light sources and/or detectors to the micromachined device in order to produce a complete system in the SWIR-MWIR range.

In comparison with silicon, III-V compound semiconductors presents important technological advantages for manufacturing Micro-Opto-Electro-Mechanical-Systems (MOEMS).

The first and decisive advantage of III-Vs is that active optoelectronic devices such as semiconductor lasers, LEDs and photodetectors can be readily integrated with the micromechanical structure.

One second important advantage is the availability of epitaxial growth techniques such as Metal Organic Vapour Phase Epitaxy (MOVPE)or Molecular Beam Epitaxy (MBE) for growing heterostructures composed of different III-V alloys with interfacial roughness at the level of a monoatomic layer and precise control of composition, thicknesses and doping level.

A third interest of III-V semiconductors is their excellent aptitude to micro-machining since conditions for very high chemical etching selectivity can be easily found between different alloys.

Proc. 32nd ESLAB Symp., 'Remote Sensing Methodology for Earth Observation and Planetary Exploration', ESTEC, Noordwijk, The Netherlands, 15-18 September 1998 (ESA SP-423, December 1998)

Thanks to these properties, free standing semiconductor membranes with extremely good optical quality surfaces can be realised using selective etching on the same III-V substrates used for making active optoelectronic devices such as light sources or detectors. Based on the optical and mechanical properties of these membranes, tunable micromachined VCEL's, LED's, photodetectors and filters have been already demonstrated [2-6]. New developments on InP substrates instead of the more investigated GaAs system open the way to the realisation of passive and active devices for applications over 1.3 μm [7-9].

2. BASIC DEVICE CONCEPTS

Tunable band-pass micro-machined filter are based on the realisation by surface micromachining of a deformable Fabry-Pérot cavity inserted between two distributed Bragg reflector (DBR) mirrors of convenient reflectivity. Figure 1 represents a schematic cross section of such a filter. Tuning is achieved by moving one of the two mirrors supported on an elastic member such as a free standing membrane or cantilever beam. It thereby adjusts the thickness of the Fabry-Pérot cavity, leading to a tuning of the device resonant wavelength. A simple electrostatic action is generally used for this actuation.. The air cavity is made by selective etching of a sacrificial layer. The DBR mirrors consist in quarter-wave stacks of alternate materials with convenient refractive index contrast. These stacks are composed classically with pairs of different lattice-matched III-V alloys (GaAs/AlAs, InGaAs/AlAs, InP/InGaAs, ...), or any of a number of other semiconductor or dielectric material pairs (Si/SiO_2 for example). Air/semiconductor stacks which can be realised by selective etching of sacrificial semiconductor layers are of particular interest regarding optical performances.

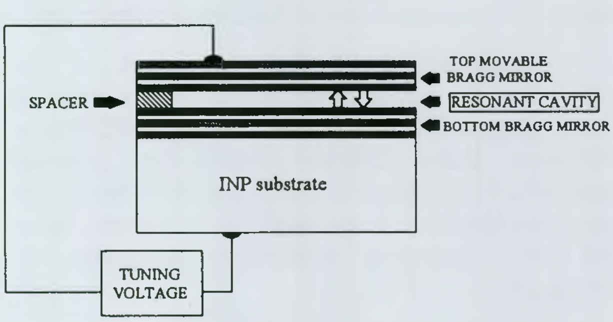


Figure 1. Cross view of a basic tunable Fabry-Pérot filter.

Such tunable band -pass filters have been mainly developed as demodulators for WDM (Wavelength Division Multiplexing) optical communication systems. In this case, selectivity performances are of prime importance.

They can also be used to realise integrated spectroscopic detectors for chemical analysis or remote sensing applications in the short to middle range Infra-Red since III-V semiconductors exhibit excellent transmission properties in this wavelength region. For this type of application, maximal tunability performances will often be required.

In order to gain full benefits from the optoelectronic properties of III-V materials, photodetectors have to be integrated monolithically with the filter. A PIN photodiode can be located outside the cavity, which is the simplest case. The photodiode can also be part of the cavity, so that the device behaves as a « Resonant Cavity Enhanced » (RCE) photodetector [10]. At the resonance wavelength, the quantum efficiency (QE) of the absorbing layer is strongly increased by the resonance factor of the Fabry-Pérot cavity. Besides this inherent selectivity, a major advantage of QE enhancement lies in the fact that the absorbing layer can be made very thin as compared to conventional photodiodes. This property may be used either to decrease response time for high speed communication systems or to decrease dark current for low level spectroscopic sensing applications.

3. TECHNOLOGICAL PROCESSING

A major advantage in favour of III-Vs surface micromachining is that standard planar processes used to realise conventional optoelectronic active devices can be easily adapted with the adjunction of few supplementary technological steps.
An example of an adapted process sequence is as follows:

1 - Epitaxial growth of the semiconductor structure,

2 - Conventional planar device process,

3 - Top Bragg mirror deposition (Si/SiO_2 for example),

4 - Reactive Ion Etching or wet chemical etching of the vertical trenches, (structuration)

5 - Wet lateral etching of the sacrificial layers, (suspension)

6 - Drying procedure.

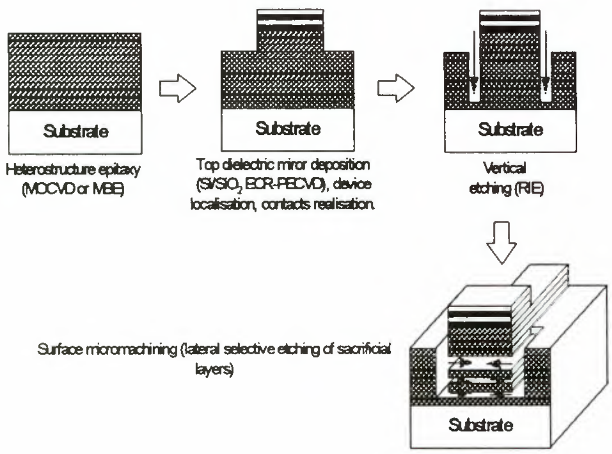


Figure 2 : Schematic micromachining process (example of a micromachined QW photodetector with bottom Air/semiconductor Bragg reflector)

Such a surface micromachining process requires control of some specific technological steps :

i) **Control of residual strain in epitaxial and top dielectric mirror layers:**

A special attention must be paid to minimise residual strain accumulated in the whole structure. Indeed, it may induce uncontrolled deformation of suspended layers leading to degradation of final devices performances.

Regarding epitaxial layers, strain may originate from imperfect lattice matching to the substrate. As a consequence, compositions must be carefully calibrated. In the particular case where both arsenic and phosphorous compounds are used in the structure, process parameters have to be optimised to eliminate natural tendency of arsenic to carry-over into phosphorous layers.

In the case of the top dielectric mirror the residual strain is directly related to deposition parameters of silicon and silicon dioxide which have to be drastically monitored and optimised.

ii) **Vertical structuration of the microstructures:**

This step consist of a vertical etching of trenches permitting access to the sacrificial layers by the lateral wet etching solutions for the suspension of the microstructures. Reactive ion etching will be preferred for vertical structuration since it allows abrupt etching over thicknesses of several microns without parasitic side etching encountered with wet chemical etching.

iii) **Lateral wet etching of the sacrificial layers:**

The fundamental parameters to take into account for the process definition are:

-the etching rate which may be driven by the low etching rate planes (kinetic mode) or by diffusion of reactive species in the solution (diffusional mode), and by the isotropic or anisotropic nature of the dissolution (related to the Zinc-blend crystallographic structure of III-V compounds). These parameters may be adjusted by temperature and concentration of the etchant solution. Mask orientation, doping level of epitaxial layers and agitation or illumination during etching are also useful to adjust the etching rate,

-the selectivity of the chemical etching between the different material used in the microstructure. In optimal cases, the chemical solution will etch only the sacrificial layers, leaving the structural layers unaffected (infinite selectivity). In some less favourable cases, the etchant and materials systems have to be optimised according to the minimal selectivity acceptable in order to keep the thickness variation of structural layers during etching of the sacrificial layer area acceptable.

iv) **Drying procedure:**

Final drying of a micromachined structure is a crucial step since during the elimination of the rinsing liquid in the etched cavities, the capillary forces may definitively stick the suspended beams to the underlying layer. Several specific techniques are used to overcome this effect: selection of very low surface tension solvents, **freeze drying** in which the solvent is frozen and subsequently sublimed without going through liquid phase or carbon dioxide **critical point drying** in which the rinsing medium goes continuously from the liquid phase to the gaseous phase through an hyperfluid state at critical pressure and temperature.

4. OPTICAL AND MECHANICAL DESIGN ISSUES

4.1 TUNABILITY

The optical thickness d of a Fabry-Pérot air cavity designed for resonance at a wavelength λ is a half-wave multiple, i.e.:

$$d = \frac{k.\lambda}{2}. \qquad (1)$$

$k \geq 1$ is an integer factor defining the cavity longitudinal mode.

The tunability of a micromachined device stands on the modulation of the cavity thickness under the mechanical action of an electrostatic field. In response to a thickness modulation of Δd, the modulation of the resonance wavelength will be :

$$\Delta\lambda = \frac{2}{k}.\Delta d \qquad (2)$$

In the simple case of a rectangular beam restrained at both ends (as illustrated in Figure 3), the deformation Δd generated by the electric potential U will be in the form :

$$\Delta d = \frac{A.L^4.U^2}{d^2 h^3} \qquad (3)$$

with A being a stiffness constant, L and h, the length and thickness of the suspended beam respectively, and d, the air cavity thickness.

From equation (2) it appears that the cavity mode k , *i.e.* the thickness of the cavity, should be the smallest possible in order to achieve the maximum wavelength shift for a given cavity deformation. Equation (3) obviously indicates that the maximum tunability will be gained for a long and thin suspended beam. Optimisation of these two conditions for a maximal tunability sensitivity relies almost entirely on the quality of technological steps (i) and (iv).

Figure 3 presents an example of design compromises for a simple low selectivity pass band filter demonstrator structure. The structure is configured at 1.5 μm with an 1 μm thick, 100 μm long InGaAs beam acting as the top DBR mirror suspended over a 0.75 μm thick air cavity and a 4.5 alternances InGaAs/InGaAlAs bottom DBR. In order to deal with capillary effects, the optical thickness of the suspended beam is configured at 9λ/4 to increase its stiffness. Final drying was simply performed using a boiling solvent.

Figure 4 shows the resonance wavelength shift as a function of electrostatic potential for this demonstrator. A resonance wavelength shift of 80 nm is obtain for a 18 V actuating voltage. In this case, it will be possible to increase the tuning sensivity by reducing the thickness of the suspended beam using a more sophisticated drying process like freeze-drying or critical point drying to prevent sticking by capillary effect.

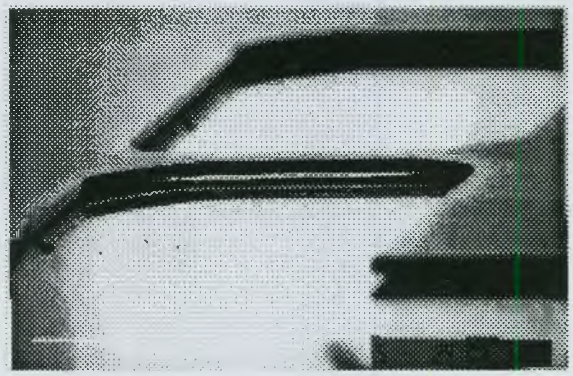


Figure 3: micromachined InGaAs/air low selectivity tunable pass band filter. [11]

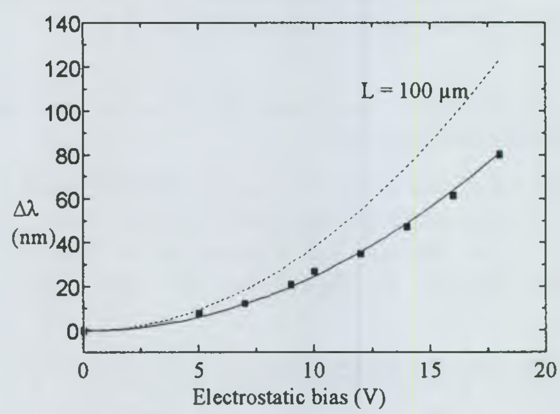


Figure 4: calculated (dotted line) and measured (solid line) resonance wavelength shift as a function of applied bias. [11]

4.2 SELECTIVITY

The selectivity of a loss-less Fabry-Pérot cavity of thickness L = k.λ/2 with end mirrors of equal reflectivity R is basically described by equation (4) below:

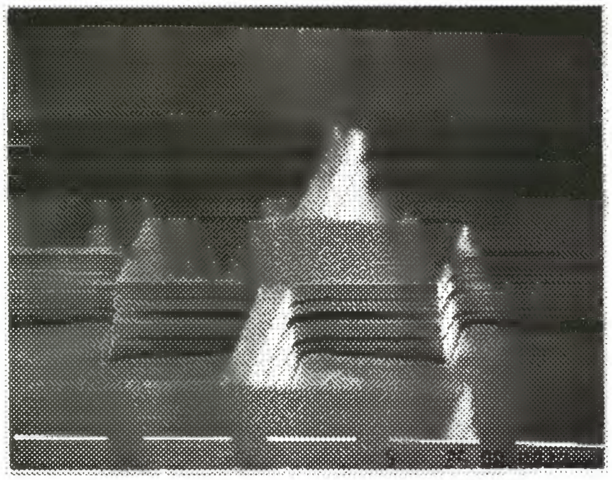
$$\Delta\lambda = \frac{\lambda}{k.\pi}.(1-R) \qquad (4)$$

This simple formula is useful for a rapid quantification of the influence of mirror reflectivity and longitudinal mode of the cavity. It shows the importance of high reflectivity Bragg mirrors to achieve a high selectivity. A drawback of III-V compound semiconductors and especially lattice-matched compounds to InP is the low index stepping between materials. It implies that all semiconductor Bragg mirrors must be grown with a large number of quarter-wave alternances in order to achieve high reflectivities. This drawback can be overcome by the use of micromachined Air/InP quarter-wave stack for the bottom Bragg mirror which yield better reflectivity as shown on table 1.

Bragg mirror	Index stepping (Δn)	Alternance number	Reflectivity (%)
GaAs/AlAs	3.37 : 2.9 (0.47)	20	99.7
InGaAs/InP	3.58 : 3.17 (0.41)	20	93.2
Si/SiO2	3.6 : 1.45 (2.15)	3.5	99.5
InP/Air	3.17 : 1 (2.17)	3	99.9

Table 1: Reflectivity of different kinds of quarter wave stack Bragg mirrors calculated at 1.55 μm. [11]

Figure 5 show an application of this improvement in reflectivity for the realisation of a selective (10 nm) pass band filter configured at 1.55 μm.



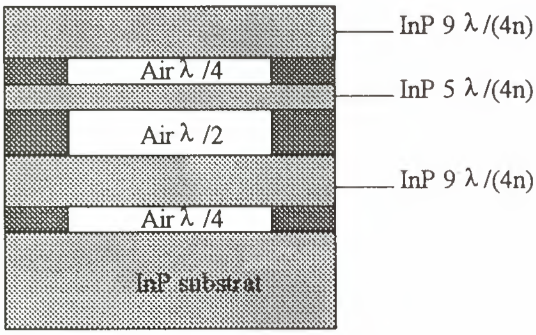


Figure 5 - Micromachined selective filter. Suspended membrane 40 μm x 40 μm [11]

The epitaxial structure is made by stacking lattice-matched InP/InGaAs layers. Then, InGaAs layers are selectively etched away to produce suspended InP beams and λ/4 or λ /2 air cavity. The central thicker Fabry-Pérot air cavity is visible at the middle of the structure with on both side the top and bottom reflectors. As the drying as been made with boiling solvent, InP beams have been configured in 9 λ/4n or 5 λ/4n to increase their stiffness and sustain capillary effects.

Microreflectivity measurement shows a spectral response in accordance with theoretical evaluation (figure 6). The full width at half maximum of the transmittance peak around 1.55 μm is 13 nm instead of a predicted value of 12 nm. The resonance position is slightly moved as compared to the expected value owing to a miscalibration of epitaxy growth rates.

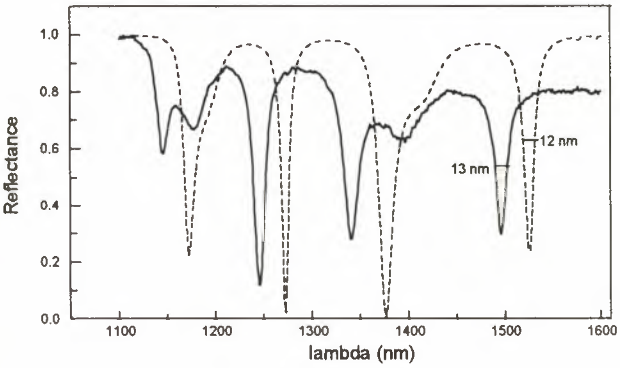


figure 6: theoretical (dotted line) and measured (solid line) reflectivity of InP/air filter.[11]

4.3 STOP-BAND

The stop-band of a reflector is by definition, the wavelength range where the reflectivity is maximum. The stop band of the Bragg reflector must be compatible with the aimed wavelength tunability range of the device. The stop band is expressed approximately by equation (5) below:

$$\frac{\Delta\lambda}{\lambda} = \frac{4}{\pi}\arcsin\left|\frac{1-n_h/n_l}{1+n_h/n_l}\right|\frac{1}{2k'-1} \qquad (5)$$

where n_h and n_l denotes optical index value of the two materials used for the Bragg reflector and k' the longitudinal optical mode in the quarter-wave layers of the reflector.

This equation shows that the stop band is inversely related to the thickness of the mirror layers by the k' term. It means that the stop-band decreases when the layer thickness is increased to sustain drying problems or residual strain. The n_h/n_k term indicates that the high index contrast obtained for micro-machined air/semiconductor stacks improve drastically the stop band as compared to all semiconductor mirrors. Table 2 illustrates those two effects.

Bragg mirror	Index stepping (Δn)	Alternances number	Reflectivity (%)	Stop band (nm)
InP $\lambda/4$-Air $\lambda/4$	3.17 : 1 (2.17)	2	95	350
Si $\lambda/4$-SiO_2 $\lambda/4$	3.6 : 1.45 (2.15)	2	95	350
InP 9 $\lambda/4$-Air 3 $\lambda/4$	3.17 : 1 (2.17)	2	95	75
InGaAsP $\lambda/4$-InP $\lambda/4$	3.4 : 3.17 (0.23)	20	95	50

Table 2 : Stop band of different types of quarter-wave stack Bragg mirrors calculated at 1.55 μm. [11]

4.4 DESIGN RULES SUMMARY

The design of the micromachined device must take into account device targeted characteristics in regards to technological limitations:

-the materials system must be lattice matched to the substrate (in order to limit the crystallographic defects density detrimental to device performances) and have convenient optical characteristics (the reflector materials should be non-absorbing at the resonance wavelength) associated with the availability of a etching solution with good selectivity,

-the reflectivity of Bragg mirror must be high enough to ensure good optical efficiency and selectivity, this may be achieved by the use of air/semiconductor alternances instead of semiconductor alternances for the bottom Bragg reflector,

-the design (layers thicknesses and length) of the Fabry-Pérot cavity and Bragg layers should give enough stiffness to sustain drying step and residual strain and on the other hand allow a sufficient cavity deformation under moderate electrostatic field to achieve the aimed tunability range. The mirror mode k' (i.e. the thickness of Bragg reflector layers) should be kept as low as possible to achieve a large stop band.

5. APPLICATION TO RESONANT PHOTODETECTORS

As stated above, fully integrated active MOEMS devices such as electrically tunable selective detectors can be fabricated by micro-machining III-V materials. Using the RCE detector concept, it is possible to use very thin absorbing layers, i.e. Quantum Wells (QW) to realise low dark current tunable detectors for the SWIR band. A sketch of a prospective resonant QW diode is drawn on Figure 7.

The bottom Bragg mirror is constituted of a quarter wave stack of InP and air layers obtained by selective under-etching of sacrificial InGaAs layers. The resonant cavity includes an air gap plus the InP layer forming the diode itself. The absorbing QW is located at the position of an electric field maximum at the resonance wavelength. The top mirror is made of a Si-SiO2 stack deposited using a plasma deposition process. The resonant wavelength of this device can be electrically controlled by actuating the thickness of the air cavity as for the passive filters described in the previous sections.

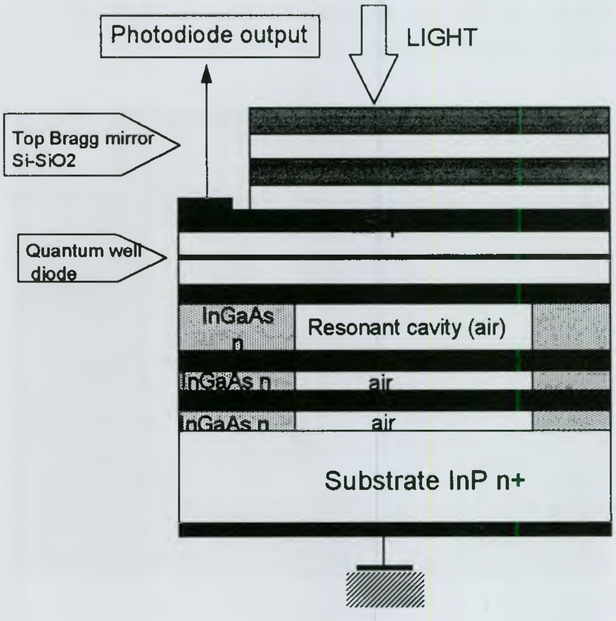


Figure 7: Resonant QW photodetector.

Using InGaAs as the material of the QW, operating wavelengths from 1.5 μm up to 2 μm and more can be obtained, depending on the In concentration of the ternary alloy.

Two critical design constraints must be considered:

i) The aspect ratio (thickness/length) of the suspended InP beams must be sufficient to avoid detrimental deformations due to residual strain in the epitaxial layers and top mirror and to prevent sticking of the layers during the drying process. Thicknesses can be increased by increments of half-wavelengths which are optically transparent. However increasing thicknesses affects the tunability range and selectivity as explained in § 4.

ii) The thickness of the diode layer embedding the quantum well must be sufficient in order to accommodate the P+ diffusion depth plus the depletion layer depth under a few volt reverse voltage.

The maximum intrinsic absorption (without any resonant effect) of the quantum well is only about 1% and will be enhanced at the resonance of the cavity. The QW enhancement factor will depend essentially on the reflectivity of the Bragg mirrors limiting the cavity so that it will be completely related to the selectivity of the device. Consequently, and keeping in mind the initial design limitation on thicknesses, an important design issue is the number of quarter wave alternances for the bottom Bragg mirror. It determines entirely the quality factor of the Fabry-Pérot resonator and consequently the quantum efficiency QE and selectivity of the resonant detection.

Due to the strong refraction index contrast between air and InP, the optical performances is a very discontinuous step function when the number of quarter-wave alternances is increased. This is illustrated in Table 3 below:

Bragg mirror	reflectivity	QE at resonance	spectral B.W.
one alternance	88 %	6 %	30 nm
two alternances	98.7 %	38 %	6 nm
three alternances	99.9 %	85 %	2 nm

Table 3. Performances of an RCE QW photodetector as a function of Bragg mirror design.

This table shows the effectiveness of QE enhancement which can be achieved for a RCE detector. These are theoretical values based on an estimated free space absorption of about 1 % for a single quantum well and optically perfect layers. Practically, a smaller value of QE is expected, due to scattering losses at the interfaces. The efficiency losses will increase as the theoretical cavity enhancement effect increase, so that the three alternances design is certainly too optimistic regarding QE. The spectral bandwidth of 2 nm will be also too low for many applications.

In this particular case, the two alternances design clearly offers the best trade-off between quantum efficiency and selectivity.

We have used experimental results from the present state-of -art control of residual strain and CO_2 critical point drying for designing the thicknesses of the beams for the demonstrator shown in Figure 8 using a two alternances bottom mirror. Numerical simulations show that the expected tunability range will be better than 10 % of the initial resonant wavelength. The QE will be 38 % and the bandwidth about 6 nm as calculated in Table 3.

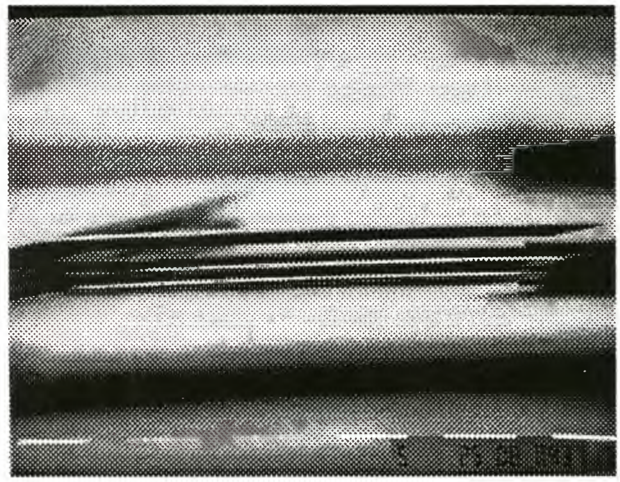


Figure 8 : Two alternances bottom air/semiconductor mirror micromachined QW photodetector structure.

6. CONCLUSION AND FUTURE POTENTIAL

We have demonstrated MOEMS devices with tunability of 80 nm. Advanced drying processes and better control of residual strain will allow to extend the tunability ranges over 200 nm or more by permitting a reduction of layers thicknesses.

One industrial advantage of tunability regarding yields is that process related resonance wavelength drift or dispersion can be corrected by setting an offset bias for each device manufactured.

The selectivity may be adjusted from less than one nanometer to 20 or more nanometers depending on the trade-off with other aimed device performances .

Tunable filters made with InP/Air layers can operate from 1 to 5 μm without parasitic absorption of the structural material. In case of photodetectors with an InGaAs absorbing QW, the maximum long wavelength operating range will be around 2.2-2.4 μm for dark current considerations.

MOEMS are expected to bring answers to a lot of applications. Narrow bandpass tunable filters will allow to select among one laser beam among others and to follow its wavelength temperature shift relaxing laser requirement for Wavelength division multiplexing data transmission, for example. Single resonnant photodiode may replace more complex grating+ moving mechanics + photodetector assembly in dedicated low cost gas sensors.

Linear arrays of such resonnant photodiodes will be the next development step. Such arrays could replace sophisticated spectroscopy instrument with corresponding cost, weight, and volume economy. Of course, reliability, performance stability and calibration techniques must be demonstrated and optical design must be adapted in order to take into account this highly reflective components. Linear arrays with 52µm pitch and more than 75% filling factor are today considered as achievable, with collective tuning bias or preferably individual ones (making possible to compensate with bias offset any manufacturing spreading) .

2D arrays seeems also possible, however filling factor will be reduced (less than 50%) and pitch limited due to the need of supporting walls at least on two sides of the microstructure. Bumps will be places on these walls. Filling factor may also be compensated in using microlens arrays on back side of the photodiode array. Individual tuning bias will no longer be possible, however, using common bias for each line and image scanning across line direction will configure the array as spectro imager with selectable spectrum sampling.

Even if there is still a long way to make such devices enter into reality, it is obvious that they will bring interesting opportunities in new instrument designs.

references :

1- M. C. Wu, « Micromachining for Optical and Optoelectronic Systems », Proc. IEEE, vol. 85, NO. 11, pp 1833-1856, 1997.

2- F. Sugihwo et al., «Low threshold continuously tunable vertical cavity surface emitting lasers with 19.1 nm wavelength range », Appl. Phys. Lett.,vol 70, pp 547-549, 1997.

3- M.Y. Li et al., « High performance continuously tunable top-emitting vertical cavity laser with 20 nm wavelength range », Electron. Lett., vol. 33, pp 1051-1052, 1997.

4- G. L. Christenson, « Long wavelength resonnant vertical cavity LED/photodetector with a 75 nm tuning range », IEEE Photon. Technol. Lett., vol. 9, pp 725-727, 1997.

5- E. C. Vail et al., « GaAs micromachined widely tunable Fabry-Pérot filters », Electron. Lett., vol. 31, pp 228-229, 1995.

6- M. C. Larson et al., « Broadly tunable resonant cavity ligth emitting diode », IEEE Photon. Technol. Lett., vol. 7, pp 1267-1269, 1995.

7- K. Streubel et al., « Room temerature pulsed operation of 1.5 mm vertical cavity lasers with an InP-based Bragg reflector », IEEE Photon. Technol. Lett. vol. 8, NO 9, pp 1121-1123, 1996.

8- K. Streubel et al., « 1.26 mm vertical cavity lasers with two InP/air-gap reflectors », Electron. Lett., vol. 32, NO. 15, 1369-1370, 1996.

9- C. Seassal et al. « Fabrication of InP-based freestanding microstructure by selective micromachining », J. Micromech. Microeng., vol. 6, pp 261-265, 1996.

10- Unlü et al, J. Appl. Phys., vol 78, N°2 , pp 607, 1995.

11- C. Seassal, « Elaboration et étude de micros-dispositifs mecano-optiques en semiconducteurs III-V », Mémoire de Thése n° 97-33, Ecole Centrale de Lyon, 1997.

SWIR FOCAL PLANES FOR REMOTE SENSING APPLICATIONS: PRESENT STATUS AND FUTURE EXPECTATIONS

X. HUGON ; O. AMORE; L.BRISSOT; S.CORTIAL; S. LACROIX; D.MOLLARET; C.PAUTET

THOMSON TCS
BP 123
38521 SAINT EGREVE CEDEX
FRANCE
Tel : 33.4.76.58.31.93
Fax : 33.4.76.58.31.05
email : hugon@tcs.thomson.fr

ABSTRACT

InGaAs photodetectors has proven to reach unbeatable performances level at near room temperature for Near Infrared spectral range covering 0,8 µm up to 1,7 µm.

Present achieved performances will therefore be summarized, based on several space programs as well as on industrial applications. Electro-optical performances such as spectral response, noise, operating temperature, detectivity will be discussed as well as multiplexing and read out techniques.

New trends are then reviewed. 2D array development status and future will be discussed and last results presented. Extended (toward longer wavelength) spectral range device feasibility, difficulties and limits will be overviewed. The different techniques involved, their own expectable performances and limitation will be exposed even as well as foreseen availability.

1. INTRODUCTION

Short Wave Infrared spectral range extends just in Silicon sensitivity spectra prolongation. i.e. : from about 1µm up to 2,5 µm basically. Consequently silicon made photodetectors are inefficient in this spectrum. Among many extensively studied materials, InGaAs, a III-V compounds has proved its incomparable aptitude to fulfill the most stringent performance requirements in SWIR spectral range.

The following will remind some basic aspect concerning the material, the resulting detectors and their operation.

InGaAs success comes from two main factors : the availability of good quality substrates and rather good epitaxial growth ability, both missing in MCT case, for example. However, a lot of development efforts have been necessary to bring InGaAs to its present industrial status outperforming any other candidate in the SWIR range.

III-V compounds, but binary alloys, cannot be grown in single crystal ingots since III or V elements are miscible in any proportion, preventing from growing single phase crystal. Thus only binary alloy substrates are available. Total miscibility means that any In_xGa_{1-x} As composition is obtainable from InAs to GaAs state. Compositional change allows to modify compound band gap (from 1.4 eV down to 0,4 eV) but also impacts lattice parameter, which in InGaAs case is highly dependent on In content. Once again stands up the inevitable substrate problem.

Fortunately, it happens that InP, another III-V binary compound, has the same lattice parameter than $In_{0,53}Ga_{0,47}As$ which band gap is 0.74eV, corresponding to 1.67 µm cut off wavelength. Afterward, epitaxial technique developments and InP substrate crystal growth improvement provide what was needed to step up InGaAs photodetector technology.

2. PHOTODIODE TECHNOLOGY

Most of InGaAs based photodetectors use photodiode as sensing elements. Photodiodes are made on an InP/InGaAs/InP heterostructure where the InGaAs active layer is epitaxially grown on an InP substrate and is covered by another InP cap layer ***(Figure1)***. Due to technological constraints, these layers are N type doped or near intrinsic . P/N photodiodes are thus obtained through Zn diffusion for P doping, using planar technology .

Proc. 32nd ESLAB Symp., 'Remote Sensing Methodology for Earth Observation and Planetary Exploration', ESTEC, Noordwijk, The Netherlands, 15-18 September 1998 (ESA SP-423, December 1998)

Therefore, InGaAs based arrays are always in a common cathode configuration. InP cap layer allows junction edge to rise up in a large gap material, reducing junction perimeter contribution to current leakage.

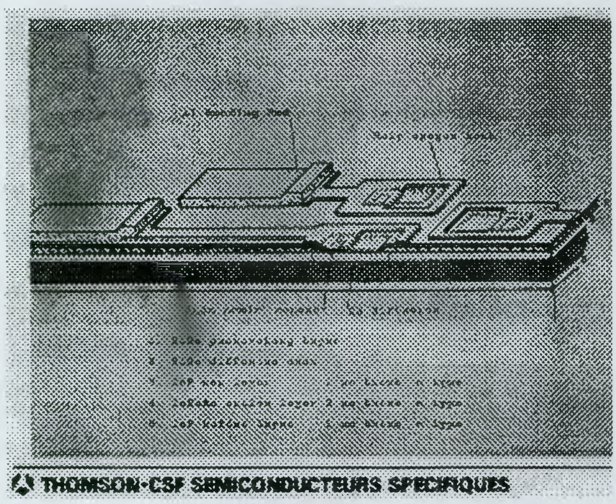


Figure 1 - Photodiode cross section

Processing III-V compound and especially InP/InGaAs faces some difficulties compared to silicon technology. Low temperature processes are required, since InP starts to loose phosphorus above 250°C, without encapsulation or phosphorus counter pressure. There is no native oxide with good electrical properties (as SiO_2 for silicon). Crystalline defects like dislocations limits manufacturing yield while InP substrates are only available in 2" diameter, (soon in 3"). Dedicated processing lines are required, which in addition to ten times higher substrate cost than for silicon leads obviously to more expensive components. Though a lot of progress have been done which allows to introduce such detectors in large scale industrial applications. As an example ***(Figure 2)*** shows how photodiode dark current performances have been improved by the past, allowing TCS to propose zero defect arrays quite commonly.

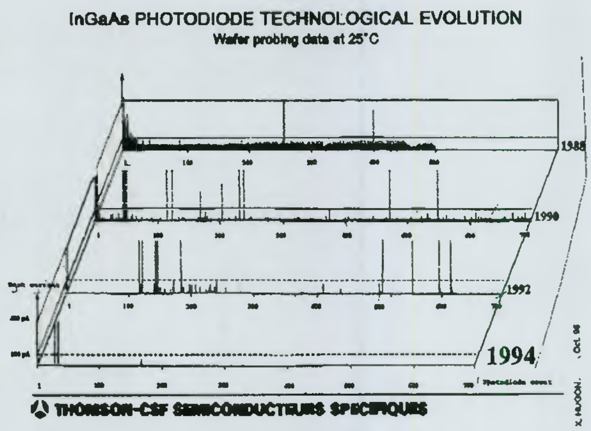


Figure 2 : Ambient temperature wafer probe results over 6 years development

This chart is a plot of dark current of all photodiodes taken along one line on a processed wafer in 1988 (small wafer size), in 1990, 1992, 1994 where it can be seen that both mean values and singular pixel density have been dropped down.

3. LINEAR ARRAYS DEVICES

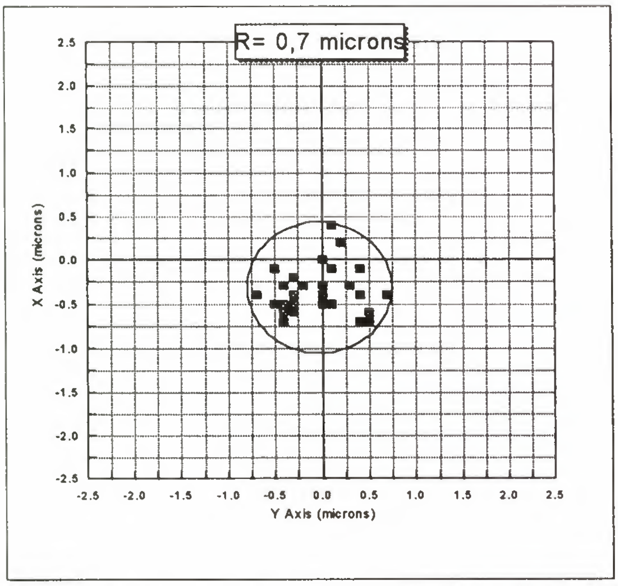
3.1 Device Concepts

Photodiode process on InP/InGaAs has required large efforts in the past ten years, however further complex functions are still unachievable with same performance level as done on silicon. InGaAs detector arrays have, therefore, to be coupled to silicon made readout circuits and hybrid concept is imposed. Moreover, considering the limited wafer size, TCS has developed a buttable elementary module hybrid concept. At first, only linear arrays were manufactured since each photodiode has to be connected to one input of a readout /multiplexing circuit using the only available interconnection technique : wire bonding , even though small pitch requirement (26µm) looked like a breakthrough. A two wire bonding layer at 104µm pitch on each side of the central photosensitive area has solved the concern. 2D arrays for this reason have been developed later on, as it will be mentioned hereafter.

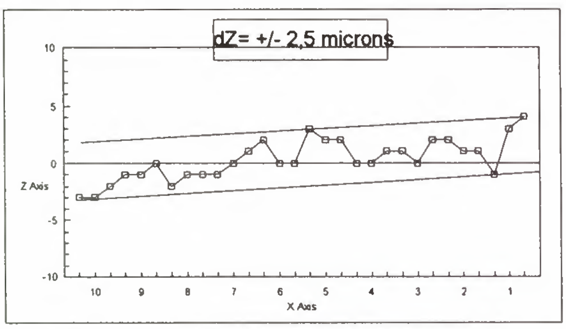
Elementary modules including one InGaAs die and its readout circuits on a ceramic substrate are manufactured, tested and sorted. Such elementary modules may be butted together so as to build long arrays.

High accuracy butting technique have been developed which allows to manufacture arrays up to 90 mm long with less than +/- 1.5 µm absolute error on any pixel positioning in X-Y direction and less than +/- 5 µm along Z direction (perpendicular to focal plane).

Figure 3 shows typical results on such an array.



Plot of X and Y direction deviation from theoretical positions of 30 pixels sampled along 3000 pixel arrays (80 mm long)



Zaxis (perpendicular to focal plane) pixel position spreading

Figure 3 : Long butted array geometrical accuracy

3.2 Readout circuits

TCS has designed a CCD type read out circuit achieving on chip photocurrent integration, collection, serialization and conversion to voltage [1], [2]. As CCD basic technology is handling electrons while InGaAs photodiodes inject holes a « vidicon mode » read out technique has been implemented. It consists in injecting known quantity of electrons (Qb) on photodiode node capacitance -Cin-. During integration time, injected holes recombine with electrons, thus in measuring the lack of electrons, at the end of integration time, photosignal is deduced.

As direct consequences of this read out technique one may notice that :

- photodiodes must be reverse biased (since charges are stored on their capacitance node), this means that even in darkness a photodiode leakage current (dark current) will be integrated as well as it's associated shot noise.
- Photodiode reverse bias varies during integration time. Moreover reverse bias is reduced until no longer electrons are available, i.e. bias has become zero (Cin is fully discharged) which, by the way, provides intrinsic antiblooming.
- At each integration cycle, an electron quantity corresponding to the saturation photosignal must be prepared and injected into Cin.

Such CCD based devices present many advantages :

- They use inexpensive process (11 mask levels compared to 24 for some CMOS technologies).
- They easily allow to read out video data in parallel with next integration cycle, a valuable feature for no lag continuous integration cycles (like for pushbroom imaging).
- Qb charge injection, and photosignal charge read out are serial processes and use the same stages for all data, which reduces spreading.
- Input stages threshold spreading effects are canceled thanks to vidicon mode.
- For each pixel is provided a reference level (floating diode level) which contains readout stage reset state information allowing, with correlated double sampling, to remove low frequency temporal read out noise.
- Operating frequency up to 10 MHz are possible even if TCS present devices are limited to 2.5 MHz.
- They provide Irradiation withstanding superior behavior.

In the other hand, such architecture brings limitations among some are the following which has pushed us to investigate other alternatives.

- Photodiode reverse biasing implies that extra charge handling capacity is implemented for dark current integration. This point could become a stringent limitation when dealing with photodiodes having large I_{sat} like extended wavelength InGaAs ones or MCT made or even for large size photodiode arrays.

- Electron injection on photodiode node capacitance process (photodiode reset) brings so called Johnson noise which can be expressed in this case as $\sqrt{kTCin}$. Cin includes read out circuit input capacitance, photodiode capacitance, bonding pad capacitance's (one on photodiode die, one on read out circuit die) and interconnection wire capacitance. As photodiode size is reduced, its contribution to Cin drop down such that for imagery application Cin is dominated by wire bonding pads capacitance's. Moreover on last generation CCD based read out circuits [3], total noise is dominated by photodiode reset noise, i.e. detector output noise is actually wire bonding pad noise! Quite a crazy situation.

The following ***Figure 4*** shows relative weight of photodiode shot noise, photodiode node reset noise (taking into account Cin in whole) and what would have been photodiode node reset without bonding pads and wire (call « photodiode alone reset noise » in figure 4). Total readout noise is also drawn for comparison. It appears clearly that until 10ms integration time photodiode performances are spoiled by the readout technique.

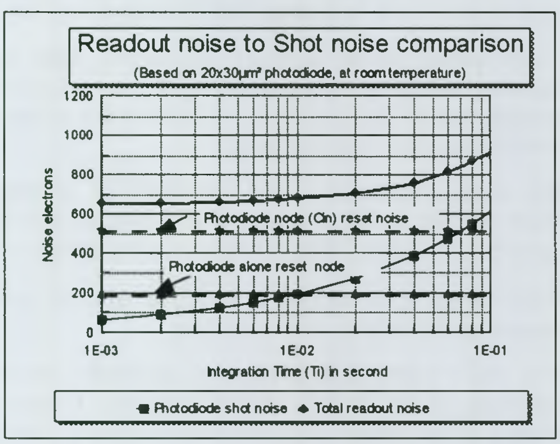


Figure 4

This last concerns is purely linked to the fact that photocurrent integration does not occur at constant photodiode bias. To implement such an operating mode one way could be to use hole handling CCD, some more ten years of development...! A more realistic way is to implement close looped input stages for each photodiode. As CCD technology (NMOS) is not able to provide such function, CMOS technology must be used.

3.3 CMOS readout circuits

In the selected architecture, each photodiode is interconnected to a CTIA input (Capacitance feedbacked transimpedance amplifier). Therefore, it is operated at a constant bias, its current (darkcurrent + photocurrent) is integrated on a feedback capacitor C_{int}. Once constant voltage operation is achieved, zero bias operation becomes obvious to avoid dark current contribution to output signal. Using zero bias strategy changes photodiode critical parameter into Shunt Resistance (Rsh). Rsh spreading distribution is multiplied by input offset spreading distribution and results in DSNU (Dark Signal Non Uniformity). The reason while as low as possible input offset spreading is mandatory, since photodiode Rsh uniformity may be difficult to drastically improve. Thus we have implemented an automatic offset zeroing system on each CTIA, which is activated at each integration cycle. Offset spreading of less than 0,3 mVrms are obtained on devices in development. Care must then be taken not to exchange CCD $\sqrt{kTCin}$ noise with other CTIA noise sources, such as input offset noise..

Preliminary measurement results on TCS CMOS linear 128:1 readout/multiplexing circuit are presented in ***Table 1***.

The input pitch is 52 µm, compatible with 26µm pitch linear array manufacturing. Correlated double sampling and some logic utility functions are included on chip. These circuits are also buttable. High impedance output capability is implemented in order to allow single output devices whatever is butted module count

Performance item	Unit	Present design	Future trend	Comment
Pitch	µm	52	26	
Input number		128	300	
Output number		1	1	With chaining capability
Conversion factor	µV/e	0.16/0.016 switchable	1	
Saturation level	V	2,2	2,2	
Input noise	µV	50	40	
Input offset non uniformity	Mvrms	0,2	0,15	
Non linearity	%	< 1	<1	Over 3 decades
Total read out noise with 20x30 µm photodiode	µV e	180 1100	500 500	

Table 1 : CMOS readout circuits performances

As it can be seen , noise figure is rather disappointing. The reason is that once again input node capacitance Cin becomes a limitary factor through CTIA input noise voltage, Iin, amplification. e_{in} acts in two ways. First it results in a photodiode noise current :

$$I_{in} = \frac{e_{in}}{rsh}$$

which is integrated as well as photodiode shot noise current. In addition, eout, the CTIA output noise, with a pure capacitive input source can be expressed as :

$$e_{out} = e_{in}\left(1+\frac{C_{in}}{C_{int}}\right)$$

pointing out Cin impact.

Cin will be difficult to drop down so most work are focused on e_{in} reduction, however, within a particular technology, few solutions are possible, i.e. no miracle is to be waited for.

Concerning Cin, it will be difficult to cut it down drastically, unless interconnection technique is changed. Bump technology may be an alternative in the future.

Therefore, it must be retained that such CMOS multiplexors are not necessarily a better option than CCD one.

A CMOS readout circuit represents a good alternative when photodiode shot noise is predominant, ie for large size photodiode or high current density ones, since zero bias operation is possible while ultra low offset spreading have been demonstrated and large charge handling capacity are possible (large Cint).

New generation SWIR linear array could be based on these new readout circuits, providing easier operation as well as better performance level on large size spectroscopy dedicated as well as on extended cut off wavelength InGaAs photodiodes.

3.4 Linear arrays performance

Available linear arrays are based on CCD readout circuits. They provide high performance level with spectrum coverage ranging from 0.9µm up to 1.7µm. For historical reason operating temperature are set to +15°C and integration time to 3 ms. Table 3 gives a synthesis of typical features measured on 3000 20x30µm² pixels with 26µm pitch in line arrangement. Implemented conversion factor from injected electrons to output voltage is 0.26µV/e i.e. 1µV corresponds to about 4 electrons readout.

ITEM	FIGURE	UNIT	COMMENTS
Mean dark voltage	2	mV	
Mean noise	175	µV	
Quantum efficiency	0.7		
Conversion factor	0.26	µV/ e	
Absolute Response	9.5	Vcm²/ µJ	[1.57µm- cutoff]
PRNU	+/-10	%	
Cutoff wavelength	1690	nm	At 50% of Rmax
Cutoff spreading	10	nm	Over whole array
Dynamic range	2	V	
Non linearity	+/-1	%	Over 1.5 Volt
MTF	0.7		Across track
	0.4		Along track
NEP	40	fW	3ms integration
D*	$8\ 10^{11}$	$CmHz^{0.5}/W$	3ms integration

Table 2 - Typical linear array figures

Performance in darkness can be better appreciated in looking to Figure 5 histograms, where are presented darksignal, noise, and darksignal stability over 2 hours. This last figures are obtained in acquiring output signal in darkness Vd, continuously over two hours, and to plot for each pixel the difference between the higher acquisition and the lower one.

Figure 5 points out the high quality level achievement of these arrays: no singular nor dead pixel, about 0.1mVrms in DSNU, 175µV mean noise with 12µVrms in noise spreading from pixel to pixel, less than 20µV in mean darksignal stability with all pixel below 160µV stability.

Response spreading or Photo Response Non Uniformity (PRNU) is also shown in Figure 5. PRNU figures include cut off wavelength spreading effects (from module to module) since measurement are done using a 1.57µm cut on high wavelength pass filter, so that 1nm in cutoff spreading leads to near 1% in PRNU.

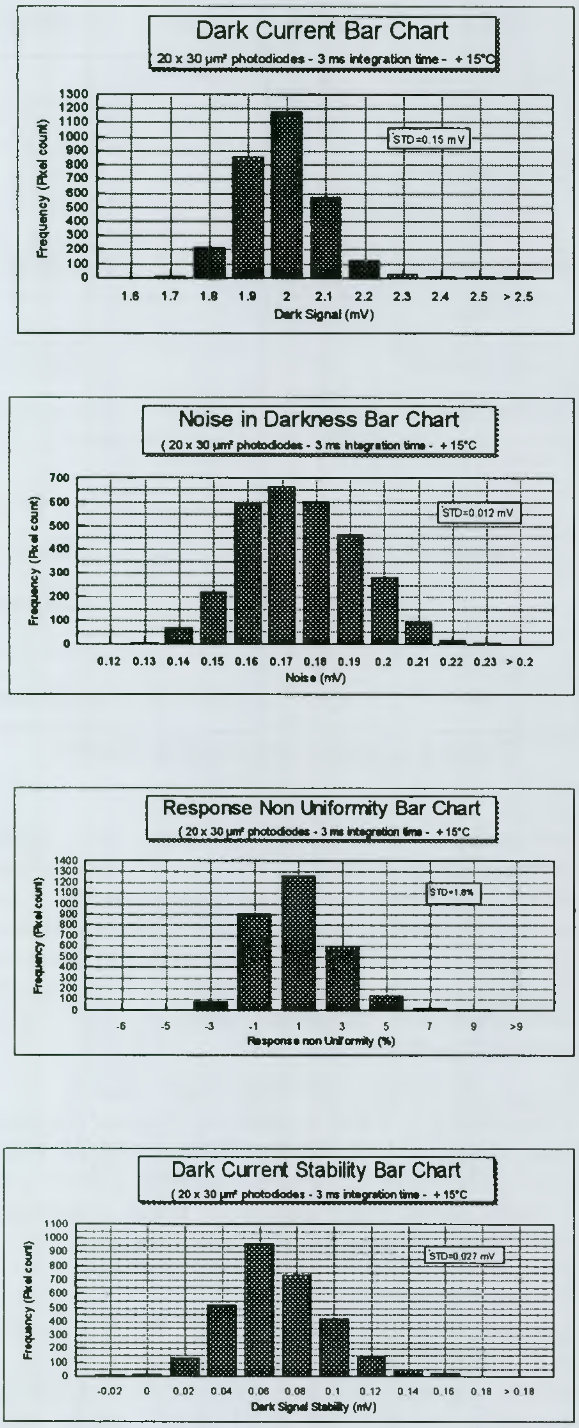


Figure 5 **: 3000 pixel linear array performance**

Taking into account noise is mainly so called "KTC" noise coming from photodiode reset, it appears obvious that Noise Equivalent Power (NEP) as well as specific detectivity (D*) will improve with integration time, until photodiode shot noise become predominant. This is shown in ***Figure 6*** for two different photodiode sizes.

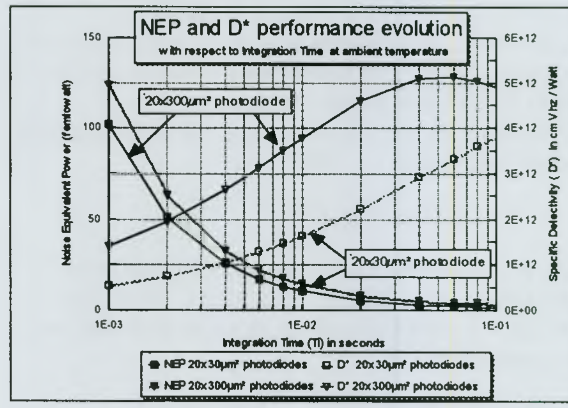


Figure 6

3.5 Linear arrays new trends

New high resolution linear InGaAs arrays are under development. Pitch will be 13µm, 13x13µm² photodiodes will be staggered along two rows separated by 26µm. They will use new 26µm pitch CMOS readout/multiplexing circuits with improved noise performances (less than 500 electrons in total noise is expected) . Interconnection will be done through four wire layers on each side of the photosensitive area. Wire bonding pitch on each layer will be kept to 104µm as before. A dedicated technology will be used to reduce bonding pad capacitances. A 6000 pixel array is then expected, including not less than 6760 wires in a single package. Such space qualified devices are expected in early 2001.

4. 2D ARRAYS

InGaAs 2D arrays' manufacturing assume the capability of interconnecting each photodiode, laid on a matrix organisation, to a multiplexor input. This is obtained through so called "bump" interconnection technique. It consists in processing on one of the substrates small balls made of low melting point metal (commonly In or SnPb) attached to electrical contacts. The other substrate is positioned so that its electrical contacts face the balls then, the balls are fused.

In this configuration photodiodes are back side illuminated (through InP substrate). Bump size must, of course, be lower than the pitch therefore photodiode node capacitance (Cin) is drastically reduced since interconnection pad (now located over photodiode area itself) no longer adds extra capacitance.

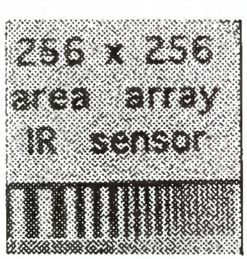
TCS has designed, manufactured and tested a 256x256 2D array with 48µm square pitch. It is based on a new CCD 2D readout multiplexing circuit using the same "Vidicon" mode readout technique. CCIR operation is possible and Achieved performances are presented in ***Table 3***. hereafter.

Pixel count	256 x 256
Pixel pitch	48 µm
Filling factor	100 %
Number of output	1
Max readout frequency	10 Mhz
Conversion factor	0.9 eV/e
Mean dark voltage	20 mV
Mean response	100 Vcm²/µJ
Response uniformity	3 %
Saturation voltage	2 V
Noise voltage in darkness	500 e-
CTF at 10,4 fp/mm	> 0.6
Singular pixel	< 1 %

Table 3 - 2D array present performance at +5°C and 20 ms integration time

Bumping basic technology is no longer a concerns and pretty good yield are achieved. But as InGaAs photodiodes exhibit more than 10^{12} ohms shunt resistance in reverse bias conditions, electrical insulation between bumps must, at least, be in the same range. Leakage between two inputs can introduce signal mixing and, due to threshold difference from input to input, will allow preload charge (Qb) exchange from one input with higher threshold to another with a lower one. This last effect could result, for example, in having negative signal in darkness on some pixels (at photodiode reset one could find more electrons on a pixel than injected with Qb while in the same time another pixel will be found with extra darksignal –lack of electrons- of the same quantity). For this reason bump process has to be carefully chosen.

Figure 7 shows dark signal image of such an array and an image without and with dark signal correction.



Non treated image

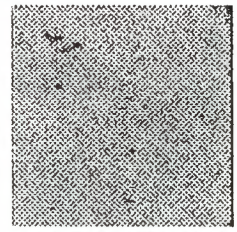


Image in darkness (used for previous image treatment)

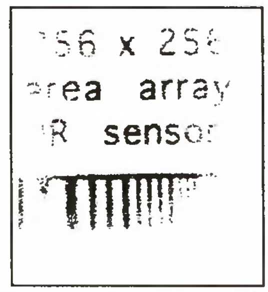

Treated image (Darksignal substraction)

Figure 7: 256x256 2D array image

5. SPECTRAL EXTENSION

As exposed earlier, InGaAs compound's success comes from its providential InP substrate, however lattice parameter matching to InP leads to about 1.67µm cut off wavelength. Any increase in cut off wavelength needs for In richer compound resulting in lattice mismatch with InP substrate. TCS commonly use $3°/_{00}$ mismatched heterostructures in order to reach 1.7µm cut off, however mismatch grow up rapidly with cut off wavelength it reaches 2% for 2.5µm cut off. Such lattice parameter mismatch means that every 50 atoms one is in excess! As it can be imagined compressive strain related energy accumulation in growing layers increases so rapidly with thickness that, in the best case, it reaches after few atomic layers dislocation generation energy threshold, otherwise so called "3D" growth takes place which results in micro island structures with major defects at their boundary junction – a preliminary step before poly crystallinity- . Highly perturbed crystals are then obtained, unable to be processed. Therefore InGaAs spectral coverage extension is rather more a matter of crystal growth technology.

Up to now a lot of works have been reported aiming to overcome these concerns. Three main techniques are competing :

- Metamorphic techniques the oldest one, and the more advanced. It consists in introducing in between InGaAs layer and InP substrate a so called "buffer" structure. The most common buffer structure is designed so as to progressively extend lattice from the substrate level up to expected active layer's one within buffer thickness which by the way must be quite large (10µm or more) -a kind of energy dilution-.

- Paramorphic techniques which consist in growing thin mismatched InGaAs layer (commonly some tens Angstroms) on a sacrificial lattice match layer, then to remove the sacrificial layer by wet selective etching and to have it stuck again on the underlying substrate. In the mean time InGaAs layer will have been free standing for a while, during which strain release occurred. Finally it remains a thick substrate with on top few layers of desired lattice parameter and good crystalline quality.
- Compliant substrate techniques which principle is to try to reach dislocation energy threshold within the substrate before it happens in the growing active layer. Physically this situation could be obtained with very thin substrate (some tens atomic layers). However, as such thin membrane cannot be handled all works are dealing with the way to stick these substrates on thick ones with mechanical uncoupling.

None of these techniques have yet demonstrated their total achievement though the metamorphic way, has been industrialised and used for manufacturing up to 2.6µm cut off wavelength even for space applications [4].

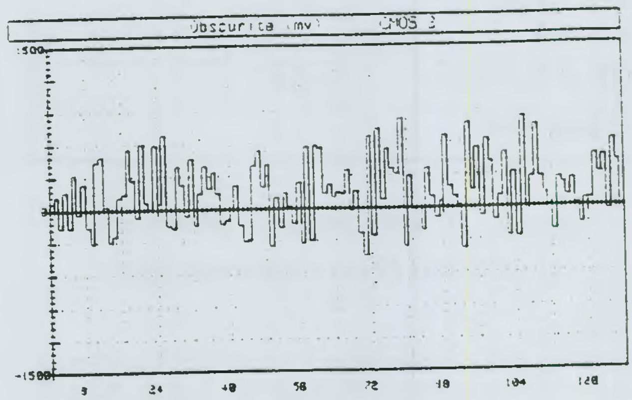
However practical concerns prevent from obtaining as good active $In_{0.8}Ga_{0.2}As$ layer as $In_{0.53}Ga_{0.47}As$ (lattice matched to InP) ones. Dark current density is a good revealing parameter: from 10-7A/cm² (at 1Volt reverse bias and ambient temperature) for lattice match InGaAs photodiodes, it rises up to 10-5A/cm² for 2.1µm cut off InGaAs and reaches 10-3A/cm² for 2.6µm cut off. Knowing that band gap related dark current density increase is expected, in the worst case, to about x1000 lights up at least one decade improvement still achievable. Moreover, crystal growth difficulties consequences impact greatly darkcurrent uniformity as well as darkcurrent temperature dependence from pixel to pixel within the same array. This last effect is, at present, the main drawback of this technology since cooling down is most of the time required (in order to reduce photodiode shot noise and thermal noise) and thereby hits severely DSNU (dark signal nonuniformity). The new techniques are expected to eliminate at least DSNU problems.

TCS has developments undergoing on all techniques however the first device generation will be based on metamorphic technique and on our new CMOS readout circuits in order to use these photodiode arrays under zero biasing conditions.

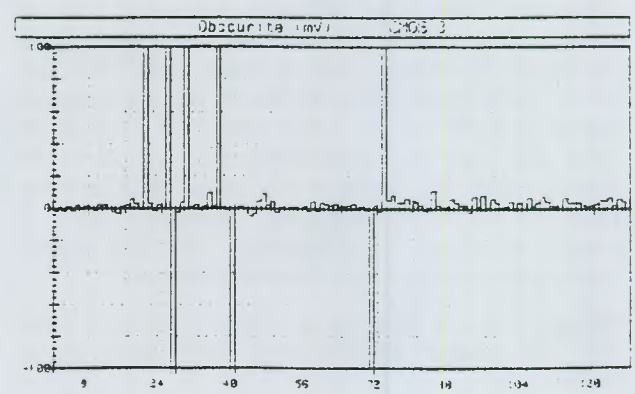
Preliminary results concerning a 2.6µm cut off array, measured at ambient temperature are presented in ***Figure 8***. as well as the same measurement at -20°C.

It can be seen that DSNU is drastically improved thanks to very low offset spreading CMOS readout circuit. Then even if Rsh spreading is not reduced with cooling down (due to cristal defects, traps etc...) it is sufficient that its mean value has been dropped down to remain within acceptable absolute DSNU.

Quantum efficiency up to 60% are obtained in the spectral bandwidth which extends from 1.4µm up to 2.6µm. Spectral response is limited in low side due to cap layer which no longer can be made of InP (for lattice mismatch reasons) but of lattice match InAsP.



Darksignal mapping at ambient and 2.24ms integration time
(verticaly: mV , horizontaly: pixel number)



Dark signal mapping at -20°C and 2.24ms integration time.
(verticaly : mV , horizontaly: pixel number)

***Figure 8* -128 pixel 2.6µm cutoff linear array darksignal**

CONCLUSION

We have reviewed most of present achievements regarding InGaAs photodetector technology. It confirms the high performance level of such devices in the 0,9 µm-1,7 µm spectral range. Read out circuit strategy has been discussed through TCS CMOS multiplexor new development, which concludes that CCD is still an interesting technique for low size photodiode, high speed read-out/multiplexing.

TCS first promising trials or extended wavelength InGaAs photodiode arrays have also been presented, which are expected using CMOS multiplexor advantages to allow good quality near room temperature operated photodetectors with up to 2,6 µm response.

References :

[1] ***CCD for SWIR photodiode multiplexing***
JP. Moy, J. De Cachard, S. Chaussat
Presented at the International Symposium on the Technologies for optoelectronics (16-20 November 1987, Cannes) SPIE

[2] ***3000 InGaAs photodiode multiplexed linear array for the Spot 4 SWIR channel***
JP. Moy, X. Hugon, J. Chabbal, J. De Cachard, C. Lenoble, D. Mollaret M. Villard, B. Vilotich

[3] ***Near room operating temperature SWIR InGaAs detectors in progress***
Proceeding of the SPIE. The international Society for Optical Engineering Conference Title : Proc. SPIE-Int. Soc. Opt. Eng. (USA) vol. 2552 pt.2 p. 738-47
X. Hugon, O. Amore, S. Cortial, C. Lenoble, M. Villard.

[4] ***High performance 1024 element InGaAs hybrid focal plane arrays from 1 to 2.6 µm***
OPTO 94 p 307, 1994
K.R. LINGA et al.

LASER RADAR IN SPACE

C. Werner

DLR – Lidar Group, PO Box 1116, D-82230 Wessling, Germany
Tel: +49 8153 282573 Fax: +49 8153 281608 Email: Christian.Werner@dlr.de

Abstract

Laser Radar as a sensor for ranging, cloud and aerosol characteristics , wind component determination and gaseous concentration measurements was discussed in the space agencies since the mid 70´s. Ranging devices are in orbit since the mid 80´s, the newest systems monitor the Martian surface.
Small and efficient lasers were developed for this purpose. A few systems were tested as Space Shuttle versions. LITE was a spectacular example. Now the application in a specific mission is the main concern. Atmospheric radiation and atmospheric dynamics are two missions including laser radars ESA has selected. Both laser radar sensors, ATLID and ALADIN, will be discussed with respect to the ongoing activities in the other space agencies. The acceptance of a sensor depends on the user requirements. The sensor development, its technologies and the feedback with the user community plays an important role.

1. Introduction

Lidar technology and applications are well established (Krichbaumer et al. 1993). A backscatter lidar technology experiment was tested in space in 1994 (Winker et al. 1994). Scientists need global information on wind, clouds, and aerosol layers. On a spaceborne platform only a limited amount of power is available for a lidar system; therefore, a compromise is necessary between the possibilities and the requirements.
A) There is an urgent need for the development of a global observing system for climate measurements in order to
- monitor the present climate state, detect natural climate variability and the anthropogenic impact on climate change,
- to provide data for initialisation and validation of global climate models.

Radiation and the wind fields were identified as most important parameter (Houghton, ESA-SP 1143,1991).

The following climate-relevant parameters can be measured in principle by a backscatter lidar in space:

1) Height of thick low, midlevel, and upperlevel clouds, cloud phase near cloud top
2) Climate-relevant parameters of thin tropospheric clouds: height, optical depth and cloud phase
3) Polar Stratospheric Clouds: their type, extent, geometrical,and optical thickness
4) Tropospheric and Stratospheric Aerosol Layers: their height, extent, geometrical, and optical depth
5) Height of Planetary Boundary Layer
6) Effect of thin layers of aerosol and cloud on radiation balance and surface fluxes

B) Beside this, the accurate ice and topography map of the earth is important. Accuracies in the order of 10 cm for a 70 m footprint are sufficient.

C) The estimation of energy convergence and divergence asks for the precise measurement of wind with the accuracy and resolution shown in table 1:

Table 1 Requirements for a spaceborne Doppler lidar (ALADIN 1989)

	Stratosphere	Troposphere
horizontal resolution	100 km (50 km)	100 km (50 km)
vertical resolution	3 km	1 km between 2 - 15 km 0.5 km below 2 km
accuracy	2 - 3 m/s	1 - 2 m/s
frequency per day	4	4

One needs the wind information in a grid size of 100*100*0.5 km in the troposphere with an accuracy of 1-2 m/s four times a day.

Proc. 32nd ESLAB Symp., 'Remote Sensing Methodology for Earth Observation and Planetary Exploration', ESTEC, Noordwijk, The Netherlands. 15-18 September 1998 (ESA SP-423, December 1998)

2. Laser Radar Principle

Figure 1 shows the principle of space lidar operation.

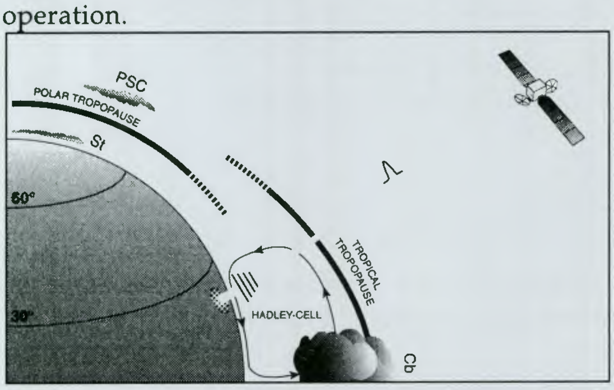


Figure 1 Principle of space lidar operation

A lidar system consists of a powerful laser, a receiving optic, and a detector with signal electronics. A laser pulse of duration τ (10 ns for the backscatter lidar or a few microseconds for the Doppler lidar) and with power P_t is transmitted into the earth atmosphere. The aerosols or clouds at range R and the earth surface scatter the light back to the receiver (telescope area A) in space. The received power P_r depends on the transmission through the atmosphere τ and the backscatter ß. The transmission can be used to identify trace gas concentration by the proper choice of two wavelengths within and outside of a specific absorption line i.e Differential Lidar or DIAL (Boesenberg et al. 1998). The lidar equation in its simplest form is:

$$P_r = P_t \frac{K\,A}{R^2}\, ß(R)\, \tau^2$$

where:

range R is related to the speed of light: $\frac{c\,\Delta t}{2}$, with c as speed of light and Δt =time between transmitting the laser pulse and receiving the information (for example 1 $\mu s \approx 150$ m)

K = system constant
A = receiver area
ß(R) = backscatter coefficient of molecules, aerosols or clouds
or $\beta = \rho$ as reflection from a hard target

$$\tau^2 = \exp(-2\int_0^R \sigma dr)$$

with σ as extinction coefficient (absorption and scattering)

In principle, P_r is proportional $\frac{1}{R^2}$ and this is the most important factor for a spaceborne laser radar. As valid for every lidar system, the eye-safety rules have to be met.
Caused by the Doppler effect produced by the moving aerosols with the wind and the platform motion versus the scatterers the received light is shifted in frequency. Figure 2 shows the principle.

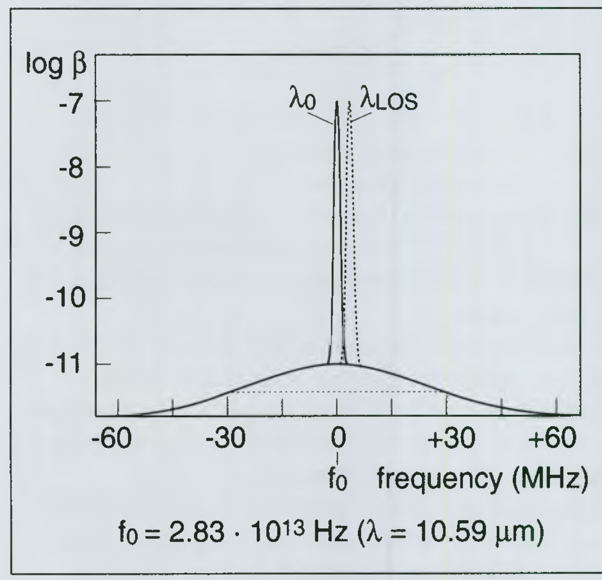


Figure 2 Schemetic of the frequency (wavelength) dependency of the backscattered signal in a range gate at 10.59 μm wavelength

Backscatter is composed by molecular and aerosol contributions. If there is aerosol in the scattering region, a delta function representing the laser linewidth appears with a broad molecular contribution. The delta function is shifted by the Doppler effect as well as the broad molecular return.

2.1 Backscatter measurements

If one detects the integrated backscatter information shown in figure 2 using a normal optical filter which is in the order of a few hundred MHz, there is no influence on wind velocity or platform velocity. This is the integrated backscatter lidar principle. One can measure cloud top height and all the parameters listed in the introduction. If one uses polarisation sensitive detection devices,one can distinguish between ice and water clouds. If one uses different receiver apertures, one can detect cloud microphysical parameters like drop size distribution and extinction based on the multiple scattering effect.

2.2 Direct Doppler Detection Technique

The lidar return spectrum is the summation of the molecular scattering and particular or aerosol scattering. Incoherent techniques measure directly the return center frequency either with a multi-channel spectrometer or by filter transmission variation. The molecular spectrum (figure 2) has a width - in units of wind velocity - of about 600 m/s and the aerosol spectrum that of 40 m/s.
The measurement of the center frequency may be performed either with multi-channel spectrometer or with photometric principles (Korb et al. 1994). As presented in figure 3 , the multi-channel fringe imaging spectrometer measurement principle, the first principle aims at sampling the received spectrum with a resolution compatible with the spectrum width. A centroid computation then provides the location of the spectrum centre.

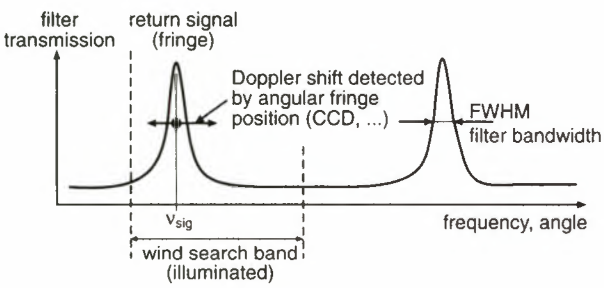


Figure 3 Multi-channel fringe imaging spectrometer measurement principle

At first order, the best obtainable accuracy with such receiver is:

$$\Delta v = \frac{c\;\delta\lambda_{rms}}{2\lambda\;SNR}$$ in m/s on LOS winds,

where c is the light velocity, $\delta\lambda_{rms}$ is the rms width of the detected return spectrum, SNR is the direct detection Signal-to-Noise Ratio.
On the other hand, the photometric principle (figure 4) utilises a filter whose response varies versus wavelength. The principle works on the edge of an interferometric filter for the aerosol signal as well as for the molecular signal the so called edge technique.

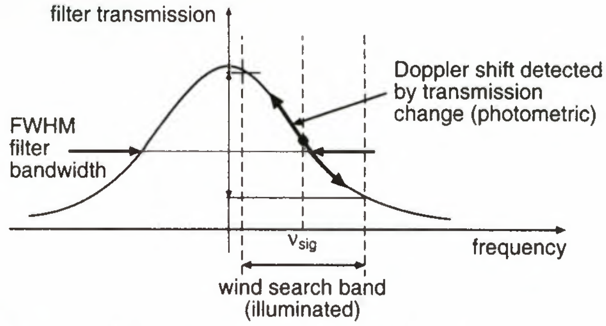


Figure 4 Photometric detection (edge technique, double edge technique) measurement principle

The obtainable accuracy in this case is similar to the above for optimised filters and is given to:

$$\Delta v = \frac{c\;\frac{\partial\lambda}{\partial\Gamma}}{2\sqrt{2}\lambda\;SNR}$$ in m/s on LOS winds,

where $\frac{\partial\lambda}{\partial\Gamma}$ is the filter spectral resonse or spectrum slope.
For both principles, the accuracy is mainly limited by return spectrum linewidth and the direct detection SNR is bounded by technical factors as optical/ quantum efficiency and detection noise. An intrinsic loss factor in incoherent lidar is the optical filter reflection loss related to the wind search bandwidth.

2.3 Coherent Detection Technique

A coherent Doppler lidar consists of a pulsed, frequency-controlled laser transmitter (L1), a continuously scanning transmit and receive telescope, a heterodyne detector (D) where the local oscillator radiation (LO) is mixed with the Doppler-shifted backscattered signal, and a signal processing system. Figure 5 shows the scheme. The laser pulse with a pulse length of a few microseconds and an optical carrier frequency f_0 is sent out via the transceiver telescope into the region of investigation. Some of the radiation is backscattered by small aerosol particles which move with the prevailing windspeed through the laser focus volume (Figure 5) (Post M.J.and R.E. Cupp 1990). The windshifted Doppler frequency directly determines the line-of-sight (LOS) component (V_{Los}) of the wind vector. At CO_2 laser wavelengths (λ) of around 10.6µm a velocity component of 1 m/s corresponds to a frequency shift, Δf_D of 189 kHz.

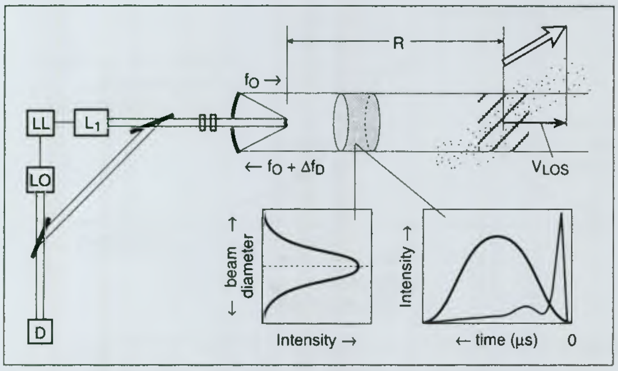


Figure 5 Scheme of a Doppler lidar.

This is obtained from the equation

$$\Delta f_D = 2\frac{V_{LOS}}{c} \cdot f_0$$

where c is the speed of light and f_0 is obtained from f=c/t.

By measuring at an azimuth angle θ and an elevation angle φ one gets a radial (line-of-sight) contribution V_{Los} which depends on the wind vector components u, v, and w given by

$$V_{LOS} = u \sin\theta \cos\varphi + v \cos\theta \cos\varphi + w \sin\varphi$$

3. Applications

3.1 Ranging

Laser Ranging is a special discipline for geodynamic applications. The first measurements were performed in the 70's using ruby lasers. Satellites with retroreflectors on were launched (Cohen and Smith 1985).Apollo astronauts placed retroreflectors on the Moon. A world-wide scientific community is established and they develope excellent sophisticeted laser radar techniques (J.Degnan and J.McGarry 1997). An example for the information content of an altimeter is shown in figure 6 (Nelson et al.1984).

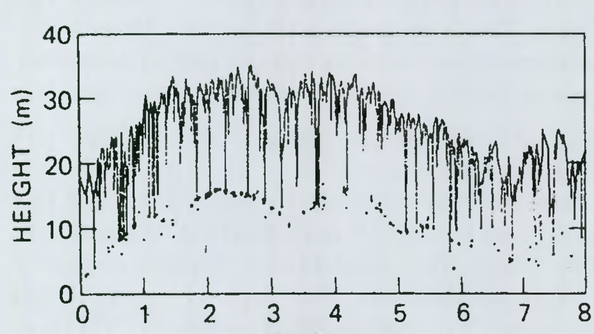


Figure 6 Airborne laser altimeter data (NASA-LASE report 1987)

With a high repetition rate laser ranging system this kind of information, tree canopy characteristics and land surface structure, can be measured also from space with an accuracy in the order of a few centimeters.

3.2 Clouds and Aerosols

Cloud and aerosol measurements were mentioned in the introduction. The results look similar like the signals shown in figure 6 but with another height scale. Different spaceborne lidar systems demonstrated the structure of clouds and - by averaging over larger scales - that of aerosols (Winker and McCormick 1994) . Two European systems will be described here.

3.2.1 ALISSA

ALISSA (l Atmosphere par LIdar Sur SAliout) is the French-Russian cooperation with a flashlamp-pumped Nd:YAG laser to measure on Priroda the backscatter information from clouds and aerosols at night (Chanin et al.1998).
ALISSA was a multi-agency project, it took a long time for development and in the final test on the space station MIR it gave results similar to that of the system LITE. The main problem on the space station was mentioned with the position accuracy. A special procedure was necessary to correct the determined range to the clouds into exact data for geoposition and height above ground (Balin et al.1994).

3.2.2 Backscatter lidar ATLID

Since several years, the ESA Earth Observation Preparatory Programme (EOPP) together with the ESTEC Technology Research Programme has given priority to ATLID (Matra, 1990, Battelle 1990), for reasons of all-up potential system feasibility and associated costs, whilst continuing to pursue ALADIN (1989) in knowledge of the very high application interest.

Based on the preparatory activities (ESASP-1108, 1989)., two parallel ATLID phase-A studies were initiated, with the work being performed by Battelle Europe and Matra. The teams came up with two different instrument concepts, one with a fixed telescope/scanning mirror and one with a scanning multiple-telescope configuration. Both studies confirmed the feasibility of the backscatter lidar instrument, however, they identified some critical technology areas which would require further development before the instrument could be considered to be employed at the First Polar Platform (later named Envisat-1).

The instrument design (Morancais 1994) is shown in Figure 7. The concept trade of led to the preferred architecture, that is the single-axis sinusoidal scanning telescope configuration. The required radiometric performance could be satisfactorily achieved by reduction of the telescope diameter to 0.6 m and an optimized arrangement of the various instrument sub-units. The required radiometric instrument performance could be fully achieved, despite the reduced receiver aperature. This was feasible thanks to the improved signal averaging and the reduced receiver field-of-view (FOV), resulting from the decreased swath-width and the limited vertical observation range, respectively. The values for mass, power and dimensions could be brought down to 210 kg, 470 W (end of life) and only 1.6 m x 1.3 m x 1.1 m (Table 2).

These budgets are now well comparable with the demands of instruments currently foreseen to be flown on the ENVISAT-1 platform.

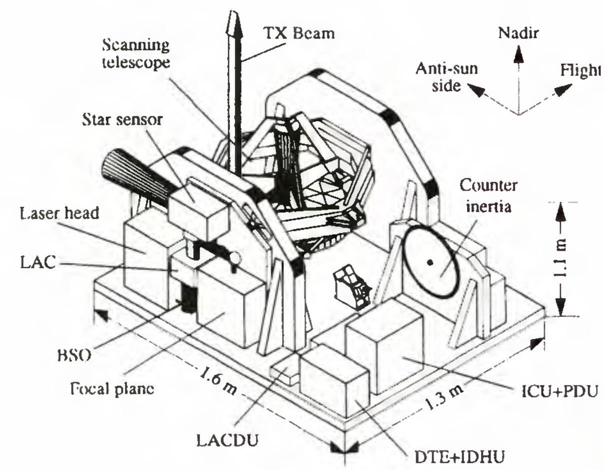


Figure 7 Current ATLID instrument architecture (Morancais 1994)
(TX.. Transmitter, LAC.. Lag-angle compensator, LACDU--LAC drive unit, BSO.. Beam shaping optics, DTE.. Detection electronics, IDHU.. Instrument data handling unit, ICU Instrument control unit, PDU.. Power distribution unit.)

3.3 Wind profiling

The Doppler lidar measures with an appropriate pulse repetition rate and a scanning mechanism (to cover a wide swath) a height resolved line-of-sight velocity component in the atmosphere. If one uses the laser Doppler system onboard an aircraft or in space, the speed of the aircraft (spacecraft) modifies the Doppler shift with

$$v_{LOS} = v_{LOS}(\text{wind}) + v_{LOS}(\text{carrier})$$

where $v_{LOS}(\text{carrier})$ is the carrier speed with respect to the laser line-of-sight.
The problem mentioned for the cloud measurements with the position accuracy increases dramatically for the wind measurements. The platform speed is about 7 km/s and the measurement accuracy required is 1 m/s!

3.3.1 ESA Aladin

Concerning the Doppler wind lidar (DWL), two pre-phase A studies for the Atmospheric Laser Doppler Instrument (A. Marini and M. Rast 1993, A. C. Lorenc et al.:1992) have been initiated end of 1993.
A demonstration experiment of a Doppler lidar on a Shuttle or a space station was proposed about 10 years ago (Bilbro et al.1987). Now the International Space Station is a place where again a precursor experiment can be to proposed. Figure 8 shows the station and the installation of a Doppler lidar on the TRUSS structure.
Caused by the platform characteristics like variable orbit height and pointing stability (caused by maneuvers and man-in-space activities), the Doppler lidar operation can overcome the arising difficulties (flexible lag angle etc. necessary) only if a step-scan instead of a full conical scan is performed(Johann et al. 1997). Therefore the pyramide in figure 8 stands for four beams which were send to the atmosphere alternately (Werner and Brand 1995).

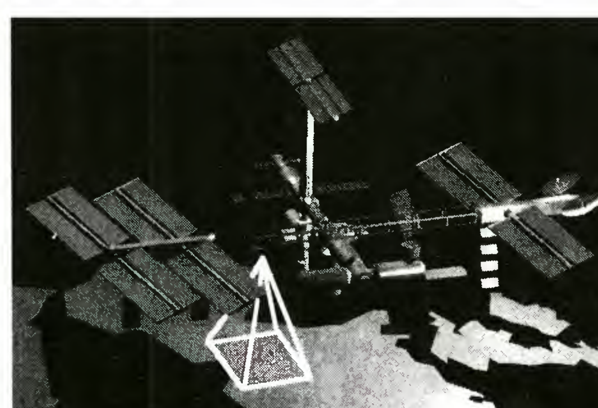


Figure 8 International Space Station and viewing directions (Pyramide) of a Doppler lidar with coverage over Europe.

3.3. 2 NASA SPARCLE

SPARCLE stands for Space Readiness Coherent Lidar Experiment and its expected to fly on the Space Shuttle in 2001 (Kavaya 1992). Figure 9 shows the artist view. It is one of NASA´s New Millenium Earth science experiments.
The first global wind measurement from space is made possible by a major breakthrough in laser technology in 1997 by U.Singh (J.Yu et al.1998). They demonstrated for the first time an all solid state, room-temperature, diode pumped

Ho:Tm:YLF laser, which prodices an eye-safe, 2.05 µm laser output of 600 mJ at 10 Hz. The lidar will fly in the cargo bay of the Space Shuttle and will attempt to measure winds from just above the surface of the Earth to a height of about 10 miles. This is a key region for a better understanding climate relevant events.

Figure 9 SPARCLE artists view

Researchers hope to obtain approximately 50 hours of wind data. Two Get-Away Specials (GAS) cans will be used. This allows scientists to put their optical and electronic systems in a pressurised environment. It saves a lot of money for such a demonstration of the technology if one dosnt need to ruggedise the system for space.
If successful, the SPARCLE experiment could lead to a large laser facility with a significant larger aperture (25 cm for SPARCLE) , and a polar orbiting weather satellite in 2007.

3.3.3 Development of Scans for the Doppler Wind Lidar

During the development of the wind lidar in space the proposed scan techniques to cover a wide swath were modified many times. The following steps 1-5 demonstrate the development:

1 - Conical Scan
2 - Shot Management to get signal pairs
3 - Step-Scan
4- Single LOS
5- Targeted Observation

Based on the requirement in the early 80´s a horizontal wind should be measured using a conical scan (1) known from ground based instruments. To met these requirements a shot management strategy was developed to combine laser pulses at the same position on Earth (using the forward and backward viewing direction) by variing the laser pulse repetition rate adequately (2). This laser pulse variation causes severe problems in the laser design. A step-scan was proposed during pre-phase-a study to overcome also another technical problem,the lag angle compensation(3) . In the meantime the weather services were convinced to use any kind of wind information available in good quality, they could accept single line-of-sight component (4). The next step (5) is in discussion now to focus on - for the weather service - sensible target areas.

3.4 Gas Concentrations

Measurements of the atmospheric pressure, the profiling of temperature, humidity, ozone concentration or concentration of other minor gases were successfully demonstrated in airborne lidar versions. A proposal ORACLE exists for a spaceborne ozone lidar experiment (Dudelzak et al.1998, Browell et al.1993).

3.5 Mars observation

What is possible on Earth, should be possible on Mars. Altimeter, cloud measurements and wind profiling is in orbit or under development. Figure 10 shows the principle layout of the Mars Observer altimeter (Bufton and Abshire,Web pages: http://denali.gsfc.nasa.gov/sla).

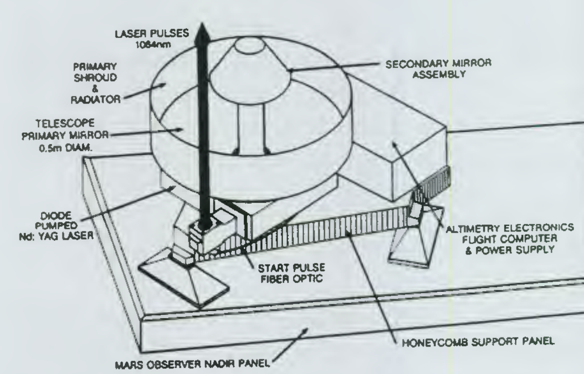


Figure 10 Mars Observer Laser Altimeter MOLA

A Mars lander (ML) with a lidar is provided. The system should be capable to measure clouds and dust, and humidity using a laser diode system with two wavelength in the 935 nm region.
Another system using laser diodes, frequency stable for wind profiling, was proposed by Menzies (Menzies et al. 1998).
The techniques are similar for Mars observations as seen in the figures 7 and 10 for the backscatter lidar and the altimeter. The altimeter on Mars

detected also clouds with a very good resolution. The experiences of the laser ranging scientists could contribute to the lidar community and vice versa. The parameters of the two systems MOLA and ML are shown in table 2 for comparisons with the other sensors.

3.6 Comparisons

Table 2 shows for comparisons different systems with the parameters wavelength, transmitted pulse energy,receiver area, range resolution and the two main parameters for a spaceborne system, the power consumption and the weight.

Table 2: Comparison of the parameters of different lidar systems

	System						
Parameter	ALISSA	ATLID	SPARCLE	DWL	MOLA	ML	LEM
λ (μm)	0.523	1.064	2.05	?	1.064	0.935	0.9
P_r (mJ)	5 (40)	70	100	?	45	?	0.0003
Dia (cm)	40	60	25	70	50	20	5
ΔR (m)	150	50	150	1000	0.3	?	0.05
P_e (W)	?	225	405	226	34	5	2
W (kg)	?	170	185	300	26	2	2.5

The systems ALISSA, ATLID, SPARCLE, MOLA, and ML are mentioned in the text. DWL stands for the ALADIN version. It is today no decision which kind of instrument will be further developed, therefore the question marks. LEM is the laser rangefinder and speedometer, the german police uses for comparison with the other systems. One can see that the systems for use on Mars have similar weight and power consumptions like LEM. SPARCLE and ALISSAon the Shuttle or MIR have no problems with power consumption but for ATLID and DWL a further reduction should be possible if more efficient lasers were on the marked.

4. Synergy

For further reduction of weight and power for spaceborne systems a synergy with other sensors is necessary. Two examples are given to explain the possibilities. One is the use of the single level wind observations over oceans using SAR or microwave scatterometers, the other is application oriented, the active participation of the user in the selection of the target.

4.1 Wind over sea measured by SAR and Doppler lidar

Wind field over sea can be determined with ERS-satellite by Synthetic Aperture Radar (SAR) with an accuracy of about 2 m/s and 5 degree using the ESA CMOD4 algorithm. Wind profiles were measured by the airborne Doppler lidar ADOLAR. A first comparison in a few levels (surface up to 1200 m above sea) was performed. During an experiment near the island Rügen a synophic SAR/ADOLAR dataset was acquired. An analysis of the dataset comparing to ground truth and the DWD analysis was performed.

Figure 11 gives a summary of the results. Wind speed versus altitude at the DM gridpoint with the following information: SAR, stations Arkona and Putbus, the analysis model for the grid point and ADOLAR wind (circles). The wind speed measured from the aircraft by a 5-hole sonde mounted on a nose boom are also displayed (circles). Normally these data are available with 10 Hz repetition rate. For the purpose of comparison the measurements of the nose boom are averaged over 10 s and are also calculated for the same estimation of the platform speed as used for ADOLAR (Rahm 1995). This means, that the data for the ground speed from the IRS normally used for the estimation of the wind speed of the nose boom are replaced by the more accurate ground speed processed from the sources IRS, GPS, and ground return together. With this improvement the wind estimates of the nose boom and the Doppler lidar data combine to a homogeneous wind profile which correspond well with the forecasting profile 1 of the DWD model the measurement until 200 m altitude. At higher altitudes the profile 1 from the DWD model is underestimated with about 2 m/s. As another result it can be seen that SAR is about 1.5 m/s lower than the land stations.

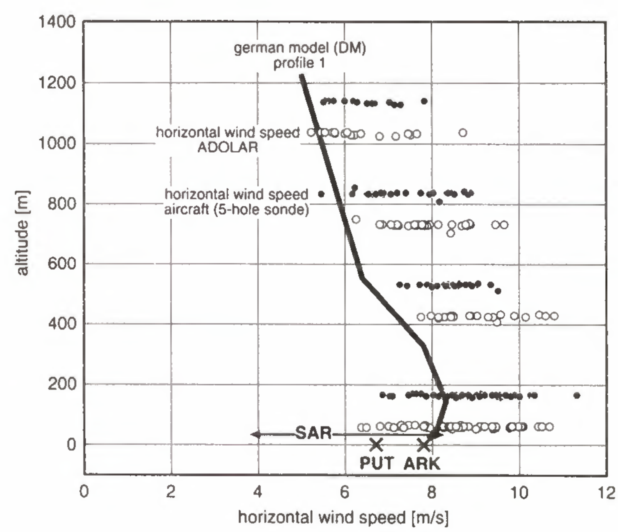


Fig. 11 Summary of the windspeed measurements on Nov. 21, 1996 for the flight track over the Baltic Sea(Werner et al. 1998)

For the rectangular pattern over Rügen the average wind speed for two altutudes are available: 12.3 m/s at 202 m from the nose boom, and 11.3 m/s at 98 m from ADOLAR. This applies good to the profiles of the DM model which predict about 10 m/s in 100 m altitude, and 11 m/s in 200 m altitude.

4.2 Estimating the impact of lidar-winds on Numerical Weather Prediction (Wergen and Cress 1998)

The DLR-impact study is now established. If the Doppler lidar measures a LOS wind component in the real atmosphere, what is the impact on the weather forecast? This question is the driver of this study. A virtual instrument or sensor will be developed to interact with the weather service. There is a cooperation between DLR, the German Weather Service, and the Institute of Atmospheric Optics in Tomsk.
In order to estimate the potential benefit of the addtitional LOS lidar windprofiles on the numerical weather forecasts, an impact study was initiated. It consists of two parts. In the first study, the importance of wind profile data in general was assessed. In the second part, the impact of the additional lidar-winds will be estimated.

Role of wind observations

The basic prognostic variables of current Numerical Weather Prediction (NWP) models are geopotential, humidity and wind. In the extra-tropics, there exists a strong coupling between the geopotential and the wind field, the so-called geostrophic relation. It allows to derive approximate wind components from geopotential gradients. It is therefore important to find out whether these geostrophic winds are sufficient or whether observed wind profiles result in a more realistic definition of the initial state for NWP.
To answer this question, a parallel data assimilation with the Global-Model (GM) of DWD was perfomed in which the wind profile observations from radiosondes and aircraft over the United States and Canada were excluded. By comparing the forecast quality between the runs using or not using the wind data, an estimate of their importance can be given. By tracing back the differences in the forecasts to the intial difference, sensitive areas can be identified.
Fig. 12 shows the mean anomaly correlation for the 250 hPa geopotential, averaged between 20° and 90°N as a function of forecast length. Naturally, forecast quality deteriorates with forecast time and forecasts are usually considered of little value if the correlation falls below 60%. The control forecast (full) using all observations crosses the 60% line close to 192 hours into the prediction. By contrast, the forecasts not using wind profiles over the North American continent (dashed) crosses the 60% limit at 168 hours, thus resulting in a degradation of almost 24 hours in useable forecast length. This clearly highlights the importance of having wind observations for defining the initial state also in the extra-tropics.

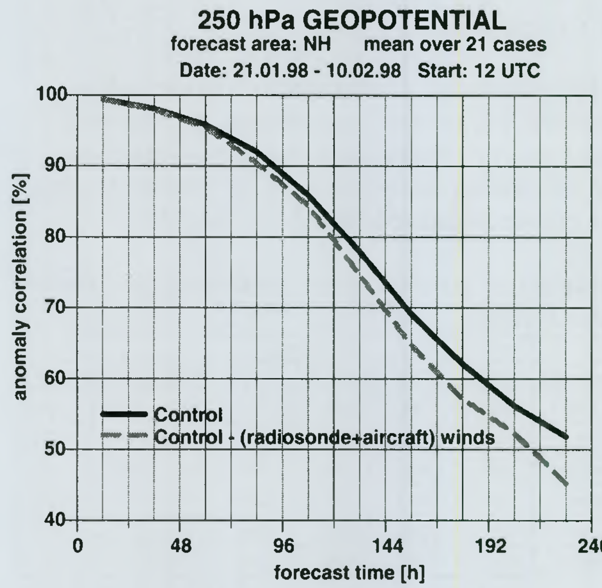


Figure 12 Anomaly correlation of the 250 hPa geopotential height field, averaged over 21 cases and for the area between 20° and 90°N as a function of forecast length. The full curve is for the control using all observations, dotted for experiments not using radiosonde and aircraft wind data over the US and Canada.

5. Spin-offs

Spin-offs of the spaceborne sensor development can be seen in smarter devises for industrial and personal application. But this is not the main part for spin-off. This development is user oriented. Ranging is the main application field. Lidar for visibility detection on airports, highways and from the car itself to warn the driver for a fog situation in advance is in the test phase.

A real spin-off can be seen in the comparison of the Doppler lidar parameters (DWL) and the Mars Lander (ML) parameters in table 2. If such a sensor is developed, a similar system can be expected for wind profiling on Earth. Figure 13 shows an artists view of such a device. The size is in the order of a laser cloud ceilometer . Instead of a conical scan a three axis system is used for wind profiling. Three single components can be combined to a 3D wind vector.

Figure 13 Artists view of a small , portable wind profiler

6. Summary

This paper gave an overview - starting from simple ranging over cloud top height measurements up to gas cocentration and wind measurements - of spaceborne laser radars. Each of these techniques had their own development.
Ranging reached accuracies in the order of a few centimeters. The user community is now trying to get millimeter resolution. The detection technique was improved as well as the laser efficiency.
For the backscatter lidar the situation is more complex. Complex means compared to the range resolution of one single range information on the top of the surface or the trees. For the backscatter lidar one needs accurate information in all range gates from the stratosphere down to the atmospheric boundary layer.
The rapid laser development helped to lower the required mass and power consumption and made it possible to include a lidar in atmospheric missions.
An intense discussion with the users was the next step to lower the technical needs (see chapter 3.3.3, scanning for example). For the Doppler Wind Lidar DWL, which is also an excellent backscatter lidar, the first demonstration will be made in 2001 on the Space Shuttle.
A synergy with other sensors will increase the acceptance and will produce a new quality of information. Flux measurements are one example. The potential benefits for the community are for example:

- weather forecast improvement,
- optimal flight planning for the airlines to reduce fuel,
- better understanding of climate and environment protection parameters.

7. References

ALADIN, Atmospheric Laser Doppler Instrument - ESASP-1112, June 1989.

Balin, Ju.S., V.E. Mel'nikov, A.A. Tikhomirov, I.V. Znamenskii, S.V Samoilova, and V.E. Zuev: Spaceborne Aerosol Lidar BALKAN-1. SPIE 2310,144-151, 1994.

Phase-Astudy of the Atmospheric lidar (ATLID), **Battelle** Europe, Final Report, ESAContract No. 8436/89, 1990.

Bilbro J.W., R. Beraneck, D. Fitzjarrald, and J. Mabry, (Eds.), Shuttle coherent atmospheric lidar experiment; Final report, MSFC TM 100307, June 1987.

Boesenberg J.: Ground-based differential absorption lidar for water-vapor and temperature profiling:methodology. Appl. Opt. 37,3845-3860,1998

Browell E.V. et al. : Ozone and Aerosol Changes Observed during the 1992 Airborne Arctic Stratospheic Expedition. Science 261, 1155-1158,1993.

Chanin, M.-L., A. Hauchecorne, D. Nedeljkovic, J.-E. Blamont, M. Dubois, Y. Tulinov, V. Melnikov: Premiers resultats du lidar ALISSAembarque a bord de la station MIR, Europto Symposium in Remote Sensing, London, Sept.24-26,1998
ESA - SP1108, 1989, Laser Sounding from Space, ESAspecial report.

Robert J. Curran et al. :LASA - Lidar Atmospheric Sounder and Altimeter, NASAEOS Vol.IId,1987

S.C.Cohen and G.R.Smith: LAGEOS scientific results. Geophys. Res. 20,9217,1985

J.J.Degnan amd J.F. McGarry: SLR 2000:Eyesafe and autonomous single photoelectron satellite laser ranging at kilohertz rates,Proceedings Laser Radar Ranging and Atmospheric Lidar Techniques, EUROPTO Series Vol 3218, 1997

A.E.Dudelzak et al: ORACLE (Ozone Research with Advanced Cooperative Lidar Experiment) Joint NASA-CSA development of a space based ozone DIAL, Proceedings of the Remote Sensing Conference in Barcelona,EUROPTO Series ars02 1998

ESA - SP 1108, 1989, Laser Sounding from Space, ESA special report.

Hall,F.F.Jr.; R.M.Huffaker, R.M. Hardesty, M.E.Jackson, T.R. Lawrence, M.J.Post, R.A.Richter, and B.F. Weber: Wind Measurement Accuracy of the NOAA Pulsed Infrared Doppler Lidar. Appl. Opt. 23, 2503,1984

Houghton, J.: Monitoring Climate and Climate Change . in ESA SP-1143,page 74-81, Report of the Earth Observation User Consultation Meeting, Noordwijk, May 29-31,1991.

Johann, U. : ESA ALADIN Pre-Phase A Study, Daimler Benz Aerospace Report 1996.

Kavaya, M.J.: Projected Technical Specifications for a 2 µm Space-based Coherent Lidar Wind Sensing System. Proceedings 16th International Laser Radar Conference, Boston,MA July 1992- NASA Publication CP-3158

Korb, C.L., B.M. Gentry, and S.X. Li: Spaceborne Lidar Wind Measurements with the Edge Technique. SPIE-2310,206-213,1994.

Krichbaumer, W., H. Herrmann, E. Nagel, R. Häring, J. Streicher, Ch. Werner, A. Mehnert, Th. Halldorsson, S.Heinemann, P Peuser, N. P. Schmitt: A diode-pumped Nd:YAG lidar for airborne cloud measurements, Optics and Laser Technology, 25, 283-287,1993

Lorenc A. C. , et al.: „Preparation for the use of Doppler wind lidar information in meteorological data assimilation systems" - Final report of ESA study contract No. 9063/90/HGE-I, February 1992.

Marini, M. Rast:" The ESA Atmospheric Laser Doppler Instrument (ALADIN): Requirements for a coherent Doppler wind lidar" - Proc. of the 7th Conference on coherent laser radar, July 1993, Paris, France.

Phase-A study of the Atmospheric lidar (ATLID), **Matra**, Final Report, ESA Contract No. 8436/89, 1990.

R.T.Menzies, G.Cardell, C.Esproles, S.Forouhar, H. Hemmati,and D. Tratt: Dvelopment of miniature lidar using narrow linewidth semiconductor lasers for Mars boundary layer wind and dust opacity profiles. Proceedings of the 19th ILRC, Anapolis, July 6-10,1998 (NASA report)

Morancais, D. : Compact version of the atmospheric lidar instrument (ATLID). Proceedings Satellite Remote Sensing, Vol. 2310, 43-50, 1994.

R.L. Nelson, W.B.Krabill, and G.A. Maclean: Determining forest canopy characteristics using airborne laser data. Remote Sensing Environment 15, 201,1984.

Post M.J., Cupp R.E.: Optimizing a pulsed Doppler lidar; Appl. Opt. 29 (4145-4157) 1990.

Rahm, S.: Measurement of a wind field with an airborne continuous-wave Doppler lidar. Opt.Lett. 20, 216-218, 1995.

Werner, Ch. and B. Brand: Doppler Lidar on the International Space Station. SPIE Europto conference "Lidar Technique for Remote Sensing" Paris, Sept. 25-29, 1995

Werner, Ch.,Werner Wergen,Alexander Cress,Victor Banakh , Igor Smalikho,Ines Leike and Juergen Streicher: ALADIN Impact Study, Europto Symposiun on Remote Sensing, Barcelona, Sept. 21-24,1998,

W. Wergen and A. Cress: Impact study (Private Communication 1998) see also Werner et al. Impact study.

Winker, D.M. and M.P. McCormick: Aerosol and Cloud Sensing with the Lidar in-Space Technology Experiment (LITE). SPIE-2310, 98-105,1994. (see also M. P. McCormick, 1995: `Lidar In-space Technology Experiment - Overview and Early Results', in *Optical Remote Sensing of the Atmosphere*, vol. 2, OSA Technical Digest Series (Optical Society of America, Washington DC 1995), pp.16-18. , and D. M. Winker, 1995: `Aerosol and Cloud Observations from the Lidar In-space Technology Experiment', in *Optical Remote Sensing of the Atmosphere*, vol. 2, OSA Technical Digest Series (Optical Society of America, Washington DC 1995), p. 24.)

J.Yu, U.N.Singh, N. Barnes, and M.Petros: 125 mJ-diode pumped injection-seeded Ho:Tm:YLF laser, Optics Letters 23, 780-782,1998

LOW NOISE FAR-INFRARED DETECTION AT 90 K USING HIGH-T_C SUPERCONDUCTING BOLOMETERS WITH SILICON-NITRIDE BEAM SUSPENSION

M.P. Bruijn, M.J.M.E. de Nivelle, P.A.J. de Korte
Space Research Organization Netherlands, Utrecht, The Netherlands

S. Sánchez, M. Elwenspoek
MESA Research Institute, Enschede, The Netherlands

T. Heidenblut, B. Schwierzi
Institut für Halbleitertechnologie und Werkstoffe der Elektrotechnik, Hannover, Germany

W. Michalke, E. Steinbeiss
Institut für Physikalische Hochtechnologie, Jena, Germany.

ABSTRACT

High-T_c $GdBa_2Cu_3O_{7-\delta}$ (GBCO) superconducting transition edge bolometers with operating temperatures near 90 K and receiving area of 1 mm^2 have been made with both closed silicon-nitride membranes and patterned silicon-nitride (Si_xN_y) spiderweb-like suspension structures. To enable epitaxial growth of the GBCO layer, a thin monocrystalline Si layer is prepared on the silicon-nitride base, using fusion bonding techniques. By pattering the silicon-nitride supporting membrane the thermal conductance G is reduced from 20 to 3.5 μW/K. The noise of both types of bolometers is fully dominated by the intrinsic noise from phonon fluctuations in the thermal conductance G. The optical efficiency in the far infrared is about 75% due to a gold black absorption layer. The optical noise equivalent power (NEP) is 1.8 pW/√Hz, and the detectivity D^* is 5.4×10^{10} cm√Hz/W. Time constants are 0.1 and 0.6 s, for the closed membrane and the spiderweb like bolometers respectively. We have observed an empirical limit for the NEP for this type of bolometers. The effective timeconstant can be reduced with a factor of 3 by using an electronic feedback system or by using voltage bias. A further reduction necessarily results in an increase of the NEP due to the $1/f$ noise of the superconductor.

1. INTRODUCTION

Detection of far-infrared radiation (wavelength between roughly 30 and 300 μm) is an important driver for definition of space-based scientific missions. Far-infrared radiation is emitted by relatively cool objects in space, like planets and interstellar dust clouds. Because of absorption in the earth's atmosphere, observation must be done from space. Observation of these very faint structures requires the use of detectors with the highest possible sensitivity. Thermal detectors and photovoltaic (photon) detectors at temperatures below 10 K are normally the primary choice. For a review, see Richards (1994). For investigation of chemical processes in the earth's atmosphere, detection of far-infrared radiation plays an important role as well. Hydroxyl radicals (OH) for instance play a role in the catalytic cycles, that control the stratospheric ozone budget: the most important Odd-Hydrogen cycle and the NO_x and ClO_x cycles (Wennberg et al. 1994). The OH radical can be discriminated from other players in these cycles most easily in the far-infrared part of the spectrum (Chance et al. 1994). Because of the higher radiation levels, compared to space astronomy, less sensitive detectors can be of use. This enables the use of higher operating temperatures, which require less complicated, light weight and low power consuming cooling facilities. At liquid nitrogen temperatures, thermal detectors posses a better performance for detection of far-infrared radiation than doped Si or Ge photon detectors (Richards 1994).

Amongst the thermal detectors, high-T_c transition edge bolometers are the most promising. The very sharp resistance change enables changes in temperature to be detected with high sensitivity. During the last few years, much effort has been put into development of this type of detector (Verghese et al. 1993, Richards 1994, Brasunas & Lakew 1994, Johnson et al. 1994, Berkowitz et al. 1996, Méchin et al. 1997). Members of the present collaboration have been involved in this development since 1995, starting with bolometers on Si-membranes (Neff et al. 1995). Soon thereafter the noise equivalent power (NEP) was improved by one order of magnitude by replacing the Si-membrane by a silicon-nitride supporting layer (at the cost of time contant) (de Nivelle et al. 1997). Bolometers from this development will be flown on a balloon mission SFINX (Hoogeveen 1997). The present work aimed at further improvement of the detectivity $D^*=\sqrt{Area}/NEP$ by exploiting the limits of reduction of the thermal conductance G through the use of larger membranes and structuring of the supporting membrane (like the "spiderweb" bolometers of Bock et al. 1995). Furthermore the reduction of timeconstant was investigated (theoretically) by the use of other biasing schemes than regular current bias.

Proc. 32nd ESLAB Symp., 'Remote Sensing Methodology for Earth Observation and Planetary Exploration', ESTEC, Noordwijk, The Netherlands, 15-18 September 1998 (ESA SP-423, December 1998)

In this publication we describe the results obtained with three different types of high-T_c bolometers. The first type contains a closed SiN membrane, but now with a larger membrane size. The two other types are bolometers with a patterned SiN membrane resulting in eight SiN beams suspending the sensing area of the membrane with a remaining foil of PtO_x between these beams, and bolometers without this foil which have completely freestanding SiN legs.

2. PRODUCTION TECHNOLOGY

The fabrication process of bolometers with a closed silicon-nitride membrane was described before by Sánchez et al. (1998). The present fabrication processes form an extension of this process. The processes of the bolometers with patterned membranes is illustrated in Figure 1. From the various routes that were investigated, two routes are presented here. The first part of both routes is the same.

The starting substrate is a so-called silicon-on-nitride (SON) wafer. The SON wafers are made by bond-and-etch-back of two wafers: A commercially available silicon-on insulator (SOI) wafer from SOITEC and a Si wafer on which a 1 μm silicon-rich silicon-nitride ("SiN") layer has been grown. Previous to the bonding the SiN is chemically-mechanically polished (CMP), using a standard procedure from integrated circuit planarisation technology. Si and SiO_2 from the SOI wafer is removed by wet etching. The SiN layer at the backside of the Si wafer is used as a mask for etching the membrane windows. The etching is stopped when the SiN membrane is still supported by about 25 μm of Si. The resulting membranes are strong enough for the following processes.

The next step is epitaxial growth of a 40 nm double bufferlayer of yttrium stabilized zirconia (YSZ) and CeO_2 on the Si top layer using molecular beam epitaxy. The layer growth is monitored in-situ using low energy electron diffraction (LEED). Then the meander pattern for the high-T_c layer is etched in the bufferlayers and the Si layer by argon sputter etching and reactive ion etching (RIE) respectively. This step will define the area of the GBCO layer that will be superconducting (inhibition patterning). A second resist mask is used to define the suspension beams in the SiN layer. Etching is done by RIE. It is stopped when there is about 250 nm SiN left between the suspension beams.

From this point on the two routes will differ. The substrates are diced into pieces of 1×1 cm^2. In route a) (see Figure 1) the GBCO layer is deposited by hollow magnetron sputtering. On top of the GBCO a passivation layer of 200 nm platinum oxide (PtO_x) is deposited by planar magnetron sputtering. Then, using a special designed protection holder, which seals off the front side of the substrates, the remaining 25 μm of Si is removed

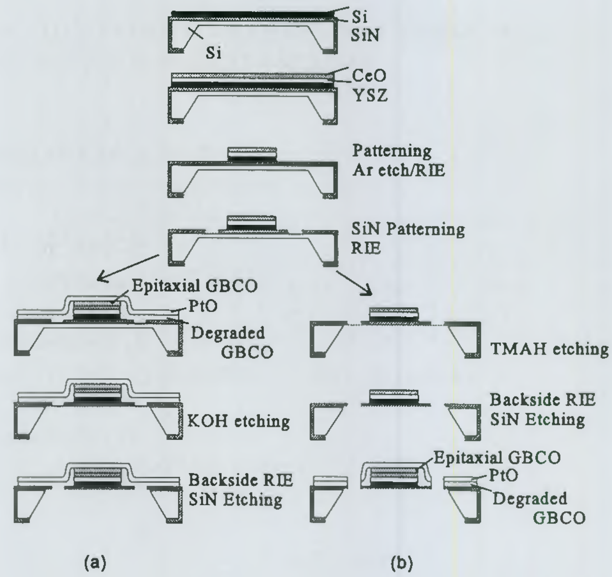


Figure 1. Schematic overview of the production routes of HTS bolometers. The individual process steps are clarified in the text. Route a) results in bolometers with a foil of degraded GBCO and PtO_x between the suspension legs. Route b) results in freestanding spiderweb-like bolometers.

from the backside of the membranes by KOH etching. The 250 nm thick SiN layer between the beams of the web structure prevent the KOH from reaching the front side. This SiN layer can be removed by RIE. The resulting bolometers have a remaining GBCO/PtO_x layer between the SiN beams. This GBCO is degraded and will not become superconducting. To get rid of the GBCO/PtO_x between the SiN beams it is required to follow route b): The Si at the backside of the membrane and the 250 nm of SiN between the beams have to be removed before deposition of the GBCO.

Finally on top of the bolometers an absorption layer of gold-black is deposited through a shadow mask with a pinhole of 1.1 mm diameter. The thickness is between 25 and 50 micrometer, with a filling fraction of 0.003. The resulting absorption efficiency for radiation with wavelengths around 100 μm is approximately 90% (Becker 1996).
Figure 2 shows a photograph of the backside of a bolometer with 8 SiN supsension beams and 60+200 nm GBCO/PtO_x between the beams. The PtO_x is buckling due to its compressive stress. In the center of the bolometer the superconducting meander can be seen through the SiN layer. In Figure 3 a bolometer is shown with really free-standing SiN beams, prepared along route b) in Figure 1.

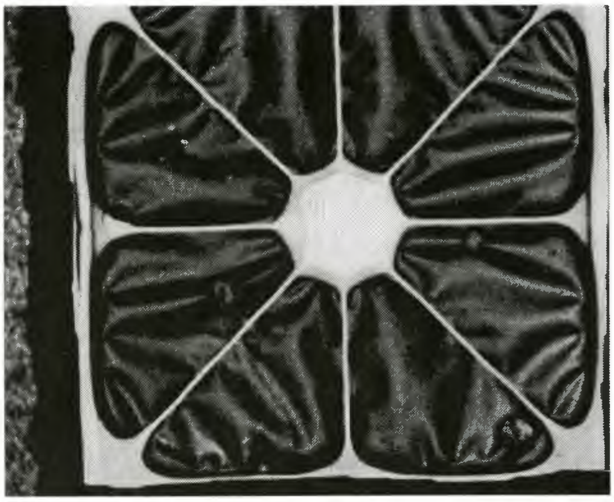


Figure 2. Photograph of the backside of a bolometer with a 4.5x4.5 mm membrane size. Between the 8 SiN suspension beams the residual layer of GBCO/PtO$_x$ can be seen, which is buckling due to its compressive stress. The beams are 40 μm wide and 1.6 or 2.2 mm long.

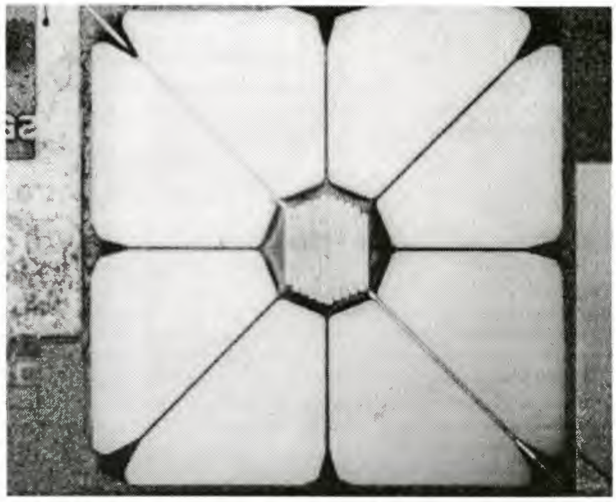


Figure 3. Photograph of a bolometer with freestanding SiN beams. The dimensions are the same as in Figure 2.

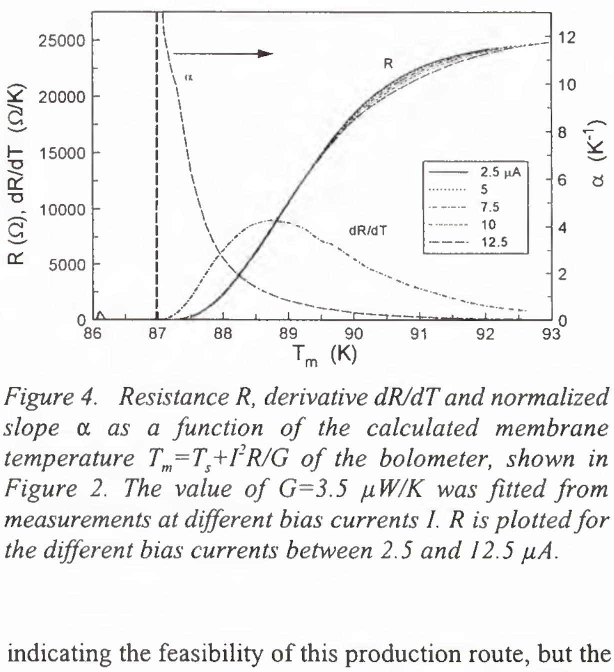


Figure 4. Resistance R, derivative dR/dT and normalized slope α as a function of the calculated membrane temperature $T_m = T_s + I^2R/G$ of the bolometer, shown in Figure 2. The value of G=3.5 $\mu W/K$ was fitted from measurements at different bias currents I. R is plotted for the different bias currents between 2.5 and 12.5 μA.

3. BOLOMETER PERFORMANCE

The bolometers were characterized in a vacuum cryostat with liquid nitrogen cooling. The bolometer temperature can be controlled with temperature fluctuations smaller than 5 μK (on a timescale of 100 s). The normal operating mode is current biasing. To reduce the low frequency noise of the electronics, a square wave current modulation at 1 kHz was used together with phase sensitive amplification.

3.1. Resistance vs Temperature Measurements

The thermal conductance G was detemined by measuring the superconducting transition at different current levels, and fitting the R-T_m curves on top of each other. Here $T_m = T_s + (I^2R+P)/G$ is the calculated temperature of the membrane, T_s is the temperature of the substrate and P is the radiation load on the bolometer. The results for the bolometer shown in Figure 2 are given in Figure 4. It was found that G=3.5 μW/K. The power load P could be neglected during this measurement. The normalised slope of the transition $\alpha = 1/R \, dR/dT$ at the point where the slope is at maximum is 1.1 K^{-1}. For all bolometers typically values between 1 and 2.5 K^{-1} were found.
An overview of the specifications of three typical bolometers with different geometries including the parameters R_{mid}, T_{mid} and α_{mid}. is given in Table 1. The subscript *mid* indicates the midpoint in the transition where the slope is at maximum.
The bolometer of Figure 3 was only tested sofar in a cryostat with unknown and high radiation load. The R-T measurement showed a superconducting transition, indicating the feasibility of this production route, but the data was unsufficient reliable to be incorporated in Table 1. The yield of fabrication of bolometers along route b) (Figure 1) was largely enhanced by removal of half of the top Si-layer by oxidation followed by wet chemical etching of the oxide. Presumably the top part of the Si-layer contained damage that prevented growth of high quality buffer layers and high T_c layers.

3.2. Optical Responsivity

The absolute optical response of the bolometers has been determined with two radiation sources with different power levels. The sources were black bodies at 30 and 50 °C. Their radiation was filtered with a well-characterized filter at a temperature of 90 K and then collected on the bolometers with a compound parabolical concentrator (CPC or "Winston cone"). The resulting spectrum covers

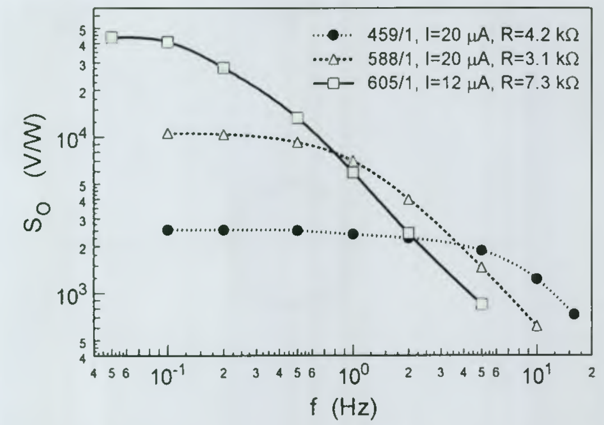


Figure 5. Measured optical responsivity of the three bolometers listed in Table 1. as a function of the signal frequency for radiation around 100 μm wavelength. The bias points are indicated in the figure. The boxes (605/1) represent the bolometer of Figure 2.

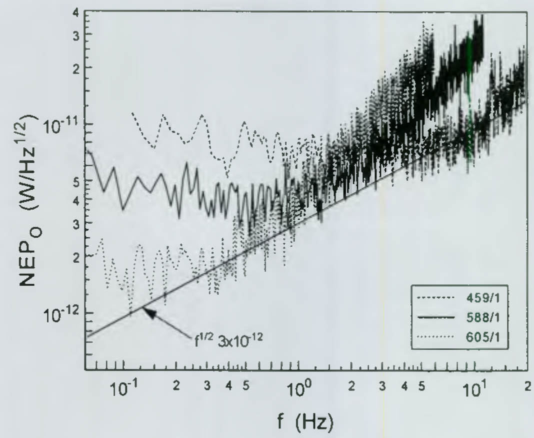


Figure 6. Measured NEP spectra of the three bolometers listed in Table 1. for radiation around 100 μm wavelength. The dashed data represent the one with 2x2 mm closed membrane. The solid line represents the one with 4x4 mm closed membrane and the dotted data belong to the spiderweb bolometer (Figure 2).
The straight line indicates the observed limit: $NEP_O=3x10^{-12}\sqrt{f}$ $W/\sqrt{Hz}$.

a band with wavelengths from about 70 to 200 μm with a maximum at 85 μm.

The frequency roll off of the bolometer was determined by measuring the response to a modulated light emitting diode. Figure 5 shows the resulting response spectra at indicated bias points of the three bolometers listed in Table 1. By comparing the measured optical reponse with the electrically determined responsivity we could conclude that the efficiency of the bolometers in the test setup is between 0.7 and 0.8. Careful analysis of the effects of defocussing at the exit opening of the CPC and the temperature gradient over the membrane indicate a true absorption efficiency of the detector close to 0.9, a value that can be calculated for the gold black absorption layer.

3.3. Noise Equivalent Power & Detectivity

The noise equivalent power (NEP) was determined by dividing the noise spectra by the optical responsivity. The result for the three bolometers is shown in Figure 6. For all bolometers there is a range of bias temperatures and bias currents for which the NEP spectra are similar and for which the NEP is minimal. Usually the current and the temperature can be chosen such that R is around R_{mid} and the bias parameter $L_0=I^2R\alpha/G$ is between 0.1 and 0.3. In the frequency range where the NEP is constant the NEP is dominated by the phonon noise in the thermal conductance G. The observed levels are in agreement with the theoretical level for the phonon noise given by:

$$NEP_{phonon} = \frac{\sqrt{4kT^2G}}{\eta} \quad (1)$$

in which η is the optical absorption efficiency.
In Figure 6 a line is drawn which represents the experimental limit for the NEP of these bolometers. This limit $NEP_{limit} \approx 3\times10^{-12}\sqrt{f}$ W/√Hz. By decreasing the thermal conductance of the bolometers the NEP becomes smaller while at the same the bolometers become slower. The limiting relation between the NEP and the frequency

Table 1. Overview of the performance of three typical bolometers with different geometries.

#	R_{mid} (kΩ)	T_{mid} (K)	α_{mid} (K^{-1})	G (μW/K)	τ (s)	η	NEP_O ($pW/Hz^{1/2}$)	D^* ($cmHz^{1/2}/W$)	f (Hz)	Description
459/1	3.4	87.6	2	45	0.03	0.7	6.9	$1.4x10^{10}$	0.7-5	$2x2mm^2$ closed membrane
588/1	8.0	90.2	1.5	20	0.14	0.75	3.8	$2.6x10^{10}$	0.4-0.6	$4x4mm^2$ closed membrane
605/1	8.0	88.7	1.1	3.5	0.55	0.80	1.8	$5.4x10^{10}$	0.09-0.3	$4.5x4.5mm^2$ spiderweb (Figure 2)

is determined by the heat capacities of the bolometers which are all roughly 2 μJ/K. This can be seen by substituting for G in formula (1) the relation $G=C/\tau\approx 2\pi f C$. Formula (1) transforms into:

$$NEP_{phonon} \approx \frac{\sqrt{8\pi f\, kT^2 C}}{\eta} \qquad (2)$$

C could be made smaller by reducing the thickness of Au-black, PtO_x, Si or Si_xN_y. Reducing the Au thickness will cause the efficiency to decrease. The PtO_x is needed for protection against environmental attack. Reducing the Si thickness could be done by oxidation and etch back. The relative gain is only small. The only way is reducing the nitride thickness. This will cause the bolometers to become more fragile. We have not explored the limit in nitride thickness.
Another way to improve the NEP is by reducing the operating temperature. This also causes C to decrease. The problem is to find a suitable material with lower T_c and good noise properties. The aim of this route would be to fill in the gap between liquid-He cooled detectors and liquid-N_2 cooled high-T_c detectors, where cryocoolers can be used.

It has been argued that by using voltage bias with strong electrothermal feedback the effective timeconstant of the bolometers can be decreased (Mather 1982, Irwin 1995). Consequently, by reducing the thermal conductance the NEP can be decreased, while keeping the effective speed of the bolometer constant. Also biasing schemes can be designed in which a feedback loop reduces the fluctuations of the bolometer resistance by electrically compensating variations in the power of absorbed radiation. Thereby temperature excursions are reduced and the response speed is increased.
However, an unexpected noise term prevents this strategy to be effective with high-T_c bolometers: The $1/f$ noise. When electrothermal feedback is taken properly into account, the Johnson and $1/f$ noise can be written as (De Nivelle et al. 1997):

$$NEP_{Johnson} = \frac{1}{\eta}\sqrt{\frac{4kTG}{\alpha L_0}}\,|1+i\omega\tau| \qquad (3)$$

and

$$NEP_{1/f} = \frac{1}{\eta}\sqrt{\frac{\gamma_H 2\pi}{n_c E\omega}\,\frac{G}{\alpha}}\,|1+i\omega\tau| \qquad (4)$$

in which γ_H is the Hooge parameter (Hooge 1981), n_c the number of charge carriers and E the volume of the film. These formulae are independent of the biasing scheme. Using voltage bias, the loopgain L_0 can be made large because --unlike with current bias-- there is no limit set by the thermal runaway criterium. A large loopgain suppresses the NEP due to the Johnson noise, but the $1/f$ noise is not suppressed. At high frequencies, i.e. $\omega>>1/\tau$, the NEP due to the $1/f$ noise becomes:

$$NEP_{1/f} = \frac{1}{\eta}\sqrt{\frac{\gamma_H f}{n_c E}\,\frac{2\pi C}{\alpha}} \qquad (5)$$

Typical values for our bolometers are: $\gamma_H=1$, $n_c=2\text{x}10^{21}$ cm^{-3}, $E=2.5\text{x}10^{-8}$ cm^3, $C=2$ μJ/K, $\alpha=2$/K, $\eta=0.75$. Substituting these values yields $NEP_{1/f} = 1.2\text{x}10^{-12}\sqrt{f}$ W/√Hz.
Adding this in quadrature with the phonon noise, the minimum attainable NEP for our high-T_c bolometers at a given frequency, using voltage bias, will be $NEP_{limit} = 1.7\text{x}10^{-12}\sqrt{f}$ W/√Hz. In terms of detectivity this can be expressed as $D^*=\sqrt{\text{Area}}/NEP = 1.5\text{x}10^{11}\sqrt{\tau}$ cm√Hz/W. This is only a small improvement over the present limit with current bias. Alternatively, at a given NEP the device can only be made a factor 3 faster by using voltage bias.

4. CONCLUSIONS

By using a new technique to decrease the thermal conductivity, i.e. structuring the supporting membrane, we have developed high-T_c bolometers with an optical NEP of $1.8\text{x}10^{-12}$ W/√Hz at an operating temperature of 90 K. With the present technology we have approached the limit of the attainable detectivity at a given time constant of the device, and using current bias. These bolometers have the highest detectivity for time constants between 0.01 and 1 s, compared to those reported in present literature. The empirical limit in detectivity for bolometers on a 1 μm thick nitride membrane $D^*=8.4\text{x}10^{10}\sqrt{\tau}$ cm√Hz/W, when using current bias. Voltage bias will only result in improvement of detectivity with a factor 1.8, or time constant with a factor 3, compared to current bias, because of the effect of $1/f$ noise in the high-T_c films. The empirical limit for our bolometers in current bias is a factor 5 higher than Haven's empirical limit for room temperature thermal detectors (Rogatto 1993).
Really freestanding spiderweblike bolometers have been fabricated, but no conclusive data can be presented yet.
Shifting the limit for high-T_c bolometers further upward will require the heat capacity and/or the operating temperature to be lowered. We consider the feasibility of reduction of the thicknesses unlikely.

ACKNOWLEDGMENTS

Part of this work was performed under funding from the Earth Observation Preparatory Programme and the Technology Research Programme of the European Space Agency. The other part was financed by the Dutch Organization for Scientific Research (NWO).

REFERENCES

Becker W., Fettig R., Gaymann A., Ruppel W., 1996, Phys. Status Solidi B194, 241.

Berkowitz S.J., Hirahara A.S., Char K., Grossmann E.N., 1996, Appl. Phys. Lett. 69, 2125.

Bock J.J., Chen D., Mauskopf P.D., Lange A.E., 1995, Space Sci. Rev. 74, 229.

Brasunas J.C., Lakew B., 1994, Appl. Phys. Lett. 64(6), 777.

Chance K., Wijnbergen J.J., Schneider W., Burrows J.P., 1994, SPIE Proc. 2159, 21.

De Nivelle M.J.M.E., Bruijn M.P., De Vries R., Wijnbergen J.J., De Korte P.A.J., Sánchez S., Elwenspoek M., Heidenblut T., Schwierzi B., Michalke W., Steinbeiss E., 1997, J. Appl. Phys. 82(10), 4719.

Hooge F.N., Kleinpennig T.G.M., Vandamme L.K.J., 1981, Rep. Prog. Phys. 44, 479.

Hoogeveen R.W.M., Wijnbergen J.J., De Nivelle M.J.M.E., De Jonge A.R.W., Goede A.P.H., De Korte P.A.J., SPIE Proc. 3117, 19.

Irwin K.D., 1995, Appl. Phys. Lett. 66(15), 1998.

Johnson B.R., Foote M.C., Marsh H.A., Hunt B.D., 1994, SPIE Proc. 2267, 24.

Mather J.C., 1982, Appl. Opt. 21(6), 1125.

Méchin L., Villégier J.C., Bloyet D., 1997, J. Appl. Phys. 81, 7039.

Neff H., Laukemper J., Khrebtov I.A., Tkachenko A.D., Steinbeiss E., Michalke W., Burnus M., Heidenblut T., Hefle G., Schwierzi B., 1995, Appl. Phys. Lett. 66, 2421.

Richards P.L. 1994, J. Appl. Phys. 76(1), 1.

Rogatto W.D. (ed.), 1993, The Infrared & Electro-Optical Systems Handbook, Vol. 3, 204, copublished by Infrared Information Analysis Center & SPIE Optical Engineering Press.

Sánchez S., Elwenspoek M., Gui C., De Nivelle M.J.M.E., De Vries R., De, Korte P.A.J., Bruijn M.P., Wijnbergen J.J., Michalke W., Steinbeiss E., Heidenblut T., Schwierzi B., 1998, J. Microelectromechanical Syst. 7 (1), 62.

Verghese S., Richards P.L. Sachtjen S.A., Char K., 1993, J. Appl. Phys. 74, 4251.

Wennberg P.O., Cohen R.C., Stimpfle R.M., Koplow J.P., Anderson J.G., Salawitch R.J., Fahey D.W., Woodbridge E.L., Keim E.R., Gao R.S., Webster C.R., May R.D., Toohey D.W., Avallone L.M., Proffitt M.H., Loewenstein M., Podolske J.R., Chan K.R., Wofsy S.C., 1994, Science 266, 398.

STATUS AND TRENDS IN DETECTOR DEVELOPMENTS IN SWIR, MWIR AND LWIR WAVEBANDS

P. Tribolet, A. Manissadjian, P. Costa, Ph. Chorier

SOFRADIR, 43-47 rue Camille Pelletan,
92290 Chatenay-Malabry, France
Phone: 33.1.41.13.45.30 Fax : 33.1.46.61.58.84

&

G. Destefanis, J.L. Tissot

LETI/CEA.G-DOPT/LIR
17, rue des Martyrs
38054 Grenoble Cedex 9, France
Phone : 33.4.76.88.44.00 Fax : 33.4.76.88.51.67

ABSTRACT

Earth observation and planetary observation require infrared detectors with the constraint to operate in various wavebands and under specific non adapted flux conditions.

A second constraint is the goal of simplifying the payload concept by using passive coolers.

In order to answer these needs, Sofradir proposes HgCdTe infrared focal plane assemblies (IRFPAs) operating at high temperature with outstanding performance. In addition, Sofradir cooperated with CEA-LETI (Laboratoire InfraRouge LIR) to propose uncooled IRFPAs based on amorphous Silicon microbolometer technology.

The goal of this paper is to present the performance of Sofradir IRFPAs for SWIR, MWIR and LWIR, as well as the trends in detector developments.

SWIR waveband :
A deep analysis of dark current versus operating temperatures will be described for SWIR components. Then performance results of Sofradir IRFPA will be presented. Silicon readout circuit constraints and future trends will be discussed.

MWIR waveband :
Large staring arrays are proposed by Sofradir at operating temperatures up to 150 K for full performance and up to 200 K with decreased performance. Analysis of existing Sofradir staring arrays in terms of performance versus operating temperatures will be presented, and future trends for large staring arrays will be discussed.

LWIR waveband :
Two approaches can be proposed by Sofradir for LWIR. The first one is based on HgCdTe technology and provides high performance with operating temperature up to 110 K. New results for HgCdTe LWIR staring arrays will be presented and future trends will be discussed.
The second approach is based on amorphous Silicon microbolometer technology, operating at ambient temperature and developed at LETI/CEA.G (LIR). This new technology propose medium performance large staring arrays with very large wavelength range, and is adapted to medium and high fluxes conditions. Performance of first demonstrators will be presented and future developments will be discussed.

1. INTRODUCTION

Earth observation and planetary observation require infrared detectors with the constraint to operate in various wavebands and under non adapted flux conditions.

Moreover the use of passive coolers is more and more required to move to a simple and lightweight payload concept.

In order to answer most of these needs, Sofradir proposes HgCdTe infrared focal plane assemblies (IRFPA) operating at high temperature with outstanding performance. As the high performance IR detectors segment is small regarding quantities, Sofradir approach consisted in developing a simple cooled IRFPA technology capable of answering most of the high performance needs, to increase IR detectors quantities and to reduce production cost.

Therefore, Sofradir IR detectors are based on a technology (licensed from LETI/CEA.G) which uses Mercury Cadmium Telluride (HgCdTe / MCT). This semiconductor material is unique since as a ternary compound its composition can be tuned to detect wavelengths between 1 µm and 14 µm, at 80 K for a 10.5 µm cut-off wavelength as well as at 200 K for a 4.2 µm cut-off wavelength. In each case the same technology is used without the need to duplicate equipments or to have dedicated clean rooms.

One of the main advantages of HgCdTe material is its ability to operate at high temperatures with respect to other candidate materials and therefore to reduce the cooling constraints (size, cost, power consumption,...) by using small Stirling-cycle cryocoolers or thermoelectric coolers (TEC).

Based on this approach and on the vast experience acquired with the variety of development programs, Sofradir mastered the production of second and third generation IR detectors and demonstrated the producibility and affordability of these components in the SWIR, MWIR and LWIR main IR wavebands for different applications including space applications.

Proc. 32nd ESLAB Symp., 'Remote Sensing Methodology for Earth Observation and Planetary Exploration', ESTEC, Noordwijk, The Netherlands, 15-18 September 1998 (ESA SP-423, December 1998)

Furthermore, Sofradir proposes other detector types adapted to new application requirements and in particular Sofradir cooperates with LETI/CEA.G-DOPT/LIR to propose uncooled IRFPAs based on amorphous Silicon microbolometer technology.

In the following, Sofradir's IRFPA performances in different wavebands SWIR, MWIR and LWIR are reviewed as well as the trends in detector developments to meet space application requirements.

2. SOFRADIR COOLED IRFPAs

2.1 Technology overview

Sofradir has a Vertically Integrated Development and Production Facility (VIPF) dedicated to MCT IRFPA production and dewar-cooler electronic integration.

The key points of Sofradir/LETI-LIR technology are :

- growth of Cadmium Zinc Telluride (CdZnTe) ingot using a Bridgman method, and growth of HgCdTe layer using a Liquid Phase Epitaxy (LPE) method (Te-rich corner and slider technique) which yields a very accurate material composition and thickness;
- processing of photovoltaic (PV) diodes using a well-controlled ion implantation technique. This technique allows accurate control of the pixel size for very good Modulation Transfer Function (MTF), and allows the use of a totally planar technology, which is the easiest way to obtain the best passivation and therefore the best long term stability;
- hybridization of PV diodes and silicon readout circuit using a unique Indium bump reflow technique (100 % bonding yield). This technique minimizes stress to the diodes and therefore improves results in long term stability and reliability. In addition, this unique process provides a self-alignment procedure making possible multichip hybridization step on silicon wafer [1] [2] (see figure 1). This leads to a batch process up to IRFPA selection by 80 K wafer test, and therefore to a significant producibility improvement.

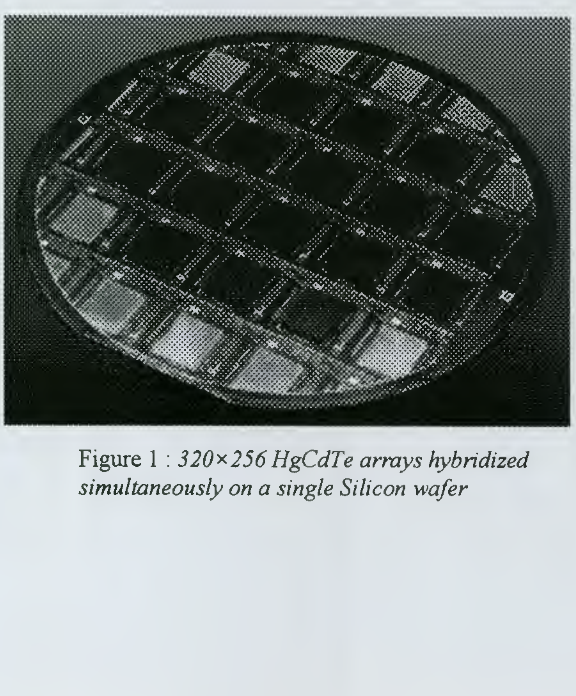


Figure 1 : *320×256 HgCdTe arrays hybridized simultaneously on a single Silicon wafer*

Thanks to this multichip hybridization process, Sofradir and LIR can propose cost efficient solution of butted arrays (9) in order to fabricate very large focal plane arrays for space or reconnaissance applications (see figure 2).

Based on these main technological steps, Sofradir/LETI-LIR technology provides high yield, high performance and low fixed pattern noise, for numerous (λ_c , T_{FPA}) working points and can develop most of the IR detectors needed for IR applications.

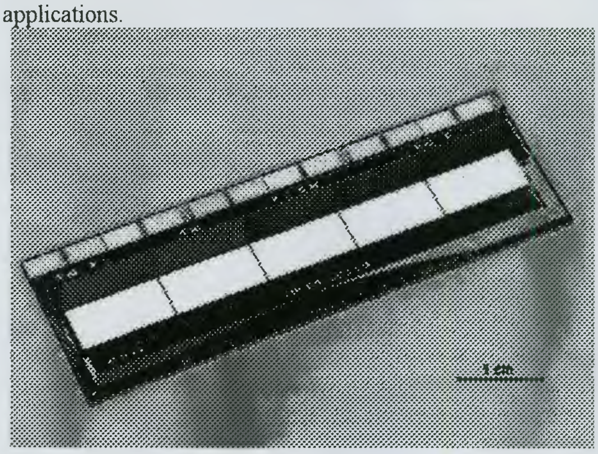


Figure 2 : *1500×1 element linear MWIR HgCdTe butted array*

2.2 Short Wave IR band

The SWIR waveband from 1 to 3 µm has many applications and is required for space applications with earth observations (military needs, agriculture monitoring, meteorological studies,...) and astronomy applications.

For the first set of applications, performances rely on key features such as low dark current, high quantum efficiency and low readout noise. Indeed, low optical fluxes are often involved in this spectral band and then quantum efficiency must be as high as possible to increase the detector sensitivity, and at the same time, the dark current is mainly the limiting factor.

Devices based on different IR materials have addressed this spectral range with some limitations. For instance, InGaAs exhibits high performance for material whose composition is matched to InP (cut-off wavelength close to 1.7 µm) but its performance decreases a lot for higher cut-off wavelengths (due to mismatch-induced defects with the substrate).

Widely used for MWIR and LWIR spectral bands, HgCdTe can also be used for SWIR applications in order to meet applications needs.

Based on studies performed at Sofradir and LETI/CEA.G (LIR), measurement results have been obtained for SWIR applications in terms of dark current evolution with FPA temperature. This is the case for 2.2 µm and 3.16 µm cut-off components for instance.

Measurements on Sofradir photodiodes at 200 K show that HgCdTe SWIR detectors exhibit very good performance: for 2.2 µm cut-off, measurements give extremely high shunt resistance (typical value : 10^{11} to 10^{12} Ω) high quantum efficiency (above 60-70 % typical), responsivity uniformity < 8% and cut-off wavelength uniformity ≤ 0.01µm.

In parallel, a dark current model has been worked out and validated in the 3 to 13 µm cut-off wavelength range. This model has been extrapolated for cut-off wavelengths below 3 µm and compared with the experimental results. This analysis is presented in figure 3.

The good agreement between theoretical and experimental results for SWIR high cut-off wavelengths (close to 3 µm) can be noticed. For lower wavelengths, experimental results obtained at 2.2 µm fit with a prediction at 2.4 µm. Therefore the accuracy of dark current versus cut-off wavelength modeling is about 0.2 µm, for very short wavelengths.

As an example, it is possible to calculate the maximum FPA operating temperature in order not to degrade the signal to noise ratio, assuming that the dark current must be less than 0.3pA for 1000µm² diodes (dark current density < 100nA/cm^2). Then, one can see that for high performance SWIR detection, HgCdTe can be used above 180 K for λ_c ≤ 2.7 µm, using thermoelectric coolers (TEC). This allows significant reduction in size, power consumption, and cost, in comparison to other detector materials candidates for SWIR applications.

As part of a study contract with ESA, Sofradir will characterize a 320 x 240 staring array with 30 µm pitch and a cut-off wavelength close to 2.5 µm. This device will be operated in the 170-230 K operating range in order to demonstrate the ability of Sofradir MCT technology to answer SWIR requirements for earth observation coupled with the use of passive coolers.

The last subassembly which is critical for SWIR applications is the readout circuit structure. As a matter of fact, the readout circuit is very often a limiting element of the global performance of an IRCMOS device, in particular for low flux applications in which input stage is very critical for two major parameters : bandwidth and noise. Sofradir already experienced and validated two coupling methods in order to increase the bandwidth of the readout circuit input stage : buffered direct injection and transimpedance amplifiers.

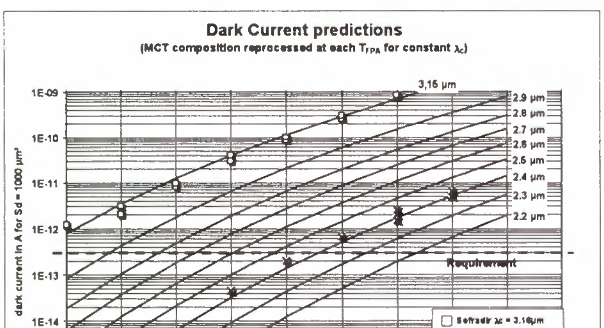


Figure 3

2.3 Medium Wave IR band

The MWIR waveband is of highest interest for many applications, due to atmospheric transmission behavior.

Two main sub-bands are used by most of the applications : 3.0 to 4.2 µm, and 3.7 to 4.8 µm, each of them having its own advantages for the systems.

Thus the great advantage of MCT is its ability to finely tune for each sub-band the cut-off wavelength in order to fit with the application need, resulting in a minimization of the dark current versus the photonic current.

2.3.1 High performance applications

In the 3.0 to 4.2 µm bandwidth, the MCT detectors are typically tuned at 4.2 to 4.3 µm cut-off. The induced dark current at 80 K is thus about 100 times smaller than the dark current for a detector with a 5.0 µm cut-off or about 1200 times smaller than the one for a detector with a 5.5 µm cut-off (see figure 4). The induced performance is therefore much better than the ones with higher cut-off materials, and the possibility is given to the system user to increase the operating temperature without any degradation of the performance.

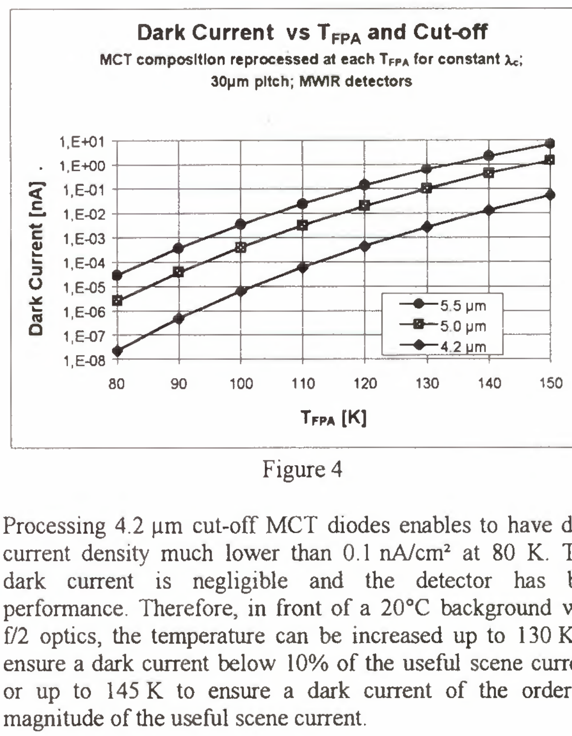


Figure 4

Processing 4.2 µm cut-off MCT diodes enables to have dark current density much lower than 0.1 nA/cm² at 80 K. This dark current is negligible and the detector has blip performance. Therefore, in front of a 20°C background with f/2 optics, the temperature can be increased up to 130 K to ensure a dark current below 10% of the useful scene current, or up to 145 K to ensure a dark current of the order of magnitude of the useful scene current.

In comparison, applications using Indium Antimonide (InSb) detectors must have an operating temperature below 100 K, and are highly limited by the dark current which is very important compared to the useful flux, due to the fact that InSb cut-off can not be tuned and is set to around 5.5 µm, even if the useful bandwidth stops at 4.2 µm.

In the 3.7 to 4.8 µm bandwidth, InSb, for which cut-off and performance are adapted to these applications, but at T_{FPA} < 100 K, compete with MCT. However, Sofradir has demonstrated that its stabilized and well mastered MCT process enables to produce high performance, large 2D arrays detectors which take the market lead. For instance, the production for a European missile program has been granted to the Sofradir 320×256 MCT FPA.

High performance and reproducibility are demonstrated at Sofradir. Figure 5 shows the performance of 320×256 FPA, 30 µm pitch, at f/2 field of view, 20°C background, accumulated on more than 100 detectors. The Noise Equivalent Temperature Difference (NETD) presents a very nice gaussian profile with low number of defects.

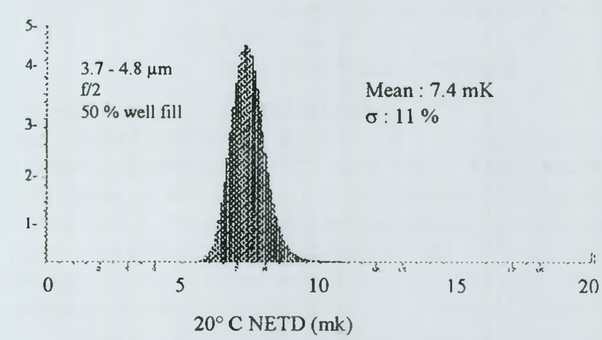


Figure 5 : *NETD of cumulated 320×240 detectors, 3.7-4.8 µm, f/2, 20°C*

For 3.7 – 4.8 µm bandwidth, NETD is constant up to 120 K for f/4 applications and up to 130 K for f/2 applications, while InSb can not stand this stability for T_{FPA} higher than 90 K.

Systems have therefore great benefits to use MCT at higher temperature: all systems get shorter cooldown times; missile seekers increase their Joule-Thomson cooling autonomy; FLIRs can run their Stirling-cycle cooling engine at lower speed for increased lifetime or power savings without loss of performance. For example, the cryogenic power of a rotary micrococler needed to cool down an FPA at 120 K is only 55 % of the one needed to cool it down at 77 K, under 20°C ambient, and is only 50 % under 71°C ambient.

Systems working around 150 K will allow Sofradir to propose either a very cost effective approach by reducing the constraints on the microcoolers, or to propose very long lifetime solutions by using new type of coolers like Pulse Tube.

2.3.2 Medium performance applications

Thermoelectric coolers (TEC) can be used to operate MCT MWIR detectors at FPA temperature above 180 K with no noise, no vibration and long lifetime.

T.E. cooled detectors present high interest for space applications as well as commercial systems for thermography, surveillance or portable applications.

At these FPA temperatures, InSb does not work at all, while MCT continues to be sensitive and presents good enough performance.

When useful flux is low compared to the dark current, versatile current skimming can be implemented in the read-out circuit (auto-zero mode) to offset the dark current and integrate only the useful current .

With a λ_c processed at 4.0 µm at 200 K, an f/1.2 aperture and a 20°C background, typical performance are NETD = 45 mK without skimming (integration time = 0.9 ms) and NETD = 20 mK with skimming (integration time = 6 ms).

Sofradir detectors have demonstrated these very good performance at 200 K [3], are today in production for third generation 200 K FPAs used in commercial cameras and can also be proposed for medium performance space applications in order to simplify payload concept thanks to the TEC used.

2.4 Long Wave IR band

The LWIR or 8 to 14 µm waveband is used in a lot of military applications like FLIR, IRST, missile seekers (LWIR is indeed well adapted for army applications such as tank FLIRs and anti-tank weapons), and also in commercial or space applications (surveillance, microscopy, spectrometry,...).
The systems using LWIR detectors can be divided in three categories corresponding roughly to three different wavelength bands :

- systems using linear or staring arrays sensitive up to 12 to 14 µm;
- scanning systems, using 288×4 to 480×6 linear arrays sensitive between 8 and 10.6 µm and up to 11.5 µm;
- staring systems, using 128×128 to 320×256 staring arrays sensitive between 8 and 9 µm and up to 9.5 µm.

2.4.1 12 to 14 µm cut-off bandwidth

Many scientific applications, such as spectrometers, IR microscopes or space systems often require detection cut-off between 12 µm and 14 µm with high IR detector performances. Sofradir/LIR MCT technology can offer a lot of tradeoffs for system applications.

The choice for the (λ_c, T_{FPA}) operating working point will highly depend on the system specifications, in terms of ambient fluxes, field of view and defect criteria for instance.

Above 12 µm, the (λ_c, T_{FPA}) tradeoffs will consider the advantage of λ_c increase when T_{FPA} is lowered, or will consider the highest acceptable λ_c (and therefore maximum acceptable dark current) when T_{FPA} is kept at standard cryogenic values.

On one hand, spectrometers and IR microscopes applications can be proposed in standard integrated detector dewar cooler assemblies (IDDCA) for cryogenic temperatures between 60 K and 77 K.
In order to answer these applications requesting improved intrinsic performance without lowering the FPA temperature, efforts have been made at LETI/CEA.G to improve the PV detector RoA products for high λ_c [4] [5]. The obtained gain on the RoA product enables a Signal to Noise Ratio increase for the same FPA temperature.

On the other hand, space systems often have very specific needs or configuration, and specific coolers will provide very low FPA temperature (below 60 K) to ensure highest λ_c and/or highest signal to noise ratio (when astronomy useful fluxes are extremely low for instance).

As an example, LETI/CEA.G and Sofradir demonstrated the good performance at 55 K to 60 K of long linear array beyond 12.5 µm cut-off for space application, with the High Resolution Thermal Infrared Radiometer (HRTIR) [6].

2.4.2 8 to 11.5 µm bandwidth

Sofradir has a very large experience in the manufacture of high performance linear arrays sensitive in the LWIR range between 8 µm and 10.6 µm and up to 11.5 µm, and especially for TDI linear arrays.

The production experience on the 288×4 detector is around 1000 units already delivered all around the world. Almost all the second generation FLIRs which are currently developed or produced are based on the Sofradir 288×4 detector, which is already in production for four thermal imagers (DNTSS from Texas Instruments, IRIS from Sagem, Sophie from Thomson-CSF Optronique and recently a FLIR from Officine Galileo). Moreover, Sofradir 480×6 detector is proposed for the US next generation FLIRs (SADA II and SADA I programs), and is ready for production.

2.4.3 8 to 9 µm or 8 to 9.5 µm bandwidth

Staring arrays are more and more desired for infrared applications. However, many IR applications need relatively high frame rate operation and thus need specially adapted high performance IR detectors.
Sofradir offers today high performance MCT LW staring arrays adapted to these applications in the formats 128×128 with 50 µm pitch and 320×256 with 30 µm pitch.

As discussed in [7], for high frame rate staring systems, three parameters limit the performance which are : the maximum amount of charges storable in the readout circuit, the integration time linked to the frame time, and the dark current of the sensitive material. For given frame time and maximum storable charge, the way to increase the LW staring systems performance is to reduce the dark current, in order to get photon limited system performance. For the reduction of dark current, two solutions can be envisaged :

- to decrease the focal plane operating temperature;
- to decrease the cut-off wavelength by adjusting the MCT composition.

The second solution allows to increase the focal plane operating temperature without degrading the performance.

Compared to a 7.7 to 10.5 µm staring array at 80 K, a 7.7 to 9.5 µm staring array can have its focal plane operating temperature increased up to 87 K with the same performance, as shown on figure 6.

For a constant focal plane operating temperature, the performance are higher for λ_c = 9 µm due to the fact that dark current is lower and integration time can be increased. As a consequence, focal plane operating temperature for λ_c = 9 µm can be increased up to 90 K with the same performance (NETD < 18mK for the given example) than for λ_c =10.5 µm at 77 to 80 K.

A way to increase more the focal plane operating temperature is to use the LETI/CEA.G (LIR) improved photodiode technology [4] [5]. Thus, the focal plane operating temperature could be as high as 105 K for 7.7 to 9 µm detectors, and 102 K for 7.7 to 9.5 µm detectors (for NETD < 18mK, for the given example).

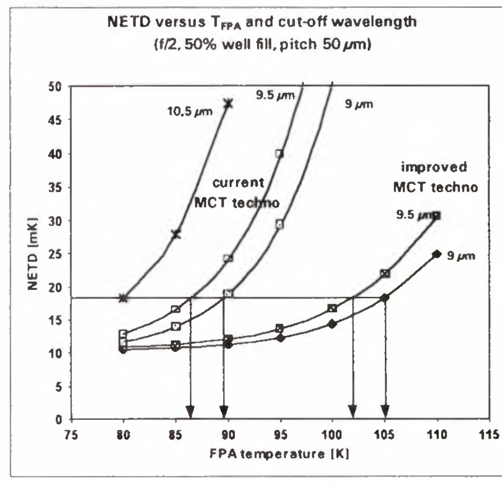


Figure 6

Furthermore, for some specific applications, by using additional focal plane processing functions like automatic current skimming (like for 200K MWIR arrays) to overcome readout circuit maximum storage limitation [3], it is possible to eliminate dark current and increase even more the focal plane operating temperature with maintained performance.

In figure 7, a new result of a 320 x 240 IRFPA, 30 µm pitch operating above 80 K and with a cut-off of 9.5 µm, is given. This LWIR staring array exhibits very high performance for f/2 optics and high uniformity.

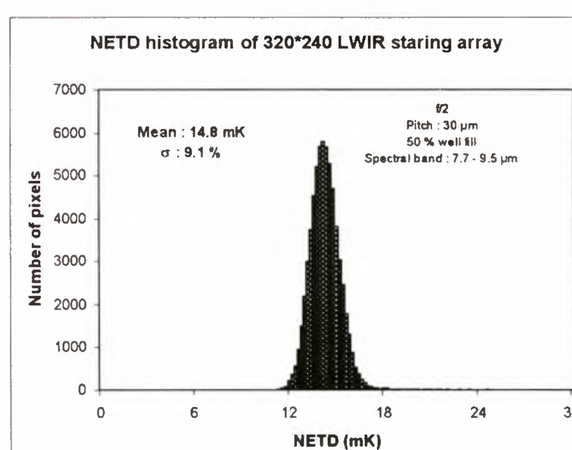


Figure 7

The previous analysis shows that MCT material is well adapted to high performance and high frame rate LWIR applications. Thanks to the ability of MCT material to tune the cut-off wavelength versus material composition, it is easy to fix the right (λ_c, T_{FPA}) operating point suitable to the application, in order to reduce system constraints.

2.5 Future trends regarding cooled IRFPAs

New MCT MWIR IRFPA developments at Sofradir and LETI/CEA.G are based on the following approaches [8].

- Detector cost reduction : producibility improvements ;
- Increase in detector format (TV format and more) by reducing detector pitch (20 µm and less). The challenge of pitch reduction is to keep performance constant. Regarding TV format LETI/CEA.G already developed in 1997 a 640 x 480 staring array in 3 to 5 µm and with 25 µm pitch [10] and as presented in figure 8, this 640 x 480 exhibits high performances.
- Development of new Silicon readout functions in order to increase detector performances and to simplify system interfaces.

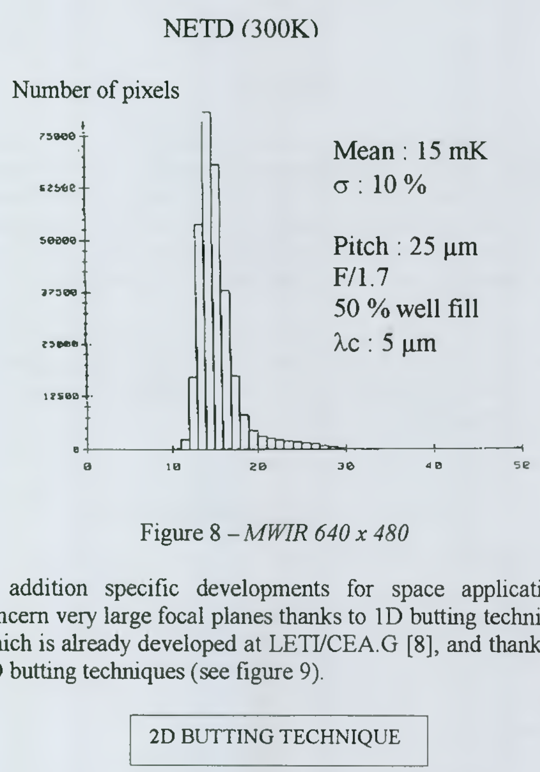


Figure 8 – *MWIR 640 x 480*

In addition specific developments for space applications concern very large focal planes thanks to 1D butting technique which is already developed at LETI/CEA.G [8], and thanks to 2D butting techniques (see figure 9).

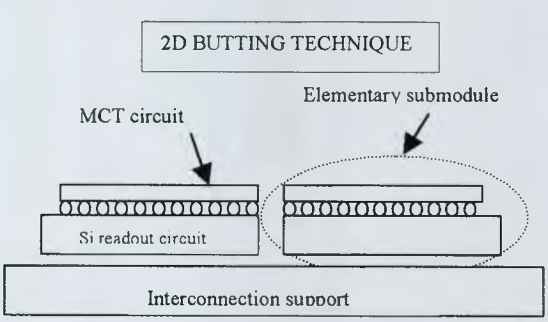


Figure 9

BUTTING OF SUB MODULES

3. UNCOOLED MICROBOLOMETER DEVELOPMENT

3.1 Introduction

Work on uncooled infrared detectors is currently showing rapid growth as a result of developments in silicon technology which pave the way for production of low-cost, high-performance detection arrays.

LETI/CEA.G developed a device with a monolithically integrated structure, over a fully completed readout integrated circuit (ROIC) specially designed for this application, from a commercially available 0.8 µm design rules CMOS line. This strategy avoids long and costly specific IC developments for bolometer integration within CMOS flow, and lends itself to high fill factors, because the entire pixel area is available for detector implementation.

Regarding the encapsulation, Sofradir developed packaging for uncooled detector based on its large experience regarding the use of TEC coolers for 200 K operation.

3.2 Pixel design and fabrication

The bolometer comprises a very thin microbridge thermometer (0.1 µm) of doped amorphous silicon with no extra supporting layer or membrane provided with an IR partially absorbing arrangement, supported by two legs anchored over the silicon substrate by metal studs (figures 10 & 11). The microbridge is built on a sacrificial polyimide layer, and is freed in a final step when the polyimide is ashed away to achieve a vacuum quarter wave cavity of 2.5 µm between microbridge and readout circuit in order to enhance 10 µm wavelength IR radiation.

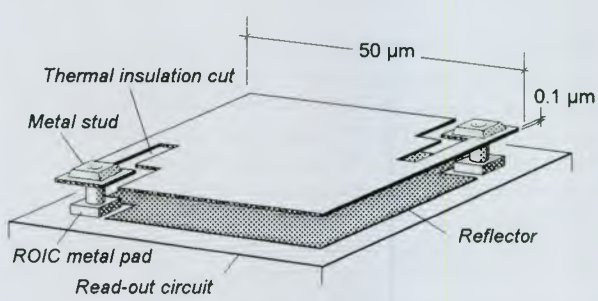


Figure 10 : *Schematic structure of pixel*

Figure 11 : *SEM view of a microbolometer array*

The thermal resistance (Rth) between the amorphous silicon (a-Si) thermometer and the isothermal readout circuit participates to a great extent in the final performance of the bolometer, so design of the support legs is of prime importance. Standard microbolometers are designed with two 3 μm wide, 10 μm long legs, leading to ~ 1.2×10^7 K/W thermal resistance which is twice the conventional structure developed in USA. Due to the very thin layer used for the microbridge, the thermal capacity (Cth) of the suspended parts amounts to 0.35 nJ/K, the thermal time constant Tth = Rth.Cth is then ~ 4.2 ms, largely compatible with 25Hz frame rates or higher.

3.3 Results

Electro-optical tests were performed on 256 x 64 arrays with a pitch of 50 μm under standard conditions. The resulting characteristics are summarized in table 12.

Field of view	f/1	Noise (0.125Hz, 25Hz)	700 μVrms
Operating temperature	295 K	NETD	70 mK
Frame rate	25 Hz	Responsivity	10 mV/K
Integration time	150 μs	Operability	> 99 %

Table 12 : *Characteristics and performances*

The LIR 256 x 64 microbolometer FPAs demonstrate a responsivity of about 10 mV/K with a non uniformity (std/mean) of 8 %. The resulting NETD is about 70 mK. It is noteworthy that the operability of these FPAs is over 99 %. This technology provides with a very responsive bolometer equal to the structure developed elsewhere. Furthermore, several technological improvement studies are currently being carried out in order to lower NETD figures towards 50 mK or less for military or space applications.

This demonstrator was integrated in a Sofradir package and exhibits high performance in tactical conditions.

3.4 Development trends

Important technological developments are being made worldwide in uncooled infrared imaging technology. These developments are likely to have a major impact in the use of thermal imaging systems both for military and civil applications providing they really result in the expected low cost of manufacture. Consequently, from the set of parameters which have guided the choices made at LETI/CEA.G, our priorities lie in simplification of construction of the microbolometer and its thermal insulation to ensure quick development of the basic technology and easy industrialization in the near future. The first results demonstrate the potential of the technology and support the choices made.

Next steps of work concern the pitch reduction down to 30 to 35 μm with increased performance, and the easiness of use of the microbolometer by implementation of functions at the level of the readout circuit (for suppression of the thermal regulation, offset correction, ...).

Finally, development of uncooled infrared technology shall not be limited to the focal plane; the study of encapsulation constraints (as thermal regulation and isolation...) is critical and is carried out by Sofradir/LIR but a lot of efforts are still to be done, especially for space applications like radiometry. As a matter of fact, specific design features may be necessary, like warm shield allowing to reduce extra scene incident flux and thus enhance the detector performance.

4. CONCLUSION

Regarding cooled technology, one of the main advantages of the Sofradir/LETI/CEA.G (LIR) technology is its ability to propose attractive systems tradeoffs thanks to focal plane temperature increase. In the SWIR waveband, this allows the use of TEC systems; in the MWIR waveband, it allows the use of all kinds of coolers; and in the LWIR waveband, it allows the use of small low power microcoolers. This leads to cryogenic and/or cost constraints releases which will be necessary for future space payload concept.

The simple and well-mastered HgCdTe planar homojunction technology used at Sofradir answers the present system needs in all SWIR, MWIR and LWIR wavebands. This technology is at production level in MWIR and LWIR, and exhibits outstanding results in terms of yield, uniformity and performance. In addition, this technology has been qualified for space program as HELIOS II in France.

And regarding uncooled technology, Sofradir will propose medium performance IRFPA based on LETI/CEA.G microbolometer technology which is well adapted to offer cost effective tradeoffs for space applications.

5. REFERENCES

[1] Collective flip-chip technology for HgCdTe IRFPA, J.L. Tissot, F. Marion, SPIE Proc. Vol.2894, Photonics China, 1996.

[2] New optoelectronic multichip module for infrared: module design and flip-chip hybridization requirements, F. Marion, M. Ravetto, J.P. Chamonal, C. Cigna – San José 1994, ITAP and Flip-Chip Proc., pp. 29-31.

[3] Versatile current skimming ROIC and its application in a TEC DCA, N. Reeb, F. Lembert, P.Maillart, C.Minassian, Sofradir, SPIE San Diego 1998, (to be published)

[4] Large improvement in HgCdTe photovoltaic detector performance at LETI, G. Destefanis and J.P. Chamonal, CEA/LETI Grenoble, France (Journal of Electronic Materials, Vol. 22, No. 8, 1993)

[5] High performance LWIR 256×256 HgCdTe FPA operating at 88K, G.L. Destefanis – SPIE Proc. Vol. 3061, 1997.

[6] A 12.5 μm infrared detector for earth observation, J.P.Chamonal & al., CEA/LETI, Sofradir, (ESA SP-356, December 1992)

[7] High performance MCT LWIR and MWIR staring array for high frame rate applications, P. Tribolet, V. Compain – Sofradir; R. Boch – CEA/LETI, SPIE Proc. San Diego 1995.

[8] Progress in HgCdTe homojunctions infrared detectors , P. Tribolet, J.P. Chatard, P. Costa, A. Manissadjian, Sofradir, France (Journal of Chrystal Growth 184/185, 1998, 1262-1271).

[9] 1500 element linear MWIR and LWIR HgCdTe arrays for high resolution imaging, E. Mottin, P. Audebert, P. Medina, M. Ravetto, J. Deschamps, M. Griard, J.P. Chatard, SPIE Proc. San Diego July 1997.

[10] 640 x 480 MCT 3 to 5 μm snapshot focal plane array, P. Audebert, D. Giotta, E. Mottin, P. Rambaud, F. Marion, LETI/CEA.G (France), SPIE Proc., Volume 3379, Orlando 1998.

GEOMETRIC CALIBRATION OF SPACEBORNE IMAGING SPECTROMETERS

P. Martimort * , M. Rast ** , E. Savaria *

* ALCATEL - 100 Bd. du Midi - BP 99 - 06156 Cannes la Bocca Cedex -France
** European Space Agency - ESTEC - Postbus 299 - NL 2200 AG Noordwijk - Nederlands

ABSTRACT

The extraction of geophysical information from remote sensed data set traditionally requires different kinds of data fusion strategies : multi-spectral, multi-directional or multi-temporal images exploitation.
Pixel to pixel data fusion algorithms require a highly accurate geometric registration of these images, and their performance is very sensitive to this parameter.
The object of this paper is to present the results of a study, supported by ESTEC, dealing with the identification and the performance assessment of on-ground image processing, devoted to multi-spectral and multi-temporal images registration.
The general principle of this algorithm is, first, to construct a geometric model of the image, associating ground point coordinates to each pixel (intersection of elementary line of sight with an appropriate Earth representation). This model is derived from the ancillary data describing the image acquisition process : orbit estimation, attitude (AOCS measures), instrument and focal plane geometry. This model has a physical parametric expression, allowing to optimize the parameters values, in order to reduce the difference between the predicted (modeled) and the measured ground point coordinates. This process assumes the acquisition of external data, as a reference. This constitutes the second step. Finally the optimized geometric model is used to resample the raw data set in the reference frame.
Considering a multispectral data set, assumed to have been acquired by different instruments installed on the same platform, we present a strategy for the geometric calibration and resampling of a secondary spectral band with respect to a reference band. The relative geometric model between the two bands is enhanced thanks to tie points acquisition, which are pixel pair (in reference and secondary image) viewing the same ground point. Finally the secondary spectral band is resampled in the reference band geometry.
In the time series exploitation case the algorithmic scheme is to register each image of the series to a common cartographic representation. This requires to gather cartographic Ground Control Points (GCPs), from map, GPS acquisition or a previously ortho-rectified image. These GCPs are visually or automatically located in the current image, and this information is used in the optimization process, leading to a registered geometric model. Then this model is used, in conjunction with a Digital Elevation Model (DEM), for the resampling operation. The influence of the DEM precision on the resampling efficiency has been addressed, as well as the radiometric interpolation strategy needed in the resampling process.
These processes allow to get sub-pixel registration, and permit to imagine a relaxation of the on-board geometric specifications, such as the multi-spectral co-registration and the absolute geo-location. This leads to an actual on-board/on-ground optimisation of a remote sensing spaceborne system. This has been verified during this study, by applying the algorithm to representative simulated images. The related results will be presented and intensively illustrated in the current paper.
Moreover, we will justify that this strategy, used in an advanced form in MSG program, can be applied in any situation following this general observation scheme, whatever the instrument (telescope, SAR), the platform (airborne, probe) or the planet (Earth, solar system planet, comet or asteroid) is.

Keywords: geometric calibration, spaceborne imaging spectrometer, multispectral, images registration, resampling, correlation, geolocation.

1. INTRODUCTION AND SCOPE

This report presents the synthesis of the results of the study of geometric calibration for the PRISM (Processes Research by an Imaging Space Mission) pre-phase A study achieved by ALCATEL (formerly Aerospatiale) under ESA contract n° 12057/96/NL/CN.
PRISM is a high resolution spectro-imager candidate to Earth Explorer, post ENVISAT mission.
PRISM provides high resolution images (50 m in nadir view) in several spectral bands, ranging from blue visible (0.45 mm) to thermal IR (12.3 mm). Its mission primarily aimed at the investigation of interaction processes between the land and the atmosphere.

The goal of the study is to develop a highly accurate methodology to co-register the data of a spaceborne spectroradiometer encompassing the optical range of the electromagnetic spectrum to any available and well defined mapping reference system. The approach includes correction of the platform pointing errors, instrument misalignments, and effects due to the influence of the topography and viewing angles which causes changes in pixel sizes.

Here is the Study Plan:

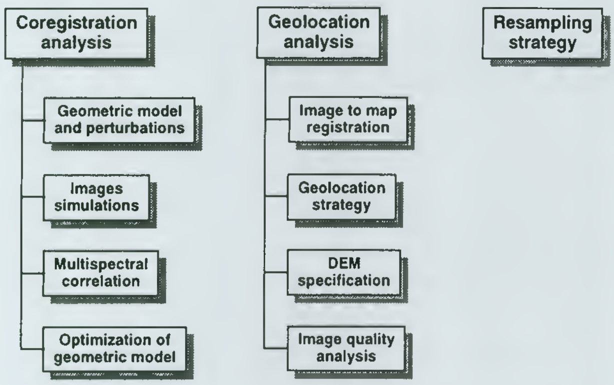


Figure 1: ***Study Plan***

The study is composed of two major tasks: « Coregistration Analysis » and « Geolocation Analysis », plus a task entitled « Resampling Strategies » which interacts with the two previous ones.
The goal of Coregistration Analysis is to develop a methodology allowing to co-register the images taken in different spectral bands - VNIR, SWIR and TIR - so as to get them in the same geometry.
The goal of Geolocation Analysis is to develop a methodology allowing to register the images w.r.t. ground so as to get them in a ground reference frame and make them exploitable for scientific applications : BRDF measurement, input to sub-mesoscale GCMs and SVAT models.
The goal of Resampling Analysis is to investigate resampling strategies and algorithms allowing to best preserve image quality for the registered images.

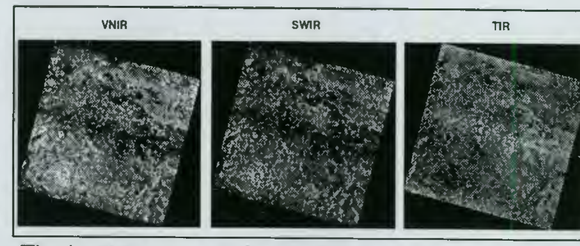
The complete results of the study -including description of algorithms, software tools, reference documentation - are extensively presented in the Final Report (Ref. 1).

2. GEOMETRIC CALIBRATION OF PRISM IMAGES

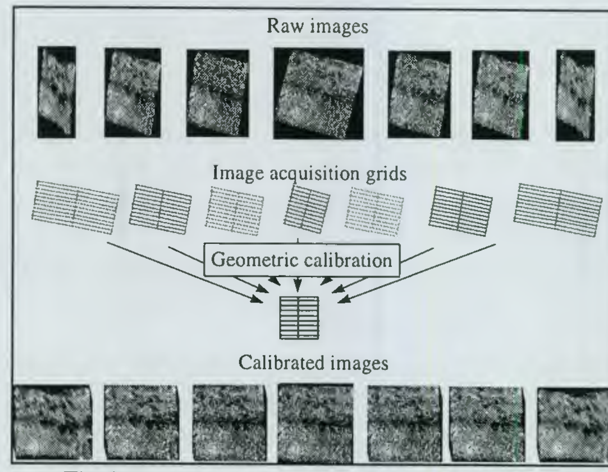
2.1 Introduction

One of the major task of the ground processing is to elaborate the data products that will be delivered to the scientific users. This involves the processing of the raw data to perform both radiometric and geometric calibration.
The current study has focused on the second aspect, since it is an important and innovative feature of the PRISM images processing, due to both the multi-spectral and multi-directional characteristics of the information acquired by the satellite.

2.2 The issues of geometric calibration of PRISM images



The images are acquired in different spectral regions, and might be mis-registrated



The images can be acquired from very different directions, leading to strongly different acquisition spatial grids and spatial resolutions (example of BRDF acquisition).

Figure 2: ***Issues of PRISM images geometric calibration***

2.3 The geometric calibration processes

Several processes are involved in the geometric calibration :
- **Correlation** : automated measurement of relative shifts between two images by feature recognition ;
- **Geometric modeling** : building of a dynamic image acquisition model so as to correct geometric image distortions and determination of an optimized geometric model taking into account the correction of these distortions ;
- **Resampling** : projection of the image in a new reference frame, usually a ground reference co-ordinate system,
- **GCP (Ground Control Points) extraction** : determination of the location of control points between an image and a ground map.

The current study of PRISM geometric calibration aims at investigating and evaluating the geometric calibration processing chain of PRISM images.

3. COREGISTRATION ANALYSIS

3.1 Coregistration overview

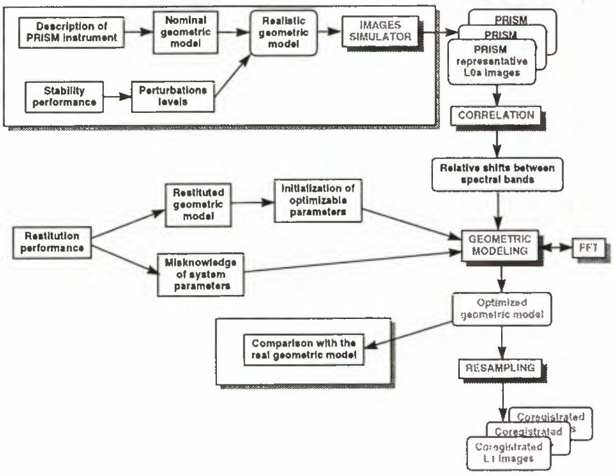


Figure 3: ***Coregistration overview***

From the description of PRISM instrument, we build a **geometric model** of PRISM image acquisition. It expresses the relation between each pixel of an image and the corresponding line of sight vector (position of the satellite center of mass, and viewing direction). This way we have a nominal geometric model of image acquisition.

In reality, there are **perturbations** on the orbit, the attitude angles and the position of detectors in the focal plane. These perturbations may change from one image acquisition to another and are not predictable. For a preliminary study, the characteristics of these perturbations can be estimated thanks to the stability performance budget of the satellite, including orbit and attitude stability performance, and focal planes alignment accuracy.

So, we get a « realistic » geometric model that is used for the simulation of PRISM representative images in three channels: VNIR, SWIR, TIR.

In the operational phase of the satellite mission, we know the characteristics of the nominal geometric model for each image acquisition and our objective is to co-register the images taken in the three spectral channels, i.e. to get them in the same geometry.

Unfortunately, this nominal model does not allow a satisfying geometric calibration (otherwise the study existence would have no justification).

In order to improve the model precision, we need supplementary information: these will be gathered by **correlation process** between two channels. Correlation aims at measuring the relative distance between two images by feature recognition.

An **optimization** process is required to match a new set of model parameters values with these supplementary data. In output we get the most suitable physical model matching with the observed image distortions: we call it optimized geometric model.

To test the validity of the optimization process, we compare this optimized model to the real model used for the images simulations: the differences will be a combination of correlation errors and optimization errors.

Eventually, with the optimized geometric model, we are able to resample the SWIR and TIR images in the VNIR image geometry. Thus, we get coregistrated images.

3.2 Geometric model

3.2.1 Mathematical form of the geometric model

A geometric model is not an image but a set of geometric parameters describing: position of the arrays of detectors in the focal plane, attitude angles, orbital motion of the satellite, time of image acquisition. Thus, for each pixel (i, j) of an image, we associate the corresponding line of sight vector (position of the satellite center of mass, and viewing direction). **If we combine one image and its associated geometric model, we are able to resample this image in any given reference frame.**

We have used **polynomials** (resp. sinus functions) to mathematically represent these geometric parameters, in order to focus on static perturbations (resp. dynamic perturbations).

3.2.2 PRISM designs

Two configurations are considered:

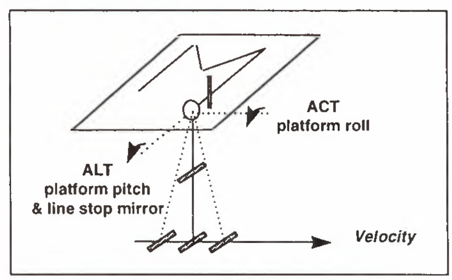


Figure 4: ***« In-plane » Design #1***

ACT depointing is fully performed by platform roll.

ALT depointing is performed by platform pitch and a line stop tilting (in-plane) mirror.

O

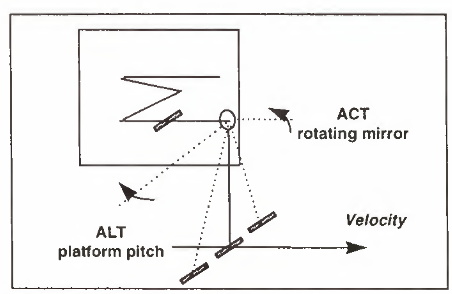


Figure 5: ***« In-beam » Design #2***

ACT depointing is fully performed by a rotating (in-beam) mirror.

ALT depointing is performed by platform pitch.

These concepts are respectively detailed in Aerospatiale (Ref. 5) and DASA (Ref. 4) pre-phase A studies.

3.2.3 Optimizable parameters

For the **coregistration analysis**, the optimizable parameters are: coefficients of the polynomials describing the position of the SWIR and TIR linear arrays of detectors in the focal plane, and coefficients of the polynomials (or sinus functions) describing the temporal laws of roll, pitch and yaw.

3.3 IMAGES SIMULATIONS

3.3.1 Selection of input data

The principle of image simulation is to get first an <u>input image</u> of higher quality than the one of the simulated system.
As input image we have used a set of MIVIS images, provided by CNES and INRA. The images cover a 35 × 10 km area, with 5 m ground spatial resolution. This is a low relief region.
<u>Remark :</u> Owing to the wide range of PRISM spectral bands (VNIR, SWIR, TIR) and to its spatial resolution and large swath width (50 km in nadir view), it was necessary to process smaller images because no other input data were available.
For the future phases of the program, if simulations are needed (for instance to make trade-offs between image acquisition modes), it would be necessary to plan a PRISM dedicated image acquisition campaign with a suitable airborne sensor (DAIS or MIVIS for instance).

3.3.2 Definition of study cases for Coregistration Analysis

<u>3.3.2.1 Spectral bands</u>

We have selected three MIVIS bands which are close to PRISM ones: 0.433 to 0.453 µm (VNIR), 2.320 to 2.328 µm (SWIR), 8.20 to 8.60 µm (TIR).

<u>3.3.2.2 Satellite configurations</u>

Yaw steering is used in order to get rectangular images by compensating for Earth rotation.
We have introduced **4 study cases** : case 1, dealing with both designs (since there is no mirror rotation), cases 2 & 3, specific to « in-plane » design, and case 4, specific to « in-beam » design.

Case 1 represent the nominal pointing configuration (nadir view), Case 3 is the extreme pointing configuration (30° ACT and 50° ALT), and Case 2 is an intermediate configuration (30° ACT).
Case 4 also corresponds to 30° ACT and 50° ALT, but ACT depointing is provided by in-beam mirror which induces image rotation that is compensated by yaw steering.

<u>Remark :</u> This case is an extreme pointing configuration for the instrument (30° is the maximum rotation angle for the in-beam mirror) and for the platform (50° pitch is the maximum value) but this is not the worst case for the image geometry.
Indeed, case 4 does not lead to equivalent ground elevation angles compared with case 3.

<u>3.3.2.3 Orbital motion</u>

The orbit is heliosynchronous with 772 km altitude and 11:00 LTDN.

<u>3.3.2.4 Focal plane geometry</u>

The geometry of the focal plane is (see figure 6):
- Regular distribution of the detectors in the VNIR region, considered as our reference band.
- Rotation of the SWIR array of detectors with respect to the VNIR array of detectors so as to get a 0.2 pixel shift at the extreme detector. This corresponds to a 0.2 mrad rotation.
- Translation of 30 pixels in both x and y directions, plus rotation (same as above but in the reverse direction) for the TIR array of detectors with respect to the VNIR one.

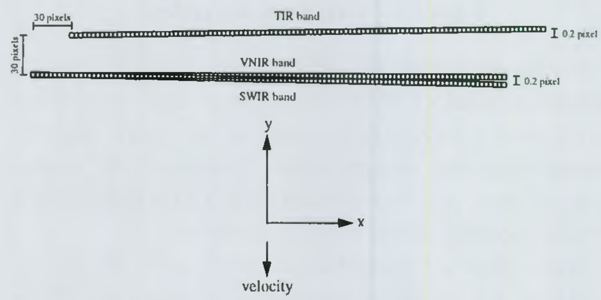


Figure 6: ***Focal plane geometry***

The rotations are within ESA specifications (0.2 pixel misalignment between SWIR and VNIR) and the translation is a conservative value based on optical alignment accuracy between TIR and VNIR focal planes

We have also assessed another focal plane geometry:
- Regular distribution of the detectors in the VNIR region, considered as our reference band.
- Rotation of the SWIR array of detectors with respect to the VNIR array of detectors so as to get a 10 pixel shift at the extreme detector. This corresponds to a 10 mrad rotation.
- Translation of 30 pixels in both x and y directions, plus rotation (same as above but in the reverse direction) for the TIR array of detectors with respect to the VNIR one.

The rotations are beyond the specifications but we feel that it is interesting to test our coregistration process in severe conditions. So, we will be able to **determine whether alignment specifications could be relaxed or not**

<u>3.3.2.5 Perturbation levels</u>

We have focused on **AOCS pointing error** (low frequency, periodic):
1> amplitude = 10 µrad on roll and pitch axes, at 1 Hz frequency, in phase quadrature.
2> amplitude = 4 µrad on roll and pitch axes, at 5.5 Hz frequency, in phase quadrature.

N.B. The above-mentioned values are intended to be **worst case values**.

And yet, we have also considered worse values: amplitude of 100 µrad and 40 µrad.

3.3.2.6 Point Spread Function

MTF of optics, detectors and smearing are taken into account.

3.3.2.7 Noise Model

The radiometric noise NEdL is taken into account in the simulation process.

3.3.2.8 Summary

During the definition of the study cases, we have introduced two perturbation levels:

- **Nominal perturbations**: rotations of 0.2 pixels between the focal planes, amplitude of 10 µrad and 4 µrad for attitude pointing errors,
- **Maximum perturbations**: rotations of 10 pixels between the focal planes, amplitude of 100 µrad and 40 µrad for attitude pointing errors.

3.4 Multispectral correlation

3.4.1 Selection of correlation algorithm

Correlation aims at measuring the relative shifts between two images. This is performed by feature recognition and requires an interpolation algorithm often named "correlation algorithm".
We have chosen the algorithm based on **Shannon polynomial interpolation**, which does not make hypotheses about the similarity of radiometric levels between the images.
References about correlation algorithms can be found in (Ref. 14) and (Ref. 15).

3.4.2 Statistic performance budget

Nadir view, Nominal perturbation levels:

		Correlation				Reality	
		Mean shift	Sigma	Abs Max shift	Rejected points	Mean shift	Abs Max shift
SWIR	column	0,02	1,00	8,98	3,7%	0	<0,01
	row	-0,14	1,04	9,00		-0,1	0,2
TIR	column	0,03	1,89	8,98	16,5%	0	0,21
	row	-0,07	1,78	8,99		0,1	0,41

The results for TIR/VNIR already include a mean 30 pixel shift in row and in column.
The results for nominal perturbation levels clearly show that **correlation does not work as well as for monospectral images**. Indeed, the real shifts in row and column should be under 0.5 pixel, whereas the maximum shift measured by correlation is nearly 9 pixels, and the dispersion (sigma) is about 1 pixel.
This is confirmed by the **histogram** of correlation shifts which show dispersion of correlation values:

- For the SWIR band, in nadir view, nominal perturbation levels:

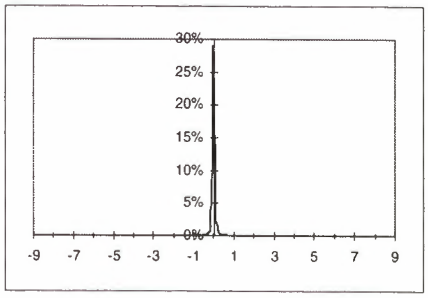


Figure 7: ***SWIR band, Shift in column***

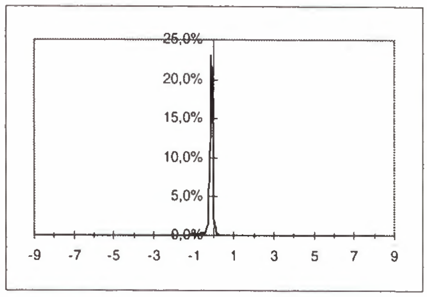


Figure 8: ***SWIR band, Shift in row***

- For the TIR band, in nadir view, nominal perturbation levels:

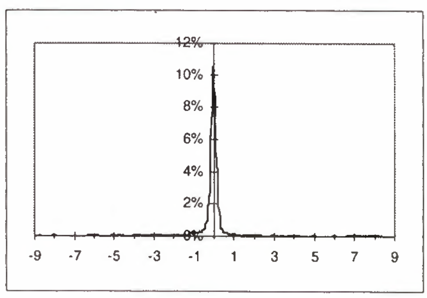


Figure 9: ***TIR band, Shift in column***

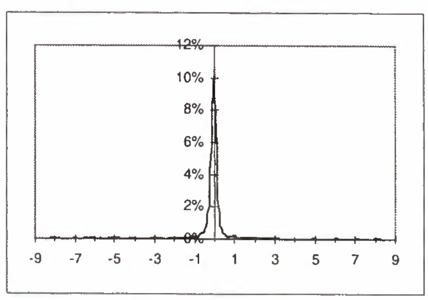


Figure 10: ***TIR band, Shift in row***

The results are the same with maximum perturbation levels: the proportion of bad correlation points is important.

3.5 Optimization of PRISM geometric model

3.5.1 Hypotheses

The influence of relief is **small in all cases**. This is a consequence of the moderate ground resolution of PRISM (50

m) and of the short distance between the two focal planes (about 30 pixels).

Each linear array of detectors is considered as a straight line: there is **no warping**. So, the shifts between focal planes will be a combination of translations and rotations.

We try to determine drifts, accelerations or sinus terms for the attitude angles. This does not correspond to real attitude laws, which usually include noise and transient vibrations, but it is sufficient for a preliminary study. Moreover, any periodic perturbation can be decomposed as a sum of sinus functions (Fourier series), and so the results obtained for sinusoidal perturbations will be potentially applicable to any periodic perturbation.

3.5.2 Influence of image size

For our images simulations, our ground truth was provided by a mosaic of MIVIS images. The problem is that the size of the mosaic is lower than the size of a PRISM image, even in nadir view. So, we have been able to simulate only a part of a full PRISM image.

3.5.3 Optimization strategy

There exist two major concepts to achieve coregistration of images: one based on the sheer exploitation of the correlation results (see Refs. 6, 7, 8) and another consisting in building a physical model of image acquisition (see Refs. 9, 10).

The first concept uses a deformation model between the two images (usually polynomial), but there is no physical link between the coefficients of this deformation model and the parameters of image acquisition. Moreover, this concept leads to interpolations of the correlation shifts in the areas where the correlation do not work, whereas a physical model gives the position of all pixels, without any interpolation.

The optimization is based on the minimization of a « **distance** » **criterion** between the current geometric model and the reference model.

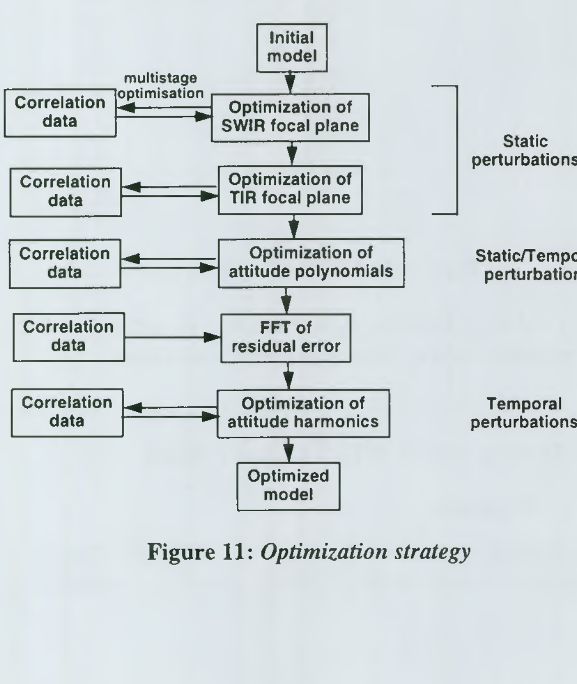


Figure 11: *Optimization strategy*

The optimization falls into several steps:

- Determination of focal plane misalignments (position of SWIR focal plane and then TIR focal plane),
- Determination of attitude static perturbations (drift and acceleration in roll, pitch, yaw),
- If needed, determination of attitude periodic perturbations, performed by FFT processing of the residual errors.

At each step a **multi-stage optimization** has been used in order to discriminate between good correlation points and bad correlation points: first we compute an « optimized » geometric model, taking into account the whole correlation results; then, we eliminate the dubious points, i.e. points that are far from the mean deviation.

3.5.4 Optimization results

<u>3.5.4.1 Comparison between the optimized model and the real model</u>

<u>In nadir view, nominal perturbation levels:</u>

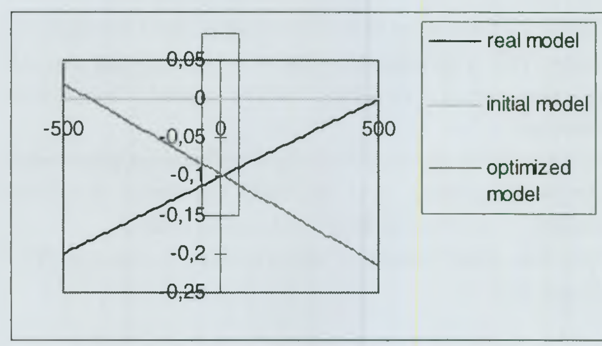


Figure 12: *SWIR focal plane*

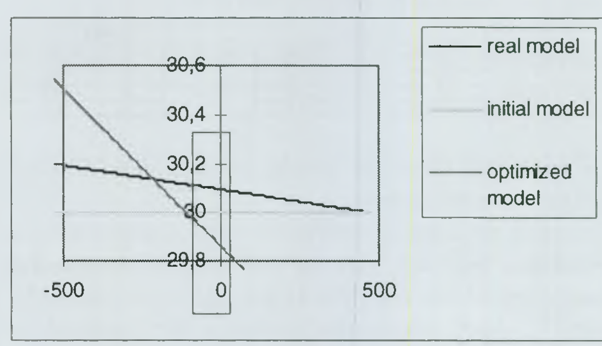


Figure 13: *TIR focal plane*

N.B. The X axis is parallel to the VNIR array of detectors and the numbers (-500 to 500) represent the index of the detectors. Both X and Y axes are graduated in pixels. The box represent the accessible area, i.e. the area where correlation shifts are available.

The rotations are not found because their amplitudes over the accessible area (115 pixels) are very small (**0.02 pixel**) compared to **correlation errors** (about **0.1 pixel**). Moreover, the maximum error, calculated in the accessible area, is relatively small: less than 0.2 pixel.

If the objective is to retrieve such tiny rotations, two solutions are possible: either develop a new correlation algorithm suited for PRISM multispectral images, or coregister many sets of PRISM images and compute the mean rotation, provided a statistical trend emerge.

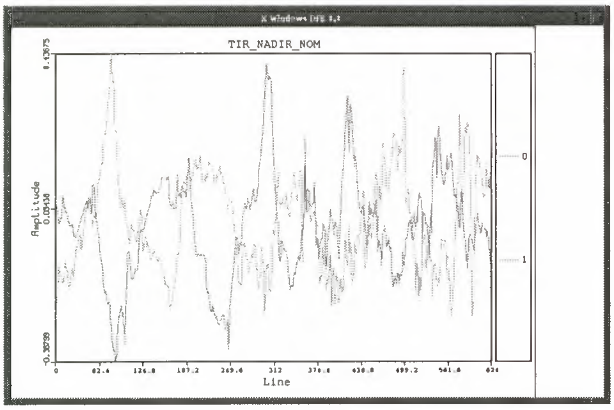


Figure 14: ***Residual errors***

N.B. Index 0 stands for row and index 1 stands for column. One can guess a sinus term through a background noise.

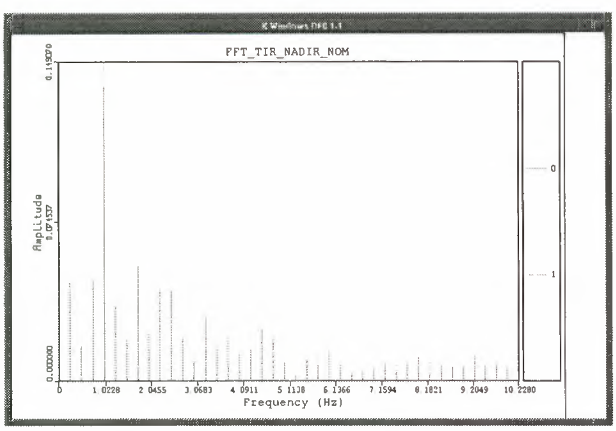


Figure 15: ***FFT of residual errors***

We get a perturbation of frequency close to 1 Hz, i.e. the frequency of one of our attitude perturbations.
The FFT algorithm used here (Bidimensional Fast Fourier Algorithm) is described in (Ref. 13).

Attitude angles:

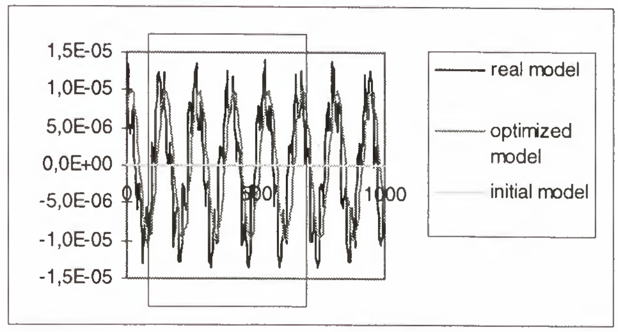


Figure 16: ***Attitude angles, roll***

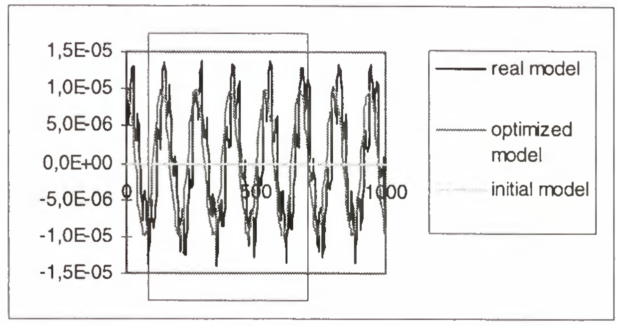


Figure 17: ***Attitude angles, pitch***

The optimized attitude laws are close to the real attitude laws. The mean error on the attitude angles is lower than 0.4 μrad, i.e. 4% of the amplitude of the sinus term (10 μrad).

<u>In nadir view, maximum perturbation levels:</u>

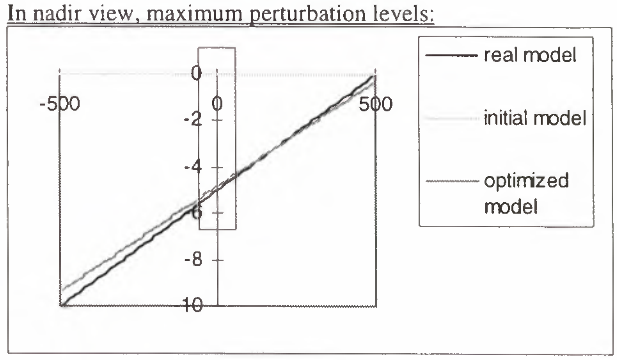


Figure 18: ***SWIR focal plane***

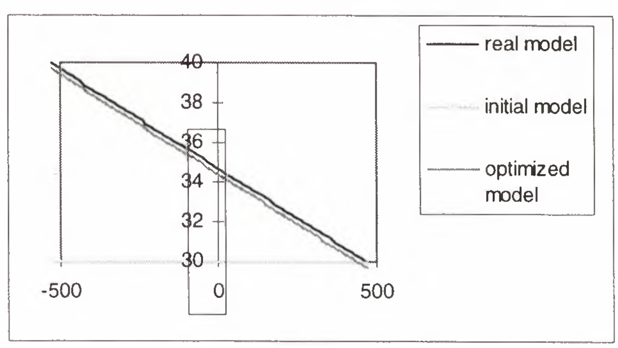


Figure 19: ***TIR focal plane***

The rotations are more precisely determined because their amplitudes are higher than correlation bias.

The error is slightly smaller for the SWIR focal plane (0.21 pixel) than for the TIR focal plane (0.3 pixel) because the attitude perturbations are less important between VNIR and SWIR focal planes. Both translations and rotations are found because their amplitudes over the accessible area (115 pixels) are higher (1.15 pixel) than correlation errors.
For the determination of attitude temporal perturbations, the results are the same as for nominal perturbations level: the sinus term is retrieved and the difference between the optimized model and the real model is small.

3.5.4.2 Influence of off-nadir viewing

These are the results for case 2 (30° ACT) for SWIR/VNIR coregistration, for maximum perturbation levels :

	Parameter	Unit	Real value	Initial value	Variance	Optimized value	Difference	Diff. (%)
SWIR	Translation along X	pixel	0	0	10	0,085	**0,085**	
	Translation along Y	pixel	0	0	10	0,118	**0,118**	
	Rotation	pixel/115 pixels	**1,15**	**0**	**10**	0,909	-0,241	**-20,9%**

The errors on the translations between focal planes are about 0.1 pixel and are lower than for case 1 (0.2 pixel). This could be a consequence of the fact that an error on focal plane position is more easily retrieved when there is depointing.

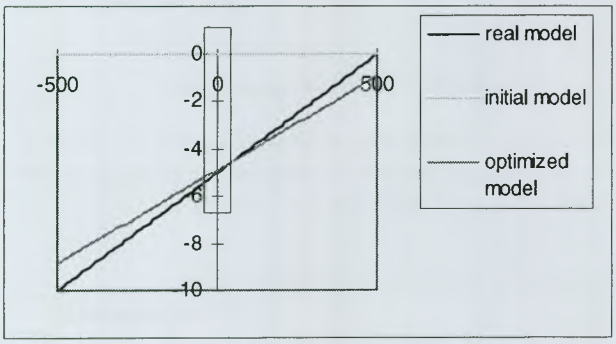


Figure 20: ***SWIR focal plane***

The error on the retrieval of the rotation is doubled, because the amplitude of the rotation (0.6 pixel) is divided by two within the accessible area (60 pixel) and is closer to the amplitude of correlation errors.
If the sizes of the images were bigger, the retrieval of the rotation between focal planes would be more accurate.

For the attitude angles, the results are similar to the results obtained in case 1 (nadir view).

3.5.5 Operational scenario

Before launch, a rough estimate of the shifts between focal planes (+/- 10 pixels) is necessary.
The correction of static perturbations (misalignment between focal planes) is necessary during the early flight check-up. Then, this may be done once or twice a year, for calibration purpose.
The correction/determination of temporal perturbations could be used for technological purpose (determination of vibrations), if required.

The coregistration process is composed of 3 sub-processes: correlation, optimization and resampling, if required. The most time-consuming process is the optimization, that represents more than **90%** of the overall running time. The overall running time is estimated to **one hour processing time** for full PRISM images. So, all things considered, Coregistration process will need time and cannot be achieved in real-time as a routine mode for all images.

4. GEOLOCATION ANALYSIS

4.1 Geolocation overview

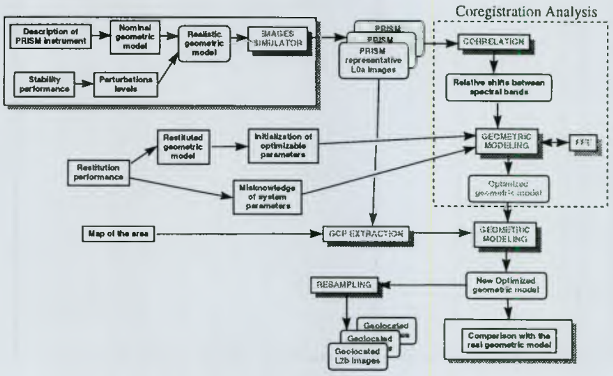


Figure 21: *Geolocation overview*

The principle is the same as for Coregistration Analysis.
In input we have a geometric model of PRISM image acquisition (derived from description of PRISM instrument, stability performance budget of the satellite). We can use the information provided by the Coregistration Analysis - focal plane misalignments, attitude drifts, attitude periodic perturbations - so as to refine this geometric model. This way, there will be a synergy between Coregistration Analysis and Geolocation Analysis.

In order to improve the model precision, we need supplementary information: these will be gathered by determination of « control points » (often called Ground Control Points or GCPs) between the image and the map : **GCP extraction**. The Ground Control Points are located within the map and within the image. Instead of a map, it is also possible to use another image that has been already geolocated : the « control points » are determined between the two images and are named tie points. This GCP extraction can be either an automated process or a man managed process.

An **optimization** process is required to match a new set of model parameters values with these supplementary data. In output we get the most suitable physical model matching with the observed image distortions: we call it optimized geometric model.
To test the validity of the optimization process, we compare this optimized model to the real model used for the images simulations: the differences will be a combination of GCP extraction errors and optimization errors.
Eventually, with the optimized geometric model, we are able to resample the VNIR image in the geometry of the map.

4.2 IMAGES SIMULATIONS

4.2.1 Selection of input data for Geolocation Analysis

We have used a set of data (image+map+DEM) provided by the IGN (French Geographical Institute). The area is a **high relief area**, as requested by ESA: about 1100 m elevation difference. The horizontal resolution of the DEM is 16 m and the area covered by the ortho-image and the DEM is about 15*15 km.

N.B. The relief of the 20 sites selected for LSPIM is usually moderate (less or equal to 300 m elevation difference) except for 3 sites with high relief (about 1000 m elevation difference): Anchorage, Ushuaia and San Sebastian. So, the selected case corresponds to a **worst case w.r.t. LSPIM standards**.

4.2.2 Definition of study cases for Geolocation Analysis

The main differences w.r.t. Coregistration Analysis are:

<u>4.2.2.1 Spectral bands</u>

The VNIR spectral band is close to SPOT panchromatic channel:0.4 to 0.7 µm, 0.3 µm wide.

<u>4.2.2.2 Perturbation levels</u>

In the definition of the study cases, we have used two sets of perturbation levels, depending on the use or not use of **GNSS/GPS based systems**:

- With GNSS/GPS based systems: 50 m error on latitude, longitude and radius (**orbital position restitution**) and 1 ms error on time of image acquisition (**datation restitution**),
- Without GNSS/GPS based systems: 500 m error on latitude, longitude and radius (**orbital position restitution**) and 10 ms error on time of image acquisition (**datation restitution**).

4.3 Image to map registration

4.3.1 Extraction of GCPs

A Ground Control Point (GCP) is a tie point between an image and a ground map.

For our study, the extraction of GCPs was a visual man managed process, but in the operational phase, this process should be automated by using GIS databases.

Within our high relief area (about 1100 m elevation difference), we have selected 26 GCPs at a wide range of altitudes (between 80 and 1085 m) and in areas where the map is geometrically reliable: forest edges, river banks, small crossroads, hilltops, bridges.

The amount of GCPs corresponds to a reasonable value for a visual man managed GCP extraction over a high relief area in a reasonable time (about one hour). If more GCPs were required, then it would be necessary to use an automated GCP extraction process.

A lower number of GCPs could probably be accepted for low or moderate relief areas.

4.3.2 Accuracy assessment

This is a tentative error budget related to the GCP location:

<u>Errors related to the map :</u>

- Specific error due to the map inherent misaccuracy: 30 m, value provided by IGN for 1/25000 scale maps,
- Location error while reading the map: 2 mm on the map, i.e. 50 m on ground.

<u>Errors related to the image :</u>

- Location of the GCP in the image: between 2 pixels (case 1 : nadir view) and 15 pixels (case 3 : ALT+ACT depointing, « in-plane » design), i.e. between 100 m (case 1) and 750 m (case 3). The difference is due to the poor contrast restitution of the image in case 3 : the image is blurred.

So, if we combine linearly the two kinds of errors, the GCP location accuracy shall range between **180 m** (case 1) and **830 m** (case 3). These **worst case values** will be consolidated in section 6 with a real quantitative GCP extraction error budget. One can notice that most of the error is related to the location of the GCP within the image, especially for extreme pointing configurations.

<u>Remark :</u> The errors committed while reading the map and while locating the GCPs in the image could be reduced thanks to an automated GCP extraction process, based on the correlation between the image and a digitized map or a GIS database or another previously geolocated image. The comparison between a visual man managed GCP extraction process and an automated GCP extraction process could be carried out in phase A studies.

4.4 Geolocation strategy

4.4.1 Optimization strategy

The principle is the same as the one used in the Coregistration Analysis. The optimization is based on the minimization of a « distance » criterion between the current geometric model and the real geometric model. The main difference is that this function (called here **mean residual error**) uses GCP location instead of correlation data.

The optimizable parameters are: bias in roll, pitch, yaw, position of VNIR focal plane, time of acquisition, orbit.

The uncertainties on the values of these parameters have been estimated:

- Attitude bias: 1 mrad for roll and pitch, 3 mrad for yaw,
- Position of VNIR focal plane: 0.2 pixel,
- Time of acquisition: 10 ms without GNSS, 1 ms with GNSS,
- Orbit: 500 m on co-ordinates x, y, z without using GNSS, 50 m with GNSS.

For reasons detailed in [RD 3], we have focused on the **attitude bias**.

4.4.2 Optimization results

4.4.2.1 Results

The real (resp. initial, resp. optimized) mean residue is the mean residual error on the location of the GCPs obtained with the real (resp. nominal, resp. optimized) geometric model of PRISM image acquisition.

The following graphs synthesize the results for the 4 study cases, with/without GNSS, with 26 or 10 GCPs :

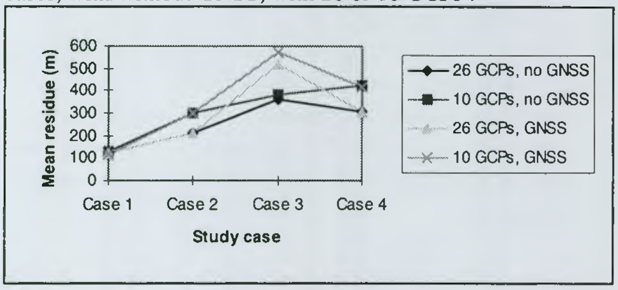


Figure 22: ***Mean residue after optimization at ground level (metres) = Geolocation accuracy in metres***

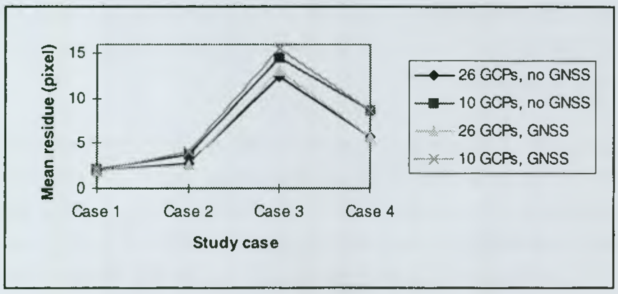


Figure 23: ***Mean residue after optimization at image level (pixels) = Geolocation accuracy in pixels***

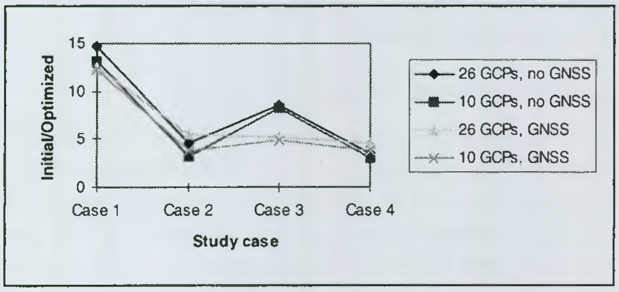


Figure 24: ***Optimization gain =Geolocation accuracy improvement***

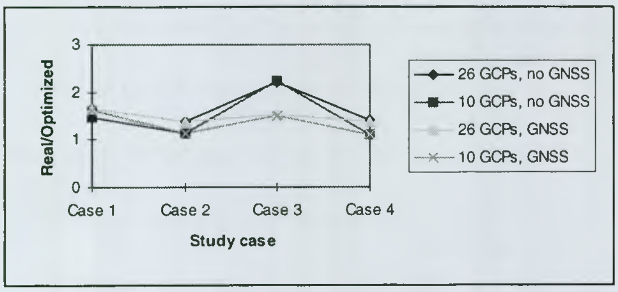


Figure 25: ***Comparison between optimized residue and real residue***

The real mean residues are coherent with the rough estimates given in the **GCP extraction accuracy** budget in section 4.3.2. Indeed, for case 1 we obtain 199 m instead of 180 m and for case 3 we obtain 793 m instead of 830 m.

The **geolocation accuracy** ranges between 120 m and 570 m, and between 2 and 15 pixels. The mean geolocation accuracy is roughly 300 m.

The geolocation error (expressed in metres or in pixels) increases when we pass on from nadir view (case 1) to moderate pointing configurations (case 2) and extreme pointing configurations (case 4 and case 3).

There is no significant difference in terms of geolocation accuracy with or without **GNSS**.

The geolocation error (expressed in metres or in pixels) decreases when we pass on from 10 GCPs to 26 GCPs, especially for high depointing configurations. Indeed, the more numerous **GCPs** are, the more tie points we have and the more precise the geolocation will be.

The **geolocation accuracy improvement** allowed by the optimization process varies between 3,5 (cases 2 & 4) and 14 (case 1).

The geolocation accuracy (=optimized residue) is of the same order of magnitude as the GCP extraction accuracy (=real residue) : the ratio real residue/optimized residue ranges from 1.1 to 2.2.

This means that the optimization works.

4.4.3 Comparison with the real geometric model

4.4.3.1 Results

The difference between the optimized value of attitude bias and the real value of attitude bias represents the **attitude bias estimation error**.

The following graphs synthesize the results for the 4 study cases, with/without GNSS, with 26 or 10 GCPs :

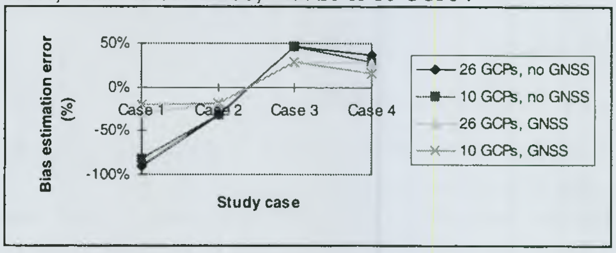


Figure 26: ***Comparison with the real geometric model: Roll bias***

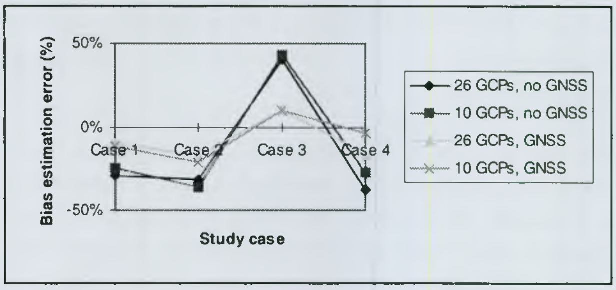


Figure 27: ***Comparison with the real geometric model: Pitch bias***

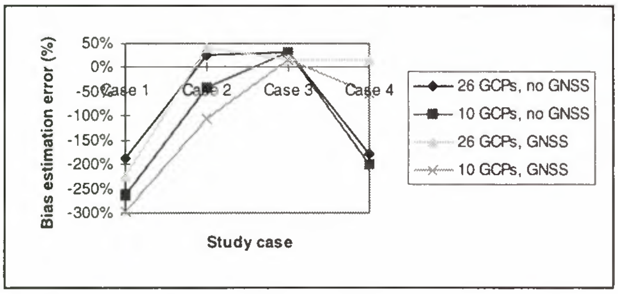


Figure 28: ***Comparison with the real geometric model: Yaw bias***

4.4.3.2 Comparison of roll, pitch, yaw

The roll and pitch bias estimation error are always lower than 90 % whereas the yaw bias estimation error exceeds 150% 6 times out of 16. Moreover, for a given pointing configuration, the yaw bias estimation error is often higher than the roll and pitch bias estimation errors.

This is a consequence of the fact that the simulated images are small (about 150×150 pixels) compared to a full PRISM image (1000×1000 pixels) and correspond to the centre of the linear array of detectors. Indeed, in the centre of the field, a yaw variation has a low impact on the image geometry whereas in the edge of the field, a yaw variation has an impact comparable to the impact of a roll variation or a pitch variation.

4.4.3.3 Influence of the pointing configuration

The roll and pitch bias estimation error does not vary significantly w.r.t. the pointing configuration. Nevertheless, the yaw bias estimation error is higher for nadir view (case 1) than for off-nadir view and can exceed 200 %. This is a consequence of the small amount of GCPs and of the fact that the simulated image is small : 160*160 pixels at PRISM resolution in nadir view.

4.4.3.4 Influence of GNSS

The attitude bias estimation error is significantly lower with GNSS/GPS based systems, especially for roll and pitch, except yaw in case 1 (nadir view). Indeed, with GNSS, the accuracy on orbital position restitution (50 m) and datation restitution (1 ms) is high so that the only significant biases to be recovered are attitude biases. On the other hand, without GNSS, The uncertainties on the orbit position (500 m on the 3 axes) and on the time of acquisition (10 ms) are mixed with the attitude bias and are considered as attitude biases in the optimization phase.

4.4.3.5 Influence of the amount of GCPs

The attitude bias estimation error is not significantly better with 10 GCPs or with 26 GCPs.

To notice an improvement of the results, we think that the amount of GCPs shall be higher (for instance 100 to 200 GCPs) in order to recover tiny attitude biases and to discriminate between orbit position errors, datation errors, roll bias, pitch bias, yaw bias. This requires an automated extraction process of GCPs in the map and in the images.

4.4.3.6 Comparison with Coregistration Analysis

For Geolocation Analysis, the optimized model is not as close to the real model as for the Coregistration Analysis. This is a consequence of the following factors:

- GCP location error: several pixels instead of 0,2 pixel for the coregistration analysis,
- Small amount of GCPs: 26 instead of roughly 100000 for the coregistration analysis: we lack statistical data,
- Several triplets of roll, pitch, yaw can closely correspond to a given line of sight,
- The uncertainties on the orbit position (500 m on the 3 axes without GNSS) and on the time of acquisition (10 ms without GNSS) are taken into account in the attitude bias.

4.4.4 Influence of image size

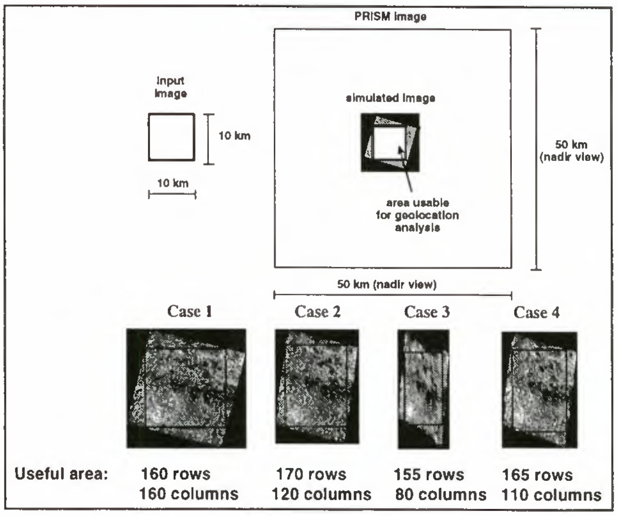


Figure 29: ***Influence of image size***

For geolocation purpose, the fact that the simulated images are smaller than real images is not as critical as for the Coregistration Analysis. Indeed, geolocation can be carried out over a small area provided there are enough tie points between the map and the image.

Nevertheless, a small image does not allow to accurately determine attitude biases and to refine the geometric model of image acquisition, as noted in the previous section.

So, in the future phases of the program, simulations of full PRISM images would be necessary in order to assess the amount of GCPs needed to geolocate the full images and to determine attitude biases with good accuracy.

4.5 DEM specification

4.5.1 Experimental study

A set of raw simulated images have been resampled using the associated optimized model (output of Geolocation Analysis) and different DEM (with increasing resolutions: 16, 32 and 64 m) over a high relief area (1100 m elevation difference).

ALCATEL has analyzed the residual geometric distortion (RMSE and SNR) between the resampled image and the original orthophoto, for study cases 1 and 2, with the 3 DEM resolutions:

RMSE

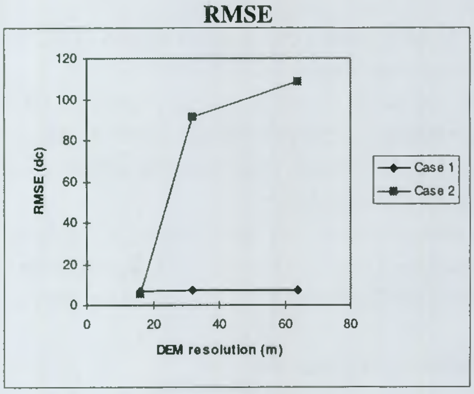


Figure 30: ***RMSE***

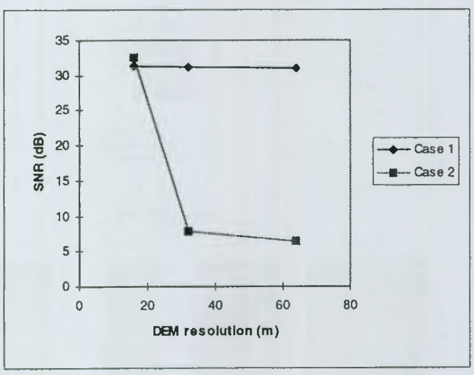


Figure 31: ***SNR***

In nadir view, since the field of view (FOV) is limited, the resolution of the DEM has a small impact on the quality of the resampled image. Indeed, the FOV of our simulated images (115 pixels) is small: about 0.55 deg around nadir direction. The SNR is higher than 30 dB and the quality of the images is good.

On the other hand, with off-nadir depointing, the impact of relief is important and the resolution of the DEM has a large impact on image quality. Indeed, the SNR obtained with a 32 and 64 m DEM resolution are poor : less than 8 dB, i.e. the ratio signal/noise is only 10^(8/20)=2,5. The resampled image is very degraded.

The RMSE and SNR obtained with a 16 m DEM resolution are nearly the same for case 1 and case 2. This is a consequence of the fact that the original DEM used for the simulation had a 16 m resolution. So, the only error that is measured is a resampling error and since the nadir image and the ACT depointing image are not very different from a radiometric point of view, the resampling error shall be nearly the same for the two images.

Remark :** **Simulations with a more precise DEM (for instance 1 or 2 m) would be necessary in the future phases of the program in order to carefully assess the impact of DEM for resolutions ranging between 5 and 20 m and to generalize the results of the current preliminary study.

4.5.2 Recommendation

We recommend a **20 m DEM resolution** in order to get resampled images of good quality **over high relief areas with off-nadir depointing**. This is really a **worst case** for image acquisition. It is clear that for low relief areas or for nadir view a DEM with a coarse resolution (roughly 50 m to 100 m) would be sufficient enough.

Since most of LSPIM sites correspond to moderate relief area, it would be interesting to assess the influence of DEM resolution on image quality for such kind of area (typically with 300 m elevation difference). Thus, we could make recommendations for nominal cases of image acquisition.

4.6 Impact of off-nadir viewing on image quality

4.6.1 Image quality parameters

- Image location: accuracy on the location to image pixels w.r.t. a ground reference frame, e.g. a map of the area.
- Length distortion: alteration of the distance between two ground points, measured in the images (in pixels) and on ground (in metres).
- Modulation Transfer Function: Fourier Transform of the impulse response of the instrument and is a good index of image quality, especially contrast restitution.

4.6.2 Performance budget computation

		Study case			
Parameter	Unit	Case 1	Case 2	Case 3	Case 4
Image location	m	1766,3	2415,4	8722,6	3528,3
Length distorsion	pixel	0,340	0,547	0,914	0,627
Along Track MTF	-	0,332	0,250	0,000	0,014
Across Track MTF	-	0,522	0,522	0,522	0,522

These results are based on analytical formulae and do not take into account the ground processing that represent the core of the study (coregistration, geolocation).

For the extreme pointing configurations (case 3 and case 4), the image location error is important and the need for ground processing is obvious.

The fact that case 4 seems better than case 3 is related to the pointing configurations (difference between platform maneuver and in-beam mirror rotation, cf. section 4.6.3).

The length distortion varies significantly with the pointing configuration but it is lower than 1 pixel/300 pixels, i.e. 0.3%.

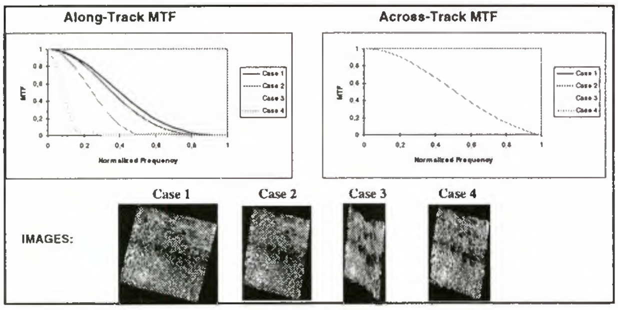


Figure 32: *MTF*

The MTF is significantly degraded for the extreme pointing configuration (case 3) : the corresponding image is blurred. There exist two ways to correct it: either adapt the sampling frequency to the ground size of the pixel, or control satellite pitch during the image acquisition (pitch steering).

4.6.3 Comparison of the two designs

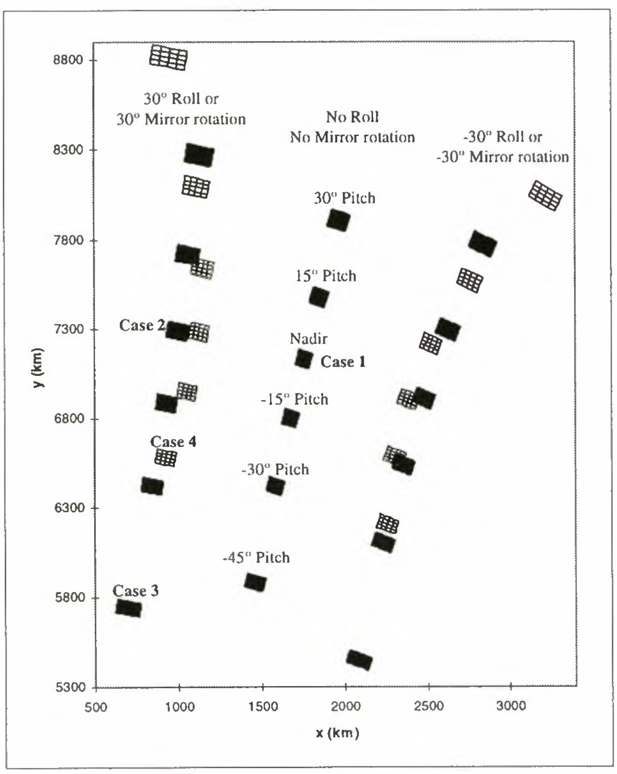


Figure 33: ***Image geometries in Mercator projection for the full range of depointing angles***

- ACT depointing : -30°, 0°, 30° either by platform roll (in-beam design) or mirror rotation (in-plane design),
- ALT depointing : -45°, -30°, -15°, 0°, 15°, 30° by platform pitch.

N.B The black boxes are related to in-plane design #1 whereas the hachured boxes are related to in-beam design #2.

Note that the study cases used in the current study (cases 1, 2, 3 and 4) are represented on the diagram.

With design #2, in-beam mirror rotation induces high image rotation (at ground level) which is compensated by yaw steering but this implies that the spotted area is not the same as for design #1. This means that our study case 4 (30° ACT mirror depointing, -50° ALT platform pitch) is not the worst case in terms of image geometry. If we look at the diagram, the worst case is the following one : 30° ACT mirror depointing, 30° ALT platform pitch.

So, in order to carry out an unbiased comparison between the two designs, this supplementary case should be studied. This could be done in PRISM Phase A.

5. RESAMPLING STRATEGIES

5.1 RESAMPLING STRATEGY

PRISM has many spectral bands and shall perform off-nadir viewing. So, the geometry of the images may vary significantly from one image to another. Image resampling is needed to get all these images in the same geometry.
Since image resampling entails distortion in the image, we recommend to make only ONE resampling for each spectral channel.
Indeed, thanks to the algorithms detailed in the Coregistration Analysis and in the Geolocation Analysis, it is possible to build a geometric model of image acquisition for each PRISM image in each spectral band.
By combination of the geometric model and of a DEM over the area, we can resample each image in the geometry of the DEM, so that all the images are superposable and can be easily exploited.

5.2 PRESENTATION OF THE RESAMPLING ALGORITHMS

Image resampling consists in computing the image radiometric value at fractional pixel positions. This radiometric value estimation is performed by an interpolation of the nearest neighbours of the reconstructed pixel.

We have identified 4 resampling algorithms: Nearest neighbour, Bilinear interpolator, Bicubic interpolator, windowed truncated Shannon interpolator - also named Hamming or Hanning - (see Refs. 12, 16, 17).

5.3 PERFORMANCE COMPARISON

5.3.1 Theoretical comparison

The optimal interpolator is a Shannon filter (cardinal-sine), but it cannot be implemented for images because it requires an infinite spatial kernel whereas the image size is finite.
So, the aim consists in trying to approximate this cardinal-sine interpolator by using a Finite Impulse Response filter, applied in a finite neighbourhood of the reconstructed filter. The more the filter is close to cardinal-sine filter, the best the algorithm is.

The following graphs represent both spatial and temporal characteristics of the filters associated to these algorithms :

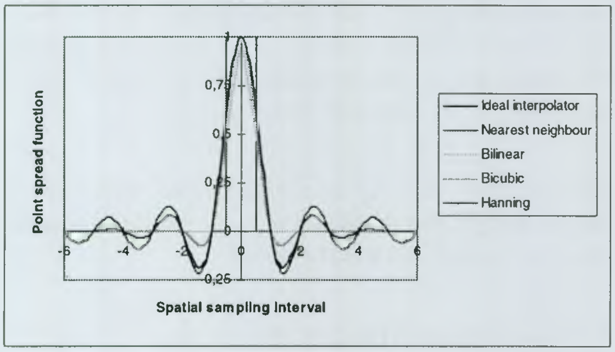


Figure 34: ***Spatial : Point Spread Function (PSF)***

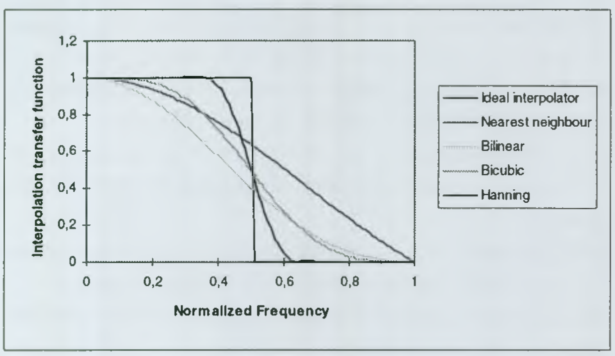


Figure 35: ***Frequency: Modulation Transfer Function (MTF)***

The order is: 1: Hanning, 2: Bicubic, 3: Bilinear, 4: Nearest neighbour.

5.3.2 Experimental comparison

We have compared these algorithms for PRISM representative images so as to get experimental results.

The following graphs represent the variation of the RMSE (in digital counts) and SNR (in dB) between the resampled image and the « real » image, for different fractional shifts, between 0 and 1 pixel by step of 0.1 pixel:

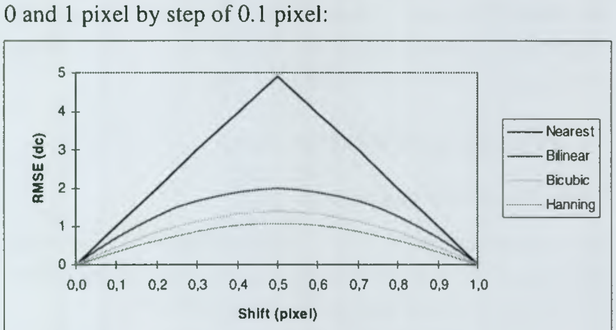


Figure 36: ***RMSE***

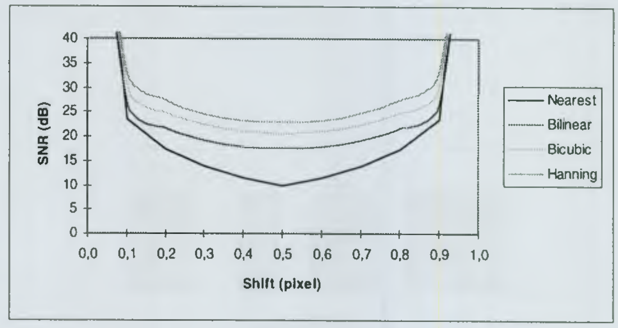


Figure 37: ***SNR***

The experimental results confirm the theoretical study.

Remark : RMSE and SNR for Hanning interpolator are close to values obtained for DCT/JPEG lossy compression of the same PRISM representative image at compression ratio close to 3. This means that resampling has an impact on image quality comparable to other processings such as lossy compression/decompression.

5.4 SELECTION OF RESAMPLING ALGORITHMS

ALCATEL recommends the two following resampling algorithms: 1. **Hanning windowed truncated Shannon resampling algorithm**, 2. bicubic resampling algorithm.
The running time associated to these algorithms is significant (compared to nearest neighbour), but it is not a major problem since the resampling is carried out in the ground segment facilities.

6. MAIN DIRECTIONS FOR FUTURE ACTIVITIES

The definition and analysis of the PRISM ground processing chains have been initiated during the study of PRISM geometric calibration carried out by ALCATEL for ESA and have been partially tested. Nevertheless, several trade-offs shall be made in the consolidation of the above processings :

- Selection of a correlation algorithm suited for multispectral images : choice of window size, similarity function, interpolation algorithm.
- For BRDF mode, two strategies shall be envisioned : separate geolocation for each image of the set of BRDF images or simultaneous geolocated of the whole set of BRDF images.
- Compromise between platform manoeuvre (image acquisition modes : oversampling, optimised MTF) and ground processing
- Selection of a resampling algorithm adapted to PRISM images, taking into account the MTF of the sensor (optics, detector, smearing)
- Selection of the optimisation algorithm to be used in the geometric tool
- Selection of a GCP extraction strategy : man managed or automated, amount of GCPs, map accuracy, ...
- Assessment of DEM accuracy (horizontal resolution and vertical resolution) on L2 processing performance

All these trade-offs shall be performed involving close discussions with ESA during phase A studies.

Moreover, planning a PRISM dedicated image acquisition campaign with a suitable airborne sensor (e.g. DAIS or MIVIS) would allow to get full PRISM images so that geometric calibration processes could be tested in a more operational way.

7. CONCLUSION

During the study of PRISM geometric calibration, we have performed the analysis of two key elements of the geometric calibration ground processing of PRISM images leading from raw data to exploitable data products : Coregistration and Geolocation.

We have demonstrated the ground processing capability to **co-register** images acquired in different spectral bands (region 1 : VNIR/SWIR and region 2 : TIR) by using a physical parametric model, with the following performance:

- Accuracy of **0.2 pixel** (i.e. 6 μm) for SWIR/VNIR focal plane relative position, in translation as well as in rotation,
- Accuracy of **0.3 pixel** (i.e. 9 μm) for TIR/VNIR focal plane relative position, in translation as well as in rotation,
- Accuracy of **5%** for attitude angles sinusoidal perturbations in some particular cases (amplitude ranging between 10 and 100 μrad, frequency close to 1 Hz).

We have demonstrated the possibility to assess static perturbations as well as temporal perturbations, at least in some particular cases.
The remaining errors are a combination of optimization and correlation errors.

An operational scenario for achieving coregistration of multispectral images during the satellite mission has been devised.

These performances correspond to worst case perturbations. And yet, they are compliant with PRISM geometric requirements:

- Spatial misregistration within regions 1 and 2 (e.g. for SWIR/VNIR) < 0.2 SSI,
- Spatial misregistration between regions 1 and 2 (e.g. for TIR/VNIR) < 4 SSI.

So, the spatial registration within and between regions 1 and 2 can be achieved by ground processing.
This possibly allows a **relaxation of requirements on the focal plane assembly accuracy**, which are very severe so far and constraining for the design and test of the focal plane assembly.

We have also demonstrated the ground processing capability to **geolocate** images by using a physical parametric model and control points (GCPs), with the following performance:

- Geolocation Accuracy ranging between **120 m to 800 m**, depending on the satellite pointing configuration (nadir view, combination of 50° along-track and 30° across-track depointing).
- Mean geolocation accuracy of **300 m**.

The remaining errors are a combination of optimization errors and GCP extraction errors.
Moreover, the geolocation strategy proposed in the study might be used to refine the geometric model of image acquisition, provided there is:

- The use of an automated extraction process of a large amount of GCPs
- The use of an accurate orbit determination system such as GNSS.

These preliminary results shall be consolidated during phase A and trade-offs shall be made.

The main trade-offs that shall be addressed during the phase A have been proposed.

Besides, it is important to note that **all the methods and algorithms presented in the current study are generic** and could be applied to other Earth Observation satellites.

REFERENCES

- **Study Reports** related to the current study of PRISM geometric calibration:

1. P. Martimort et al., PRISM : Study of Geometric Calibration of Spaceborne Imaging Spectrometers, Final Report, Aerospatiale, SE/TNQ 62/98, Issue 2, 03/06/98

2. P. Martimort et al., PRISM : Study of Geometric Calibration of Spaceborne Imaging Spectrometers, Executive Summary, Aerospatiale, SE/TNQ 88/98, Issue 1, 03/06/98

- Other **pre-phase A studies** related to PRISM:

3. JS Llorente et al., GNSS-based Attitude and Orbit determination and mission and vehicle management for low-Earth Observation systems, GMV, GMVSA 2220/96, 20/01/97

4. W. Posselt et al., PRISM pre-phase A CCN-2 Final Review, Dornier Satellitensysteme GmbH (DSS), PRI-DSS-HO-14, 06/02/97, 11/12/96

5. JY. Labandibar et al., PRISM pre-phase A CCN-2 Engineering Final Presentation, Aerospatiale, SE/TNI 38/97, 06/02/97

- Documents related to **geometric registration** of images:

6. J.P. Djamdji et al., Geometrical Registration of images : the multiresolution approach, *Photogrammetric Engineering & Remote Sensing* Vol. 59 No. 5, May 1993

7. J. Le Moigne, Parallel registration of multi-sensor remotely sensed imagery using wavelet coefficients, NASA, *SPIE* Vol. 2242 Wavelet Applications, 1994

8. P. Blanc et al., Validation protocol applied to an automatic co-registration method based on multi-resolution analysis and local deformation models, *Proc. of ISPRS Commission II Symp. Data Integration: Systems and Techniques*, Cambridge, 1998

9. I. Tannous et al., Parametric modeling of spaceborne SAR image geometry ;application : Seasat/SPOT image registration, *Photogrammetric Engineering & Remote Sensing*, 1991

10. L. Renouard et al., Geocoding SPOT products with ERS-1 SAR geometry, *Proc. Second ERS-1 Symp. Space at the Service of our Environment*, Hamburg, 1994

11. L. Polidori, Atelier logiciel de photogrammétrie analytique (ALPhA) : Cahier des charges fonctionnel et spécifications générales, Aerospatiale, SE/TOT 215/95

- Documents related to **numerical methods**:

12. W. Press et al., Numerical Recipes in C, Second Edition, Cambridge University Press, 1992

13. D. Kahaner et al., Numerical Methods and Software, 1992

- Documents related to images **correlation**:

14. P. Martimort, Corrélation d'images: évaluation des algorithmes disponibles à la Division QTIS, CNES, CT/SH/QTIS/PO n° 643/96

15. F.W. Leberl, Radargrammetric image processing, Artech House Remote Sensing Library, 1990

- Documents related to images **resampling**:

16. B. Pourcelot, MSG: IQGSE Resampling, Aerospatiale, MSG-ASC-SA-TN-0572, 10/06/97

17. A. Luton, Etude sur les interpolateurs et la convolution dans les simulateurs d'images satellites, Aerospatiale, SE/TNQ 15/97, 17/06/97

COALA - THE FUTURE OF GLOBAL VERTICAL OZONE PROFILE TREND MONITORING

J. Heilimo[1], S. Korpela[1], G.W. Leppelmeier[2], E. Sihvola[2*], H. Saari[3], H. Astola[3], C. van Eijk-Olij[4], H. Visser[4], R. Hoekstra[4], J. Lundell[5]

[1]Space Systems Finland Ltd., Kappelitie 6, FIN-02260 Espoo, Finland
[2]Finnish Meteorological Institute, Vuorikatu 15, FIN-00100 Helsinki, Finland
[3]VTT Automation, P.O.Box 1304, FIN-02044 VTT, Finland
[4]TNO-Institute of Applied Physics, P.O.Box 155, 2600 AD Delft, The Netherlands
[5]Patria Finnavitec Oy Systems, Naulakatu 3, FIN-33100 Tampere, Finland

jyri.heilimo@ssf.fi, gilbert.leppelmeier@fmi.fi

ABSTRACT

As is now well known, a severe decline in the ozone UV shield over the Earth has been particularly noted in the Antarctic in the spring, with a more gradual decrease notable at some times of the year generally over the Earth, and more recently similar "holes" have been noted over Arctic regions of the globe. Finding the trends of atmospheric ozone evolution and behaviour has become increasingly important, not only in the area "greenhouse gases", but also in areas like climate modelling and weather prediction, as ozone itself is also a major absorber of sunlight.

This paper presents a new ozone-monitoring instrument called COALA (*Calibration for Ozone through Atmospheric Limb Acquisitions*). COALA measures atmospheric ozone profiles by using limb occultation technique, similar to that of GOMOS. COALA, however, is considerably smaller than GOMOS, weighting only 20-25 kg, which allows us to fly several COALAs onboard mini satellites and, thus, resulting considerably lower costs per unit. This reduction in instrument size is achieved by neglecting trace gases other than ozone, concentrating on dark limb measurements, relaxing the measurement accuracy, and using modern state-of-the-art technologies in instrument design.

While reduced instrument accuracy (for example, due to a smaller input aperture) definitely leads to a reduction in the accuracy of the ozone profile obtained from a single occultation, flying two instruments in co-ordinated orbits not only provides complete redundancy, but gives better coverage and gains back the global trend accuracy by doubling the number of occultations per day.

* Now at University of Helsinki

1. INTRODUCTION

As is now well known, a severe decline in the ozone UV shield over the Earth has been particularly noted in the Antarctic in the spring, with a more gradual decrease notable at some times of the year generally over the Earth, and more recently similar "holes" have been noted over Arctic regions of the globe. These observations, confirmed over and over again by many types of observations have led to both extensive research efforts to understand the particular nature of the disappearance, as well as political and economic efforts to bring the causes of the ozone depletion under control. One identified cause is the use of chlorine- and fluorine-containing gases in refrigerators, air conditioners, and various industrial processes. As the use of such chemicals has been phased out, the extent of ozone depletion should decrease. However, the modelling of the atmosphere that is part of the basis of such expectations is not perfect, and therefore direct monitoring of the ozone is necessary. In addition, knowledge of the distribution of the ozone in the atmosphere is becoming increasing important in both weather prediction and climate studies, as ozone is also a major absorber of sunlight, i.e. it is a greenhouse gas.

In this context the purpose of the COALA mission is to yield global vertical profiles of ozone over the entire Earth with a sufficient accuracy and reliability that even small changes in either location or density can be noted over long periods of time, namely years to decades.

Presently data on ozone in the atmosphere come from Earth Probe TOMS (NASA) and GOME on ERS-2 (ESA). While nadir measurements have the clear advantage of yielding total ozone column density, which is very useful, for example, in estimating UV exposure and following air mass movement, vertical profiles are difficult to obtain (requiring CRAY-class computers), especially on an operational basis, and calibration to the

Proc. 32nd ESLAB Symp., 'Remote Sensing Methodology for Earth Observation and Planetary Exploration', ESTEC, Noordwijk, The Netherlands, 15-18 September 1998 (ESA SP-423, December 1998)

accuracy required for long-term trend development is problematical.

Thus COALA fits into this picture as a provider of:

1. Ozone density profiles above ~15 km altitude, with an accuracy (in a single measurement) generally of the order of 2-10% and a vertical resolution of 1-2 km, At Level 2, these are sets of profiles, one profile for each occultation, distributed over the globe. Methods of using the Level 2 product to produce ozone density field maps are currently under development
2. Long-term trends in ozone profiles and column densities with an accuracy of ~0.1%/year. These will be a Level 3 product, based on use of the Level 2 profiles over time.

2. MEASUREMENT PRINCIPLE

Like GOMOS, COALA uses the stellar occultation measurement principle, measuring the spectrum of a target star as that star sets behind the Earth and its atmosphere, i.e. a limb measurement (see Figure 1). When the tangent point and the surrounding atmosphere are in night-time, the measurement is called "dark limb" and the recorded spectrum consists of the stellar spectrum, as modified by the atmosphere through which the light has passed. If the field of view includes some atmosphere that is illuminated by the Sun, then the measurement is called "bright limb" and the recorded spectrum consists of a mixture of the stellar spectrum and sunlight scattered in the atmosphere. In principle, by imaging the field of view on a CCD, one can measure the scattered sunlight (in GOMOS parlance called "background") alone and subtract it from the spectrum that includes both stellar light and background light, to obtain the stellar spectrum alone.

Herein lies one of the critical design issues. The slit must be large enough to let all the star light through, even if there is vibration during an exposure, but small enough to limit the ratio of background light.to star light and avoid saturation of the CCD. COALA is designed to obtain a useful signal from 2-3 bright stars in bright limb, per orbit, which guarantees usable data from 10-20 stars (depending on the time of year) in dark limb.

When the target star is first acquired, it is above the atmosphere, as viewed from the instrument, and the spectrum is the natural spectrum of the star, which is used in the data retrieval as the reference spectrum. As the star sets (occulted by the Earth) a spectrum is recorded every 0.1-2.0 seconds (an experimental parameter). Successive spectra are more and more affected by the atmosphere, as light is scattered (principally by N_2, O_2, and aerosols) and absorbed (principally by ozone, but also by other trace gases). The ratio of the signal spectrum to the reference spectrum is the transmission, which depends only on the line integral of the atmospheric constituents along the light path, and their cross sections. Instrument effects such as radiometric sensitivity are thus removed automatically. (Note that this process holds as long as the wavelength assignment is stable or recoverable, which requires knowing the pointing *during* each exposure.)

3. DESCRIPTION OF COALA

3.1. System Description

The COALA instrument will comprise three units: the COALA optical unit, the COALA electronics unit, and COALA pointing mechanism. Figure 2. Shows the COALA system concept.

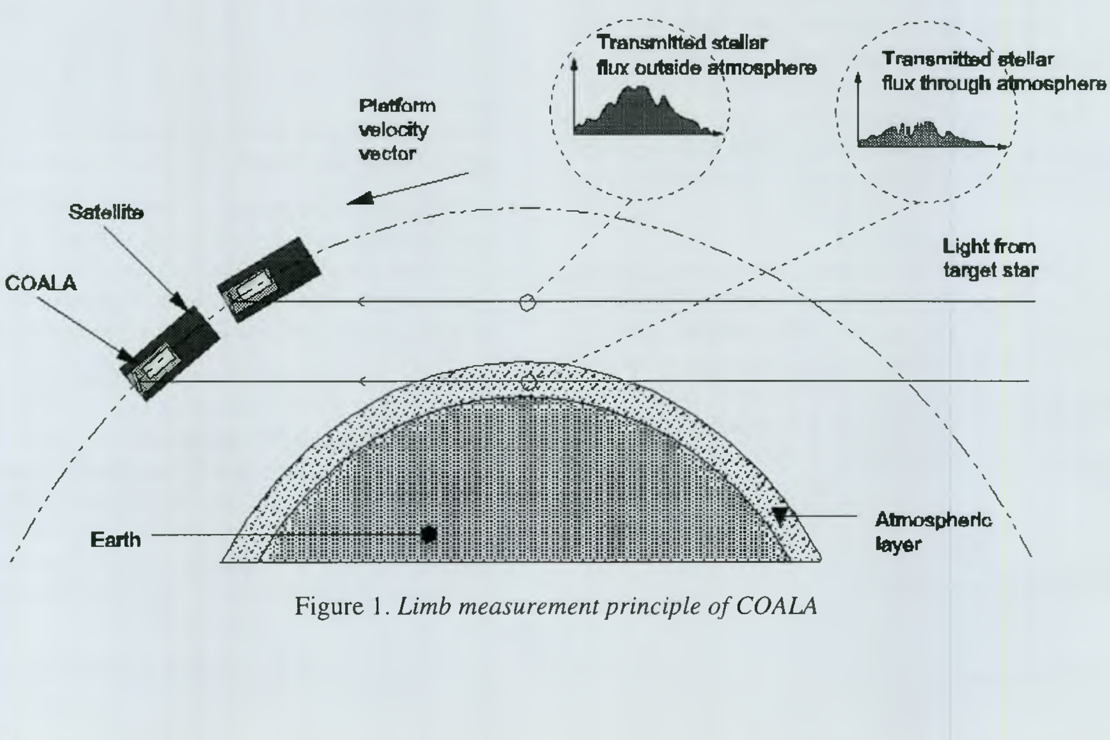


Figure 1. *Limb measurement principle of COALA*

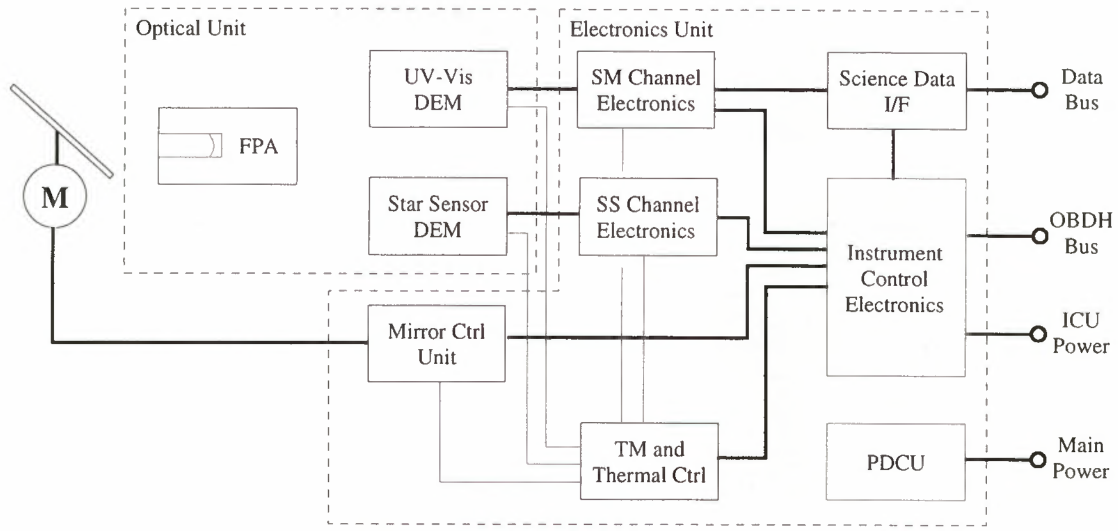


Figure 2. COALA System layout

The 2-axis pointing mechanism provides pointing capabilities of ±2.5° in elevation direction and ±30° in the azimuth direction. With not so demanding pointing requirements (w.r.t GOMOS), there will be some uncertainties to the location of the stellar image in the slit, and yet, the pointing mechanism is not tracking, which means the star image in the instrument slit and in the detector CCD is moving. The location of spectrum is calculated from the star sensor image and the CCD lines containing the spectrum and sufficient amount of background is read from the CCD (i.e. approximately 40 lines).

Aperture Slit width Slit length	72 mm 0.014° (- 0.028°) 0.17° (up to 1.0°)
Spectral Bands Spectral Resolution at FWHM Spectral Sampling	UV: 250-350 nm VIS: 420-675 nm UV: 0.5-1.5 nm VIS: 1.0 nm UV: 0.37 nm/pix VIS: 0.48 nm/pix
Mass Power Volume	22-25 kg 20 W Optics: 13 dm3 Electronics: 7 dm^3
Data rate	50 kbit/s (average)

Table 1. *Summary of COALA Technical Properties*

The detector modules are located inside the optical unit and are cooled to +5°C in order to reduce dark current. The effect of dark current is large on the dark limb performance, in which the background signal levels are lower and dimmer stars are usually used, and on bright limb measurements there is almost no effect, at all [2]. In the most important wavelength range (350 nm-675 nm) used in ozone retrieval, the SNR and ozone retrieval accuracy are improved by factor of 1.5 when temperature is lowered from +25°C to +5°C.

3.2. Optical Unit

The Optical Unit consists of Focal Plane Assembly and combined UV-visible Detector Module, and Star Sensor Detector Module. The 2-axis pointing mechanism is considered as a separate unit. In case the S/C pointing capabilities are accurate enough, the pointing mechanism can be excluded from COALA system.

The 2-axis pointing mirror is used in searching and tracking the target star and in reflecting the light into the optical unit. The optical design of COALA splits the spectral region of interest into two spectral bands, one for the UV and one for the visible. By using prisms as dispersing elements for the UV channel, a much higher efficiency in the UV can be reached than with grating. In addition, in the visible spectral channel a relatively high optical efficiency can be realised because of the possibility to use properly blazed gratings with high efficiency for the limited spectral range of the channel. High optical efficiencies allow the instrument and its telescope to get smaller and cheaper.

The optical configuration of the COALA design is shown in Figure 3. First the UV stellar image is imaged in the UV slit by using dichroic mirror and subsequently collimated, dispersed with two prisms, and finally imaged to the CCD detector. The light passing through UV dichroic is reflected by another dichroic into the visible channel, in which the light is collimated, dispersed with plane grating, and finally imaged to the same CCD detector as the UV spectrum. The light passing through the visible dichroic is reflected to the star sensor that is used in tracking star and locating the star vertical position in the CCD. The spectral resolutions of the both UV and visible channels are shown in Table **1**.

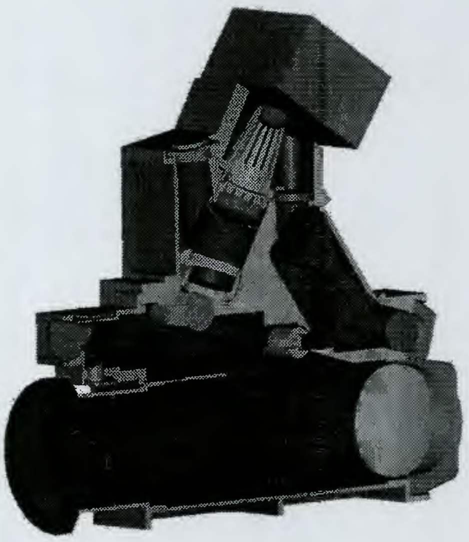


Figure 3. *Optical configuration of COALA (without pointing mechanism)*

3.3. Electronics Unit

The COALA Electronics unit consists of following sub-units:

- Instrument Control Electronics controls sequencing of instrument operations and maintains communication with host S/C,
- Science Data I/F is platform dependent interface unit to S/C data bus,
- Power Distribution and Control Unit (PDCU) regulates the supply power and distributes the power to subsystems,
- TM and Thermal Control unit that collects telemetry data and controls the Detector Module temperatures,
- Mirror Control Unit (MCU) controls the mirror actuators,
- The Spectrometer Channel (SM) controls the exposure and read-out of the spectrometer detector module, the video signal conditioning and conversion, and the serial transmission of data to the satellite via data interface,
- Star Sensor Channel (SS) have similar tasks than the spectrometer channel, except faster read-out rate, enabling the instrument control unit to detect rapid changes in the intensity of light (replacement of fast photometer) and use the stellar image position to compute star location in the spectrometer CCD.

In principle none of the COALA subsystems are redundant. The S/C interfaces are redundant, and the signal wires in the instrument harness may also be redundant. As the electronic subsystem gets larger and more complex, the increase in the amount of the circuitry tends to reduce the reliability of the system. The use of PLDs will increase the reliability due to a high degree of integration and a reduced number of ICs and package pins. The use of PLDs, off course, reduces also the mass of the instrument.

4. COALA PERFORMANCE

The key idea throughout the design of COALA has been "*How small a stellar occultation ozone measuring instrument can be made without sacrificing the scientific performance too much*". The reduction of size of COALA when compared to GOMOS is achieved by omitting the trace gases other than ozone, by concentrating mainly on dark limb occultations, relaxing the pointing requirements and radiometric performance, using high efficiency optical design, and applying modern technologies in electronics.

As a stellar occultation instrument, the vertical resolution of ozone profiles retrieved by COALA is 1-2km, which is good when compared to the profiles computed from the measurements of nadir-looking ozone monitoring instruments.

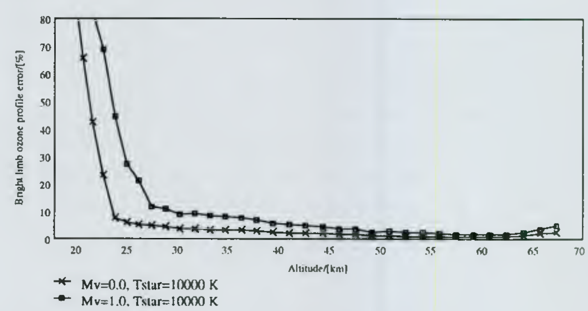


Figure 4. *Simulated bright limb ozone profile error*

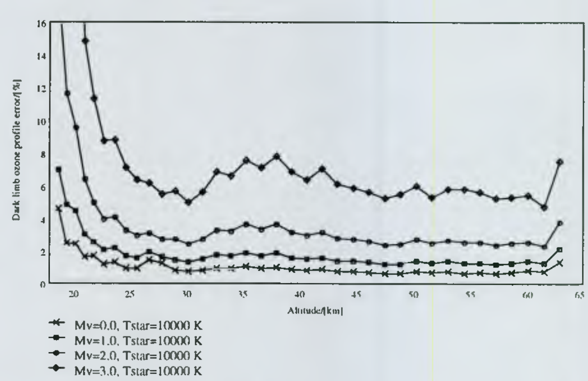


Figure 5. *Simulated dark limb ozone profile error*

In order to keep the profile retrieval error in acceptable limits the magnitude limit of the star for bright limb measurements is set 0.0 and 3.0 for dark limb. As can be seen from the Figure 4 the bright limb ozone profile retrieval error above 25 km is less than 10%, when the visual magnitude of target star is less than 0. The single measurement profile retrieval error for dark limb is less than 10% with all stars below the magnitude limit 3.0. The single measurement profile retrieval accuracy is not

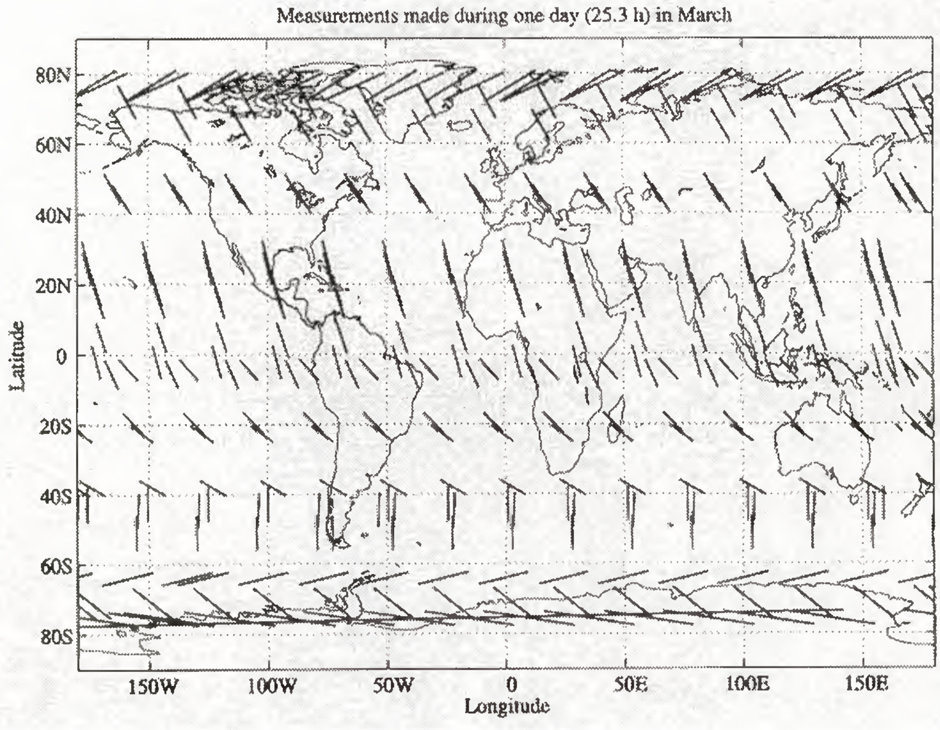


Figure 6. *Simulated coverage of limb measurements of one COALA during one day.*

considered good enough if COALA is treated as a self-standing instrument monitoring absolute ozone profiles. But in case of long-term monitoring of profiles the COALA can be used to determine ozone profile trends up to accuracy of ±0.1%.

Simulations have shown that the most critical items that have to be solved in order to achieve adequate scientific performance are:

- CCD dark current, which is solved by cooling the detector to +5°C,
- Noise of the video signal amplification, solved by using modern low noise amplifiers as a first stage after CCD, and
- High spectral transmission of the optics solved mainly by using prisms as dispersive elements in UV channel and properly blazed gratings in limited wavelength range in visible channel.

5. COVERAGE

There are some 500-600 stars that are bright enough ($M_v \leq 3$) to be potential candidates for stellar occultation measurements. Not all of these can be seen from COALA on a particular orbit at a particular time of the year, and then those that are near the Sun and those that are weak and visible in bright limb are not suitable candidates for observation. Thus a selection must be made, appropriate for a particular mission and time of year. Naturally the brighter stars are generally preferable, but also coverage and stellar temperature (which affects the ratio of near-UV to red light on the spectrum) can influence selection.

Occultation measurements form coverage "bands" in the atmosphere as can be seen from the Figure 6, in which each line represents one occultation and the length indicates the path of the light through the atmosphere when tangent altitude is ~15 km. One can see the pattern repeat from east to west as the Earth turns and the pattern of occultations is repeated on each orbital pass. The density of the measurement grid depends on the orbit selected but generally all geolocations within the "bands" are covered. Gaps between these bands are filled-in during the year, because of Earth axis rotation with respect to the Sun and different set of accessible stars in different months.

Table 2 gives the number of stars selected for a particular scenario, including nominal COALA instrument characteristics. The table shows how many occultations are typically achievable throughout the year. Clearly choosing a larger range of azimuth angle (measured from the satellites anti-velocity direction) obtains more occultations per orbit, but may be more demanding on satellite accommodation. Note that in order to obtain sufficient measurements to permit determination of long term trends to ±0.1% in 100 regions over the globe one requires ~20 occultations per orbit. Significantly fewer occultations per orbit would imply either a relaxation on the accuracy requirement or fewer regions. For example, one might argue that, since air mass motions are predominantly along latitude lines, grouping the results into latitude bands (e.g. 18 bands, 10° wide) would be sufficient.

We consider two possible two-satellite configurations: 1) both instruments on the same orbit plane (i.e. same LTDN) but in different phases, and 2) instruments on different orbit planes (e.g. LTDN separation 4-6 hours).

In the constellation option 1, the two satellites will see the same set of stars and so will have the effect of doubling the number of occultations in a day along a latitude line (E-W) and the coverage bands are filled-in faster (see Figure 7).

In the second constellation option, the two satellite will see somewhat different sets of stars, which have the effect of evening out the variances over the year, and filling in gaps in the N-S direction (see Figure 8).

Note also by comparing Columns C and E (or D and F) in the Table 2 that using orbits displaced by 90° does indeed flatten out the variation over the year. In general it is clear that the ±30° FOV is more than adequate in the case of a two-satellite scenario. The question of which option in orbits is the better would be best answered in a detailed study that considered, for example, instrument characteristics and stellar spectra more exactly.

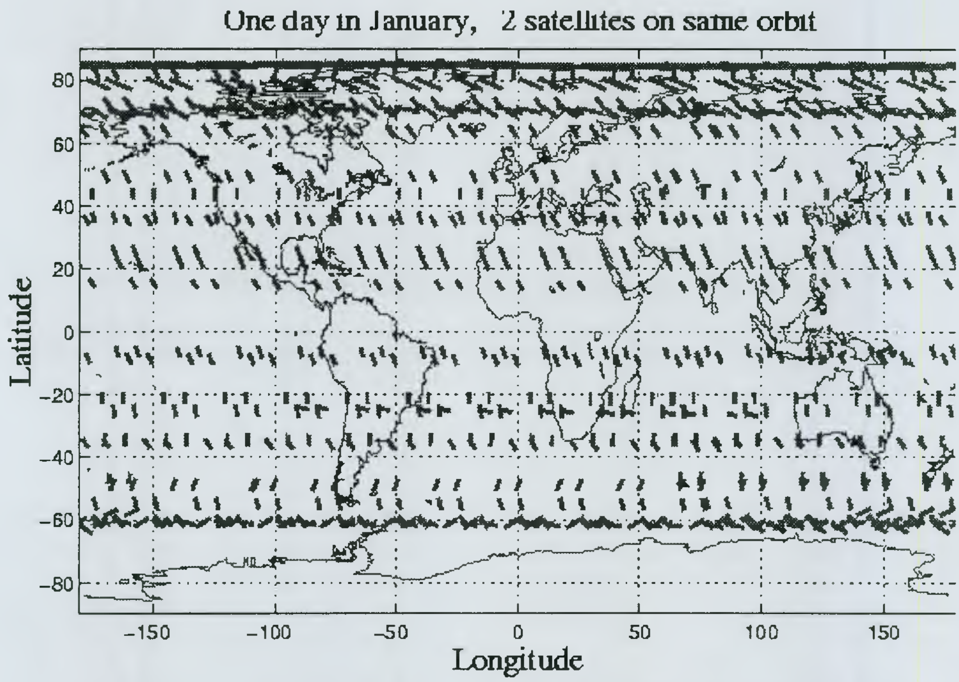


Figure 7. *Coverage achieved with constellation of 2 COALAs in the same orbit plane (Argument of perigee separation 120°).*

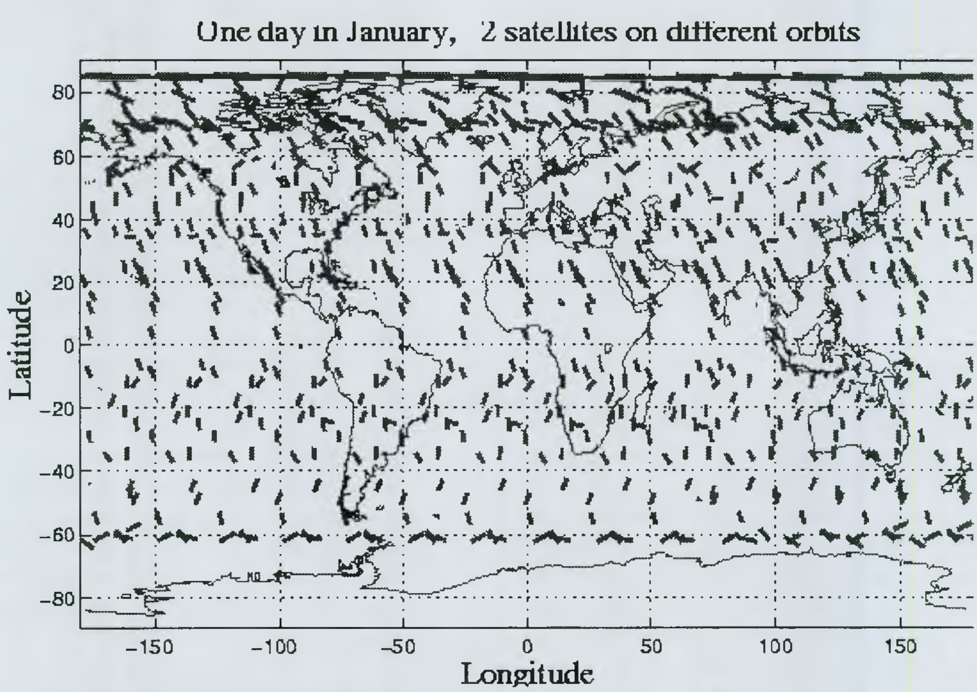


Figure 8. *Coverage achieved with constellation of 2 COALAs in two separate orbit planes (LTDN difference 6 h; argument of perigee separation 120°).*

Month	A	B	C	D	E	F
	±15° 09:00	±30° 09:00	±15° 2 x 9:00	±30° 2 x 9:00	±15° 9:00 & 15:00	±30° 9:00 & 15:00
1	21	30	42	60	33	50
2	24	27	48	54	39	47
3	10	24	20	48	20	44
4	12	20	24	40	27	41
5	15	20	30	40	25	48
6	10	20	20	40	21	40
7	15	21	30	42	26	36
8	10	28	20	56	23	44
9	11	20	22	40	23	42
10	11	15	22	30	32	45
11	13	16	26	32	37	43
12	12	22	22	36	19	41

Note: In selecting stars those closer than 40° to the Sun have been rejected in order to avoid interference.

Conditions assumed:
Column A: 9:00 am descending orbit, stars within ±15° of orbit plane.
Column B: 9:00 am descending orbit, stars within ±30° of orbit plane.
Column C: 2 satellites in a 9:00 AM descending orbit, stars within ±15° of orbit plane.
Column D: 2 satellites in a 9:00 AM descending orbit, stars within ±30° of orbit plane.
Column E: 1 satellite in a 9:00 AM descending orbit, the other in a 3:00 PM descending orbit, stars within ±15° of orbit plane.
Column F: 1 satellite in a 9:00 AM descending orbit, the other in a 3:00 PM descending orbit, stars within ±30° of orbit plane.

Table 2. *The number of occultations chosen for one orbit, month by month, under various assumptions.*

6. SUMMARY

Presently the data on ozone in the atmosphere come from Earth Probe TOMS (NASA) and GOME on ERS-2 (ESA), both using nadir measurements. The GOMOS instrument onboard ENVISAT will demonstrate the usefulness and effectiveness of stellar occultation method in atmospheric ozone profile monitoring.

The COALA instrument is an ozone-monitoring instrument, using the same stellar occultation measurement principle than GOMOS does. However COALA is considerably smaller than GOMOS, weighting less than 25kg. Its small weight and size, makes it easy to fly several COALAs onboard mini satellites, thus, resulting considerably lower cost per unit. By forming a constellation of COALAs, global ozone profile trends retrieved over the entire Earth with sufficient accuracy and reliability that even small changes in either location or density can be noted over long periods of time, namely years or decades.

Using GOMOS as reference for COALA, i.e. flying COALA simultaneously with GOMOS and comparing the results will give to COALA constellation an exceptionally solid base. The data analysis methods developed and validated for GOMOS can be applied for COALA with only slight modifications.

We see that constellation of 2-3 COALAs is the future of atmospheric ozone monitoring.

7. REFERENCES

[1] COALA Final Report, COA-VTT-TN-008. Issue 1.0, 10.12.1996

[2] COALA Parametric Analysis, COA-SSF-TN-001, Issue 1.0, dated 25.11.1996.

[3] Scientific Simulations of COALA, COA-FMI-TN-002, Issue 1.0, 28.11.1996.

[4] Bertaux, J.L., Pellinen,R. Simon, P. Chassefière, E., Dalaudier, F.,Godin, S., Goutail,F., Hauchecome, A., Letexier, H., Mégie, G., Pommereau, J.P., Brasseur, G., Kyrölä, E., Tuomi, T., Korpela, S., Leppelmeier, G., Visconti, G., Fabian, P., Isaksen, S.A., Larsen, S.H., Stordahl, F., Cariole, D., Lenoble, J., Naudet, J.P., & Scott, N. GOMOS, Proposal in response to ESA EPOP-1 A.O.1, Verrieres le Buisson: Service d'Aeronomie du CNRS, 1988, 125p.

[5] ESA, Dornier, Matra Marconi Space France. GOMOS Phase C/D Instrument Preliminary Design Review Data Package Overview PO-DP-MAT-GM-001, Report to ESA. Toulouse: Matra Marconi Space France SA, 1993. 58p.

[6] Korpela, S. A Study of the Operational Principles of the GOMOS Instrument for Global Ozone Monitoring by Occultation of Stars. Helsinki: Geological Survey of Finland, 1991. Geofysikaalisia julkaisuja 22, Doctoral Thesis, 172p.

ATMOSPHERIC TRACE-GAS MEASUREMENT WITH A BALLOON-BORNE FAR INFRARED FABRY-PEROT INTERFEROMETER (SFINX)

R.W.M. Hoogeveen, C. Bernard*, F.P. Helmich, A.R.W. de Jonge*, J.J. Wijnbergen*, and A.P.H. Goede

Space Research Organization Netherlands (SRON), Sorbonnelaan 2, 3584 CA Utrecht, the Netherlands,
**Landleven 12, P.O. Box 800, 9700 AV Groningen, the Netherlands*

ABSTRACT

One of the outstanding questions in atmospheric chemistry relates to the importance of the hydrogen catalytic cycle in the ozone balance reactions. Hydrogen radicals such as OH, HO_2, H_2O_2, and their cross products with the halogens HOCl, HCl, and HBr all have their rotational emisions lines situated in the far infrared, which thus far has eluded space measurement.

In ESA's PIRAMHYD study three instrument concepts for measurement of the atmospheric hydroxyl (OH) profiles from space were compared. SRON participated in the PIRAMHYD study with the Fabry-Perot interferometer concept. In parallel with the theoretical PIRAMHYD study SRON is developing a small (roughly a 2 : 1 scale with respect to the instrument dimensions of the PIRAMHYD study) balloon borne prototype version of a Fabry-Perot interferometer. It is due to fly on the MIPAS-B2 gondola of the University of Karlsruhe as part of the HIMSPEC project during the Third European Stratospheric Experiment on Ozone (THESEO) in the winter of 1998/99.

1. INTRODUCTION

It is well know that a quantitative explanation of the stratospheric ozone budget requires the introduction of catalytic cycles involving the HO_x, NO_x and ClO_x families of species. An ozone balance based on oxygen-only reactions, the so-called Chapman cycle, yields too high ozone concentrations compared with observation. Here the HO_x, NO_x and ClO_x catalytic cycles can provide higher ozone destruction rates. Recent observations have confirmed the important role of the HO_x cycle in the destruction of stratospheric ozone. In contrast to earlier models, the HO_x cycle was found to dominate the NO_x and the ClO_x cycle [1] in the lower stratosphere..

Within the HO_x family of species the hydroxyl (OH) radical is the most important member. Hydroxyl not only participates in the catalytic destruction of ozone, it also is responsible for the coupling of the HO_x family to members of the NO_x and ClO_x families. For example, the reaction between OH with HCl produces active chlorine, which in turn triggers the ClO_x catalytic ozone destruction cycle. Similarly, HO_2 couples directly to the NO_x cycle through the reaction with NO.

Quantitative observation of atmospheric trace gases can be carried out in the ultraviolet-visible and near infrared (Sciamachy, Sage, Toms, Gome) - the mid infrared (Mipas, Claes, Isams) as well as the far infrared and microwave regions (MLS) of the electromagnetic spectrum. Most gases are well detectable in the mid infrared by measuring their emission generated by vibrational transitions. Polar lightweight molecules show very strong rotational transitions in the far infrared and the emission of some of them, like the HOx species can exclusively be measured in that region.

Despite the importance in regulating stratospheric ozone, the OH radical has not been measured on a global basis. Few space measurements have been performed to date, notably the UV-day glow measurement from the Space Shuttle ATLAS mission[2]. However, these measurements do not cover the polar regions and neither do they measure OH in the lower stratosphere where uncertainties in the atmospheric-chemistry processes are largest.

In order to provide measurement data of global OH profiles, ESA has initiated a study of the capabilities of a space-borne limb sounding far infrared instrument for the measurement of the OH profile from space, the so-called PIRAMHYD study [3] (Passive Infrared Atmospheric Measurement of Hydroxyl). The OH rotational transition $F_1\ 7/2^+ \rightarrow 5/2^-$ at 118.455 cm^{-1} (84.42 μm; 3.551 THz) or the F1 $5/2^- \rightarrow 3/2^+$ at 83.869 cm^{-1} (119.23 μm; 2.514 THz) is employed. The underlying atmospheric spectroscopy for OH, and particularly for the 118 cm^{-1} line, is in a mature state[4,5]. Neither uncertainties in molecular parameters or in knowledge of interfering species will contribute significantly to the error budget for satellite-based measurements.

In the PIRAMHYD study three possible instrument concepts are considered:

- The Fourier Transform Interferometer (FTI). Here the spectrum is obtained from the measurement of its interferogram. Its main advantage is that within one interferogram several atmospheric species can be measured at the same time. This has to be balanced against a decreasing signal-to-noise ratio of the measurement. In order to acquire the spectrum with sufficient signal-to-noise ratio a helium cooled detector is required. The FTI provides a compromise between spectral resolution, spectral coverage and signal to noise.

Proc. 32nd ESLAB Symp., 'Remote Sensing Methodology for Earth Observation and Planetary Exploration', ESTEC, Noordwijk, The Netherlands, 15-18 September 1998 (ESA SP-423, December 1998)

- The Heterodyne Instrument (HDI). The incoming atmospheric FIR radiation is mixed with radiation from a local oscillator, and the signal at a specific difference frequency is measured. The high spectral resolution of the HDI allows full resolution of the OH emission line, and thus discriminates between emissions from different altitude regions and different molecules. A drawback of the HDI is the experimental stage of the mixer detector at 2.5 and 3.5 THz and the local-oscillator technology. A solid state local oscillator is not yet available so that a complex CO_2 -laser pumped FIR gas laser must be employed.
- The Fabry-Perot Interferometer (FPI). The concept consists of a small telescope and a FP module positioned in the parallel beam of a grating monochromator. By adjustment of the FP resonator, its frequency is tuned successively to the emission frequencies of the pertinent species. When employed with a novel high-temperature superconducting bolometer detector the required cooling can be achieved by a Stirling cycle cooler, thereby avoiding the use of stored cryogen and its limited operational lifetime in space. The FPI is a relatively simple instrument which can be made very compact and light weight. The main price for this is a limited spectral resolution and signal-to-noise ratio.

The results of the PIRAMHYD study show that the most demanding measurement scenarios could be met by the FTI and the HDI concepts and that the FPI showed a useful but lesser performance due to the relatively low signal to noise of the 90 K HTS bolometer detector (with the 4 K detector S/N is not at all an issue). However, the dominating error source for all three instrument concepts turned out to be the space craft pointing error which must be reduced from the specified 1.5 km at tangent height to less than 400 m for any of any of the instrument concepts to provide useful results.

In the present paper we concentrate on the Fabry-Perot interferometer concept. A balloon version is presently developed at SRON that will serve as a demonstration model for a possible satellite version of the FPI. The instrument is equipped with both a helium cooled Ge:Ga detector and with the experimental high-temperature superconducting bolometer detector operating at 89 K.

To summarize, the main goals of the project are three fold:
1 Demonstrate the potential of a Fabry-Perot interferometer for the detection of stratospheric trace gases.
2 Measure gases OH and HCl (and possibly HQ and H_2O_2): their vertical distribution and their temporal behavior.
3 Demonstrate the capabilities of a newly developed FIR sensor: The high-temperature superconductor (HTS) bolometer detector.

The long term goal of the project is to develop a satellite version of the FP interferometer which can fly as an add-on instrument on any platform dedicated to research on the Earth atmospheric chemistry. Earlier studies [Refs 6 - 10] have shown that such an instrument with low demand on satellite resources can meet the requirements.

2. SFINX INSTRUMENT DESCRIPTION

The SFINX instrument (SRON Fabry-perot INterferometer eXperiment) is currently being developed by SRON and serves as a demonstration model for a possible satellite-borne OH monitor. The instrument covers the wavelengths between 65 and 90 microns. With its spectral resolution in excess of 6000, it is perfectly capable of detecting the radiation emitted by rotational transitions of light-weight molecules and radicals. SFINX will be flown as a piggy-back instrument on board the MIPAS-B2 (Michelson Interferometer for Passive Atmospheric Sounding, Balloon version 2) stratospheric balloon gondola in cooperation with the Institut für Meteorologie und Klimaforschung (IMK) of the University of Karlsruhe (Ref. 11).
The principle is identical to a space-borne FPI, with differences related to balloon application (size scaled down by roughly factor of 2).

The atmospheric FIR light beam of 80 mm diameter is fed to the instrument by a scan mirror and a small 90 degrees off-axis telescope, as shown in figure 1. The spectrometer part of the SFINX instrument is contained inside a 5" helium cooled dewar, see figure 2. The 1.5 times 3.0 mm^2 horizontal slit at 300 mm from the parabolic mirror results in an acceptance cone of 5 times 10 mrad, but an internal field stop limits this to the working field-of-view of 4 times 8 mrad. Filters directly behind the entrance slit absorb the visible and thermal IR radiation, such that this does not reach the detector by higher order reflection of the grating. A spherical collimator mirror provides the parallel beam for the FP and images the scan mirror approximately at its entry stop of 24 mm. The FP has two highly reflecting surfaces of nickel mesh (1500 lines/inch) and provides the required spectral resolution of approx. 0.02 cm^{-1} at 118 cm^{-1}, working in the order 106 at a finesse of 55. Tuning of the gap distance is achieved by moving one of the mirrors up and down in the parallelogram setup over ca 2 mm, resulting a horizontal gap change of ca 0.2 mm.

After the FP, the beam is recollimated onto a reflective entrance slit for the grating monochromator. This slit is field-of-view determining. The grating of 35 times 60 mm^2, ruled in aluminum 6061, has about 18 lines per mm. The efficiency determines an optimum blaze of about 30-40^0. It selects the wavelength region of 65 to 90 μm in first order. With the 1.5 mm slit at 135 mm in the parabolic mirror focus a spectral resolution is more than 130, above

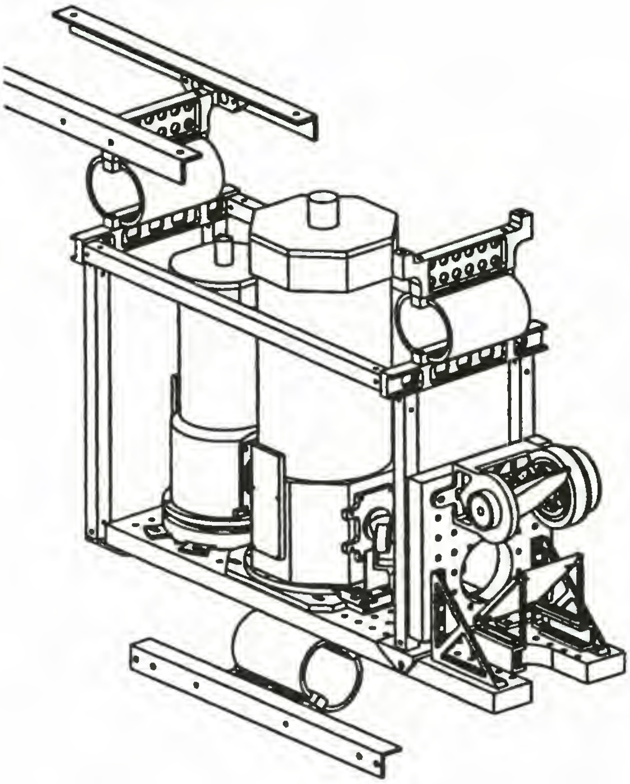


Figure 1: View of the exterior of the SFINX instrument with its suspension structure: The light is collected by a scan mirror and through a 90 degrees off-axis parabola fed to a 5" helium cooled dewar containing the spectrometer and the Ge:Ga detector. A separate 3" liquid nitrogen cooled dewar houses the high-temperature superconducting bolometer detector.

the highest FP order over the whole region of the FP. The movement of the grating is performed by a stepper motor at 4 K, resulting in no power dissipation in the stationary situation.

The FIR radiation is detected bij either a Ge:Ga photon detector at 4 K providing high sensitivity, or by a High-Temperature Superconductor (HTS) bolometer detector, placed inside a separate liquid-Nitrogen cooled dewar. The Ge:Ga detector consists of a $3\times1\times1$ mm^3 size piece of Ga doped Ge and has a sensitivity of approximately 5 A/W, resulting in a background-limited noise-equivalent power of approximately 1. $10^{-14}W/Hz^{1/2}$. The detector is used in a trans-impedance amplifier (TIA) chain mounted on the dewar floor. The TIA consists of a 100 MΩ feedback resistor at 4K and a double JFET amplifier heated to about 50 K. The detector signal is further amplified in an electronics box mounted directly to the side of the dewar to levels of a few Volts.

By tilting the grating, the beam can also be directed to a 3" liquid-nitrogen cooled dewar containing the high-temperature superconducting (HTS) bolometer detector

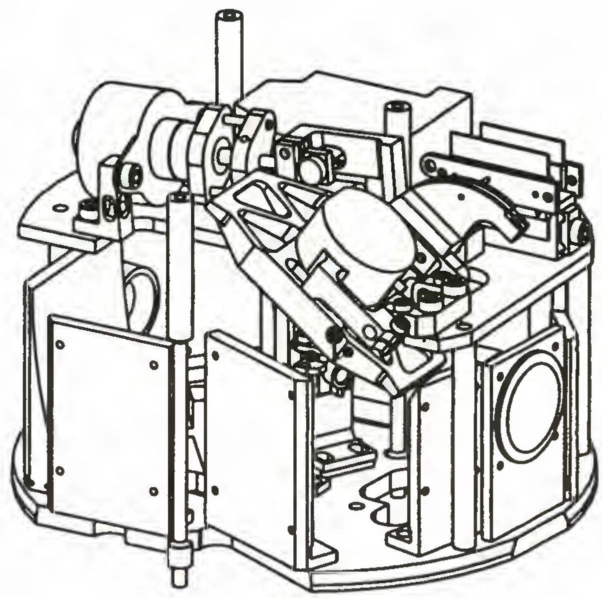


Figure 2: View of the interior of the 5" dewar containing the spectrometer and the Ge:Ga detector.

(ref. 12 and references therein). This detector consists of a 0.6 μm thick silicon nitride (silicon rich Si_xN_y) membrane of 3x3 mm^2 size with a meander of high-T_c superconducting $GdBa_2Cu_3O_{7-x}$ (GBCO). A gold black absorption layer is deposited on the bolometer to obtain efficient absorption in the far-infrared. Making use of the very high temperature coefficient of resistance at the superconducting transition at 89 K a very high detectivity over a broad band in the far-infrared can be achieved for an operating temperature within reach of liquid nitrogen cooling or small low-power mechanical cryo coolers as favoured in space-based systems. Achieved noise-equivalent power is 4 $pW/Hz^{1/2}$ with a 100 ms response time, and 1.5 $pW/Hz^{1/2}$ with a 500 ms response time on an advanced type with back-etched membrane.
Because of its limited sensitivity with the HTS detector only the relatively strong HCl line can be measured. A comparison between the signals obtained by the Ge:Ga detector and the HTS detector allow a good assessment of the performance of the HTS detector.

Calibration of the sensitivity and zero level of both detector systems is performed in-flight by two extended blackbody calibration sources. These consist of stainless steel surfaces with concentric V-shaped grooves of 60^O top angle at 2 mm distance. One black-body source is heated and other is at ambient temperature and both temperatures are monitored. They can be inserted in the dewar beam by a switchable mirror in front of the dewar entrance. The wavelength calibration as well as the spectral resolution is performed pre-flight and checked in flight on ozone or water vapour emission lines.

The elevation pointing is derived from the pointing system of the MIPAS instrument mounted on the balloon gondola. SFINX being a balloon instrument, the pointing error is of far lesser concern in the retrieval error with respect to a limb sounding satellite instrument. The attitude of the MIPAS optical platform is monitored using a gyro-based inertial reference. Because the mounting of the two optical platforms in the gondola might slowly change in time during the flight, the elevation of both optical platforms is measured using level sensors. These measure the elevation with respect to the local apparent gravity vector, and their time-averaged difference gives the relative position of the SFINX optical platform with respect to the MIPAS platform. The servo system which actuates the scan mirror with respect to the optical platform is similar to that used on the MIPAS instrument, but uses a smaller motor. Information on attitude and pointing of the MIPAS instrument is transmitted in real time to the SFINX instrument, enabling the control system to point the SFINX beam to a inertial-referenced elevation, or to slave it to the MIPAS line of sight.

3. ELECTRONICS AND DATA HANDLING

The on-board electronics and the two battery packs are housed in a 36 * 36 * 24 cm^3 box insulated by foam, and mounted separately from the optical bench. The total power consumption is estimated at 50 W, which excludes the telemetry/telecommand contribution. The instrument is controlled by a board computer, which is based on the PC104 architecture, and consists among others, of an Intel 80386 compatible CPU solid state disk, and analog to digital converters. No onboard data logging is performed.

The data acquisition rate is 64 Hz for the fast signals such as the detector AC signals, the FP gap setting, chopper position and the attitude signals; 8 Hz for the detector DC signals; and 1 Hz for low-frequency house-keeping signals. Because the telemetry comprises FP gap setting and chopper position sampled at the same frequency as the detector, synchronous demodulation of the detector signal is possible on the ground. This scheme avoids onboard demodulation and allows in-flight monitoring of detector signals and noise across the full 20 Hz bandwidth of the amplifier. Total telemetry bandwidth used by the SFINX instrument is 12288 bits/second, provided by CNES telemetry facilities. Telecommand bandwidth is 1900 bits/second, of which only a fraction will be used. Telemetry, telecommanding and instrument control are provided by the onboard Intel 80386-based computer system.

The ground station consist of three standard MS-Windows based PC's linked by their serial interfaces to the CNES telemetry/telecommand system. They will provide redundant archival of data, telemetry display during flight, and a telecommanding post. The ground station software will show provisionally calibrated spectra during operations. Data analysis after the flight will provide full calibration and recover the trace gas concentrations as a function of altitude from the observed spectral data. Retrieval of concentration profiles will be done by simultaneous fitting the concentration profile parameters of a atmospheric radiation transfer model to all of the observed data, using the optimal estimation method.

4. CONCLUSION AND PLANNING

The SFINX programme started in October 1996, and is currently (autumn 1998) entering its final system tests and calibration. The first flight of the SFINX instrument, together with the MIPAS-Balloon instrument, is foreseen in early 1999 from ESRANGE Kiruna, Sweden, as part of the THESEO (Third European Stratospheric Experiment on Ozone) campaign of the European Union.

ACKNOWLEDGEMENTS

The SFINX design team consisting of P. Schreinemakers, F. Zwart, C. van Dijkhuizen, H. Doedens, and L. van der Kaaden, as well as the MIPAS-B2 team under guidance of Dr. H. Oelhaf is kindly acknowledged. The work will be partially financed under the European Union Climate & Environment Framework 4 Program.

REFERENCES

1. P.O. Wennberg, R.C. Cohen, R.M. Stimpfle, J.P. Koplow, J.G. Anderson, R.J. Salawitch, D.W. Fahey, E.L. Woodbridge, E.R. Keim, R.S. Gao, C.R. Webster, R.D. May, D.W. Toohey, L.M. Avallone, M.H. Proffitt, M. Loewenstein, J.R. Podolske, K.R. Chan, S.C. Wofsy, "Removal of Stratospheric 0_3 by Radicals: In Situ Measurements of OH, HO_2, NO, NO_2, ClO, and BrO", Science 266, 398-404 (1994).

2. D. Offermann and R.R. Conway, "Mission studies the composition of Earth's middle atmosphere", EOS, Vol. 76, p. 337 (1995).

3. Final report of ESA PIRAMHYD retrieval study, contract number 11719/95/NL/CN, by SRON (Netherlands), IROE (Italy), RAL (United Kingdom), DLR (Germany), under management of Serco Europe Ltd. Published 1997.

4. K.V. Chance, J.J. Wijnbergen, W. Schneider, and J.P. Burrows, The OHIO Concept: A Baseline Design for Satellite-Based Measurements of Stratospheric OH, Proc. SPIE, High-Temperature Superconducting Detectors: Bolometric and Nonbolometric, 2159, 2133 (1994).

5. K.V. Chance, D.A. Jennings, K.M. Evenson, M.D.

Vanek, I.G. Nolt, J. V. Radostitz, and K. Park, "Pressure broadening of the 118.455 cm^{-1} rotational lines of OH by H_2, He, N_2, and O_2", J. Molec. Spectrosc. 146, 375-380 (1991).

6. J.J. Wijnbergen, P. de Valk, J.P. Burrows, K. Chance, W. Schneider, "The Far Infrared Fabry-Perot Spectrometer as a Satellite-Based Atmospheric Sounder", Conference on Passive Infrared Remote Sensing of Clouds and the Atmosphere 11, Rome, Proc. Europto Series SPIE 2309-287 (1994).

7. J.P.J.M.M. de Valk, J.J. Wijnbergen, K. Chance, J.P. Burrows, W. Schneider, "OH Measurements with Airborne and Satellite Instrumentation in the Lower and Upper Stratosphere", Conference on Passive Infrared Remote Sensing of Clouds and the Atmosphere ll, Rome, Proc. Europto Series SPIE 2309-279 (1994).

8. J.J. Wijnbergen, P.A.J. de Korte, M.J.M.E. de Nivelle, "Instrumental Aspects of a Satellite-Based Fabry-Perot equipped with a High-Tc Bolometer for Detection of Stratospheric OH", European Symposium on Satellite Remote Sensing Paris, Proc. SPIE 2578-306 (1995).

9. Detector Study for a Fabry-Perot Interferometer for Far Infrared Atmospheric Sounding, J.J. Wijnbergen, P.A.J. de Korte, M.J.M.E. de Nivelle, ESTEC contract Ref. JP/11/10411/dl d.d. 3 November 1994.

10 K. Chance, J.J. Wijnbergen, P. de Valk, W. Schneider, J.P. Burrows. "The OHIO Concept: Refinements on a Design for Satellite-Based Measurements of Stratospheric OH", Conf. on Atmospheric Sensing and Modelling, 29-30 Sept. 1994, Rome, Proc. Europto Series SPIE 2311-236 (1994).

11 H. Fischer and H. Oelhaf, "Remote sensing of vertical profiles of atmospheric trace gas constituents with MIPAS limb emission spectrometrs", Appl. Opt. 35 (1996) 2787 - 2796, and references therein.

12 M. de Nivelle, T. Heidenblut, B. Schwierzi, W. Michalke, E. Steinbeiss, M. Elzenspoek, S. Sanchez, M.P. Bruin and P. de Korte, "Low-noise far-infrared detection at 90 K using high Tc super conducting bolometers with Silicon Nitride beam suspension", 32e ESLAB symposium, this issue.

Active Microwave Technologies

GROUND PENETRATING RADAR FOR PLANETARY EXPLORATION.

T. Hagfors,
Max-Planck-Institut für Aeronomie,
D-37191 Katlenburg-Lindau,
Germany
Tlf (49)5556979415, Fax (49)5556979410, email hagfors@linhag.mpae.gwdg.de

Abstract.
Uses of radar for lunar and planetary exploration are briefly reviewed. From being a device for the measurement of distance and velocities radar has come to provide information on surface topography, surface structure and electrical properties of the surface material and has provided maps of reflectivity over the surface. The ability of radar to penetrate the surface has been amply demonstrated on Earth. Much thought has gone into the design of radar systems which optimize the ability to penetrate the surface to detect interfaces between dry surface regolith and layers containing ice or water. We describe a stepped frequency radar which provides the best compromise between the requirement of a high frequency to avoid ionospheric effects and a low frequency to increase the penetration depth. The stepped frequency scheme allows for a way to tune the antenna over very large relative bandwidths and for adaptively removing most of the effects of ionospheric dispersion. Similar schemes may find application in future missions to the icy Galilean satellites, to the moon and to Mercury.

1. Introduction.

Ground-based radar has developed into an important research tool in planetary investigations. Initial studies of the moon provided the first crude information on the small scale topography of the lunar surface, and lead to estimates of the dielectric properties and the surface material, the rms slope of the surface on scales much smaller than can be obtained from high resolution photographs. Further development of the delay Doppler technique, which is identical to the side-looking radar technique applied in airplanes and satellites for Earth observations. This technique has also found applications in planetary research both on the moon and on Venus, in the latter case in the projects Venera and Magellan with spectacular results.
Most of the interpretation of the data have been made in terms of a quasi-specular reflection mechanism with diffuse scattering from small scale structure super-imposed. Attempts have also been made to interpret the data with a model that involves the penetration of a low loss surface layer, and with scattering inside this layer.
Ground penetrating radar systems have found applications on Earth, particularly in areas with glaciers, or in dry desert areas. The application of ground penetrating radar to explore the planet Mars is, therefore, an obvious possible application. The reason for the interest in seeing below the surface stems from the supposition that the water, which has been flowing over the surface in the past, and which has left its imprint on the surface, must have been trapped underneath the surface in the form of ice mixed with regolith material, and even in the form of water. It is thought that there is no mechanism which can have made all the water escape into space, hence the expectation to find ice and even water below the visible surface of Mars.
Requirement for exploring the subsurface structure of the outer Galilean satellites, Mercury and the moon will no doubt surface in the future, and it is, therefore of interest to review some properties of ground penetrating radar which might prove useful particularly in Mars observations, but also for other future purposes.
The ideas presented here were developed during the preparation of a proposal for a Mars Express long wavelength radar which failed to be selected. We believe that some of the ideas to be described are of sufficient general interest to warrant presentation in this symposium. Before going on to discuss the concepts it is appropriate to list the team members who contributed to the radar concept:

V. Andrianov (IRE/RAS)
N. Armand (IRE/RAS)
Y. Barbin (CNRS/SA)
H.-M. Braun (RST)
T. Breus (IKI/RAS)
F. Costard, (Paris-Sud)
R. Grard (ESTEC)
S.-E. Hamran (NDRE)
A. Herique (CEPHAG)
W. Kofman (CEPHAG
H. Kopka (MPAe)
M. Lester (U. Leicester)
V.N. Marchuk (IRE/RAS)
M. Pätzold (U Köln)
D. Shtern (IRE/RAS)
T. Spohn (U.Münster)
H. Svedhem (ESA/SSD)
J.-G. Trotignon (LPCE)

Proc. 32nd ESLAB Symp., 'Remote Sensing Methodology for Earth Observation and Planetary Exploration', ESTEC, Noordwijk, The Netherlands, 15-18 September 1998 (ESA SP-423, December 1998)

2.0 Electrical Properties of Surface and Ionosphere.

2.1 The Surface

The reflection, refraction and absorption of radio waves depends on the dielectric constant ε ($=\varepsilon'-i\varepsilon''$) of the medium. For a homogeneous medium the reflectivity is given by $R^2 = |(\sqrt{\varepsilon}-1)/(\sqrt{\varepsilon}-1)|^2$, the damping of a radio signal at frequency f is given by $(2\pi f\sqrt{\varepsilon'}\tan\delta)/c$ where $\tan\delta=\tan(\varepsilon''/\varepsilon')$ is known as the loss tangent. The skin depth, d, for weak absorption, is given by $d = \lambda / \sqrt{\varepsilon'}\tan\delta$.

Radar measurements of Mars have given values for ε' of 3.5 (Pettengill et al., 1969) for the upper layers. The dielectric constant of the first 100 m of the Moon is thought to be related to the density, ρ (in g cm^{-3}) of the soil or solid materials by $\varepsilon' = (1.93\pm0.17)^{\rho}$ and $\tan\delta=0.006\,\rho$ (Olhoeft and Strangeway, 1975). Based on this ε' is thought to lie between 3 and 9, and the skin depth between 2000 and 600 m at a frequency of 1 MHz. The frequency dependence of both ε' and $\tan\delta$ are slight above 1 MHz and the skin depth is, therefore, approximately proportional to wavelength (Krupenio, 1980).

For pure water ice $\varepsilon' \approx 3.0$ and $\tan\delta \approx 2.1 \times 10^5 e^{\gamma T}/f$ where $\gamma=0.101C^{-1}$ and T is the ice temperature (in °C) (Paunder, 1969; Finkelshtein et al., 1977). Assuming an ice temperature of -30 C, the skin depth becomes 2800 m. In reality the ice will be mixed with rocks so that the above must be an upper estimate. Finkelshtein et al., (1977) found $\varepsilon' = 3.7$ and $\tan\delta = 0.18$ for frozen earth type mountain rocks with a weight content of ice of 0.75, at -10 C at 1 MHz leading to a skin depth on the order of 150 m. There is also a possibility that stable brines exist on Mars (Zent and Fanale, 1986; Kuzmin, 1983). This has an effect particularly on the loss tangent and the depth of the liquid-solid isotherm. The latter will occur closer to the surface than in the case of a mixture of pure ice and rocks (Andrianov, Kibardina and Kuzmin, 1993; Armand et al., 1994)

Table 1 shows the skin depth for three different upper layer compositions assuming a temperature of -60°C for the ice (there is no temperature dependence for hard rock). The two frequencies correspond roughly to the extremes which will be used by the radar for ground sensing. Clearly the skin depth can vary a lot depending on the material which means that the depth of the echoes together with their frequency dependence can tell us about the material composition. Figure 1 summarizes the penetration depth as a function of frequency for two values of ε'.

It is thought that Mars has a cryolithosphere which may be regarded in simple terms as a top layer of "dry frozen rock" above a layer of frozen icy rock (Kuzmin, 1983). The "dry" layer is thought to be 10-40 m deep at high latitudes and 300-400 m deep at low latitudes. The "icy" layer is believed to be 600-2500 m at high latitudes and 500-800 m deep at the equator. Below these layers there may be wet rock and possibly even liquid water. A radar in the MHz band should therefore be able to penetrate the dry upper layers to maybe several kilometers, and perhaps a further few hundred to thousands of meters into any "icy" layers. As a simple model of the multiple layer structure we take a uniform double layer where the top layer has a complex dielectric constant ε_1 and depth L, and the underlying layer has a complex dielectric constant ε_2. In this case the reflectivity takes the form

$$R^2 = \frac{R_1^2 + R_2^2 e^{-2\tau} + 2R_1R_2e^{-\tau}\cos\varphi}{1 + R_1^2R_2^2e^{-2\tau} + 2R_1R_2e^{-\tau}\cos\varphi} \text{ with}$$

$$\varphi = (4\pi f/c)\sqrt{\varepsilon_1'}\,L \quad \text{and}$$

$\tau = (2\pi f/c)\sqrt{\varepsilon_1'}\,L\tan\delta_1 = (\varphi/2)\tan\delta_1$ where φ is the phase path through the upper layer and back. The reflection coefficients are defined as in the case of a semi-infinite half space as: $R_1 = (\sqrt{\varepsilon_1'}-1)/(\sqrt{\varepsilon_1'}+1)$ and $R_2 = (\sqrt{\varepsilon_2'/\varepsilon_1'}-1)/(\sqrt{\varepsilon_2'/\varepsilon_1'}+1)$

One can regard the reflection as the result of interference between reflections from two layers which explains why the technique of using multiple frequencies to measure depth is sometimes referred to as interferometry. In the short pulses which we intend to synthesize the layers would show up as discrete scatterers at ranges corresponding to the depth of the layer transitions and multiples thereof.

Table 1 Values of dielectric constants with penetration depths at 0.5 and 5 MHz

	Hard rock				**Fresh ice**				**salty (2%) ice**			
Parameter / f, λ	ε'	ε''	tanδ	d [m]	ε'	ε''	tanδ	d [km]	ε'	ε''	tanδ	d [m]
f=0.5MHz (λ=600m)	5	0.5-0.05	0.1-0.01	200-2000	3.16	2.10^{-3}	5.10^{-4}	50	3.9	1.18	0.3	80
f=5 MHz (λ=60m)	5	0.5-0.05	0.1-0.01	20-200	3.16	2.10^{-4}	5.10^{-5}	50	3.3	0.15	0.04	60

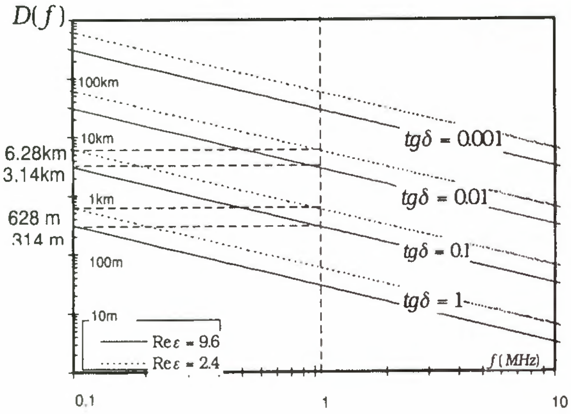


Figure 1 Skin depth versus radar frequency for some possible surface materials

For large radar wavelengths, the limiting value of R^2 is $R^2_{\lambda\to\infty} = \left|(\sqrt{\varepsilon_2}-1)/(\sqrt{\varepsilon_2}+1)\right|^2$.

For short radar wavelengths, assuming sufficient absorption, the reflection is determined by the top layer, and the reflection takes the same form, except that ε_2 is replaced by ε_1. In practice clear-cut oscillations such as implied by our model may not be observed because there may be a gradual transition of electrical properties with depth or more than one internal interface. In addition the depth of the layer or the electrical properties may vary over the lateral area sampled by the radar. Therefore, in order to measure oscillation as implied in this model it is necessary to step through the frequency band in as short a time as possible to minimize the lateral extent of satellite motion over the surface.

2.2 The Ionosphere

The frequency response functions are affected by the presence of the passage twice through the ionospheric plasma between the space craft and the ground. The phase variation with frequency will cause a pulse distortion. In a stepped frequency radar it is possible to correct for this by multiplying each frequency response by a phase factor in order to rotate the frequency response back to what it would be without the ionosphere.

The extra phase introduced by the ionosphere at a frequency f is:

$$\Delta\Phi(f) = 4\pi(f/c)\int(\mu(f,z)-1)dz$$

with the refractive index is given by:

$$\mu(f.z) = \sqrt{1-N(z)e^2/m\varepsilon_0 4\pi^2 f^2} = \sqrt{1-f_N^2/f^2}$$

where:

N(z)=electron density
e = charge of an electron
m = mass of an electron
ε_0 = free space dielectric constant

$$f_N(\mathrm{Hz}) = 9\sqrt{N(\mathrm{m}^{-3})}$$

When f is everywhere along the ray considerably smaller than the maximum plasma frequency, the phase can be approximated by:

$$\Delta\Phi(f) = -(e^2/mc\varepsilon_0 2\pi f)\int N(z)dz$$

The phase therefore in this approximation depends only on the total electron content of the ionosphere. In the radar experiment it will be necessary to make corrections for the ionosphere. The data needed for this correction can be obtained from the total contents measurements planned for the mission, by extrapolation from the topside sounder profile, by comparing the group delay in adjacent frequency bands, or by a pulse sharpening method involving the initial surface echo. An important advantage of the stepped frequency method is that a step by step phase correction can be made to compensate for the phase distortion imposed by the ionosphere. A similar correction is extremely difficult if not impossible to achieve in a pulsed system based on phase coded modulation with all frequency components represented in each pulse.

The maximum value of electron concentration in the Martian day side ionosphere is about $2\times10^{11}\mathrm{m}^{-3}$ (Kliore 1992) and $f_N \cong 4\mathrm{MHz}$. This clearly makes it close to impossible to operate a radar during the day time at frequencies low enough to allow substantial penetration of the surface and with desired range resolution of a few hundred meters. During the night the maximum plasma frequency drops to about 500 kHz, and the necessary transparency for the long wavelength radar at frequencies below 5 MHz exists. But even under these circumstances there is a serious distortion of short pulses which, our calculations show, can lead to a pulse broadening of a 1.5 MHz bandwidth pulse from the nominal 100 m to 270 m. Such pulse broadening can be corrected for in a stepped frequency radar, but not in a pulsed radar system. The conclusion to be drawn is that it is necessary to operate the radar under night time conditions in order to achieve the goals of surface penetration and detection of a layered structure there. In the stepped frequency scheme of observation it is straight forward to correct for the dispersive distortion by introducing a phase and an

amplitude factor at each frequency which corrects for the distortion. The ionospheric data required for this correction can be obtained by extrapolation of the topside profile, by making use of the total electron content measurements from the ionospheric profiler, by deriving the difference in group delay between different bands or by actually introducing adaptive filtering by peaking up on the mostly very sharp initial return from the surface.
Another ionospheric problem stems from absorption. The composition of the Martian atmosphere is such that the ionospheric absorption in relation to the electron density is considerably higher than in the Earth's ionosphere. At a frequency of 10 MHz the one way absorption during the day time may be as high as 10 dB.

2.3 Conclusion on the choice of frequency
The conclusion to be drawn from this discussion is that the frequency must be chosen as low as possible but compatible with the requirement that the ionosphere is sufficiently well behaved to allow us to correct properly for its effects. For Mars one should, therefore, strive to make observations of the surface in the band 1 to 5 MHz and design the radar system accordingly. In the case of Mars Express this is particularly important because the conflicting requirement of the surface radar and optical observations will probably make the true nighttime radar observations scarce.

3.0 The Special Radar Principle
3.1 Stepped Frequency Radar Concept
From the discussions in the previous sections it is clear that the instrument must be a long wavelength radar operating at frequencies below 5 MHz. To probe the subsurface structure with a depth resolution sufficient to resolve, or discriminate layers of 100 m thickness, as some models imply may be present, the radar bandwidth must be about 1 MHz. The radar system must be able to complete the probing in a time corresponding to a satellite motion of a small fraction of a Fresnel zone, or a few kilometers. The radar system must feed into an antenna which provides reasonable efficiency over the whole frequency band without detrimental pulse distortion. Because of the effect of the ionosphere the system must be able to compensate for its effect by an appropriate de-dispersion procedure. To reduce the major part of the clutter echoes originating from the surface it is necessary to filter the echoes to a frequency resolution which corresponds to the frequency width of the central Fresnel zone on the planet. It should also provide possibilities for cross track reduction of clutter.. To achieve the secondary goal of probing the topside of the ionosphere, it must be possible to step through the lower part of the frequency range with pulses coded to provide moderate range resolution. The observation of the daytime ionosphere makes sense in view of the possible limited time of observation during nighttime conditions
The stepped frequency radar concept may provide the only means to fulfill all the necessary requirements and have the following advantages:

- It steps through the frequency range and re-tunes the antenna frequency by frequency.
- It allows for the combination of the complex frequency responses to be combined into an equivalent pulse response of a resolution which can be chosen after the fact because the raw data will be available and transmitted to the ground.
- It will be feasible to correct for the effect of the ionospheric dispersion by applying an appropriate phase correction which can be determined from the data.
- By combining repeated frequency sweeps it is possible to remove most of the surface clutter by filtering even when there is a frequency offset due to vertical satellite velocity or large scale surface slopes.
- By operating the system from a pair of monopoles it may be possible to reduce or even eliminate cross track clutter.
- The system can be applied as an ionospheric topside sounder in a special operating mode developed in the Mars96 LWR system.

3.2 The Observation Scheme
In the radar system envisaged the stepping through a frequency sub-range containing ca. 100 frequencies can be completed in a period of 0.05 to 0.1 seconds, still allowing the re-tuning of the reactive component of the antenna impedance to match the transmitter to keep up with the frequency stepping. During each frequency sweep the satellite will have traveled much less than the distance corresponding to the size of a Fresnel zone, even at the highest frequencies. It is intended to keep and transmit to the ground the data (one complex sample per frequency) for each of these frequency sweeps, and to subsequently combine the data to synthesize a short pulse, to improve signal to noise ratio, if necessary, and to reduce by coherent integration and other means the effects of off nadir clutter which can corrupt the results.
As it is desired to keep the duty cycle of the transmitter as high as possible we assume for the sake of this discussion of principle that the duty cycle is 50%. In order to detect the echo it must be assured that the received echo falls well within one of the transmitter off periods. The repetition rate and the pulse length must be continually adjusted in order to achieve this,

and the height of the satellite over the Martian ground must be known. As the height of the satellite varies rapidly with time it is necessary to introduce a radar mode to determine this height approximately and unambiguously. This mode of operation is taken over from the Mars96 "wobbulation" technique. The procedure is explained in Figure 2.

The transmitter is on-off modulated with an interpulse period which varies continuously from 0.66 ms to 1.33 ms. The ratio of interpulse interval and pulse length is held constant at 3. In the course of the transmission of the sequence of pulses the transmitter frequency is chirped continuously and linearly during the time T_b over a frequency interval Δ. The purpose of varying the interpulse period during the time interval is to ensure that an echo is received at least part of the time between transmitter pulses. After the FFT of the decoded signal there will be a peaked frequency spectrum at a frequency f_m which can be directly translated into the height of the spacecraft over the surface. It can also be used to determine the virtual depth of the echo from the topside ionosphere. For this reason the "wobbulation mode was referred to as the PLASMA mode in the Mars96 mission In the observations where the radar is used to probe the subsurface the repetition rate can be set so that the received echo falls within a period when the transmitter is off.

Height determination by "wobbulation"

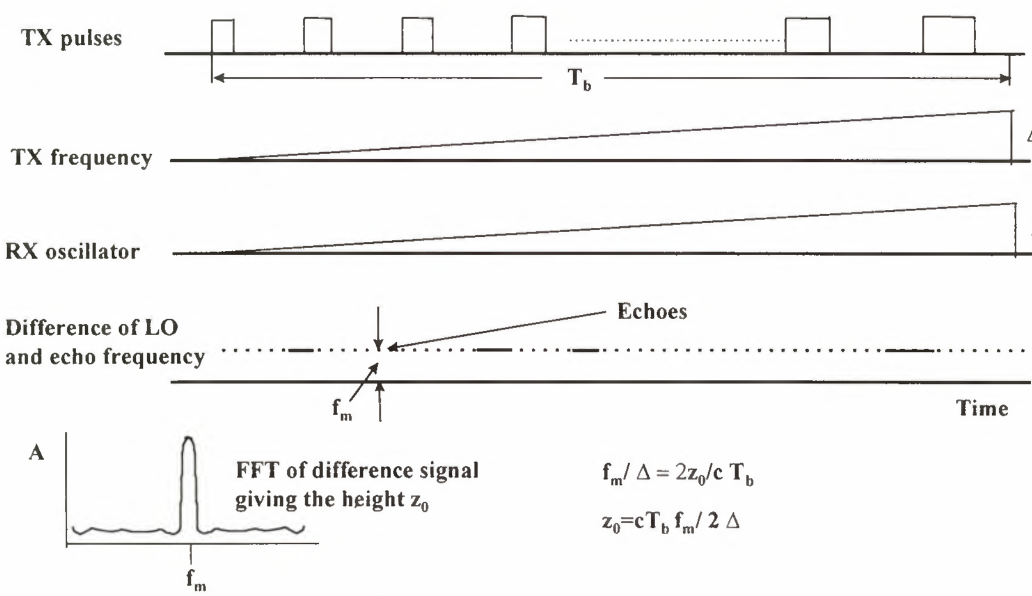


Figure 2. Determination of appropriate repetition rate.

Having determined the range the pulse length and intrapulse periods are chosen so that the receive period falls centered between transmitter pulses, as shown in Figure 3. In the figure the height is such that the first echo returns in the third intrapulse period. As represented in Figure 3 the height determination makes it possible to place the echoing region at the center of the diamond-like intersection areas in the diagram. The demodulation of the received signal takes place by cross-multiplication with an exact replica of the of the transmitter wave, delayed in accordance with the result of the "wobbulation" result. The demodulated signal, phase and amplitude together, can be interpreted as a time-dependent frequency response of a passive circuit. The frequencies of the received signals are now hovering near zero, only offset by slight Doppler shifts which may occur due to changes in altitude of the satellite. From the preliminary orbit data it appears that the maximum vertical velocity is 1.9 km/sec. This translates to a maximum Doppler shift at the highest frequency normally used (5 MHz) of about 60 Hz. It would therefore seem that sampling of the demodulated wave-form once per frequency step would be adequate to reconstruct the signals at each frequency. With as much as 100 frequencies this would create about 1000 complex samples per second (twice as many if the monopulse system is implemented), a rate which it would be possible to store and transmit to the ground station for analysis. A decision about coherent filtering to remove some of the effects of clutter along the path of the satellite can be taken after the fact and be optimized, the trade-off being between resolution and clutter rejection.

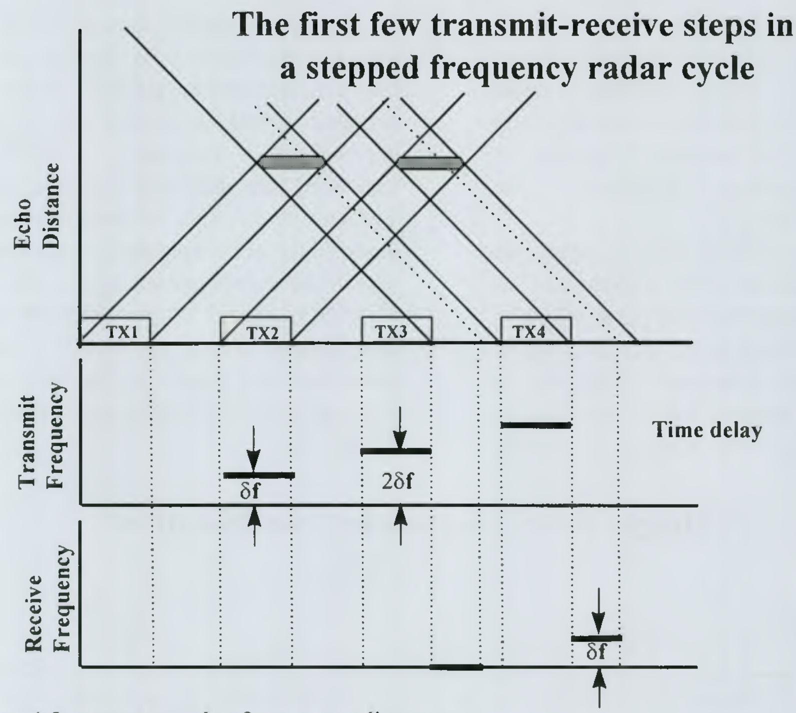


Figure 3. The stepped frequency mode after range adjustment

The equivalent pulse response of the echo can now be determined from a discrete Fourier transform with respect to the frequencies $\mathbf{f_k}$ of the frequency response functions $\mathbf{R(f_k,t)}$, possibly weighted to account for dispersion, or to improve the code side-lobes. The Doppler filtering to achieve discrimination against off nadir echoes along the satellite path is accomplished by an averaging in time over as long as is required, or as long as can be allowed on physical grounds. The time **t** must then be tagged with an index. In the next section we shall take into account the effect of the time change from frequency to frequency and from sweep cycle to sweep cycle in terms of a double index, as $\mathbf{t_{km}}$, the former linked to the frequency index, the latter to the sweep cycle number, see also the illustration for five sweep cycles in Figure 4.

Complications arise if the monopulse scheme proposed by van Zyl (private communication, 1997) is implemented. Then the amount of data to be transferred could easily be doubled, but it will still be possible to transfer the data to the ground station without processing in the space-craft.

When used as an ionospheric top-side sounder it is modulated with a 30 kHz chirp as in the "wobbulation" mode, and with the initial frequency of each chirp stepped through the range of frequencies needed to define the top-side profile (PLASMA).

3.3 Synthesis of the pulse response of the return

In the previous section it was shown how to obtain estimates of the frequency response function of the equivalent propagation circuit which describes the radar response of the Martian surface. The relation to the impulse response can be understood from the following simplified considerations. Assume that the response of the medium does not change with time. This would correspond to the (unrealistic) case of a stationary satellite. If the surface is illuminated with a sine wave at frequency f, the returned wave would be of the form:

$$g(t) = Ae^{2\pi ift} \int m(r)e^{-4\pi if\, r/c}dr = e^{2\pi ift}R(f)$$

Here $m(r)$ is a complex reflection coefficient, which describes the reflectivity, or the scattering power as a function of distance. The factor A accounts for the geometrical and other attenuation. If the received signal is multiplied by the complex conjugate of the time shifted replica of the transmitted wave, i.e., we obtain the function:

$$R(f)e^{2\pi if\Theta} = A\int m(r)dr e^{2\pi if(\Theta - 2r/c)}$$

Integrating this over frequencies, which amounts to taking a truncated Fourier transform of $R(f)$ one obtains:

$$\int R(f)e^{2\pi if\Theta}df \approx A\int m(r)dr\delta(\Theta - 2\,r/c) = Am(\Theta\, c/2)$$

From this non-rigorous calculation we see that the Fourier transform of the frequency response function gives us the "reflectivity" as a function of distance.

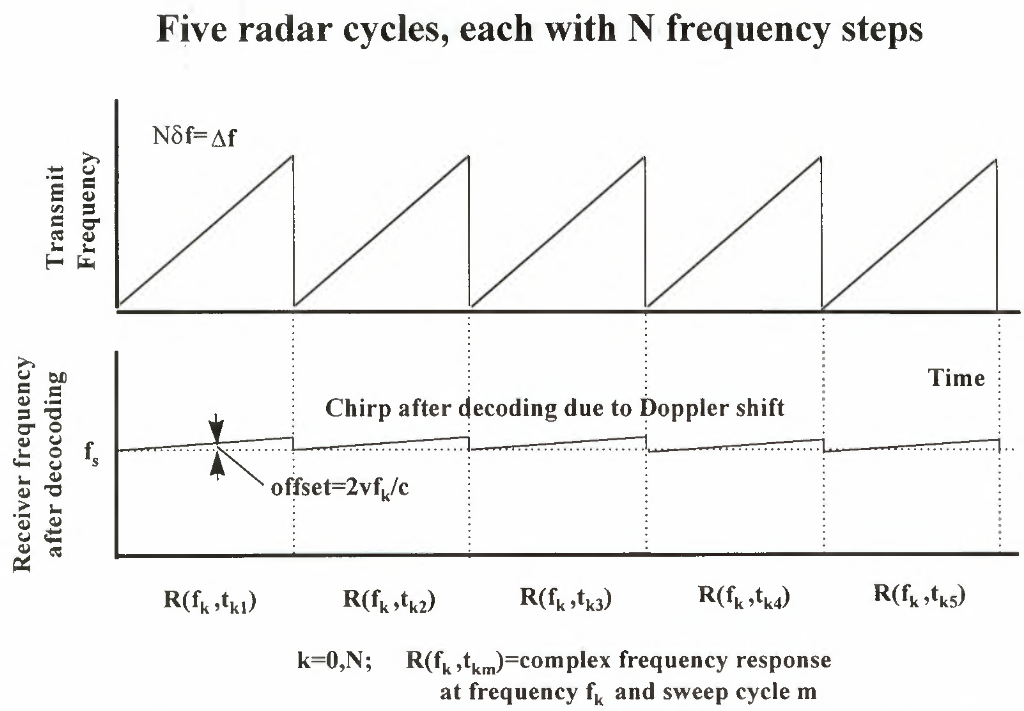


Figure 4 Several frequency modulation sweep cycles with the associated slightly Doppler shifted demodulated signal. From the received signal the time varying frequency response is determined

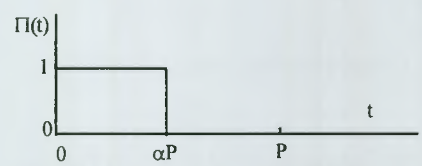
This is the essence of the stepped frequency observation method. In the LWR observations planned for Mars96 the magnitude of the frequency response was to be measured, and the periodic oscillation of the reflectivity caused by the interference of echoes from several levels was to be interpreted in term of a multi-layer model. The complex frequency response measured in the procedure here suggested contains all the data necessary to reproduce the method suggested for Mars96, often referred to as an interferometer method. The impulse response in the case of a time-varying medium, in our case caused by the motion of the satellite, can also be obtained by computing a Fourier transform of the frequency response function, but now considered as a time-varying function of frequency $R(f_k, t_{km})$. The index k pertains to the frequency, the index m to the sweep frequency interval. It is necessary to integrate over several sweep cycles, which is accomplished by summing over m, to remove clutter in the direction of travel of the satellite. This pulse response can either be determined with or without a weighting with frequency. The weighting may be useful if the propagation circuit is dispersive, and de-dispersion is required. The consideration of a dispersive medium may be applicable for Mars´ surface because of the highly frequency dependent attenuation inside the medium. We shall ignore this dispersion in our discussion of the basic principles of the analysis of the measured data. De-dispersion may also be required to correct for the effect of the ionosphere. In the ionospheric case the corrective factor at a frequency f_k is of the form $e^{-i\Delta\Phi(f)}$, where $\Delta\Phi(f)$ for sufficiently high frequencies is approximately proportional to the total electron content between the satellite and the ground.

As the frequency response is measured only at discrete frequencies, there is an ambiguity in range. If the total bandwidth spanned by the frequency steps is $N\delta f = f_N - f_0 = \Delta f$ where N is the number of discrete frequency steps each of size df, there will be an ambiguity in the determination of range which amounts to $\delta r = c/2\delta f$. In the example of Section 3.3 the width of the widest band was 1. MHz, and the number of frequency steps ca 100. This would give a frequency step of 10 kHz and therefore a range ambiguity of the impulse response of 15 km. This ambiguity is not important from the point of view of layer detection because layers as deep as 10 km will not contribute to an echo due to the high attenuation expected, particularly in the highest frequency band where the bandwidth is 1. MHz. It may, however, be important from the point of view of clutter from the surface at large ranges. However, because the surface echoes decline very rapidly with increasing angle of incidence of the transmitted wave, the clutter contribution due to the ambiguous range in most cases is expected to be unimportant. Exceptions will be those areas where there are mountains or deep valleys in the vicinity. It is possible to remove most of the effects of this ambiguity by coding the pulse at each frequency. In order to avoid complications in this explanation of principle, we shall assume that the frequency response is known at discrete frequencies. The time delay response function of $R(f_k,t_{km})$ corresponding to a single sweep through the frequency range, i.e. for m=0 only, becomes:

$$f(\Theta,v)=\int_0^{NP}dt\sum_{l=0}^{N}\sum_{k=0}^{N}\Pi(t-kP)R(k\delta f,t_{k0})e^{2\pi ik\delta ft}.$$
$$\Pi(t-\Theta-lP)e^{-2\pi i(l\delta f(1+2v/c)+f_0 2v/c)(t-\Theta)}$$

Here the amplitude function $\Pi(t)$ can be defined in terms of Heaviside unit step functions as follows:

$\Pi(t)=H(t)-H(t-\alpha P)$
P = the dwell time at any one frequency
N = the number of frequency steps
f_0 = frequency at beginning of sweep
α = fraction of on-time



The sum over l represents the multiplication by a time and velocity shifted replica of the transmission and the summation over k is equivalent to the integration over frequency to obtain the delay variation of the returned signal amplitude.

The function f defined above gives the amplitude response of the process for a single frequency sweep for the time delay θ and for the Doppler velocity υ. The separation in Doppler frequency is usually done in terms of frequency rather than by velocity. The separation by velocity is convenient when the relative bandwidth of the signal is large. With the constant weights applied to the frequency response function implied by this derivation the depth resolution function is of the form sin(Nx)/sin(x). With other weights the sidelobe level can be reduced but at some cost to the resolution.

The effective resolution in depth is $\Delta r=c/2\Delta f$. With the 1. MHz suggested in section 3.3 of this proposal for the highest frequency range, the depth resolution becomes 150 m evaluated in free space. The resolution within the medium is modified because of the slower wave velocity inside the medium, and becomes $\Delta r = c/\mu\, 2\, \Delta f$, where μ is the refractive index of the medium. We shall ignore this difference in the discussion for the time being as the refractive index is unknown. The possibly large discrepancy between the vacuum depth and the real depth must be kept in mind.

For effective clutter rejection coherent processing must be carried out on the data obtained during each frequency sweep. The response function from each sweep is therefore Fourier analyzed frequency by frequency by summing over the times t_{km} when the measurements were made. By using the thus time-averaged frequency responses the echoes can be associated with specific delay and frequency cells. With an integration time of 1 sec, as previously suggested, the frequency resolution will be one Hz, but this will be adjustable at the time of data processing, and will depend on the number of frequency sweep intervals included.

3.4 Clutter Rejection by Filtering.

The clutter which is superimposed on a desired echo at a depth h will here be assumed to originate primarily at the surface. There will also be irregularities embedded in the regolith, which can also give rise to a scattered signal which can mask the nadir echo from a subsurface layer. We shall assume that these contributions are less important than the ones from the actual surface. The clutter echoes from the surface corresponding to the desired echoes from an internal layer at a depth h and a given range resolution have their origin in a surface area bounded by two concentric circles of radii r and r+dr, centered on the nadir point below the satellite, see Figure 5. With the satellite flying at a height above the surface z_0 the angle of incidence corresponding to the circle which produces surface clutter on echoes from the depth h is given by:

$$\cos\phi = z_0/(z_0+h)$$

Figure 6 shows the angle of incidence at a range ring which corresponds to a reflecting layer depth of h. The diagram shows the relationship for two extreme satellite heights, 250 and 1000 km. The angle of incidence is an important parameter because the surface scattering law is a strong function of this angle. The radius of a range ring corresponding to depth h is given by:

$$\rho = z_0\tan\phi=\sqrt{2z_0h+h^2}\approx\sqrt{2z_0h}$$

The width of the area which contributes clutter for returns at the nominal depth h is given by:

$$d\rho = \Delta r/\sin\phi$$

and the surface ring area giving rise to the clutter is:

$$A=2\pi\Delta r(z_0+h)$$

The area of the range ring is, therefore, almost independent of h, assuming h to be much smaller than z_0. It is primarily the scattered power from this area which competes with the echo from a subsurface layer. The Doppler filtering can reduce the effect of this clutter as indicated by the sketch in Figure 7. For a given range ring on the surface, corresponding to a given depth, the maximum Doppler frequency offset occurs at the intersection of the ring and the projection of the satellite orbit and the ring. It is:

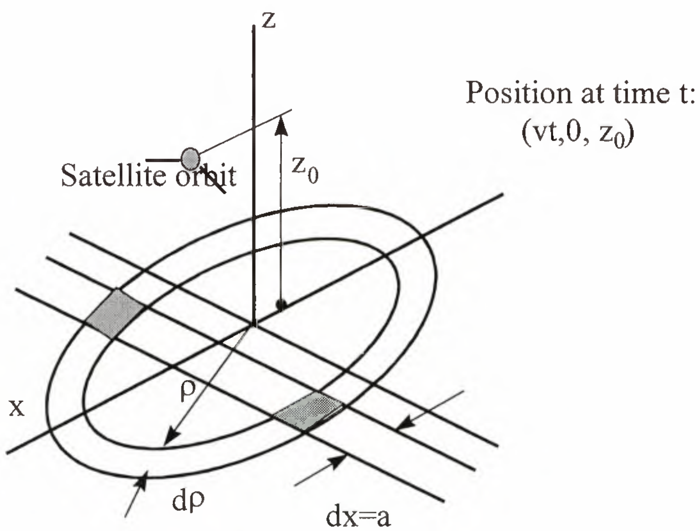


Figure .5 Sketch of satellite with surface areas limited by time delay and constant Doppler frequency lines, shaded gray in the plot

$$\Delta f_{ring} = 2vf_0 \sin\phi/c = 2vf_0\sqrt{2h/z_0}/c$$

In order for a reduction in the clutter to occur the width of the Doppler strip resulting from the filtering (coherent integration) must be less than 2 Δf_{ring}. In fact when the strip is smaller than this limit the relative reduction in the clutter component is approximately given by:

$$\text{relative clutter power} = 2\Delta f/\pi\Delta f_{ring}$$

$$= 2/\pi MNP\,\Delta f_{ring} = c\sqrt{z_0/2h}/f_0 v\,\pi MNP$$

where P is the time duration at one frequency, N the number of frequency steps and M the number of sweep cycles included in the coherent integration, f_0 the center frequency of the sweep and v is the horizontal velocity of

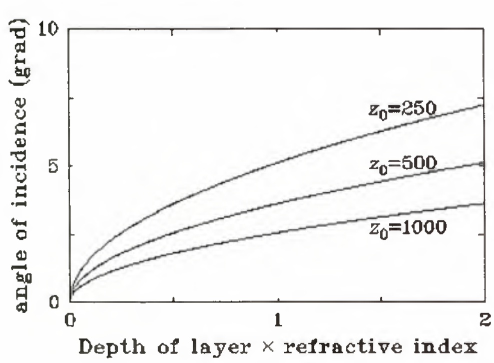


Figure 6 Relation between the angle of incidence of the clutter echoes and the depth of the echo layer the satellite.

One should note that this estimate assumes that there is no folding in of clutter due to frequency ambiguity in the intra-sweep time integration.

4.0 Conclusions

For low frequency observations such as for Mars the stepped frequency radar allows extremely large relative bandwidth to be used because the antenna mismatch can be tuned out during the sweep.

It allows for the combination of the complex frequency responses to be combined into an equivalent pulse response of a resolution which can be chosen after the fact if the raw data can be made available for analysis

It will be feasible to correct for the effect of the ionospheric dispersion by applying a phase correction which can be determined from the data.

Combining repeated frequency sweeps it is possible to remove most of the surface clutter by filtering

By operating the system from a pair of monopoles it may be possible to reduce or even eliminate cross track clutter.

The system described can be applied as an ionospheric topside sounder

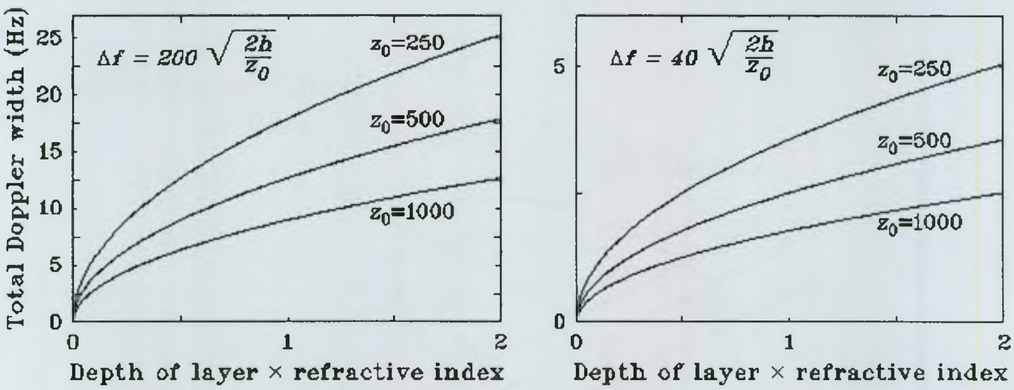


Figure 7 The total Doppler width of the echo from a clutter ring corresponding to a depth h, assuming frequencies of 5 MHz and 1 MHz, and a velocity of 3 km/s.

References

Andrianov, V.A., I.N. Kibardina and R.O. Kuzmin (1993), Deep structures of the dielectric permittivity of the Mars cryolitorphere, *Astronomicheski vestnik*, **27(6)**, 1-11.

Armand, N.A. et al. (1994), Dielectric permittivity of Mars ground at various saline compositions of ice, *Proc. of the Int. Working Meeting of Comparative Planetology (Brown-Vernadsky Microsymp. XX)*, Moscow, **20**, 11-12.

Finkelshtein, M.I., V.L. Mendelson and B.A. Kutev (1977), The Earth sub-surface radar investigation, *Sov. Radio*, 174.

Kliore, A.J. (1992) Radio occultation observations ionospheres of Mars and Venus in "Venus and Mars:Atmosphere, Ionophere and Solar Wind Interaction", *Geophys. Monograph. Ser* **66**(J.G. Luhmann et al. Eds) AGU, Washington DC 265-277

Krupenio, N.N. (1980), Radio investigation of the Moon and Earth-group planets, *Review of recent results in science and technology, Series Astronomy AN SSSR* (in Russian), **18**.

Kuzmin, R.O. (1983), Martian cryolithosphere, *Nauka Press*, Moscow (in Russian), 140.

Olhoeft, G.R. and D.W. Strangeway (1975), Dielectric properties of the first 100 meters of the moon, *Earth and Planetary Science Letters*, **24**, 394.

Paunder, E. (1969), The physics of the ice, *MIR*, *Publ. H.*, 189.

Pettengill, G. et al. (1969), Radar measurement of Martian topography, *Astron. J.*, **74**, 461.

Zent, A.P. and F.P. Fanale (1986), Possible Mars brines: equilibrium and kinetic considerations, *J. Geophys. Res.*, **91 (B4)**, D439-D445.

THE EROS TIME-OF-FLIGHT ANALYSER

S. Livi, M. Hilchenbach, N. Krupp, U. Mall, B. Wilken, and J. Woch

Max Planck Institut für Aeronomie, Max Planck Str. 2, D37191 Katlenburg-Lindau, Germany
Tel: +49-5556-979429 FAX:+49-5556-979139 E-Mail: Livi@linmpi.mpg.de

1. INTRODUCTION

The EROS time-of-flight ion analyzer determines the energy per charge, mass per charge, and direction of flight of each ion separately by combining an ion deflection system (2π field of view), an electrostatic analyzer, and a time-of-flight. The sensor is designed to accept ions from 50 to 15000 eV/e as expected to be typical of plasma observable in a planetary magnetosphere. Instrument based on this technique have been built and qualified for many space mission like for example Ampte (Rosenbauer et al, 1985), Giotto (Wilken et al, 1987), Viking, Ulysses (Gloeckler et al, 1992), Phobos, Mars96, Cluster (Wilken et al, 1997), SOHO (Hoverstadt et al, 1992), Polar, SAC-B (Orsini et al, 1992) and CRRES.

2. SENSOR DESCRIPTION

The EROS instrument is based on a combination of electrostatic filtering, which fixes the energy per charge ratio (E/q) and a time-of-flight (TOF) measurement, which determines the particle's velocity

$$v = \sqrt{2\frac{E/q}{A/q}}$$

with A denoting the particle mass in amu and q the charge in units of the electron charge.

2.1 Entrance optic

Ions enter the sensor through a pair of electrostatic deflection plates, placed in front of the aperture (a, b) (see Figure 1). These plates are, like the whole sensor, cylindrically symmetric about the z-axis. The plates are designed so that the field defined by their shape and by shielding grid (not shown) results in a close proportionality between applied deflection voltages and the resulting deflection angle.

By thus bending the effective instrument look directions within the range up to ± 45°, the sensitivity of the instrument is unaltered and the shape of the instrument function in velocity space varies only slightly. After exiting the deflection plates (a, b) and before entering the electrostatic filter (d, e), particles pass through a mechanical collimator (c). The dimensions of the collimator and the electrostatic filter determine the angular resolution and the geometric factor of the instrument. With an area of 5×5 mm^2 the resolution in the polar angle $\Delta\varepsilon$ is 6° (see MAREMF calibration results in Figure 2) and the geometric factor, with a resolution in azimuth $\Delta\alpha$ of 6° and an efficiency of 40%, will be approximately 1×10^{-3} cm^2 sr/pixel. As shown in the cross-section of the instrument in the incident particles are then further selected by a concentrical, nearly 90° electrostatic filter (d, e). On the exit side of this filter, the ion beams move on a conical surface with a geometric focus on a microchannel plate (MCP (h)Figure 1) in the TOF detection system.

2.2 Time-of-Flight section

The time-of-flight detector consists of two time-zero detectors (TZDs) in a telescope configuration. The front element (START detector) and the rear element (STOP detector) provide start and stop signals, respectively. The start-stop separation in the time domain is the TOF of the particle, the particle's velocity V follows from the measured TOF and the physical separation of the TZDs defining the flight paths. Prior to entering the TOF system the ions are post-accelerated by a potential difference U = -30 kV to increase the energy of the positive ions by the amount q·U. After post-acceleration the ions enter the conical structure of the time-of-flight detection system: the annular entry element (the START detector) provides the start signal for the time measurement and the position information in the azimuthal angle. The signal from the STOP detector in the focal point of the electrostatic filter completes the time measurement.

As indicated in Figure 1, the TOF structure is designed as an inner bubble which is kept on the accelerating potential U = -30 kV by proper isolation (AlNi ceramic) against the outer shell of the instrument which is at ground potential. The power consumption in the high voltage bubble is very low (about 2 mW for the one stage MCP (h)) and no digital or analog electronic parts are within the high voltage bubble.

The START detector is an efficient position sensitive design: a thin polymid foil with a typical

Proc. 32nd ESLAB Symp., 'Remote Sensing Methodology for Earth Observation and Planetary Exploration', ESTEC, Noordwijk, The Netherlands, 15-18 September 1998 (ESA SP-423, December 1998)

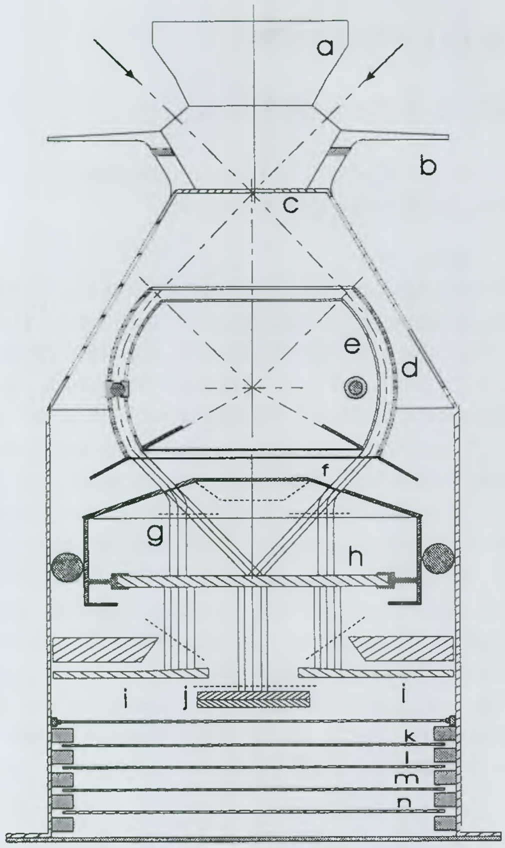


Figure 1 Schematic view of the EROS sensor (not to scale).

Entrance:
a, b aperture (ion deflection system)
c collimator
d E/q analyzer (outer sphere)
e E/q analyzer (inner sphere)

Time-of-flight (TOF) and Detectors:
f entrance foil
g grid
h decoupling, single MCP

i Start Detector (Channeltrons, 64 positions)
j Stop Detector (MCPs)

Electronic boards:
k,l,m,n analog and digital electronic boards

thickness of 2 μg/cm^2 is mounted on a slit on a conical surface such that perpendicular incidence of the E/q filtered ions is ensured for all azimuthal angles. A thin layer of aluminum on both surfaces of the foil prevents charging from particle bombardment. Passage of swift particle leads to the emission of secondary electrons (SE) from the surface of the foil, the number of which is a function of the particles mass and velocity. In the current application the mean number of emitted SE can be expected to vary between 1 and 10 electrons per incident particle. They leave the foil surface within a rather narrow energy band, which peaks at about 3.0 eV. SEs emitted from the foil in the forward direction are accelerated to about 1 keV towards the grid g, deflected and separated from the ions and transferred to the outer area of a circular MCP in the high voltage bubble (marked h in Figure 1) by means of a system of electrostatic fields. The single stage MCP is part of the high voltage bubble; the potential drop across the MCP is about 1.5 kV.

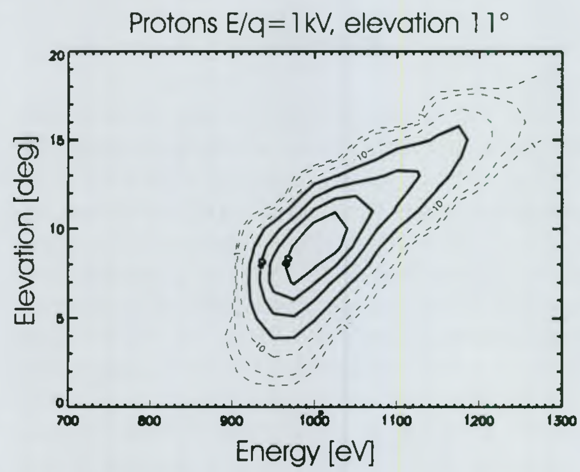


Figure 2 MAREMF calibration of elevation channel #1. Shown is a contour plot of the transmission as function of beam energy and beam direction relative to the instrument. Solid contours are 80%, 60%, 40%, and 20% of the maximum (solid lines) and 10%, 3%, and 1% (dotted lines). The MAREMF sensor, one of the predecessors of the EROS ion sensor, was calibrated at the MPAe calibration facilities. The geometry of the deflection plates and energy/charge analyzer is the same as that foreseen for the EROS sensor. Calibration results showed an angular resolution in the polar direction of 6° (FWHM) and an energy resolution of 14% (FWHM).

On the output side, after moderate amplification of about g = 1000, the electrons are accelerated by the bubble high voltage and impact on an annular detector array at ground potential. The electron optics transferring the SEs from the foil to ground potential is designed to maintain the azimuthal angle information contained in their position of origin on the foil. The annular detector array measures the azimuthal position with a resolution of 6°. As indicated by the ion and electron paths in Figure 1, the position on the START channeltrons (i) maps back to the azimuth of incidence in the entrance collimator (c).

In addition to the azimuthal information, the described START detector provides also the start signal for the TOF measurement (the electron transfer time represents a rather small and constant offset for the time measurement). Incident ions leave the foil with nearly unchanged energy and move towards the central area of the back MCP (h) in the bubble. This area of the MCP is part of the STOP detector: on ion impact an electron avalanche is started in a single MCP channel. The emerging electron cloud on the exit side of MCP (h) is accelerated by the bubble potential and focused to a high-gain MCP assembly in a chevron configuration (marked j in Figure 1) which is essentially on ground potential. The output signal represents the stop signal for the TOF measurement. Start, stop and direction informations are evaluated in the follow-on electronic circuitry, which is entirely on ground potential.

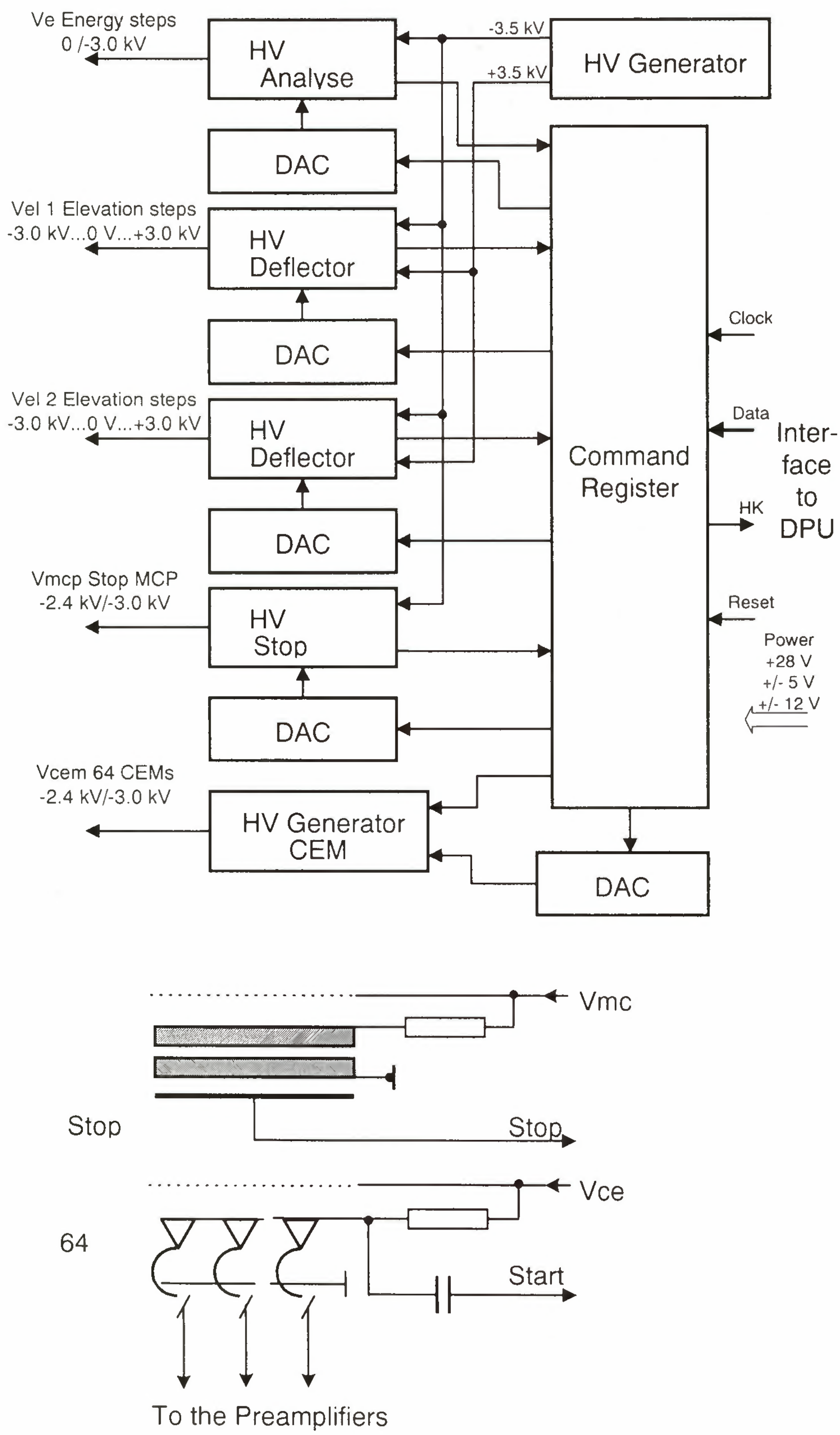


Figure 3: High Voltages Block Diagram for deflection, E/q stepping, and detectors

2.3 HV generators

HV circuits to be mounted inside the sensor box (in lower part) consists on five HV supplies of up to 3.5kV and the up to -30kV for the bubble high voltage (for reference see Figure 3):

- Vcem - common HV supply for 64 channeltrons. The value of the HV could be changed by instrument telecommand in the range from -2.4 to -3.0kV. Because of the large number of the channeltrons (64) the required power for the HV generator 2 is 800mW.

- Vmcp - supply for STOP MCP. The voltage of this circuit could be changed from -2.4 to -3.0 kV by telecommand in case of changing MCP multiplication factor. The required power is 200 mW.

- Ve, Vel 1, Vel 2 - three HV sources for supplying analyzer (voltage range is from a few V to -3.0 kV) and deflector electrodes (voltage range between ±3.0kV), respectively.

To optimize and reduce power consumption, the high voltages Vmcp, Ve, Vel 1, and Vel 2 will be generated from the same HV stabilized converter (generator 2) while each voltage is controlled independently through digital/analog converter (DAC) and HV control circuits.

2.4 Post acceleration power supply

The post acceleration power supply is identical in form and characteristics to the one successfully flown in the instrument MICS on the satellite CRRES and in the instrument CAMMICE on Polar. The MICS power supply operated correctly through the whole life of the CRRES mission. CAMMICE has operated on the highest possible voltage since start, and is still delivering excellent data. The value of the high voltage can be selected from the DPU through 4 digital lines, to allow a slow conditioning of the insulators when changing voltages.

2.5 Charge preamplifier and level detector

This subsystem will front-end the particle sensor detector by handling the 64 channeltrons devices (each corresponding to a specific input direction, see Figure 4). The first part will be composed by 64 charge preamplifiers, all integrated in Large Scale Integration (LSI) devices; furthermore, a pulse shaping & level detector will post-process the signal. The most crucial operation will consist in a hardware real time process of the true signals detection: the logic will be able to discriminate the real incoming direction even in the presence of spurious or multiple signals. The same logic will enable/disable (by experiment telecommand) one or more input channels in case of failure or sensor degradation. According to the design objective, the dynamic range of the charge detection electronics will be from 16 fC to 16 pC, and the required power will be 20 mW per channel.

2.6 Digitized direction

This subsystem will calculate the true particle input direction by delivering information on a digital word to be transmitted to the Digital Processing Unit (DPU). The design goal will be to handle more than 25.000 true direction events per unit time.

Digitized Time of Flight (TOF): The TOF electronics design will be very innovative. Considering the TOF particles range between 1 nS and 200 nS, it is extremely difficult to handle such a signal by a standard analog electronics at LSI integration level with a very low dissipated power. The new design will be based on a flip-flop chain; each cell will have a defined time-delay. The START signal coming from the sensor will trigger the chain that will count quickly up to the STOP pulse. The final count (TOF) will be carried out by digitally reading the flip-flop outputs. In order to minimize timing tolerance, the integrated LSI device will be fully temperature controlled. This system has been already tested and developed at the University of Braunschweig (Germany), and the present design phase foresees that it will be integrated with LSI technology.

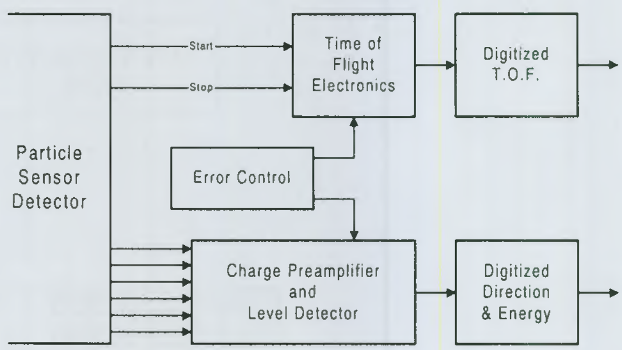


Figure 4 Front End Electronic Block Diagram

2.7 Sensor characteristics

The sensor mass resolution depends critically on the scattering in the foil. Figure 5 shows in-flight data collected in May 1997 by the instrument CAMMICE/MICS on Polar. This instrument employed the same detection technique as EROS, namely electrostatic selection of the energy per charge, followed by post acceleration and time of flight measurement. The geometry of the time of flight system is the same as the one proposed for the EROS sensor. The resolution in time of flight is equivalent to a mass resolution $m/\Delta m \approx 6$ at $m = 4$ (Helium) and $m/\Delta m \approx 4$ at $m = 16$ (Oxygen).

It must be noted that the foil employed in this experiment (5.5 $\mu g/cm^2$) was much thicker than the one proposed for EROS. Foils with 2 $\mu g/cm^2$ will improve the mass resolution of EROS as compared to CAMMICE/MICS on Polar.

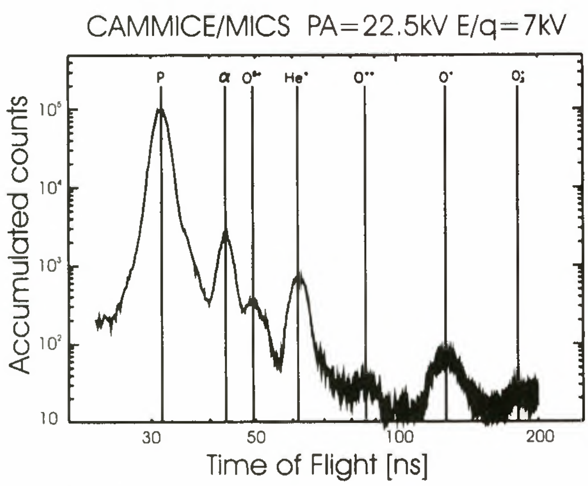


Figure 5 In-flight data collected with the instrument CAMMICE/MICS on Polar in May 1997. In the plot, the energy of the ions impinging on the foil was 30kV, equivalent to the lowest energy per charge channel of EROS.

The EROS sensor parameters, as derived from calibration values of the MAREMF and MICS sensors, are summarized in Table 1.

	Number of elements	Resolution	Technique of determination
Polar angle	16	6°	Deflection
Azimuth angle	64	6°	multiple Channeltrons
Field of view		90°×360°	
Energy/charge resolution	32	14%	Electrostatic filter
Energy/charge range		50-15000 eV/e	Electrostatic filter
Mass/charge $m/q < 10$		$m/\Delta m = 6$	time of flight
Mass/charge $m/q > 10$		$m/\Delta m = 4$	time of flight
Geometric Factor/pixel		$\approx 1*10^{-3}$ cm^2sr	

Table 1: Sensor characteristics

3. CONCLUSIONS

The very high angular and energy resolution of EROS make this sensor particularly adapted in the field of research of planetary magnetospheres. In particular, the high angular coverage (half the unit sphere) allows studies of the interaction of the solar wind with the planetary ionosphere.

The EROS sensor has been designed with particular focus at space applications, i.e. low mass and low power. The required spacecraft resources for a fully self-contained EROS instrument are summarized in Table 2.

Mass	3700	g
Power	3300	mW
Dimensions	18×18×35	cm^3
Data Rate	Max 3.3	kbit/sec
	Ave 1.5	kbit/sec
	Min 0.5	kbit/sec

Table 2: Required spacecraft resources

The combination of LSI with a compact and novel sensor head design led to an instrument that has at the same time high angular, energy, and temporal resolution and low impact on the spacecraft resources.

ACKNOWLEDGEMENTS

The authors acknowledge the enthusiastic support and the fruitful discussion with Istvan Apathy (KFKI-AEKI), Gerhard Berghofer (IWF), Maurizio Maggi (IFSI), Anatoly Remizov (IKI), and Sandor Salai (KFKI-RMKI).

REFERENCES

Gloeckler, G., J. Geiss, H. Balsiger, P. Bedini, J.C. Cain, J. Fisher, L.A. Fisk, A.B. Galvin, F. Gliem, D.C. Hamilton, J.V. Hollweg, F.M. Ipavich, R. Joos, S. Livi, R. Lundgren, U. Mall, J.F. McKenzie, K.W. Ogilvie, F. Ottens, W. Rieck, E.O. Tums, R. von Steiger, W. Weiss, and B. Wilken, The solar wind ion composition spectrometer, Astronomy & Astrophysics Supplement Series 92, 267-289, 1992

Hoverstadt, D., M. Hilchenbach, A. Burgi, B. Klecker, P. Laeverenz, M. Schoeler, H. Guenwaldt, W.I. Axford, S. Livi, E. Marsch, B. Wilken, H.P. Winterhoff, F.M. Ipavich, P. Bedini, M.A. Coplan, A.B. Galvin, G. Gloeckler, P. Bochsler, H. Balsiger, J. Fisher, J. Geiss, R. Kallenbach, P. Wurz, K.H. Reiche, F. Gliem, D.L. Judge, K.C. Hsieh, E. Möbius, M.A. Lee, G.G. Managadze, M.I. Verigin, and M. Neugebauer, CELIAS - charge and isotope analysis system for SOHO, 16 ESLAB

Symposium: Study of the Solar terrestrial system, 343, Sept. 1992

Orsini, S., M. Candidi, M. Maggi, K.C. Hsieh, C.C. Curtis, A.M. Hudor, S. Livi, B. Wilken, I.A. Daglis, E. Flamini, B. Negri, B.A. Tinsley, and M.A. Gruntman, Proposal of an Italian experiment for the mission SAC-B. ISENA: Imaging particle spectrometer for energetic neutral atoms, Proceed. of the 26th ESLAB Symp., Study of the solar-terrestrial system, Killarney, Irland, 6.-19.6.1992, ESA SP 346, Sept. 1992

Rosenbauer, H., H. Grunwaldt, S. Livi, M. Witte, and R. Zwick, The mass-separating ion spectrometer on the AMPTE ion-release module, IEEE Trans. Geosci. Remote Sensing, GE-23, No. 3, 280-286, 1985

Wilken, B., W. Weiß, W. Stüdemann, and N. Hasebe, The Giotto implanted ion spectrometer (iis): Physics and technique of detection, J. Phys. E: Sci. Instrum. 20, 778-785, 1987

Wilken, B., W.I. Axford, I. Daglis, P. Daly, W. Güttler, W.H. Ip, A. Korth, G. Kremser, S. Livi, V.M. Vasyliunas, J. Woch, D. Baker, R.D. Belian, J.B. Blake, J.F. Fennell, L.R. Lyons, H. Borg, T.A. Fritz, F. Gliem, R. Ratje, M. Grande, D. Hall, K. Kecsuemety, S. McKenna-Lawlor, K. Mursula, P. Tanskanen, Z. Pu, I. Sandal, E.T. Sarris, M. Scholer, M. Schulz, F. Sørass, and S. Ullaland, RAPID The imaging energetic particle spectrometer on Cluster, Space Science Rev. 79, 399-473, 1997

THE DEVELOPMENT OF EUROPEAN SAR AND ASSOCIATED TECHNOLOGIES AND USAGES

Eur Ing D.J.Q.Carter, M.A.B.Cohen, Dr G.G.Cox, M.D.Gibbons, M.Nunn, K.Morgan

Matra Marconi Space UK, Anchorage Road, Portsmouth, Hants, PO3 5PU
Tel: (0) 1705 704970 (8290 fax) E-Mail: davidj.carter@mmsuk.co.uk

ABSTRACT

The move of Radars into space started, from a European standpoint, in 1982 with the beginning of the European Space Agency's (ESA) first Remote Sensing Satellite (ERS-1) programme. This culminated with a launch in July 1991 of the first European Synthetic Aperture Radar (SAR).

This paper provides an overview of the systems and technologies included in the ERS and ENVISAT radars. Following this is an explanation of the CORE Radar development path, which is being followed to realise radar products capable of allowing competitive participation in world-wide markets, is provided. Finally an outline of the technologies proposed for the next generation of spaceborne radars is presented.

Key words: SAR; core radar; front end; central electronics; AMI; ASAR.

1. INTRODUCTION

The ERS project laid down many of the foundations upon which all future ESA radar programmes have been based. These include the second satellite (ERS-2) launched in April 1995, and the Advanced SAR (ASAR) being built currently and due for launch in 1999 as part of the ESA Environmental Satellite (ENVISAT-1).

Canada's Radarsat uses fundamentally the same techniques and designs as the ERS systems, but has an objective to service an operational need compared to the Scientific and Research aims against which ESA's ERS and ENVISAT-1 programmes have been developed.

The future of the Spaceborne Radar business lies in exploiting the well known principles and techniques. Such exploitation needs to produce both technically and financially efficient solutions as well as providing performance that produces a serviceable product.

2. ERS ACTIVE MICROWAVE INSTRUMENTATION (AMI)

ERS-1 was the first radar mission developed and flown by ESA. The principle objectives of the mission were oriented to oceanographic rather than land applications. The SAR applications were in their early stages, as was the scatterometer wind speed and direction usage in weather forecasting. As such the programme was very much a Research and Development project, both in terms of the instrumentation and technologies being flown but also from an applications stand point.

The AMI is a microwave package comprising 2 radars; a SAR and a wind scatterometer. The main reasons for the combining of the two into a single package were the volume and cost reductions compared to having two completely separate instruments.

Presented in Table 2-1 is the overall in-orbit radiometric performance achieved for the ERS-1 AMI SAR. The performance exceeded the requirements initially set and reflects the extensive effort and iterations in design which were undertaken during this frontier stretching programme.

Table 2-1 - In-orbit Performance Summary - AMI and ASAR.

		ERS, AMI-SAR	ENVISAT-1 ASAR			
Performance	Modes Units	Image	Image	Wide Swath	Alternating Polarisation	Global Monitoring
Polarisation	-	VV	VV or HH	VV or HH	VV,VH,HV,HH	VV or HH
Radiometric Resolution	dB	1.8	1.8	1.6	1.6	1.4
Sensitivity (NESigma0)	dB	-	-20	-20	-20	-28
Spatial Resolution	m	27.8/ 25.6	27.5/ 29.6	148.1/ 145.7	28.6/ 29.6	798.1/ 876
Radiometric Stability	dB	0.25	0.4	0.4	0.5	0.5
Radiometric Accuracy	dB	-	1.4	1.4	1.8	1.7
Swath Widths	Km	100	56-105	406	56-105	406
Localisation Accuracy (Range)	Km		0.100	0.075	0.075	0.076
Localisation Accuracy (Az)	Km	0.07	0.071	0.071	0.077	0.072

Proc. 32nd ESLAB Symp., 'Remote Sensing Methodology for Earth Observation and Planetary Exploration', ESTEC, Noordwijk, The Netherlands, 15-18 September 1998 (ESA SP-423, December 1998)

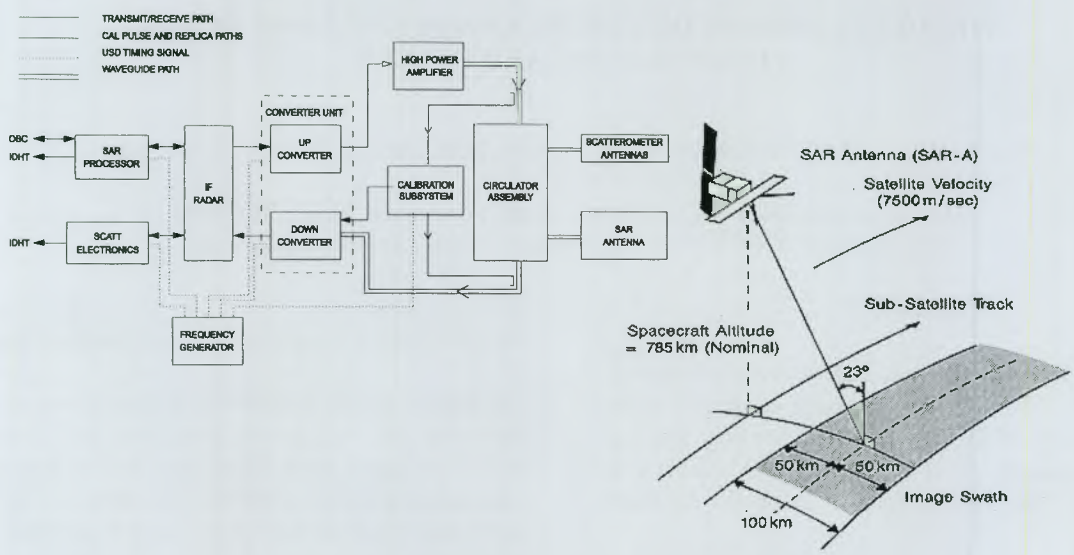


Figure 2.1 - ERS AMI Geometry and Block Diagram

Operating at 5.3Ghz, the SAR has two modes in which measurements are made; the Image mode and the Wave mode.

The former is the one which produces radar images, whereas the latter produces data from which the ocean significant wave height spectra can be derived. Figure 2-1 illustrates the geometry of the single fixed beam used in Image mode and also provides a simplified block diagram of the AMI. The main characteristics of the AMI SAR are given in Table 2-2.

Specific design features and the major technologies are summarised in Table 2-3. This is by no means an exhaustive list of all design considerations but it does draw out the more pertinent points.

The need to maintain a stable transmit and receive path in terms both of amplitude and phase errors is extremely demanding.

Table 2-2 - AMI SAR Main Characteristics

Parameter	Value
Orbit Altitude/Inclination	Nom 785 Km / 98.4°
Orbital Duty Ratio (Mission Limit)	14%
Transmit Centre Frequency	5.3 Ghz
Antenna Dimensions	10m x 1m
SAR Payload Mass	387Kg
PRF Range	1640-1720 Hz
Transmit Pulse Width	37.12μs
Incidence angle coverage	Centred on 23 deg
DC Power Consumption (Image)	1200 W
Data-rate (Image)	105Mbps

By incorporating into the design an internal calibration loop, and by running the High Power Amplifiers in saturation, combined with careful design in all equipments, the performance was achieved.

As can be determined from Table 2-3, all parts of the SAR contained design features and technology challenges. These all had to be solved in order to achieve the performance demanded.

Among the parameters causing perturbations to the image quality of the SAR, there are those which can gain advantage from a system approach. By characterisation of the properties prior to launch, correction algorithms were implemented in the image processor. This enabled a realisable space segment solution by sharing the pressure optimally across the total system.

3. ENVISAT-1 ADVANCED SAR

Following on from the highly successful ERS-1 and -2 missions, a third much more extensive mission is in preparation. This is the ENVISAT-1 mission, for which the objectives are not only to provide continuity of data sets being collected in the ERS missions, but also to increase the scope of observations.

A major part of the payload of this 8 Tonne spacecraft is the Advanced Synthetic Aperture Radar, ASAR. The requirements for ASAR were set to match the spatial resolution of ERS and to continue operating at 5.3Ghz.

The extensions afforded by ASAR compared to the ERS series are:

- Dual co and Cross polar measurements
- Incidence angle coverage increase to 15 - 60 degrees
- Wide swath imaging up to 500km

Table 2-3 - Principal Design Features and Technologies - ERS-1 SAR

Design Features	Technology Impact	Main Performance Implications
Signal Isolation Between Transmit and Receive LO Chains	High isolation interconnections (>70dB) or additional filtering.	Radiometric stability
Centralised High Power Amplification	Very high voltages and powers to be handled. Life-time and switching difficulties. Centralised heat dissipation leading to use of heat flow systems	Multipaction. Mass.
Multipaction	Slot design of primary power divider in transmit feed network. Power cable designs.	Radiometric Resolution.
Transmit Pulse Chirp Generation by use of Surface Acoustic Wave (SAW) Device	Temperature stability essential so need for ovenised environment. On-board range compression means matched-pair essential.	Increased design complexity and production risk. Overall Radiometric performance. (spatial and radiometric resolution, stability, ambiguities...)
Antenna Planarity ±4mm	Use of low CTE materials (CFRP) in radiating array plus need for light weight deployable rigid support structure.	Radiometric Stability, ambiguity and Radiometric Resolution.
Receive Signal Detection Scheme with good and stable performance Linearity ≤5% I+Q Orthogonality ≤ 5 ° I+Q gain imbalance ≤0.1dB	At the dynamic range (32dB) and bandwidths needed (16Mhz) technology dictated use of 20Mhz 8bit ADC to meet linearity and a bit stripper to meet data rate requirements (105Mb/s).	Overall image quality. Pre-launch characterisation and subsequent ground correction is a realisable option to restore quality. Radiometric Resolution and Stability.
Internal calibration scheme to monitor short and longer term gain and phase variations of the payload SAR instrument	High stability attenuators (<0.05dB), calibration couplers (<0.01dB) and ovenised delay lines (SAW devices).	Radiometric stability. Mass.

- Low data rate global monitoring operation with 1km resolution

In Table 2-1. the main in-orbit performance requirements for ASAR are given along side those for the ERS SAR. The Spatial Resolution for the Image mode was specifically limited to 30 metres to match ERS although design studies proved that finer resolutions were clearly achievable with the technology being used.

Flexibility of operation is a feature of ASAR which enables many of the extensions compared to its predecessors. Figure 3-1 illustrates the various modes of ASAR. The chief characteristics of the ASAR instrument are presented in Table 3-1

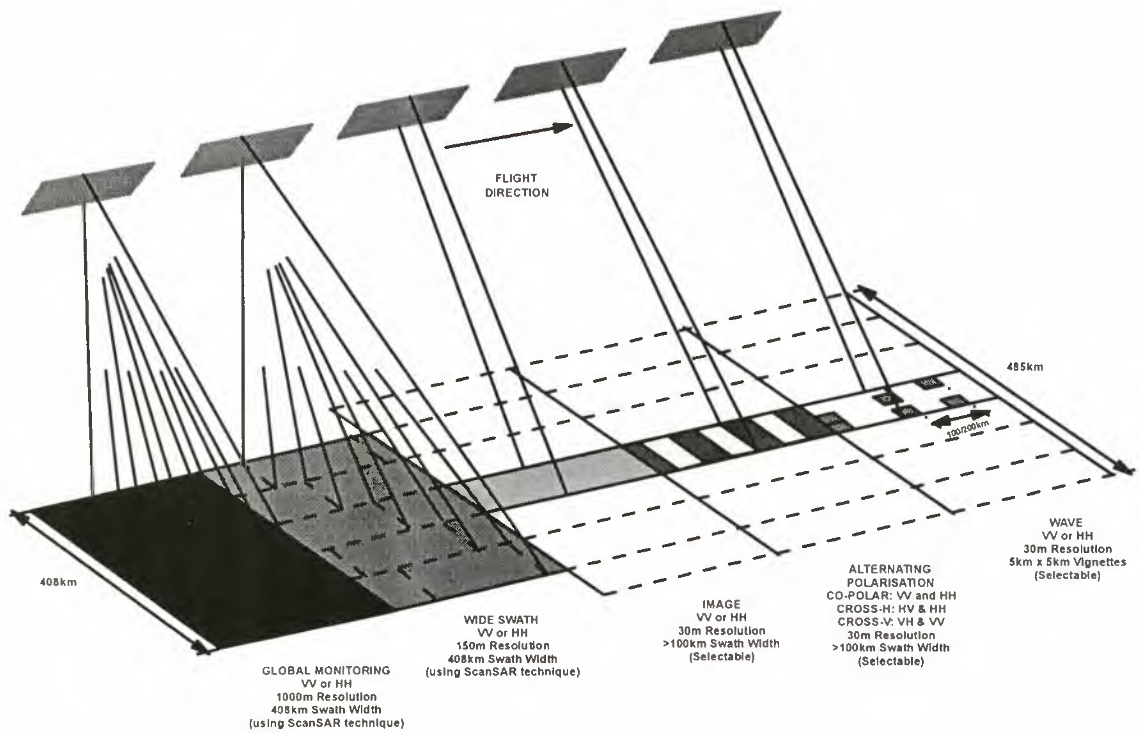


Figure 3-1 - Operation Modes of ASAR

Table 3-1 - ENIVISAT ASAR Main Characteristics

Parameter	Value
Orbit Altitude / Inclination	Nom 800Km / 98.4°
Orbital Duty Ratio (Mission Limit)	20%
Transmit Centre Frequency	5.3 Ghz
Antenna Dimensions	10m x 1.4 m
SAR Payload Mass	820 Kg
PRF Range	1580-2150 Hz
Transmit Pulse Width	20, 15, 20, 19 μs
Incidence angle coverage	15-45 / 17-43 / 15-45 / 17-43 Degrees
DC Power Consumption (mean)	1365, 1200, 1395, 712 W
Datarate (high)	100 Mbps
Data rate (low)	0.9 Mbps

* Where different by mode, order of values is Image, Wide Swath, Alternating Polarisation, Global Monitoring.

Significant advances in the design and technology were incorporated into ASAR compared to ERS AMI. The main difference is that the Antenna for ASAR is an active phased array with tri-plate annular slot radaiting array and Transmit/Receive modules. There are several other important design features in the ASAR realisation and these are summarised in Table 3-2.

Although the main technology challenges were encountered in the Antenna Subassembly, improvements in aspects of the central electronics have also been achieved. These have included the internal calibration system, the transmit signal generator and the receive signal detection quantiser where a Block Floating Point Quantiser (BFPQ) has been implemented.

4. SAR TECHNOLOGY DEVELOPMENT PATH

During the latter part of the 1980's it became increasingly evident that the then standard way of putting a spaceborne SAR system into place was heading towards untenable levels of cost.

More recently European Industry has generally seen the level of government contributions to space declining. This has brought attention to focussing developments to specific objectives. These have been more commercially orientated then previously.

Within ESA a two stream programme to satisfy the Living Planet Theme has been determined. These are the Earth Explorer and Earth Watch programmes. The latter of these is aiming at supporting industry in establishing operationally commercial Earth Observation systems.

MMS embarked on an in-house funded programme to determine the correct direction to take to ensure a future for the business.

Table 3-2 - Principal Design Features and Technology - ENVISAT ASAR

Design Features	Technology Impact	Main Performance Implications
Transmit Pulse Chirp Generation - Digital	Device speed. Method of Production. Waveform continuity at cycle boundaries. No need for specific temperature control.	Radiometric Accuracy, Data rate, Ambiguities
Blcok Adaptive Quantiser for Signal Detection	Speed of devices. Simpler overall design to ERS.	Data rate, Sensitivity
Distributed Electronics over antenna aperture	Signal and power distribution/connectivity. Thermal environment for devices/equipments Thermal dissipation	Radiometric Accuracy, overall mass. Sensitivity, operation duty ratio.
Tri-plate Antenna RF Radiator	Cross-polar purity and isolation of feeds. Ease of manufacture and repeatability.	Cross polarisation sensitivity, radiometric accuracy, mass.
Transmit/Receive modules	Close packaging of RF components in conjunction with high gaijn requirements. Accurate and stable Phase and Amplitude levels leading to temperature monitoring and compensation system.	Mass, Sensitivity, Radiometric Accuracy, Spatial resolution, overall radiometric performance.
Internal calibration scheme to monitor short and longer term gain and phase variations of the payload SAR instrument	High stability attenuators (<0.05dB), calibration couplers (<0.01dB) and ovenised delay lines (SAW devices).	Radiometric stability. Mass.

Table 4-1. Summary Applications Based SAR Requirements

Potential Applications	Crop area, boundaries, classification, condition	EEZ surveillance	Forest	General mapping	Geology	Land use, land cover	Sea-ice mapping	Law Enforcement	Topo-graphy
Frequency band	Dual L, S, C, X	Single L, C	Single L	Single C, X	Single/Dual L, C, X	Single/Dual L, S, C, X	Single/Dual L, C, X	Single/Dual S,X	Single L, S, C, X
Polarisation	Multi HH,HV	Multi VV, VH	Single VV,HV	Dual HH,VV,HV	Single HH/VV	Multi HH,VV,HV	Multi HH,VV,HV	Single VV or HH	Single VV or HH
Spatial resolution (m)	5-30	30	10-30 5 or less for boundaries	30	50	30	30-100	<3	<3
Radiometric resolution (dB)	2	2	2	1	1.5	2	1	3	2
Incidence angle (°)	20-50	20-40	10-35	30-50	25-50	30-50	25-50	20-60	25-50
Revisit time (days)	7-30	7	30-90	30-90	30-90	30-90	<3 Polar orbit required	<1	3

In Table 4-1 a much summarised version of the demands from a SAR system to cover all market types is presented. It can be seen that there is no single frequency, or indeed set of performance requirements, which can satisfy all the needs.

There are fundamentally two classes of satellites which need to be offered. One is a small satellite, in the 1 to 1.5 tonne class, which can be serviced by the lower capacity end of the Launchers and satellite platforms. This keeps the overall cost to a minimum. This is referred to as a regional system.

For wider global usage, and for multifrequency greater coverage systems, 2 tonne plus systems will still be required. Cost reductions are also necessary in these classes.

In order to realise an affordable solution to the requirements identified above, the cost of spaceborne SARs need to be addressed. Key factors which have to be satisfied are:

- Control risk
- Reduce cost Non Recurring Elements (NRE),
- Improve work practices compared to previous ESA programmes
- Utilise experience gained including from areas outside the immediate Spaceborne SAR field
- Reduce programme durations

All developments have to be focused on a need and not just an interesting technical exercise if affordable systems are the target. Once the enabling technologies and units are in place there is then no need to follow the previously lenghty ESA programme routes.

The risks associated with committing to flight design and manufacture at programme start are very low. A development programme and philosophy was established in MMS for the SAR work. This is illustrated in Figure 4-1.

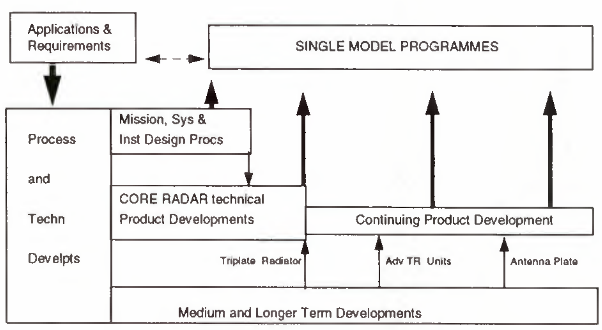


Figure 4-1: Overall Development Flow

There are two lines of development; one for the short term market opportunities, and one looking to longer term needs of the business. A lower level of funding is placed on the second line initially, but as the first is completed so sources, including initial profits are directed at the next need. A continued awareness and updating of capabilities is necessary to maintain a market lead. This is particularly important as many of the infrastructure market opportunities are coupled to technology transfer.

The main initial implementation was the CORE Radar programme which is described in the following section.

5. CORE RADAR PROGRAMME

The Core Radar programme is the MMS third generation development of radar related products intended to satisfy the requirements of a wide variety of applications and customers. Whilst the initial motivation has been aimed at spaceborne SAR missions,

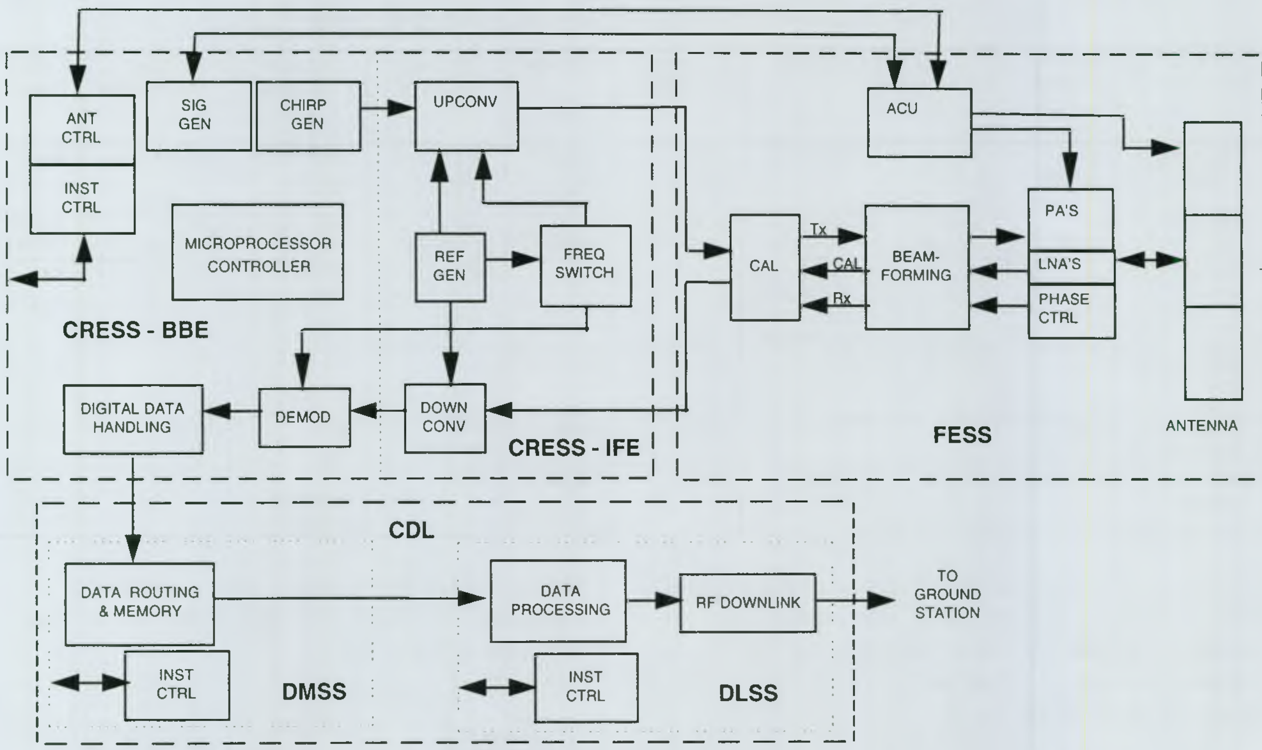


Figure 5-1: Core Radar Architecture

many parts of the system are also applicable to other instrument types and non space missions.

In addition to these 'customer' driven requirements, MMS has placed its own requirements onto the system — applicability to large and 'small' SARs, affordability, short timescales — hence the drive for Core to result in volume, mass and cost efficient generic products which can be applied in a modular, expandable and flexible fashion to many missions.

The Core Radar Product can be conveniently divided into three sections as shown in Figure 5-1:

- Core Radar Electronics Subsystem (CRESS) consisting of the Intermediate Frequency Equipment (IFE) and Baseband Equipment (BBE)
- Front End Subsystem (FESS) providing Tx/Rx Amplification / Calibration, Radiating Antenna and Antenna Control Unit (ACU).
- Core Data Link (CDL) providing Data Management (DMSS) and Down Link (DLSS) subsystems.

The Core Radar Electronics Subsystem is a group of modules and equipment's which, when selectively integrated, form the generic component of a number of radar instruments. The subsystem as a whole is therefore mission independent, able to transmit and receive Radio Frequency signals at a variety of frequencies with a number of bandwidths available. In particular, the subsystem's main flexible points are:

- Ability to operate at different frequency bands (L, S, C, X and Ku)
- Expandability to include a variable number of transmit and/or receive channels (for multiple polarisation's, multi frequency, different operating techniques etc.)
- Capability of handling wide bandwidth signals (currently up to 300 MHz)
- Ability to interface to a range of platform standards
- Ability to interface to a range of active antenna types in the Front End Subsystem

The Front End Subsystem is used to boost the low level RF signal with sufficient gain to overcome link and target losses. Phase and gain control allows beamforming and steering options. Calibration of the FESS is achieved using a three stage calibration pulse system inherited from previous successful projects. Control of all stages is achieved using the ACU. The ACU is in turn controlled by the CRESS and acts in a signal distribution role with an expansion capability to increase the scope of the generic product. Antenna deployment power control is handled by the ACU which takes its cue from the BBE. Temperature compensation is achieved at module level.
The CDL comprises the Data Management and Storage Subsystem and the Downlink Subsystem . The purpose

of the DMSS is to receive Science data from the SAR instrument and either store the data in its mass memory or pass the data to the DLSS for transmission to ground. The average data rates start 100Mbit $^{-1}$ and are likely to reach as far as 400Mbits^{-1} in the medium term. If the data is stored in memory it is then passed to the DLSS at some later time in accordance with telecommanded instructions. There is the potential for more input channels to the DMSS than output channels and, consequently, the DMSS has to perform multiplexing of parallel input data channels into output channels. The purpose of the DLSS is to receive the Science data from the DMSS and transmit it to ground.

The Core Radar Development programme has two major objectives :

- To reduce the technology risk to the point that allows a competitive single model protoflight programme to start.
- To demonstrate the critical elements of the radar in an electrical model test bed.

Within the development programme the CRESS has been completed in 1997, the FESS power amplifier and radiating element technologies have been evaluated and an X band system is being constructed in 1998 and the CDL has been investigated in terms of breadboarding in critical areas and procurement of standard products.

6. FUTURE SAR TECHNOLOGY AND DRIVERS

Future development of the CORE Radar product concept will aim to exploit recent advances in technology to achieve improved performance, lower mass and/or lower cost. This in turn will offer a choice between enhanced instrument functionality/performance and improved affordability.

The trends in the SAR application market indicate mixes of multi-frequency, multi-polar and high resolution instruments. The flexibility of the Transmit/Receive (TR) module is well recognised both for the beam-steering agility which it offers and for the improved sensitivity potential which results from the placement of transmit amplification as close to the radiating sub-arrays as possible.

As a direct result, a key area for improvement is the Transmit/Receive module, which can now benefit from recent advances in MMIC (Microwave Monolithic Integrated Circuit) and MCM (Multi-Chip Modules) technology. TR modules are now feasible in very significantly smaller packages than used on previous SAR flight programmes, with improved reliability and comparable performance. These packages can include a custom Application Specific Integrated Circuit (ASIC) which provides sufficient command and control functionality to allow direct communication with the Central Electronics. This eliminates the need for the traditional processor-based additional subsystems on the antenna (e.g. ASAR Tile Control Interface Units).

Radiator technology is another area where improvements can benefit future missions. For multi-polar instruments the further development of patch antenna sub-arrays incorporating new dielectric materials with greater structural integrity offers the potential for reducing the antenna structure mass overhead. For single-polar applications, waveguide sub-arrays may remain the preferred approach, and improvements in waveguide feed technology and again the exploitation of structural rigidity offer potential benefits. In all cases, the goal is to develop a low mass, low cost reliable and manufacturable product.

One of the issues concerning the design of the SAR system is to provide compact RF front ends that integrate well with the RF radiating assembly, the antenna array, and produce a much lower mass per unit area for the assembly. Packaging considerations for the T/R modules has led to smaller electronics where the electronic circuits can now be made comparable in cross-section to, or smaller than the radiating element (say 15 mm at X-Band) and at higher powers. However, power supplies are still seen as being too large (requiring a large capacitor), even though the available control electronics has become much smaller and more integrated. Higher integration of the electronics will lead to reduced packaging mass contributions, with issues for component screening appearing.

Some radar designs, particularly for airborne radar applications, require the electronics to stack away from the radiator, making the design of small cross-section, quite deep, but allowing circulating mediums to be used for the cooling process. For the space SAR, the emphasis probably needs to be in the direction of co-planar mounting of the electronics, so that the considerable surface area of the array can be used as the heat radiator. This not only implies that the T/R modules need to be small, but distributed in a plane making them low in height. This technology appears available today, but not in current space applications.

The antenna radiators would need to take on the role of thermal radiators (radiating towards the earth), and may even become non-planar themselves to allow better radiative cooling of the electronics. Thermally efficient antenna radiators, or non planar radiators for dual linear operation will produce their own constructional problems.

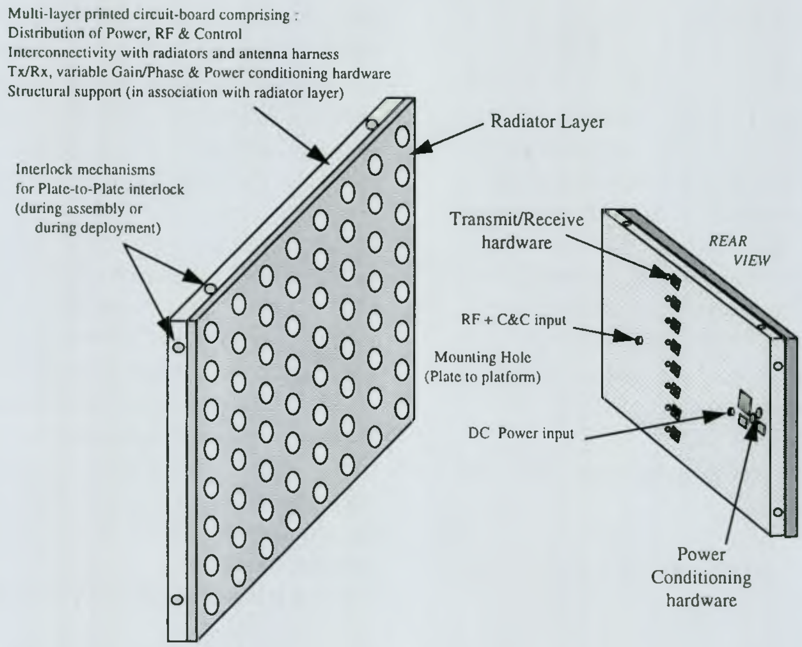


Figure 6-1: Example IFESS Building Block Concept

With the availability of thin (low height) integrated front end electronics (3 mm), the possibility of building a flexible antenna panel that could be deployed by unfolding or rolling becomes feasible. The main driver here would be to ensure planarity after deployment. Inflatable systems, and hinged deployment structures should be considered, along with applicable new material technologies.

The distribution of RF, command and control and power on the SAR antenna presents a further area where technology developments will lead to improvement. Already, optical technology offers the potential for a shared medium for the RF signals, if some what inefficiently, and for the command and control. To realise this benefit the necessary protocols and interface definitions must be developed to maximise functionality and minimise production cost.

Developments in power supply and distribution technology need to be incorporated into antenna development to complete the improvement of the functional antenna components. Such approaches as use of high voltages for power distribution to reduce harness mass and the corresponding impact on interface design need to be assessed.

The technology areas discussed above lead to a philosophy for future SAR antennas in which the functional elements are integrated, rather than modularised, to reduce packaging overhead and exploit structural rigidity and strength. The Integrated Front End SubSystem (IFESS) is the 'ultimate' goal of MMS. An example IFESS building-block is illustrated in Figure 6-1.

7. CONCLUSION

This paper has provided an overview of the systems and technologies included in ERS and ENVISAT radars for ESA, showing the evolution of the SAR requirements and the required technology. The MMS CORE Radar development path has been described with the aim of producing modular systems providing multi-frequency and polarisation at affordable prices, into wider non-institutional markets. Lastly, a view of the future SAR technology has been outlined to support the development of the next generation of spaceborne radars. The production of radars for space needs to take a new turn, incorporating technology from the communications and aerospace markets, allowing radars to be produced quickly and economically to meet the new commercial requirements.

ACKNOWLEDGEMENTS

The authors acknowledge the work of their colleagues in MMS and the support provided by the United Kingdom government and European Space Agency during the activities mentioned.

LOW-COST FAST-ACCESS RADAR ALTIMETRY SEA STATE MAPPING BY MICROSATELLITE NETWORK

Yuanxing Zheng, R.A.dc Silva Curiel, Martin N Sweeting
Surrey Space Centre, University of Surrey, Guildford, Surrey GU2 5XH, UNITED KINGDOM
Tel: (44) 1483 259143 Fax: (44) 1483 259503
Email: y.zheng@ee.surrey.ac.uk
WWW: http://www.ee.surrey.ac.uk/EE/CSER/UOSAT

ABSTRACT

Microwave radar altimeters have been proved an effective method for geographic, geodetic and oceanographic remote sensing during the last two decades. Compared with passive remote sensing methods, like CCD, altimeters have unique properties in geometric resolution independent of sensor altitude and all-weather imaging. From 1973, nine radar altimeters have been flown on USA, French and Japanese satellites, but never on a microsatellite. With the maturing of microsatellite design and the introduction of new altimeter technique, the design of light weight, low power, and highly accurate small satellite borne altimeters has now become possible. The whole system, satellite plus payload would only cost ~US$2.5M. On the base of that, in this paper a cheap but very effective altimeter network for maritime disaster migration is proposed. This network could have the ability to provide near real time altimetry measurement to any user around the world. A constellation of 12 satellites only cost for less than US$45M - a small fraction of the cost of a single conventional 'big' satellite borne radar altimeter. The small mass & volume of the microsatellite allows the whole network to be launched on two small launchers into the 800 km orbit.

I. INTRODUCTION

Satellite altimetry plays an important role in glaciology, topography study and sea state monitoring, especially for the remote areas where are difficult to arrive. The first space-borne microwave radar altimeter appeared in 1973 on SKYLAB [Mill L. S. et al, 1972], from then on nine space-borne altimeters had been launched and testified the important role of altimeters in climate related research. However most of their application were focused on the topography study, therefore they need very complex and expensive subsystems in orbit determination, attitude control, and signal processing. For example, the cost of GEOSAT, the America Navy altimeter mission, was approximately $60 million [McConathy D. R. et al 1987]!

Up to now, there does not have any altimeter that is specially designed for quickly delivering sea state monitoring altimetry results, which mainly include significant wave height (SWH) and wind speed. On the other hand, we notice there are almost 40,000 bulk carriers of 500 tonnes or more at sea today, many of which make either occasional or routine long distance, trans-ocean journeys which constantly require the altimetry measurement results for route planning and management. Therefore in this paper, we propose a Global Altimeter Network Designed to Evaluated Risk (GANDER) constellation which composes of 12 very low cost 50 kg microsatellite to maritime disaster mitigation and ship routine optimisation

Section II is a basic introduction of altimeter operating principles, section III is system consideration for a microsatellite borne altimeter, which covers a basic introduction for 15 years successful microsatellite design and research in Surrey, a new highly efficient altimeter concept – Delay/Doppler altimeter, the feasibility analysis for each satellite main subsystem. In the last section, a 12 microsatellite borne radar altimeters network constellation is outlined, and the network ability -- to provide near real-time sea state significant wave height (SWH) and wind speed measurement mapping for world wide users is discussed.

II. ALTIMETRY FOR SIGNIFICANT WAVE HEIGHT MEASUREMENT

The altimeter operation principle is conceptually simple: a nadir looking, high-resolution radar that measure the distance from the satellite to the ocean's surface with high accuracy. By tracking the half power point of the leading edge of the return signals wave-form, we could extract the wanted significant wave height (SWH) information with an accuracy of 0.5m, and by analyzing the plateau of return signals we could get the wind speed result. These two items are the most useful information that indicate the current sea state.

Proc. 32nd ESLAB Symp., 'Remote Sensing Methodology for Earth Observation and Planetary Exploration', ESTEC, Noordwijk, The Netherlands, 15-18 September 1998 (ESA SP-423, December 1998)

Figure 1 illustrates the return signal waveform for a conventional pulse limited (the area illuminated by one radar pulse is smaller than the area decided by the antenna half power angle) altimeter.

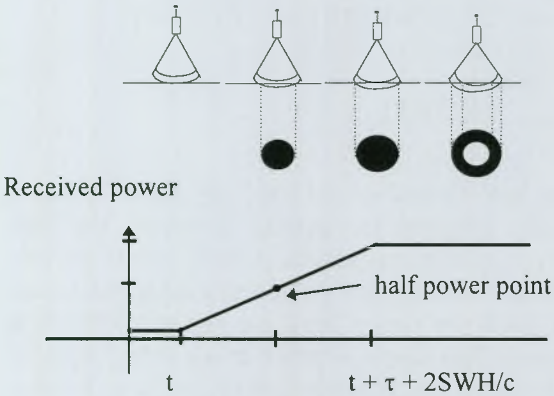


Figure 1. Power received by a satellite altimeter as a function of time

SWH corresponds approximately to the crest-to-trough height of the 1/3 largest waves in the footprint and is denoted by $H_{1/3}$. It is shown in figure 1 that the slope of the leading edge of the return signal waveform is influenced by the SWH. The larger the SWH (the worse the sea state), the longer time the return signal need to reach its peak power.

According to Brown [Brown 1977], the radar altimeter return signal wave-form is a triple-convolution of:

1) the transmitted pulse shape P_{ptr};
2) flat surface response which included effects of antenna pattern, off-nadir pointing angle Pfs;
3) the height probability density of the specular points P_{pdf}.

He used this convolutional model with assumptions common to satellite radar altimeter systems to produce a simplified close-form expression:

$$P_r(\tau) \approx \begin{cases} \eta P_T P_{FS}(0)\sqrt{2\pi}\sigma_p \left[1 + erf\left(\frac{c\tau}{\sqrt{2}\sigma_c}\right)\right] \Big/ 2, \tau < 0 \\ \eta P_T P_{FS}(\tau)\sqrt{2\pi}\sigma_p \left[1 + erf\left(\frac{c\tau}{\sqrt{2}\sigma_c}\right)\right] \Big/ 2, \tau \geq 0 \end{cases} \quad (1)$$

in here,

η ~ the pulse compression ratio,
P_T ~ the peak transmitted power,
$P_{FS}(t)$ ~ the flat surface impulse response,
σ_p ~ related to the point target 3 dB width (T) by $\sigma_p = 0.425T$,
σ_c ~ determined from $\sigma_c = \sqrt{\sigma_p^2 + (2\sigma_s/c)^2}$, where σ_s is the rms height of the specular points relative to the mean sea level,
erf(•) ~ denotes the error function,
$\tau = t - 2h/c$.

By properly tracking the height from satellite to sea surface, the algorithm could estimate the SWH. In figure 2, the SMLE obtains a measure of the half-width of the return by equalising the shaded areas in figure. Because the return waveform is a erf(•) function, therefore we can get

$$\int_0^T erf\left(t \Big/ \sqrt{2}\sigma_c\right) dt = T/2 \quad (2)$$

and further more, the SWH could be deduced by

$$SWH = 2c\sqrt{(0.6801T)^2 - \sigma_p^2} \quad (3)$$

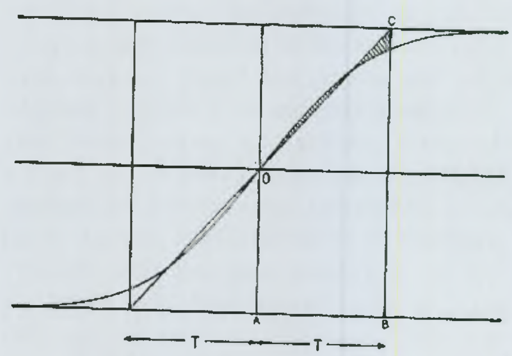


Figure 2. Schematic illustration of the area equalization procedure used by the SMLE to estimate the half-width of the altimeter return

III. RADAR ALTIMETER PAYLOAD CONSIDERATION FOR MICROSATELLITE

In the past two decades, the concept of microsatellite develops very fast. The widely use of solid state device make the satellite could achieve reasonable functions with small mass, whilst the use of the off-the-shelf commercial components decrease its cost dramatically. Microsatellite is specially useful in the remote sensing such as disaster monitoring, as it is very flexible, cheap, and quickly built. Its use in passive remote sensing has already been demonstrated by Surrey Space Centre [Sw] , the latest mission (10 July, 1998) for Thailand carry a the camera with 5 meter resolution. In the meaning while, the use of microsatellite in active remote sensing, such as the radar altimetry has also been carried on actively in Surrey.

To a microsatellite payload design, the strictest limitation is the on board dc power. For a 50kg 1000 km SS orbit microsatellite, the average available dc power is 27 W. Such a low power supply could not satisfy the payload SNR requirement completely, especially when we consider to provide enough link margin for bad weather (the situation where altimetry results are especially needed). In the following part of this section, a new altimeter technique -- Delay/Doppler altimeter is introduced which could provide 10 SNR improvement compared with the conventional pulse limited altimeter. In the end, a feasibility of main satellite sub-systems will also be discussed.

III-A. Surrey Microsatellite

The University of Surrey (UK) has pioneered microsatellite technologies since beginning its *UoSAT* programme in 1978. Surreys' first experimental microsatellites (UoSAT-1 & 2) were launched free-of-charge as 'piggy-back' payloads through a collaborative arrangement with NASA on DELTA rockets in 1981 & 1984 respectively [Sweeting M N [1982]. Since then, a further ten low cost yet sophisticated microsatellites have been placed in low Earth orbit using Ariane and Tsyklon launchers for a variety of international customers and carrying a wide range of payloads.

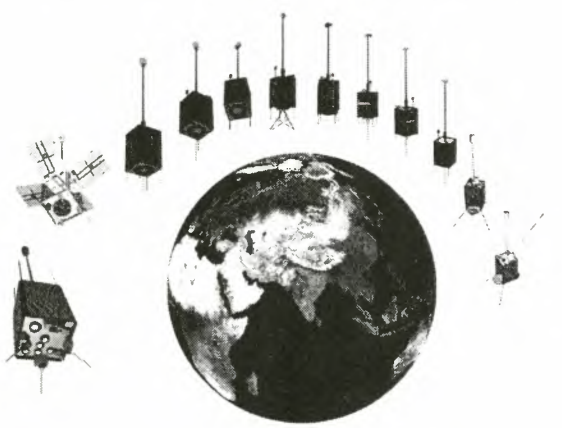


Figure 3. Surrey microsatellite missions in LEO

The latest two new missions for Chile and Thailand was launched by Zenit from the Baikonur Cosmodrome on July 1998. Three new microsatellites for Malaysia, France and the US-DoD and an experimental (350kg) minisatellite are currently being prepared for launch in 1999. Future plans are both exciting and ambitious, and range from a 350kg 'UoSAT-12' minisatellite for launch in 1999, a 5kg nano-satellite 'SNAP', to the 'EarthRise' and Lunarsat lunar missions, as well as a microsatellite constellation for world-wide disaster monitoring and mitigation.

Microsatellite	Launch	Orbit	Customer	Payloads
UoSAT-1	1984-D	560 km	UoS	Research
UoSAT-2	1984-D	700 km	UoS	S&F, EO, rad
UoSAT-3	1990-A	900 km	UoS	S&F
UoSAT-4	1990-A	900 km	UoS/ESA	Technology
UoSAT-5	1991-A	900 km	SatelLife	S&F,EO, rad
S80/T	1992-A	1330 km	CNES	LEO comms
KitSat-1	1992-A	1330 km	Korea	S&F,EO, rad
KitSat-2*	1993-A	900 km	Korea	S&F,EO, rad
PoSAT-1	1993-A	900 km	Portugal	S&F,EO, rad
HealthSat-2	1993-A	900 km	SatelLife	S&F
Cerise	1995-A	735 km	CNES	Military
FASat-Alfa	1995-T	873 km	Chile	S&F,EO
FASat-Bravo	1997-Z	830 km	Chile	S&F,EO
TMSAT	1997-Z	830 km	Thailand	S&F,EO
TiungSAT-1	*1999-?*	*650 km*	*Malaysia*	*EO, Comms*
UoSAT-12	*1999-S*	*650 km*	*SSTL & Singapore*	*EO, Comms*
PicoSAT	*1998-S*	*650 km*	*USAF*	*Military*
Clementine	*1999-A*	*735 km*	*CNES*	*Military*

Table 1. University of Surrey Microsatellite Missions
* built in Korea using SSTL platform & KAIST payload
D=Delta; A= Ariane; T=Tsyklon; Z=Zenit; S=SS18/Dnepr

All SSTL platforms are derived from a modular microsatellite bus, Microbus-70, and larger platforms benefit from the flight heritage of the microsatellite platform. The Minibus platform offers the next level of payload services, and makes available greater volumes for payloads.

III-B Delay/Doppler altimeter technique

The Delay/Doppler altimeter is proposed by Raney [Raney R.K.1995]. The main idea is delay compensation by taking use of the along track Doppler frequency, as a consequence all the along track data contribute to the track point, therefore take better use of the transmitted power compared with the conventional pulse limited altimeter.

The range processing is as follows: A (very) long linear fm modulated pulse is transmitted. Upon reception, the demodulator multiplies the received signal by a delayed replica of the same long FM pulse, and the product is low-pass filtered. The result is a cw signal corresponding to each scattering facet, and the frequency of that cw is proportional to the range difference between the delay of the scatterer and the delay of the replica. The length of the cw tone is limited by the overlap region of the signals and the replica, but by design the length is a constant of the system. The application of an FFT to a cw tone yields a pulse whose width (resolution) is inversely proportional to the length of the cw, and whose position is proportional to the cw frequency.

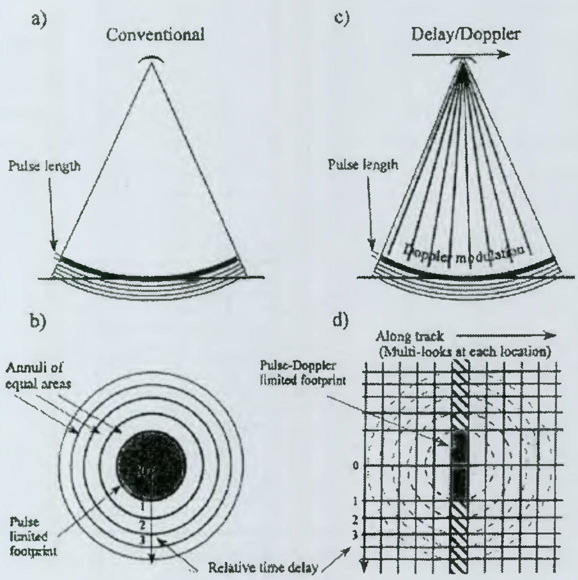


Figure 4. Conventional and Delay/Doppler altimeter response

The relative radiometric response of the pulse limited and Delay/Doppler altimeters is given by the ratio

$$\frac{P_{DD}}{P_{PL}} = \alpha_R \frac{A_{DD}}{A_{PL}} \qquad (4)$$

in here

α_R : orbital gain, expressed as $\alpha_R \approx \dfrac{R_E + h}{R_E}$.

R_E : satellite height

h : earth radius

Then we can get the received power improvement is around 12dB for a 1000 km height orbit altimeter.

The detail signal processing research is currently under developing in Surrey now.

III-C. Feasibility analysis for microsatellite subsystems

A. Oribtal determination

The primary goal of the most of the past altimetry missions is to make precise measurements of sea level for the study of global ocean circulation. In order to fully exploit the altimetric data, the satellite's radial ephemeris must be known to sub-decimetre accuracy. From the first day of altimeter began to operate, enormous energy has been put on to achieve that. For example, the TOPEX/POSEIDON (T/P) altimeter, the most complex altimeter up to now, has four subsystems been implemented to give a precise orbit determination. They were: Satellite laser tracking (SLR), Doppler Orbitography and Radiopositioning Integrated by Satellite (DORIS) tracking, global Positioning System (GPS) tracking, and Tracking and Data Relay Satellite System (TDRSS) tracking. Obviously, a microsatellite could not afford so many subsystems.

Fortunately the major aim of this altimeter is to measure the SWH and wind speed which depends on analyzing the leading edge of the return waveform, therefore the absolute height measurement is not as critical as the previous missions. Nevertheless, some orbit determination will be necessary to aid processing, and stationkeeping. In the proposed system, an on-board GPS receiver will be employed based on designs for TMSAT, TiungSAT and UOSAT-12.

B. On-board power budget consideration

On board dc power budget is very important for satellite payload and design. For the platforms under consideration, the solar panels are body mounted. This reduces mission risk, and permits a low cost gravity gradient spin-stabilised attitude mode to be employed. The size of the satellite platform is therefore directly related to the available Orbit Average Power (OAP) for the payload and spacecraft sub-systems. Nevertheless, some power savings are possible. The altimeter is only required over the oceans and consequently the required Orbit Average Power is only approximately 70%. Furthermore, operation at high latitudes (>70°) is not required, leading to an overall system operational requirement of 62%, or about 31W.

The real time altimetry data rate is very low, lower than 1kbps, if the on board signal processing is fully used. Therefore data storage requirements are modest, as a single sea state measurement per second only results in a few hundred kbytes per day, which is well within the data capacity of the standard MicroBus downlink. In order to fully capitalise on the potential power savings possible, it is necessary to use a flexible On-Board Processing system and operations regime. Fortunately, the SSTL platform permits autonomous operation of the combined spacecraft and groundstation system, and can include complex software to schedule payloads depending on location and available on-board resources such as power and data storage. It is feasible to include a task, which only operates the payload over designated areas, and return data on demand or to designated stations. For a 50kg level 1000km SS orbit microsatellite, the average available power ranges from 27 to 43 W. Refer to the power budget in table 2, we can see the whole system power requirement could be satisfied if the whole system design is very efficient.

System Components	Average DC Power (W)
GPS subsystem	5
Altimeter payload (RF)	11.5
Altimeter payload (signal processing)	3
ADCS (one wheel)	2
Broadcast payload	3.5
OBC	2
Total	27

Table 2. Microsatellite and its payloads power budget

C. Propagation effect

As we want to use this altimeter to monitor the sea state especially in the bad weather condition, the propagation effect should be considered carefully -- enough margin should be added to accommodate the propagation attenuation and other influence. In high frequency, such as Ku band, the propagation effects are quite complex, basically the effects can be divided into three classes:

a) Amplitude attenuation – at Ku band, the rain and atmospheric molecular all may degrade the transmit and return signal amplitude, especially in bad weather conditions. Sometime this attenuation may as large as 20 - 30 dB for the worst case.

b) Phase distortion – the scintillation effect of electronic content will introduce in phase distortion, which for coherent signal processing system is undesirable

c) Time delay -- which results in the increase of total propagation time, and thus increase the measurement distance from satellite to the illuminated area.

To this system, the more important considerations are the first two items. Rain has a very important and complex effect on propagation at frequencies of Ku band and higher, particular when the weather is stormy and the rain-drops are not only more numerous but large enough to be compared with the wavelength of the signal and often accompanied by ice and snow. In storm, the size, shape and attitude of raindrops reduce not only the strength of the signal but also the purity of its polarisation. Previous survey shows [Monaldo F. M. et al 1984] for some worse cases of Ku band signals, the maximum attenuation is around 9 dB for a 5km rain cell at 20mm/hr rain rate. That means at least about 10dB margin should be added in the radar payload design, otherwise large gap would appear in the downloading altimetry map for the heavey rain area whist it is the area that altimetry results are strongly needed.

IV. RADAR ALTIMETER NETWORK CONSETELLATION

We know due to the high requirement of range resolution, the transmit pulse after compression is very short, for this condition altimeter's swath width is decided by the pulse length of the compressed radar transmit pulse and the satellite altitude, expressed as:

$$Swathwidth = 2\sqrt{c\tau h} = 1.9km \qquad (6)$$

for this paper's application

$\tau = 3.125$ *ns* --- pulse length after compressed

h = 1000 *km* --- satellite altitude

$c = 3\times10^8$ --- velocity of light

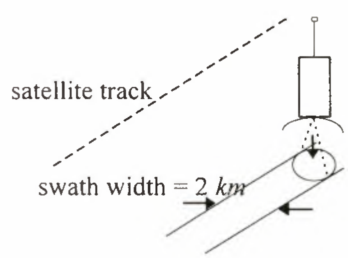


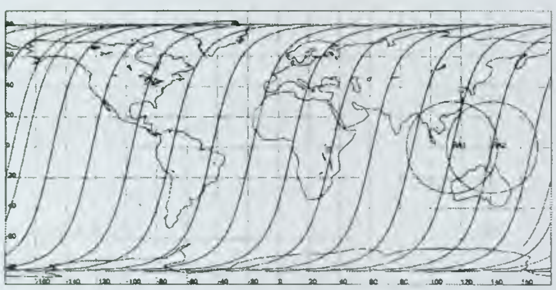
Figure 5. Radar altimeter swath width diagram

Obviously this swath width is very narrow, especially when we consider the distance between two continuous pass is roughly 3300km for a 1000km SS orbit. Therefore an altimeter network should be arranged carefully to get a meaningful enough altimetry measurement grid before it could be provided to the users.

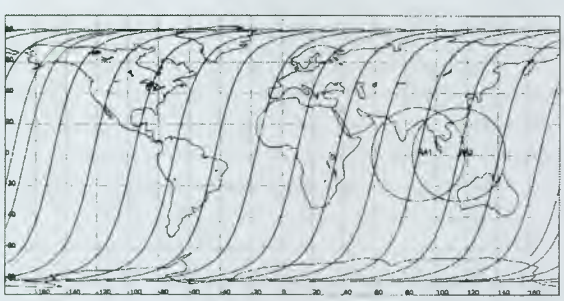
It is therefore naturally to think that we put several altimeters equally within the two continuous satellite tracks. Consider the satellite is operating in broadcast mode, and the radius for the broadcast footprint is 3360 km therefore for a user who is also in the area within the two continues pass, he would receive the real time measurement from all these altimeters, although some of the altimetry measuring place are may quite far from him. Still this will be very helpful for the ships in optimising the their shipping routes. The next question is how many satellites are needed for this application? We know the worst case is that the ship is at the middle point between two satellites track, for this condition, in equator the distance is

$$d = \frac{D}{2(n-1)}\ km \qquad (7)$$

where n is the numbers of satellites. If $n = 6$, then d is roughly 300 km, for most case it is an acceptable result. Figure 6 shows the altimeter track for satellite 1(RA1) and satellite 6(RA2).



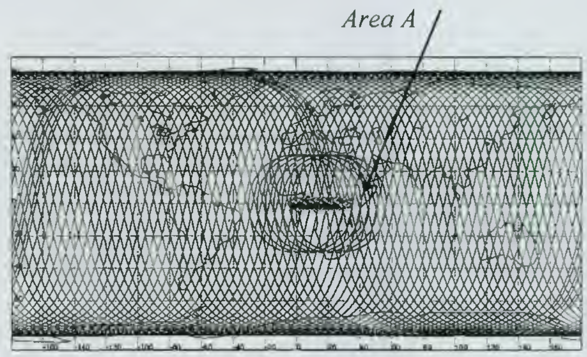
(a)

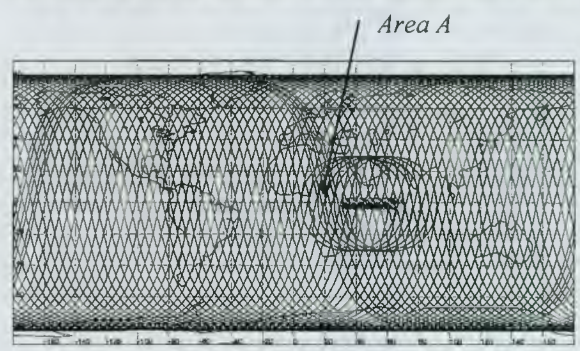


(b) one period later

Figure 6. Satellite passes for one satellite period time, in here we omit the other four satellites between RA1 and RA2, we could see RA2 always repeat the track of RA1, this give the user who is near the track of RA1 some benefit to get longer time to access the altimetry measurement result (the round area is the area that could be illuminated by the satellite broadcast antenna).

However, to get real or near real time altimetry measurement in the surrounding area, only six altimeters is not enough. Considering for the worst case that satellite 6 just passed the area, the user therefore have to wait nearly half a day before the satellites come the area again, shown in figure 6.





(b)

Figure 7. Only six satellites condition, (a) the starting position, we could see the six satellite just leaving area A, roughly longitude 40 degree (b) shows after nearly 10 hours before satellite 1 return to the area A

Such a long waiting time is unacceptable for lots of applications where near real time sea state information is strongly needed. To decrease the waiting time, another six satellites are added, shows in figure 8.

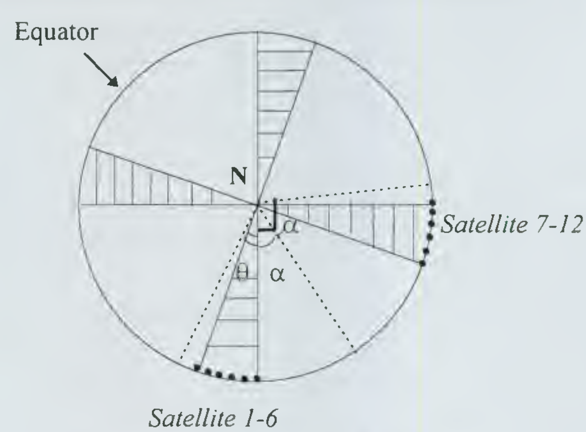


Figure 8. Constellation basic diagram

In here:

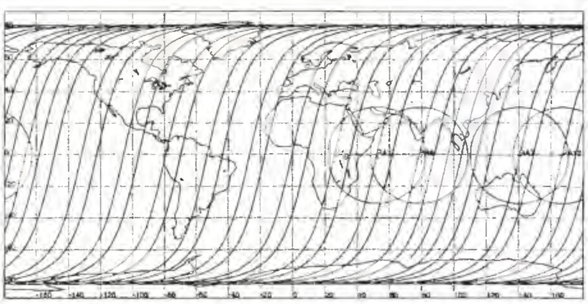
α = 31.47deg is the half of broadcast antenna illumination angle

θ = 26.28deg is the angle between satellite two continuos passes.

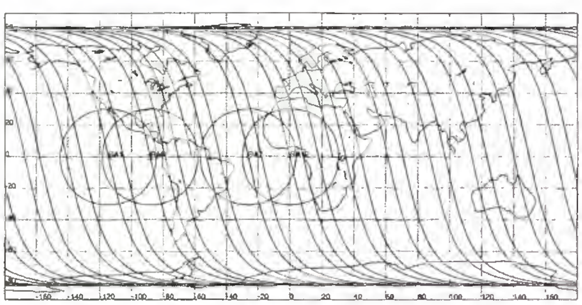
Therefore we could get:

$$\theta + \alpha + \alpha = 26.28^{\circ} + 62.94^{\circ} = 89.22^{\circ} \quad (8)$$

that means even the ship in the equator, the satellite footprint still could cover the half earth completely. Consider the worst case -- the satellites just pass by the ships, therefore the ship just have to wait another satellite period (100 minutes for 1000 km) when the satellites pass again. Figure 9 shows within one satellite period, the footprint of the whole altimeters network could cover most area of the world. In here, in order to show clearly we omit RA2 ~ RA5 & RA8 ~ RA11, from the figure we could see in descending part, the satellite footprint covers from longitude 20 ~ -160 degree, while in the ascending part the satellite network footprint covers from longitude -150 ~ 30 degree.



(a) Decending



(b) Ascending

Figure 9. Altimeter network footprint diagram within one satellite period

V. CONCLUSION

A world wide sea state mapping by 12 microsatellite borne radar altimeters network is proposed to provide near real time sea state significant wave height and wind speed altimetry measurement. This is due to the mature of low light, low dc power requirement radar altimeter design and the low-cost, rapid-response, highly-reliable, long-life-time and advanced microsatellite. The total cost of the altimeters network is only US$20M, which is much less than a conventional 'big' satellite-borne radar altimeter.

- Due to the 10dB received SNR improvement, Delay/Doppler altimeter technique is proposed for a microsatellite altimeter radar payload design.
- The longest time for the user to download at least one satellite altimetry measurement is less than 90 minutes for a 1000km SS orbit condition
- For cross track dimension, the grid of altimetry measurement varies at different longitude. In equator, the distance is the largest, around 600 km

REFERENCE

Calman J. et al [1989] Real-time satellite altimetry *Johns Hopkins APL Technique Digest, vol. 10, pp. 380-385*

Chelton D.B., et al [1989] Pulse compression and sea level tracking in satellite altimetry *Journal of atmospheric and oceanic technology vol. 6, pp. 407-428*

Fouquet M. et al [1996] Earth observation using low cost SSTL microsatellites IAF-96-B.3. pp. 215, 47th IAF Congress, Beijing China

McConathy D. R. [1987] The navy GEOSAT mission: an overview *Johns Hopkins APL Technique Digest, vol. 8, pp. 170-175*

Miller L. S. et al [1972] Objectives and capabilities of the Skylab S-193 altimeter experiment *IEEE Trans. On Geoscience electronices vol. 10, pp. 73-79*

Monaldo F. M. et al [1984] Altimeter height measurement errors introduced by the presence of variable cloud and rain attenuation *NASA Conf. Publ. No. 2303, pp. 287-296*

Raney R.K. [1995] Delay/Doppler compensation: A New Concept for Radar Altimetry *Proc. of*

International Society for Optical Engineering, vol. 2358, pp. 412-419

Sweeting M. N. [1982] UoSAT - An investigation into cost-effective spacecrafe engineering *The Journal of the Institute of Electrical & Radio Engineers, vol. 52, pp. 363-378*

RADIO OCCULTATION DATA ANALYSIS: FROM PLANETARY ATMOSPHERE SOUNDING TO OPERATIONAL METEOROLOGY

P. Silvestrin[1], P. J. Baptista[2], P. Hoeg[3]

[1]Earth Observation Preparatory Programme, ESA
ESTEC APP-JPP, Noordwijk, NL-2200AG
Tel/fax: +31 71 565 3689/4696, Email: psilvest@estec.esa.nl
[2]Electromagnetics Division, ESTEC TOS-EEP, ESA
[3]Danish Meteorological Institute (DMI), Copenhagen (DK)

ABSTRACT

The radio occultation (RO) technique uses radio signals passing through planetary atmospheres to probe these media. It has been used to study the atmospheres of terrestrial and outer planets and their moons. With the advent of satellite navigation systems like GPS and GLONASS, applications in meteorology and Earth climate research have also become possible. A proof-of-concept has been obtained with the launch of the GPS-MET experiment in 1995. In the stratosphere and upper troposphere, where humidity is low, the temperature profile can be accurately retrieved, whereas in the lower troposphere, the humidity profile can be retrieved if the temperature is given by other means. The data reduction has first been based on the approach used for planetary applications, i.e. using the Abel transform to recover refractivity from refraction angle. The vertical resolution of the retrieved profiles is diffraction-limited to the size of the first Fresnel zone, however algorithms have been developed that yield sub-Fresnel-scale resolution and cope also with multiple propagation paths. The paper reviews the principles of the technique and various aspects of the data analysis that have been the subject of recent research.

INTRODUCTION

Our understanding of the state of the Earth's atmosphere and the related capabilities for modelling and forecasting changes over short and long time-scales are currently hampered by our limited knowledge of the temperature and humidity fields in the troposphere and the stratosphere. Reliable forecasts from a numerical weather prediction (NWP) system require an accurate description of the initial state of the atmosphere, since small initial state errors can rapidly grow to dominate errors in the subsequent forecast. Sensitivity studies show that the state estimation should be accurate especially in the geographical areas close to atmospheric fronts and jet-streams. Cloud cover is often extensive in such areas, so all-weather atmospheric sounding techniques are required.

For long time-scale studies, the current climate models predict a global surface temperature increase by about 1 K over the next 20-30 years together with a rapid increase of the humidity and a corresponding cooling in the stratosphere. Both climate modelling and climate monitoring require a good description of the mean state and the variability of the atmosphere. They therefore depend on the availability of observations with high absolute accuracy and long-term consistency (that is, ideally, bias-free measurements).

The models used for both operational meteorology and climatology would most benefit from measurements with a vertical resolution appropriate to the physical scales in the troposphere and the stratosphere, so the vertical resolution of retrieved profiles should ideally be better than 1 km in the stratosphere and reach ~100 m in the troposphere.

Both meteorology and climatology suffer from the limitations of current sources with regard to geographical coverage (radiosondes are deployed only in a small fraction of the land), vertical resolution (satellite radiometers have at best a vertical resolution of 3 - 4 km), and absolute accuracy (all currently deployed systems suffer from bias problems).

The radio occultation (RO) technique uses radio signals passing through planetary atmospheres, ionospheres and rings to probe these media. When applied to the Earth's atmosphere, the technique has the potential to overcome the above limitations (coverage, resolution, and accuracy) and to provide a great amount of atmospheric information on a global scale and in a very cost-effective manner. The RO technique is a well-known one in planetary science where it has been successfully used since the early Sixties. Because of the lack of radio sources suited to meet the performance requirements of Earth weather and climate studies, it was only with the deployment of global navigation satellite systems (GNSS) such as the US Global Positioning System (GPS) and the equivalent Russian system (GLONASS)

Proc. 32nd ESLAB Symp., 'Remote Sensing Methodology for Earth Observation and Planetary Exploration', ESTEC, Noordwijk, The Netherlands, 15-18 September 1998 (ESA SP-423, December 1998)

that such applications became possible.

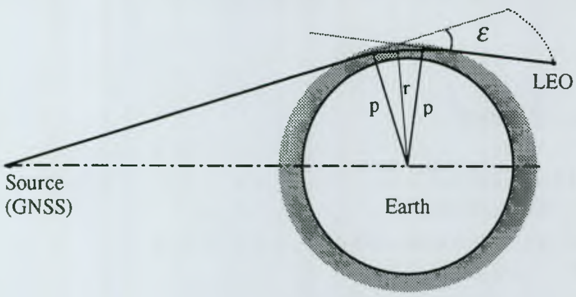


Fig. 1: basic geometry of a radio occultation

The principle of the RO technique is as follows (see Fig. 1). As radio waves pass through the atmosphere, either through a descending path (set event) or an ascending one (rise event), they are retarded and refracted through an angle determined by the refractivity gradients along the path. Compared to the free path in vacuum, these effects result in an excess path, which, for the Earth's atmosphere, is mainly due to the refraction. The phase and amplitude of the signal can be measured very precisely and used to recover the refraction angle profile through a full set or rise event. The refractivity gradients in the atmosphere depend on the gradients of air density (and hence temperature), water vapour and electron density. In the stratosphere and upper troposphere, where the water vapour is low, refractivity is dominated by vertical temperature gradients and the temperature profile can be accurately retrieved. In the lower troposphere, the water vapour effect is dominant and, if a (moderately accurate) temperature estimate is provided by a (NWP) model, the water vapour profile can be retrieved.

SOUNDING OF PLANETARY ATMOSPHERES BY RADIO OCCULTATION

The RO technique has been a prime tool for remote sensing of planetary atmospheres for over 35 years as the core part of radio science experiments designed to observe also ring structures, wind profiles, gravity waves, scintillations, electron density, and geodetic parameters. Planetary RO measurements have often resulted from 'piggyback' experiments, generally exploiting the spacecraft telecommunication subsystem, the signal being generated on the Earth, up-linked to the spacecraft and retransmitted to ground by the on-board transponder. Only more recently, with the Voyager, Galileo and Mars Global Surveyor missions, the RO objectives have resulted in the addition of an on-board ultra-stable oscillator (USO). In all cases, because of the low impact on the mission implementation, the RO measurements have been very cost-effective. The uncertainties on the chemical composition and the atmospheric structure of the bodies to be studied have often complicated the planning and design of planetary RO experiments. This was the case for instance for the Voyager observations of Titan, where a priori uncertainties on the surface pressure ranged over a factor of 1000. The planned Pluto Express mission is also tackling similar issues.

Starting with the Mars flyby by Mariner IV in 1964 [1,2], RO experiments were or are about to be carried out in the majority of planetary missions, including Pioneer (Jupiter, Venus, Saturn), Voyager (Jupiter, Uranus, Saturn, Io, Titan), Venera (Venus), Magellan (Venus), Galileo (Jupiter, Io, Europa), Cassini-Huygens (Saturn, Titan), Mars Global Surveyor (Mars), and Pluto Express (Pluto). Here we can only mention briefly a few examples of the results obtained. Voyager (1 and 2) [3] collected many RO measurements in the outer solar system. In the case of the Titan observations, they supplied the first accurate estimate of the radius of Titan and its atmospheric structure and chemical composition. The Voyager data allowed also the evaluation of the shape of Saturn, Jupiter and Uranus through the determination of iso-pressure levels. Using an analysis technique that provides estimates of the different signal components based on the extinction and scattering signatures, the data was also used to study the rings of Saturn and Uranus.

Several RO measurements of Venus's atmosphere were obtained first with the Venera and Pioneer series of satellites and more recently with Magellan [4]. The signal attenuation measurements of Magellan allowed the derivation of important information on the composition of Venus's atmosphere. A remarkable feature of this atmosphere is its refractivity vertical structure that prevents RO measurements below a height of ~33 km, at which critical refraction occurs (the radius of curvature of the propagation path exceeds that of the planet, so that no signal can exit).

With the exception of the measurements of planetary rings, the reduction of the RO data obtained from planetary missions has always been performed by means of an integral inversion approach based on the Abel transform, as described later. When possible, an analysis of the scintillation signal was also carried out, which provided information on the dynamics of the atmosphere, scintillation being related to turbulence or to rapid fluctuations of the refractive index. This approach allowed for instance the identification of vertically propagating gravity waves in Titan, having ruled out the possibility of turbulence in a very cold and slowly rotating body.

RO measurements are being carried out or are planned with new planetary missions. Galileo is taking measurements of Jupiter's atmosphere and rings, and its observations have already allowed recognising the presence of an atmosphere in Jupiter's moon Europa. The Mars Global Surveyor mission has carried out RO measurements of Mars's atmosphere from January to April 1998 and further data will be collected for an entire Martian year during the mapping phase of the mission, starting in March 1999.

RADIO OCCULTATION OF THE EARTH'S ATMOSPHERE

Proposals for applications of the RO technique to the Earth's atmosphere were advanced shortly after the first planetary RO results became available [5,6], however the results of the early RO experiments of the Earth's atmosphere did not reach the performance required for meteorology and climate studies. A major obstacle was the absence of suitable spaceborne radio sources of opportunity, which lasted until the deployment of the GPS and GLONASS systems. Each of these systems consists of a constellation of 24 satellites in highly inclined circular orbits at ~20,000 km altitude. Each GNSS satellite transmits signals generated from an on-board USO at frequencies of approx. 1.6 GHz and 1.2 GHz. In 1987 Russian scientists proposed to use GPS signals for RO measurements and pointed out the potential of a constellation of Low Earth Orbiters (LEO) with RO sensors for meteorological observations [7]. The first demonstration with GPS signals was achieved in 1995, when an U.S. consortium launched the GPS/MET experiment on the Microlab-1 satellite. This triggered a successful international effort to analyse the GPS/MET data and to study the issues expected in future operational applications. It is now widely accepted that GNSS-based RO data can provide profiles of density, pressure and temperature in the middle atmosphere and the colder troposphere, as well as profiles of water vapour in the warmer regions of the troposphere, that meet the operational meteorology requirements (summarised in Table 1).

	Temperature	**Humidity**
Vertical Domain	500 to 10 hPa (5-30 km)	> 300 hPa (0-10 km)
Vertical Resolution	0.5-1.0 km	0.5 km
Absolute Accuracy	< 1.0 K	< 10 % or < 0.2 g/kg
Timeliness	< 3 hrs	< 3 hrs

Table 1: observation requirements for meteorology

The features that make GNSS-based RO data attractive for meteorology and climatology include:

- coverage: ~1000 profiles per day per sensor, globally distributed thanks to the distribution of the GNSS sources;
- all-weather capability;
- high vertical resolution (better than 1.5 km);
- high accuracy of retrieved temperature profiles (1 K in dry regions);
- inherent long-term consistency: the measurements are essentially time-delay observations, which are close to ideal in a metrological sense;
- cost-effectiveness: only a (specialized) receiver has to embarked on the LEO.

MEASUREMENT PRINCIPLE

The basic geometry of a RO measurement is shown in Fig. 1, where a ray representation of the radio wave path is assumed. This is usually valid at least in the stratosphere and upper troposphere, where the wave propagation can be adequately described by geometrical optics (GO). In Fig. 1, ε is the refraction angle and p is the ray impact parameter, defined as the distance from the Earth's centre to the ray asymptotes. This description implicitly assumes that the atmosphere refractivity possesses a local spherical symmetry. If this is not the case, different impact parameters should be defined for the ray entering the atmosphere and for that exiting it.

The measurements provided by the sensor are the amplitude and phase of the received GNSS signals, sampled at a 10-100 Hz rate throughout each RO event and precise to ~1 degree and ~0.2 dB, respectively. The sensor measures also other GNSS signals that do not cross the atmosphere so as to determine very precisely the LEO velocity and position. This is required to remove the doppler shift caused by the orbital motion and derive the excess doppler shift. The required accuracy, particularly for the along-ray velocity component that must be known to ~0.1 mm/s, imposes the use of differential GNSS techniques that rely on simultaneous GNSS signal tracking data collected on ground.

The first step of the inversion requires deriving from the sensor measurements the excess phase and the corresponding doppler shift caused by the path modification in the atmosphere. In addition to the removal of orbital motion effects, it is necessary to correct for USO errors, which is also accomplished by differential GNSS techniques, and, when the path crosses the lower atmosphere also for diffraction and multipath effects.

The overall data reduction scheme is outlined in Fig. 2. From the time series of excess doppler, the refraction (bending) angle ε is retrieved as a function of the ray impact parameter p, assuming local spherical symmetry.

From this angle profile, the refraction index profile n(r) is recovered, where r is the ray distance to the Earth's centre, as described in a later section.

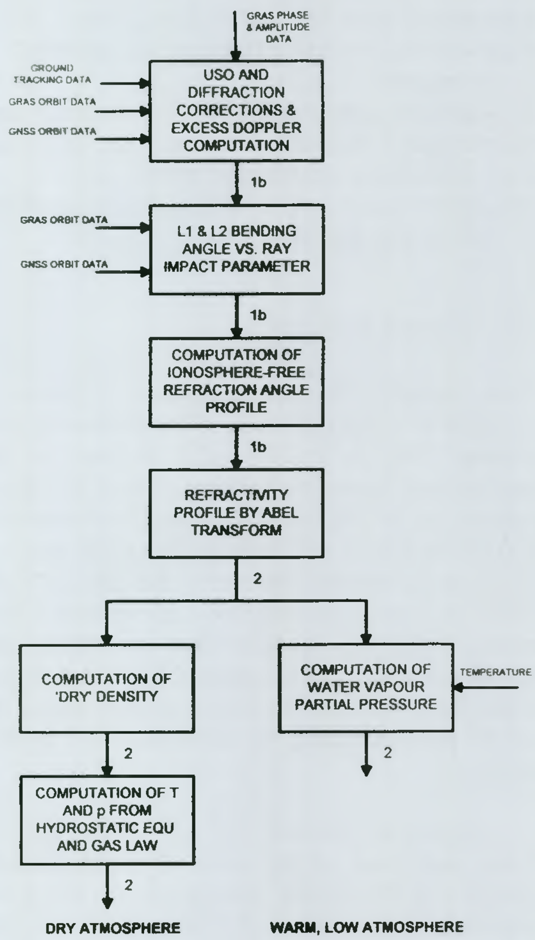


Fig. 2: outline of data reduction scheme

Once the refraction index profile is known, the determination of atmospheric parameters stems from its dependence (or that of refractivity $N=(n-1)10^6$) on atmospheric pressure P [mb], temperature T [K] and water vapour partial pressure P_W [mb]:

$$N = 77.6\frac{P}{T} + 3.73\cdot 10^5 \frac{P_W}{T^2} - 4.03\cdot 10^7 \frac{n_e}{f^2} + 1.4W$$

The additional dependence of N on electron density n_e [m^{-3}] and liquid water content W [gr/m^3] are viewed here as perturbations, although in other applications n_e is the geophysical parameter of interest. The error caused by the ionospheric crossings within the propagation path is corrected for using signals at two frequencies, while that caused by liquid water scattering is generally negligible. The computation of the ionospheric correction gives better results if performed on the two refraction angle profiles obtained for the two frequencies rather than on the phase measurements (as usually done for other GNSS applications). This allows proper handling of the different refraction at the two frequencies (ray splitting) caused by the ionosphere [8,9].

Using a standard atmosphere model, the contributions by the 'dry' term (no water vapour) and by the 'humid' term (water vapour included) in a tropical atmosphere and in a dry one are as shown in Fig. 3. Note that for the dry atmosphere the humidity contribution is negligible.

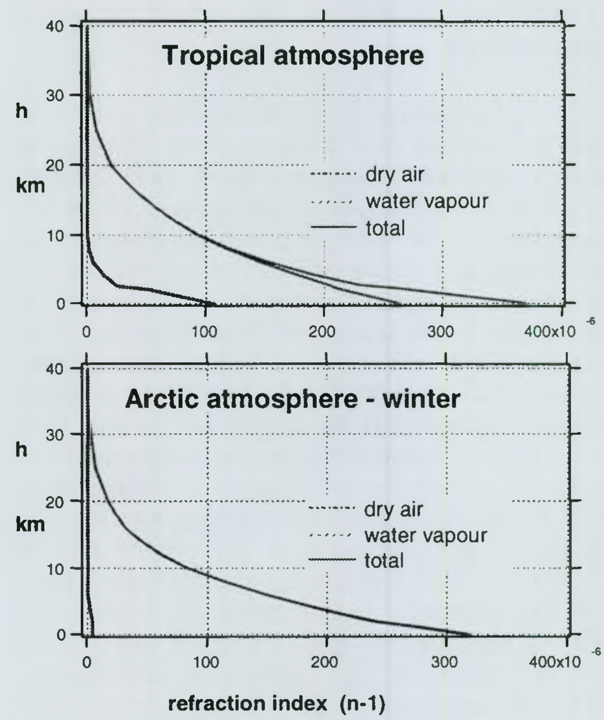


Fig. 3: contributions to the refractive index versus height (h) in humid and dry atmospheres

If water vapour is negligible, the refraction angle is found to be about 1 degree at grazing incidence. Considering a RO event in which the rays lie on the LEO orbital plane, the refraction angle is approximated by $\varepsilon \cong \Delta v / v$, where Δv is the velocity equivalent to the excess doppler shift and v is the transverse velocity of the transmitter-receiver line in the plane of propagation. In the stratosphere Δv can be ideally determined to about 1 mm/s and v is ~3 km/s, therefore ε can be determined to ~$0.3\,\mu rad$. In practice, the accuracy is limited by thermal noise and the precision for ε is ~$1\,\mu rad$, which is the value of the refraction angle at altitudes of 60 - 70 km. As measurement errors and refraction angle values are comparable, the retrieved profiles in the upper stratosphere show large errors. These errors can propagate down to the lower stratosphere when a ray tracing or an integral inversion method is used. A way to solve this problem is to combine (in a statistically optimal sense) the refraction angle profile obtained from the RO measurements with a profile obtained from a (good) model of the atmosphere. This provides a smooth transition from the measured profile in the height region where the measurement errors are relatively small, to the model profile in the region above where the errors are comparable with the difference of the measured and model profiles.

As indicated in Fig. 2, using the gas law linking pressure, density and temperature, the refraction angle profile provides directly the density profile in dry areas of the atmosphere. The pressure is then derived by integrating the hydrostatic equilibrium equation from a high level in the atmosphere down to ground. Once both pressure and density profiles are known, the temperature profile is obtained using again the gas law. In humid areas, the relation $P_W = c_1 NT^2 + c_2 PT$ is used to solve for water vapour, provided that the temperature is known to ~2 degrees, which is typically provided by a weather analysis.

REFRACTIVITY PROFILING BY RAY-TRACING INVERSION

Ray tracing methods to determine the refractivity profile have been successfully used in the early RO sensing of planetary atmospheres. The so-called 'onion peeling' ray tracing method assumes that the atmosphere is composed of a large number of spherical layers. Local spherical symmetry can be assumed to simplify the computations, but is not mandatory. The refractivity profile is computed starting with the top ray that passes only through the upper layer. The refractivity needed within this layer to provide the correct ray bending is computed, and the process is iterated, so that the refractivity in any layer is recovered using the knowledge of refractivity in all layers above it.

To avoid the need for a huge number of layers and consequently of doppler observations, the refractivity in each layer is not considered constant, but the refractivity gradient is. Continuity in refractivity is imposed at the layer boundary. An a priori model of refractivity N(r) is defined and the relation among the small deviations in refraction angles $\delta\varepsilon_i$ and in refractivity gradients is written as $\delta\varepsilon = \mathbf{A} \cdot \delta(\nabla N)$ where A_{ij} represents the partial derivative of the refraction angle with respect to the refractivity gradient in the j-th layer. **A** is a lower-triangular matrix that can be sequentially inverted (from the top layer down). The vertical profile N(r) is then recovered from top to bottom by integrating the recovered ∇N vector.

The ray tracing method has the advantage of not requiring spherical symmetry, however it does require a large computational effort, which currently makes the Abelian integral inversion the favourite approach for operational applications.

REFRACTIVITY PROFILING BY ABELIAN INTEGRAL INVERSION

In general the local refractivity field is close to a spherically symmetric one, because the field variations along a limb path are dominated by the vertical gradient of atmospheric density, which implies a radially directed refractivity gradient. Referring to Fig. 1, if **r** is the vector to a generic position P on the ray path and ϕ is the angle between **r** and the tangent to the ray path in P, then Bouguer's rule (i.e., Snell's law for spherically layered media) gives $n(r) r \sin\phi = p = n(r_o) r_o$, where r_o is the ray perigee and r is the length of **r** [11]. From this relation, one can derive the refraction angle as a function of the impact parameter:

$$\varepsilon(p) = 2p \int_{r_o}^{\infty} \frac{1}{\sqrt{n^2 r^2 - p^2}} \frac{d\ln(n)}{dr} dr$$

To obtain the refraction index profile, the above relation is inverted using the Abel integral transform [12], giving:

$$n(p) = \exp \int_{p}^{\infty} \frac{\varepsilon(p)}{\pi\sqrt{b^2 - p^2}} db$$

The corresponding ray perigee is then r = p/n. The refraction angle and the impact parameter need to be computed in the reference frame originating in the centre of the local curvature of the Earth surface, so that the ellipticity of the atmosphere can be corrected for. The systematic retrieval errors caused by lack of spherical simmetry and lack of co-planarity of the ray paths have been studied in detail [10,13]. Horizontal inhomogeneities with scales larger than the horizontal resolution of the technique (in the range 100 – 300 km, see following section) have been found to give negligible errors. If the scales are smaller, then a priori information about the refractivity field must be included in the retrieval to minimize the loss of accuracy.

SPATIAL RESOLUTION OF PROFILES

A high vertical resolution of the retrieved profiles is a very valuable property for atmospheric studies, such as those of gravity waves, tropopause structure and distribution, and marine boundary layer. Assuming a GO description, with λ denoting the wavelength and D the LEO distance from the ray perigee, the vertical resolution is limited by the vertical size d_F of the first Fresnel zone:

$$d_F = 2\sqrt{\frac{\lambda D}{1 - D\frac{d\varepsilon}{dp}}}$$

The typical profile of vertical refractivity gradients results in a refraction angle profile such that d_F decreases from ~1.5 km in the stratosphere to ~200 m in the troposphere. Local increases in refractivity gradients, for instance at temperature inversion layers, imply an even lower resolution and require a correction of diffraction effects in the processing. The lateral resolution remains close to ~1.5 km because the lateral refractivity gradients are very low.

There is no generally agreed definition for the horizontal resolution ΔL, which is obviously much larger than the vertical one because of the limb sounding geometry. A useful measure is given by the length of the ray segment enclosed between two concentric spherical layers spaced by d_F, with the ray tangent to the inner layer. This gives $\Delta L = 2\sqrt{2rd_F}$, which ranges from ~300 km in the stratosphere to ~100 km in the troposphere. These are acceptable values, particularly if the retrieved profiles are assimilated into NWP models that take into account this measurement characteristic. The NWP systems are based on a data assimilation approach, which includes the optimal estimation of the initial state of the atmosphere in a physically and statistically consistent way achieved by combining all the observations (ground-based, airborne and spaceborne) in an accurate model of the atmospheric processes.

ERROR ANALYSIS AND GPS/MET DATA EVALUATION

A number of error analyses have been performed, in particular for the retrieved temperature profiles [10,13,18-21], based on theoretical modelling and on simulations. For the latter, the EGOPS (End-to-end GNSS Occultation Performance Simulator) software developed for ESA at University of Graz, DMI and CRI (Copenhagen) has been extensively used. Such analyses take into account a range of errors sources, including: instrument (thermal noise, local multipath, USO drift), orbit determination, transmitter USO correction, ionosphere correction, spherical asymmetries in the atmosphere, and humidity effects. Such analyses agree quite well in predicting an accuracy better than 1 K between ~45 km and a lower height limit that, because of humidity, depends on the geographical location but is generally well below 10 km.

The processing of actual GPS/MET data has been performed by various groups. The validation of the results has been quite a challenge because of the intrinsic quality of the RO profiles. Comparisons with results from other instruments (radiosondes, lidars, microwave limb sounders) have been performed, which agree with the above theoretical prediction but also point out the difficulty of interpreting the profiles in areas where humidity effects are not negligible. More important perhaps are the statistical comparisons made between sets of RO profiles and NWP analyses, since the latter are based on the assimilation of a large number of observations. Fig. 4 shows the results of a statistical comparison of a set of ~600 RO profiles with NWP analyses from the UK Meteorological Office [15], divided for regions (tropical, north and south). This provides an upper estimate of the errors, which are found to be lower in the northern hemisphere than in the southern one or in the tropics. This is easily explained by the scarcity of observations taken in the southern hemisphere or in the tropical region and which negatively affect the quality of the NWP analyses. Further results are provided in [15], including a statistical comparison using a set of 4500 GPS/MET profiles.

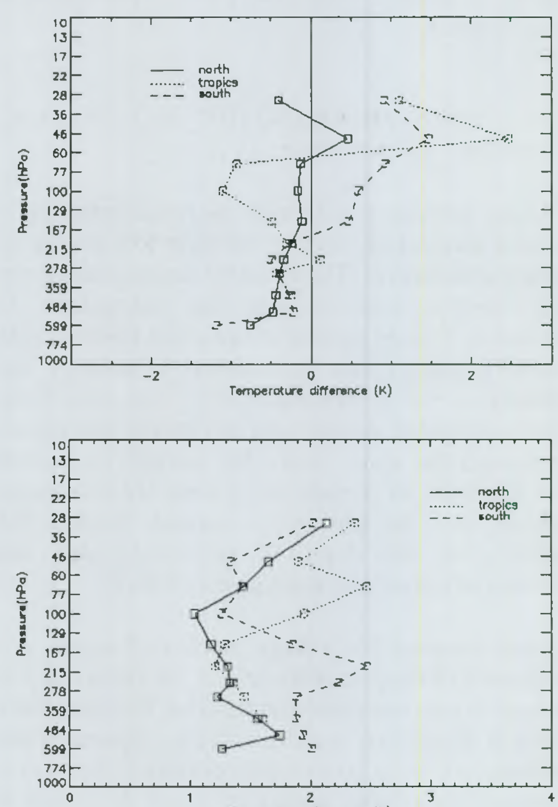


Fig. 4: mean (top) and standard deviation (bottom) of temperature differences (retrieved profiles minus NWP analyses) for tropical, northern and southern regions.

DIFFRACTION AND MULTIPATH CORRECTIONS

Large retrieval errors are found when the processing described above is applied to the RO data collected during the crossing of the lower troposphere (below approx. 8 km). This is due to the large and rapid fluctuations of the refraction angle mainly caused by temperature inversion layers and by the complex structure of the humidity. In the lower troposphere the atmospheric inhomogeneities have often a characteristic vertical scale smaller than d_F and large local refractivity gradients caused by such inhomogeneities result in multiple propagation paths (multipaths). In such a case the atmosphere behaves like a focussing lens, rather than a defocussing one, during parts of a RO event and the sensor measures the combined electromagnetic (EM) field resulting from the interference of multiple waves, as indeed observed in GPS/MET data. Under extreme conditions, critical refraction can also occur, causing absence of signal for short intervals.

The GO description is insufficient to derive algorithms valid in such a situation and the relation between refraction angle and ray impact parameter can no longer be used in a straightforward way as for the upper part of a RO event. Physical optics (of which the GO constitutes a special case) has therefore to be resorted to as the basis for the data reduction and indeed the scalar diffraction theory can be used to handle properly the multipath and diffraction effects so as to recover meaningful refractometric data.

Two basic approaches have been developed and tested with both simulated and real data: the 'phase screen' (PS) method [14,15,17] and the 'aperture synthesis' (AS) method [15,16]. The methods are based on the idea that multipath and diffraction effects can be eliminated or made negligible if the sensor is close enough to the region where the atmospheric inhomogeneities occur. This is equivalent to re-compute the EM field from the actual data on a virtual observation surface close to the minimum height region where no multipath exists. An attractive feature of both methods is that they can enhance the vertical resolution of the retrieved profiles to a sub-Fresnel scale of ~100 m, which exceeds the requirements of all applications. Both methods require measuring the complex EM field at the LEO with sampling rates of 50 to 100 Hz. Once the field on the observation surface is known, the processing can proceed as described before, namely with the computation of doppler shift from the phase, followed by refraction angle profile and geophysical parameter determination.

The PS method assumes that the atmosphere can be modelled with one or more thin phase screens. Each PS modulates only the phase of the EM wave going through it. As the wave propagates further, an interference pattern develops, which affects both phase and amplitude. When a sequence of many phase screens (several hundreds) is used, the multiple PS approach gives good results for all heights [14], but becomes prohibitive from a computational standpoint. This fact presently makes this approach impractical for the recovery of geophysical parameters from RO data. The approach remains nevertheless very useful for the simulation of RO measurements in presence of multipath and diffraction.

The theory for the single PS method shows that the complex EM field at the PS can be recovered by an inverse Fresnel transform of the RO data [15]. The assumptions about the single PS geometry, in particular the distance of the PS to the LEO, are however critical for the final accuracy. A proper placement of the PS has been found to be difficult to achieve. This difficulty is mainly due to the shift of the ray perigee during a RO event, which is large (70–80 km) during the tropospheric crossing, just where the method is needed. For this reason a modified inverse Fresnel transform has been developed that includes an empirical model of the PS motion as a function of height [15,17]. In this way, both theory and results from the inversion of GPS/MET data show that a good accuracy is obtained down to a height of ~5 km. Below this height, the empirical correction of the PS placement is insufficient and the errors increase substantially. Fig 5 shows an example of temperature profile at 52 deg N, 12 deg E recovered using the Fresnel transform inversion and compared to ECMWF analysis. The results agree quite well and the retrieved profile shows significant vertical temperature structure (derived assuming dry air) thanks to the improved vertical resolution. Further results can be found in [15].

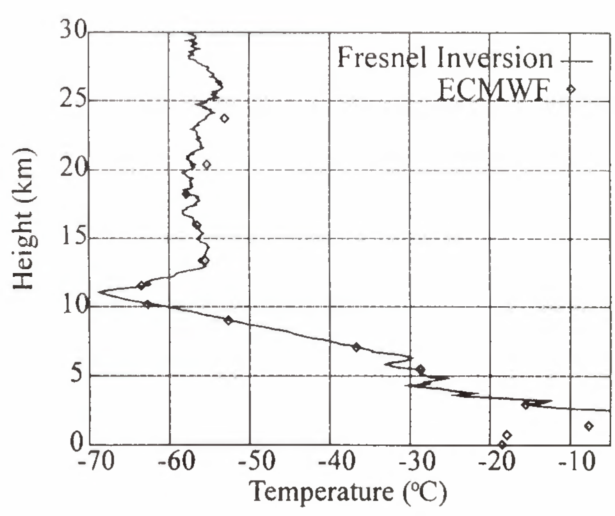


Fig. 5: temperature profile at 52 deg N, 12 deg E recovered using the Fresnel transform inversion

With the AS synthesis method, the complex EM field is re-computed (back-propagated) on an auxiliary plane close to the atmospheric region where inhomogeneities cause multipath and diffraction effects. While in the previous method the PS is assumed to be in vacuum, with the AS method the auxiliary plane is within the atmosphere (i.e., no attempt is made to model the entire atmosphere with one or more PS). Fig. 6 depicts the basic geometry for applying the AS method and emphasizes the occurrence of multipath. The back-propagation is an inverse diffraction problem, solved using the Helmholtz equation (for the bidimensional case, since diffraction and multipath effects are significant only in the occultation surface) [11]. From a physical point of view, this is equivalent to back-propagating the interfering rays along straight lines, so disentangling the multipath structure. The sensor data constitute the boundary conditions for this problem, but needs to be known only on a source curve that encloses a small domain of the LEO orbit and has a length of a few Fresnel zones. Once the complex EM field on the auxiliary plane is known, the phase information is used as before to derive refraction angle and refractivity profiles. Fig. 7 and 8 show two examples of refraction angle retrieval obtained with the AS diffraction correction method [15]. Note that the function $\varepsilon(p)$ is ill-defined before the diffraction correction (for a single value of the impact parameter p it gives more than one angle), while it returns meaningful results after the correction (different values of the impact parameter can result in the same angle, e.g. when multipath occurs). With this method, the placement of the auxiliary plane is not critical and improved results are obtained for the profile in the lower troposphere. This diffraction correction method allows one to retrieve the small-scale structure of the refractivity and improve the (sub-Fresnel) vertical resolution to $\lambda D / S$ where D is the distance from the atmospheric region to the LEO, i.e., to ~100 m for S= 6 km.

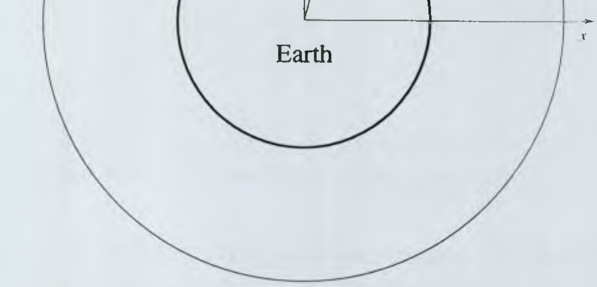


Fig. 6: geometry for multipath propagation and diffraction correction

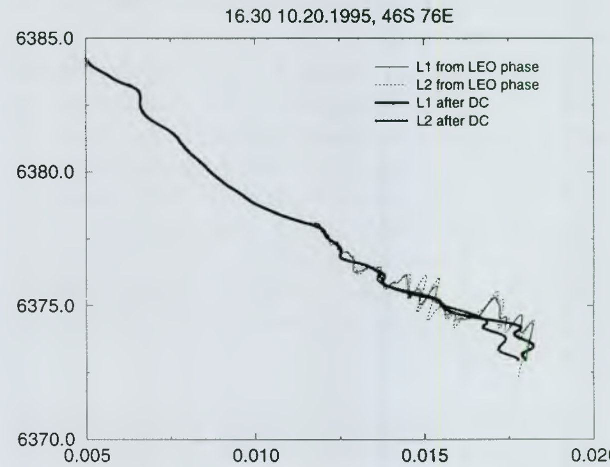


Fig. 7: tropospheric refraction angle in rad (abscissa) versus ray impact parameter in km (ordinate) for a GPS/MET RO at low latitude (L1: signal at 1.6 GHz; L2: signal at 1.2 GHz; DC: diffraction correction)

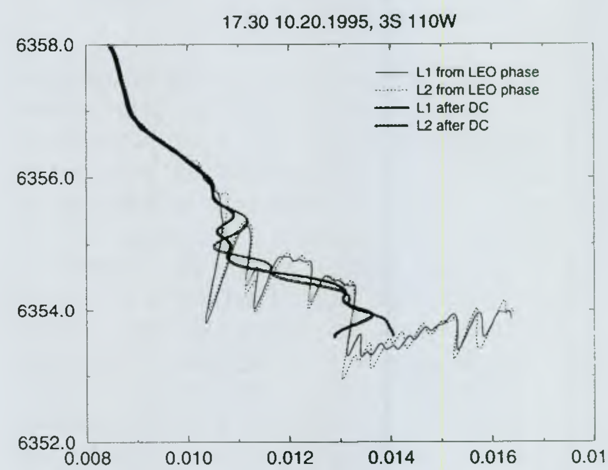


Fig. 8: tropospheric refraction angle in rad (abscissa) versus ray impact parameter in km (ordinate) for a GPS/MET RO at low latitude

CONCLUSIONS

The RO technique for atmospheric profiling has been described, with an emphasis on the main concepts used in the data analysis, particularly with regard to the refractivity profile derivation and the diffraction correction problem. The use of this technique, which will be started in an operational scenario within the ESA MetOp missions in 2003, will represent a major advance for meteorology and climate studies. One important item omitted in this review is the methodology for the assimilation of the RO data into NWP systems, which will be the key for the exploitation of the technique in operational meteorology. The progress in this domain can be found in a recent study [15], where the combination of the AS diffraction correction method with the processing required for NWP data assimilation is described in detail.

REFERENCES

1) A. Kliore et al., Occultation experiment: Results of the first direct measurement of Mars' atmosphere and ionosphere, Science, Vol. 149, No. 3689, pp. 1243-1248, Sept. 10, 1965
2) G. Fjeldbo, V. R. Eshleman, The atmosphere of Mars analyzed by integral inversion of the Mariner IV occultation data, Planetary Space Science, Vol. 16, pp. 1035-1059, 1968
3) G.L. Tyler, 'Radio Propagation Experiments in the Outer Solar System with Voyager', Proc. IEEE, Vol. 75, N.10, October
4) P.G. Steffes et al., Radio occultation studies of the Venus atmosphere with the Magellan spacecraft, Icarus, Vol. 110, 71-94, 1994
5) B. Lusignan et al., Sensing the Earth;'s atmosphere with occultation satellites, Proc. IEEE, Vol. 4, pp. 458-467, 1969
6) F. Fishbach, A satellite method for temperature and pressure below 24 km, Bulletin American Meteorological Society, Vol. 9, pp. 528-532, 1965
7) A. S. Gurvich, T.G. Krasilnikova, Navigation satellites for radio sounding of the Earth's atmosphere, Issled. Zemli Kosmosa., No. 6, 1987
8) V.V. Vorobev, T.G. Krasilnikova, Estimation of the accuracy of the atmospheric refractive index recovery from Doppler shift measurements at frequencies used in the NAVSTAR system, Physics of the Atmosphere and Ocean, No. 29, 602, 1994
9) H.P Ladreiter, G. Kirchengast, GPS/GLONASS sensing of the neutral atmosphere: model-independent correction of ionospheric influences, Radio Science, Vol. 31, No. 4, July-August 1996
10) W.G. Melbourne et al., The application of spaceborne GPS to limb sounding and global change monitoring, JPL Publication 94-18, 1994
11) M. Born, E. Wolf, Principles of optics, 6th ed., Pergamon Press, Oxford, UK, 1993
12) R.N. Bracewell, The Fourier transform and its applications, 2nd ed., McGraw-Hill, 1986
13) E.R. Kursinski et al., Observing the Earth's atmosphere with radio occultation measurements using the Global Positioning System, Journal of Geophysical Research, Vol. 102, No. D19, 20 Oct. 1997
14) E. T. Karayel, D.P. Hinson, Sub-Fresnel vertical resolution in atmospheric profiles from radio occultation, Radio Science, Vol. 32, No. 2, pp. 411-423, 1997
15) M. Gorbunov et al., Synergies and impacts of Global Navigation Satellite Systems for operational meteorology and climatology, Final report of ESA contract 11930/96/NL/CN, 1998
16) M. Gorbunov, A.S. Gurvich, L. Bengtsson, Advanced algorithms of inversion of GPS/MET satellite data and their application to reconstruction of temperature and humidity, Report 211, Max Planck Institute for Meteorology, Hamburg, 1996
17) M.D. Mortensen, P. Hoeg, Inversion of GPS occultation measurements using Fresnel diffraction theory, Geophysical Research Letters, Vol. 25, No. 13, pp. 2441-2444, 1 July 1998
18) K. R. Hardy et al., Accuracies of Atmospheric Profiles Obtained from GPS Occultations, International Journal of Satellite Communication, vol. 12, pp. 463-473, 1994
19) J. P. Luntama, Total bending error estimates for NWP applications, presented at Seminar on Atmospheric Sounding by Radio Occultation, ESTEC, 15-16 January 1998
20) I. R. Linscott, D.P. Hinson, G.L. Tyler, K.R. Hardy, Earth atmospheric profiles using GPS occultation, Proc. of International Technical Meeting of the U.S. Institute of Navigation, 1993
21) P. Hoeg et al., GPS Atmosphere Profiling Methods and Error Assessments, DMI Scientific Report 98-7, Copenhagen, 1998

MONITORING THE EARTH WITH SPACEBORNE SCATTEROMETERS

Volkmar Wismann

Institute for Applied Remote Sensing, Klarastraße 81, 79106 Freiburg, Germany
T: +49-(0)761-2851717; F: +49-(0)761-2851719; E: ifars@compuserve.com

ABSTRACT

Spaceborne scatterometers are active microwave instruments for measuring the normalized radar cross section (NRCS) of the Earth surface. These measurements are independent of cloud coverage and illumination by the sun and provide a global coverage within 3 to 4 days. Thus they are well suited for a wide range of operational monitoring tasks. Presently, the scatterometer is exclusively dedicated to the determination of the wind speed and direction over the oceans. It is increasingly acknowledged that, despite the coarse resolution, also a variety of geophysical parameters can be measured and monitored over land surfaces and sea ice. This paper presents an overview of recent developments for obtaining information on, e.g., sea ice, snow properties, thawing of soils, soil moisture, and vegetation from spaceborne scatterometer data.

INTRODUCTION

The first spaceborne scatterometer was operated during the American Skylab mission in 1973 and 1974, demonstrating the feasibility of spaceborne scatterometers. Between June and October 1978, the American SEASAT-A Satellite Scatterometer (SASS) proved that accurate wind field measurements could be made over the oceans from space. Since August 1991 the scatterometers aboard the European ERS satellites (ERS-1 and ERS-2) measure operationally the normalised radar cross section (NRCS) of the Earth's surface at C-Band (5.3 GHz). A global map of the NRCS is depicted in Figure 1.

Figure 1 : Map of the mean C-band NRCS at 40° of incidence angle.

During its short lifetime between August 1996 and June 1997 the Japanese ADEOS satellite provided the same type of measurements with the NSCAT system but for K_u-Band (14 GHz) at horizontal and vertical polarisation. In the near future further scatterometers will be flown mainly for operational wind measurements over the oceans. This is the reason why these systems are often called windscatterometers or even named *SeaWinds*. Table 1 provides an overview over past, present and future systems.

Satellite	Instrument Radar Band	Operation Launch
SEASAT	SASS - K_u-Band	Jun '78 - Oct '78
ERS-1	AMI - C-Band	Since Jul. '91
ERS-2	AMI - C-Band	Since Mar. '96
ADEOS	NSCAT - K_u-Band	Aug. '96 – Jun. '97
QuikScat	SeaWinds - K_u-Band	24. Nov. '98
ADEOS II	SeaWinds - K_u-Band	2000
METOP	ASCAT - C-Band	2003

Table 1 : Overview over past, present, and future spaceborne scatterometer systems.

The scatterometer measurements are independent of cloud coverage and illumination by the sun and provide typically a global coverage within 3 days. Presently, the scatterometer is dedicated only to determining the wind speed and direction over the ocean by measuring the normalised radar cross section of the sea surface. Despite its coarse resolution it was recognised that the spatial and temporal variability of a variety of geophysical parameters can be measured and monitored over sea ice and land surfaces. (Frison and Mougin, 1996a, 1996b, 1998; Kennett and Li, 1989a, 1989b; Kerr and Magagi, 1994; Long and Hardin, 1994; Magagi and Kerr, 1995; Mougin et al., 1994, 1995a, 1995b; Rott and Rack, 1995; Wiesmann and Mätzler, 1994; Wismann and Boehnke, 1994; Wismann et al., 1993, 1994a, 1994b, 1995, 1996a, 1996b).

Most of the data examples presented in this paper originate from the Active Microwave Instrument (AMI) flown on ERS-1 and ERS-2. This instrument operates at C-band (5.3 GHz) and uses three antennae with vertical polarisation (VV) in the scatterometer mode. They are looking 45 degrees forwards (Forebeam), sideways (Midbeam), and 45 degrees backwards (Aftbeam) with respect to the satellite flight direction. The incidence angle of the radar ranges from 18 to 57 degrees, illuminating a 500 km wide swath on the right hand side of the satellite track. The along-track and cross-track spatial resolution is 50 km (ESA, 1992). The measurement geometry is depicted in Figure 2.

Proc. 32nd ESLAB Symp., 'Remote Sensing Methodology for Earth Observation and Planetary Exploration', ESTEC, Noordwijk, The Netherlands, 15-18 September 1998 (ESA SP-423, December 1998)

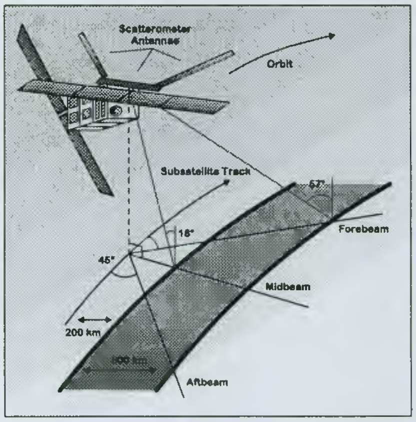


Figure 2 : Schematic of the ERS-1 spacecraft and the scatterometer measurement geometry.

Besides the strength of the radar backscatter the dependence of the NRCS on the incidence angles (slope) provides additional information on surface properties, allowing to discriminate between volume scattering, e.g., within a dense canopy, or surface scattering, e.g., from bare soil. Figure 3 depicts the NRCS as a function of incidence angle for two sites on Greenland and Siberia. Data are plotted for winter and summer to indicate also the seasonal variation. A global map of the slope is shown in Figure 4. Although this map looks similar to the NRCS map in Figure 1 there are obvious differences, which are related to surface characteristics.

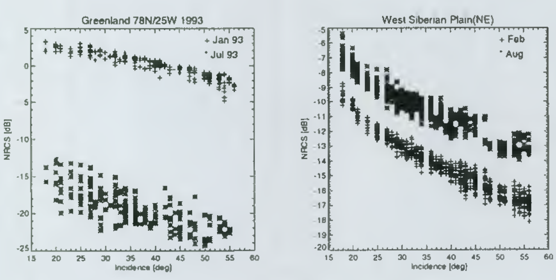


Figure 3 : Incidence dependence of the NRCS for two locations (Greenland and Siberia) and winter and summer, respectively.

*

Figure 4 : Map of the mean slope (defined as the derivative of the NRCS with respect to incidence angle).

The excellent calibration and maintenance of the ERS instruments guarantee high quality data, which, for the first time, allow a precise evaluation of the spatial and especially the temporal variability of the NRCS of the Earth's surface. Figure 5 depicts a time series of the C-Band NRCS for a location on Greenland. The measured *seasonal variability* of 0.3 dB indicates the accuracy of the instrument while the overall temporal stability reflects the excellent calibration and maintenance of both scatterometers. It is noteworthy that there was a switchover from ERS-1 to ERS-2 in May 1996, which is not visible in the data.

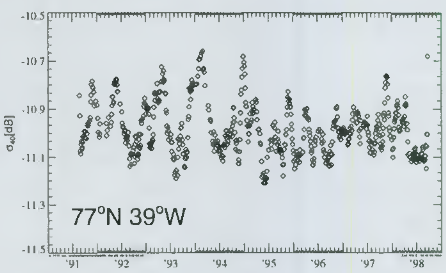


Figure 5: Time series of the C-Band NRCS at 77°N; 39°W on Greenland.

WIND MEASUREMENTS OVER THE OCEANS

Radar backscattering from the ocean surface is determined primarily by the short-scale surface roughness which is created by the local wind. Thus it is possible to infer sea surface winds from the ocean's NRCS measured by a scatterometer. This NRCS depends on wind speed and wind direction relative to the antenna look direction (azimuth angle), on radar frequency, radar polarisation, and on the incidence angle of the radar beam. In the past these dependencies have been studied extensively within the framework of the SEASAT and ERS-1 missions. Figure 6 depicts the dependence of the NRCS on incidence angle, wind speed and radar look direction with respect to the wind direction.

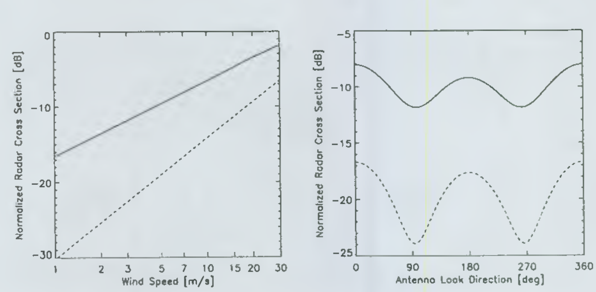


Figure 6 : Dependence of the NRCS at 30° and 50° of incidence angle (solid and dashed curves, respectively) on wind speed (left) and radar look direction (right). The antenna is looking into the wind at 0° and the wind speed is 7 m/s.

Theoretical as well as empirical model functions have been developed and the wind retrieval techniques were improved during the last years of ERS operation. Today the wind product from the ERS scatterometer is assimilated operationally into numerical weather prediction models by all major meteorological services. For a review of wind measurements over the ocean by means of spaceborne radar it is referred to the special section "Advances in oceanography and sea ice research using ERS observations" of the Journal of Geophysical Research (JGR, 1998) and the literature cited their.

SEA ICE MONITORING

The formation and melt of sea ice are spectacular manifestations of the heat exchange between the ocean and the atmosphere. Sea ice, once formed, inhibits exchanges across the air-sea interface and considerably modifies the climate of polar regions. Although first-year ice is seldom more than one meter thick, multi-year ice, as observed in the Canadian Arctic, may reach a thickness of five meters. Ice ridges, formed by the buckling of the ice sheet due to compression imposed by the wind drag, may reach heights from keel to crest greater than twenty meters. Understanding and monitoring the mechanisms of sea ice formation and dynamics is therefore important both for scientific and navigational applications.

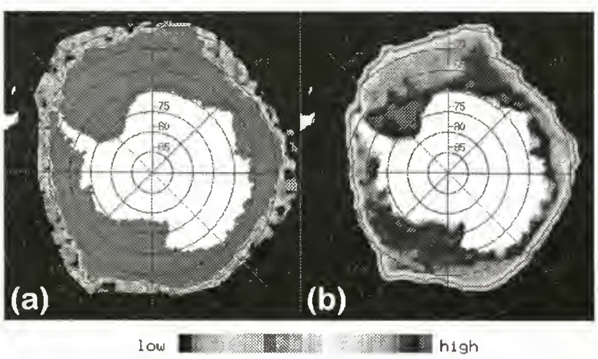


Figure 7 : Ice products from active and passive microwave sensors. (a) Ice mask derived from ERS-1 scatterometer data and (b) ice concentration deduced from SSM/I data.

Scatterometer data can be used to distinguish different ice types, to provide ice edge information, to monitor sea ice drift, and to detect summer melting. Microwave radiometer data appear to be better suited for determining ice concentrations than scatterometer data,. The scatterometer is capable of furnishing the ice edge information necessary for climatological purposes with a precision comparable to that of passive microwave instruments. Figure 7 depicts a comparison of the ice edge masks for the Antarctic derived from the scatterometer data and the respective ice concentration derived from passive microwave SSM/I data[1].

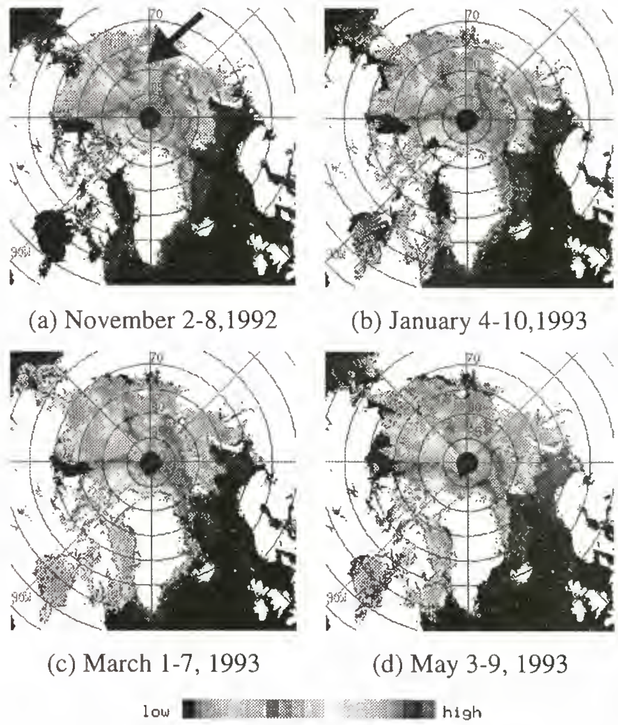


Figure 8 : Ice drift through the Arctic Ocean, Nov.'92 to May '93. Note the multi-year ice feature drifting from the Siberian Archipelago towards the North, marked by the arrow in map (a).

Multi- and first-year ice can be discriminated by their NRCS and thus be monitored using the ERS scatterometer data. Figure 8 shows a series of images of backscatter at 40° of incidence angle over the Arctic Ocean, which illustrates the potential to distinguish ice types and to monitor ice drift. Low backscatter near the Siberian coast corresponds to first year ice, high backscatter near the North American continent to multi-year ice; a mass of ice is detected drifting from the Novosibirskiye islands northward at a speed of about 50 km/month. The detectability degrades at the beginning of summer, when surface melting hides the backscatter gradient between different ice surfaces. The structures in the sea ice can be distinguished again at the beginning of fall when the surface re-freezes. Further examples are given by Gohin et al. (1998). For a detailed review of the radar backscatter from sea ice it is referred to the special section on "Electromagnetic properties of sea ice" in IEEE Transaction on Geoscience and Remote Sensing (IEEE, 1998) and the literature cited their. Today the French Processing and Archiving Facility CERSAT produces operationally sea ice maps from ERS scatterometer data.

---

[1] Figure 7 and Figure 8 were produced by A. Cavanié (IFREMER) and taken from Wismann et al. (1996b)

FROZEN SOIL MONITORING

Based on ERS scatterometer data a method was developed to monitor the state (frozen/thawed) of the upper layer of the soils in arctic to temperate climate regions (Boehnke and Wismann, 1996a, 1996b, 1997a, 1997b). When the vegetation cover is sparse or absent, the NRCS at C-band depends mainly on the moisture content of the soil, the dielectric constant, the penetration depth and the surface roughness. The radar backscatter increases with soil moisture and with surface roughness. When liquid water in soil freezes, the dielectric constant of the soil decreases dramatically. This process is reversed in spring when the soil thaws. From NRCS time series it can be seen that in spring, during snow melt, the radar cross section first decreases by up to 5 dB and shortly later, when the soils start thawing, increases dramatically.

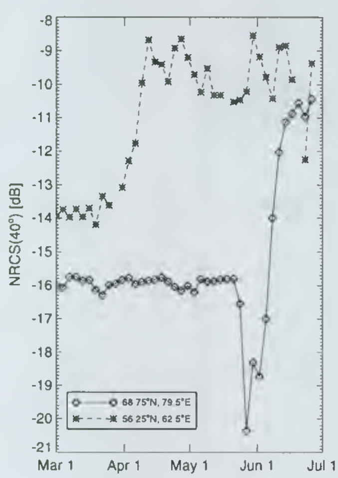


Figure 9 : Time series of σ_{40} for two grid points in spring 1993.

Figure 9 shows two typical time series of the NRCS at 40° of incidence angle (σ_{40}) for two grid points in Siberia, their geographical positions are given in the figure. Note the extreme temporal stability of the radar cross section during early spring and the large variations (>5 dB) associated with the onset of thawing. This is delineated for begin of April and June for the grid points located south and, respectively. As can be seen from the σ_{40} time series of the northern grid point, spring snow melting first leads to a strong short-term decrease of 4 dB. This can be explained by the increasing wetness of the snow leading to an enhanced absorption of the microwaves in the wet snow cover and the subsequent low radar return from melt water ponds formed on the frozen grounds when infiltration is still blocked.

The application of a change detection algorithm to the scatterometer data over the Siberian test site revealed isochron-maps of thawing for a region extending from 50° N to 80° N and from 42° E to 172° E. The signatures on these maps correlate well with the geographical distribution and the interannual variability of air temperature[2] and snow cover[3]. These isochron-maps are shown in Figure 10 for 1992 and 1995. Each colour represents the area thawed before the date marked under the colour scale. Each of the 16 colours corresponds to a period of 8 days between March 1 and June 28. Superimposed are contour lines for elevations of 100, 250, 500 and 1000 m (Lee and Hastings, 1995).

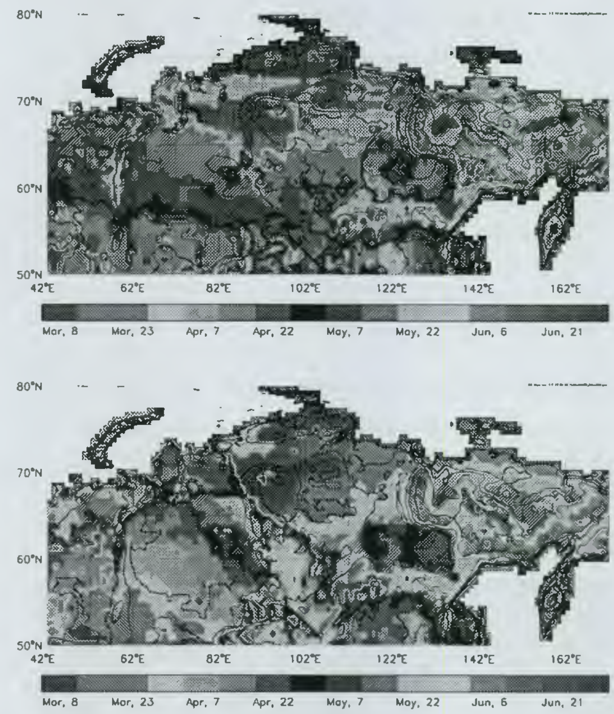


Figure 10 : Maps of Siberia, depicting isochrons of the "onset of spring" derived from scatterometer data for 1992 and 1995 (top and bottom).

The temporal and spatial evolution of the onset of thawing differs significantly over the years. In 1992, the onset of thawing is late. After the southern part of the test site starts thawing in April, thawing gradually moves towards the north coincident over the complete range of longitudes. By the middle of May, thawing has reached most of Siberia with the exception of the Central Siberian Plateau. For 1995, thawing commences early in the southwest and progresses towards the northeast somewhat slower than in 1992, so that by the middle of May in both years, approximately the same area has thawed (green to grey colours). In both years, the Central Siberian Plateau starts thawing in June. On a regional scale the onset of thawing follows orographic particulars, e.g., mountainous regions like Ural, Werchojansk, and Stanowoi thaw later than regions of

---

[2] Global temperature data of the NCAR/NCEP reanalysis were provided by NOAA Climate Diagnostics Center.

[3] Ralph Ferraro, Microwave Sensing Group, NOAA/Satellite Research Laboratory, provided global snow cover and precipitation data derived from SSM/I radiometer data.

lower elevation. It is remarkable how well some isochrons are in accordance with the elevation contour lines, especially in eastern Siberia.

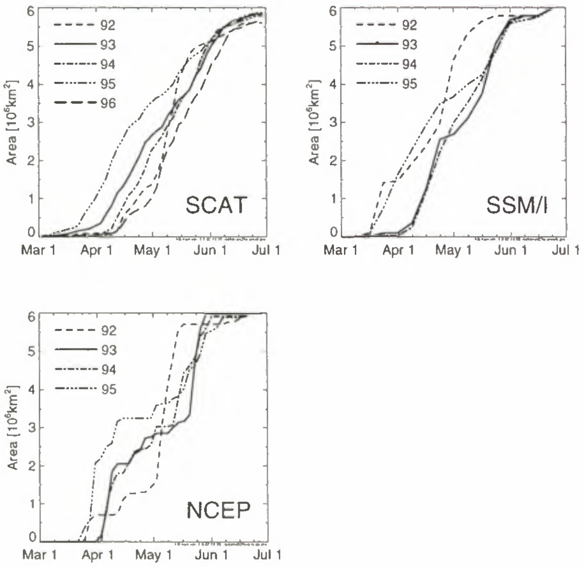


Figure 11 : Time series of the extent of the area detected as thawed (SCAT), depleted of snow cover (SSM/I), and with mean air temperatures above 0°C (NCEP).

For a sub-region of the Siberian test site encompassing the West Siberian Lowland and the Central Siberian Plateau (55°-70° N; 50°-120° E), the interannual variability of soil thawing was estimated. The thawed area was derived from the scatterometer data, the SSM/I snow cover information, and the air temperature data for time steps of 3 days. The slope of the area versus time curves is rather constant for all the years except for 1992 (see Figure 11). In the years 1993 to 1996, the thawed area gradually increases throughout spring, whereas in 1992, the onset of thawing is late, but then the majority of the Siberian sub-region thaws much quicker than in the other years. In 1995 thawing begins very early and in 1996 very late. The overall large-scale interannual variability is on the order of one month.

Analogous time series plots are shown for the SSM/I snow cover data and the air temperature data in Figure 11. The characterizations of "spring" from the scatterometer data, the SSM/I data and the air temperature data agree well. The early progress in 1995 and the rapidity of the "onset of spring" in 1992 are consistent. Discrepancies are found for 1992, where the "onset of spring" in the SSM/I data (the area no longer covered with snow) is detected very early. The step-like increases in the NCEP-plot arise from the large grid size of the data being 2.5° by 2.5°.

PREDICTING INUNDATION

The Pantanal in South America is one of the largest tropical wetlands. It is located in the upper Paraguay River basin, covering an area of 137,000 km², located mainly in Brazil, only in the West, small parts are in Bolivia and Paraguay. Presently, the main river system is still unregulated. The climate of the Pantanal is tropical with a marked wet season between November and March. The annual rainfall across this region is generally 1000-1700 mm/a. Maximum inundation occurs in February in the northern regions and due to the delayed drainage of the river system as late as June in the southern part. The total area inundated can vary from 11,000 to 110,000 km². Hamilton et al. (1995) demonstrated that microwave radiometer data can be used to monitor the inundation in the Pantanal area.

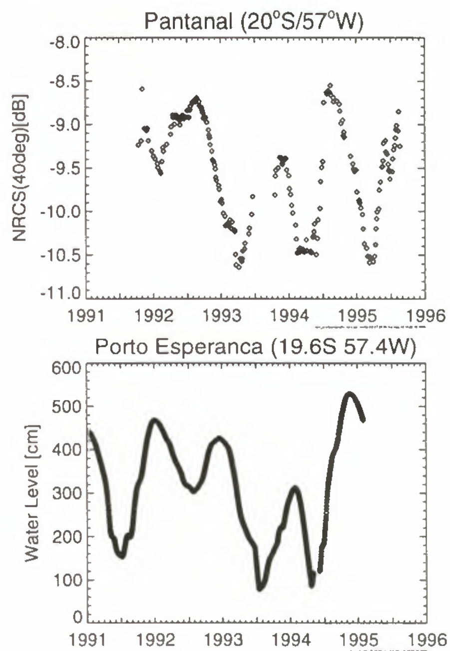


Figure 12 : Time series of the NRCS (top) and of the water level at Porto Esperanca (bottom).

Wismann and Boehnke (1996a) analysed ERS-1 scatterometer over the Pantanal region acquired between November 1991 and August 1995. During this period the inundation extent varied considerably, the relatively wet year 1992 was followed by dry years 1993 and 1994, while in 1995 a maximum inundation was encountered. The waterlevel at the station Porto Esperance[4] is representative for the inundation extent in this area. The time series of the water level and of the NRCS for the respective area is shown in Figure 12.

The high correlation between the water level and the NRCS also reflects the large interannual variability.

[4] Water level data for the Paraguay River were provided by Senhor Vinicius and Senhor Avila of Departamento Nacional de Águas e Energia Elétrica (DNAEE), Brazil.

Furthermore a phase shift between both time series of 1-2 months was found, whereby the radar signal leads the inundation extent. The radar cross section reacts to the onset of vegetation growth in the rainy season which happens timely between the actual precipitation in the north and the flooding in the south. However, there is no satisfactory explanation for this phase between the water level and the NRCS. But, the possibility of predicting the inundation extent by means of scatterometer measurements at least 1 month in advance might have a great impact on various ecological and economic aspects related to the rivers Paraná and Paraguay.

MONITORING SNOW PROPERTIES ON GREENLAND

Greenland's snow and ice surface can be divided into four shell-like zones: the central dry snow zone at high altitudes is surrounded by the percolation zone, the wet-snow zone, and the ablation zone. These facies result from different diagenesis of the snow and ice cover which is determined by the amount of snow-accumulation and melt and, therefore, on the local climate at the respective elevation (Benson, 1962). A large lateral shift of the borderlines between the different snow and ice facies will result even from a slight climate change due to the very gentle slope of the ice shield. Therefore, monitoring these borders can provide information on climate change. The capability of discriminating different snow and ice facies by radar remote sensing was shown for the C-band SAR of ERS by Fahnstock et al. (1993) and for the SEASAT K_u-band and the ERS C-band scatterometers by Long and Drinkwater (1994a, 1994b).

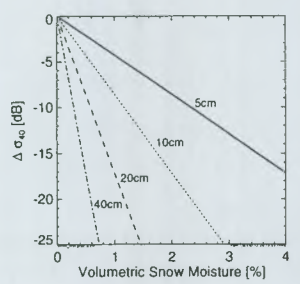


Figure 13 : Change in σ_{40} as a function of volumetric snow moisture and depth of the wet snow.

SUMMER SNOW MELTING

Due to the large difference between the dielectric constants of ice and water, the radar is very sensitive to the free liquid water content of snow. The change in radar cross section as a function of snow moisture and depth of the wet snow layer is depicted in Figure 13 (Winebrenner et al., 1994). Although from ERS scatterometer measurements it is not possible to discriminate between an increase in snow wetness within a thin snow layer and an increase in depth of a wet snow layer, relative spatial and temporal changes, in particular the intensity of snow melting processes, can be monitored.

A typical time series of σ_{40} at the border between the percolation and wet snow zone is depicted in Figure 14. The dominating signal is a strong short-term decrease of several dB in radar cross section every year during the months June to August. These *"summer-dips"* coincide well with the occurrence of air temperatures above 0° Celsius (Wismann and Boehnke, 1997).

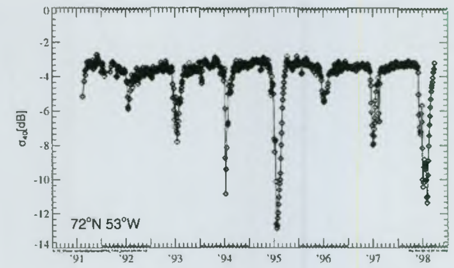


Figure 14 : NRCS time series at 73°N; 53°W.

The time series in Figure 15 summarises the main results of a spatial analysis of the snow melting. The extent of snowmelt increased dramatically between summer 1992 and 1995. Concomitant, a similar trend can be observed for the mean decrease in NRCS and the mean elevation of the melt area. This indicates that the summer melting of snow has intensified from 1992 to 1995. However this trend did not continue in 1996 to 1998. In 1998 the duration of the summer melt was remarkably long compared to the other years. This temporal behaviour can be seen also in the NRCS time series in Figure 14.

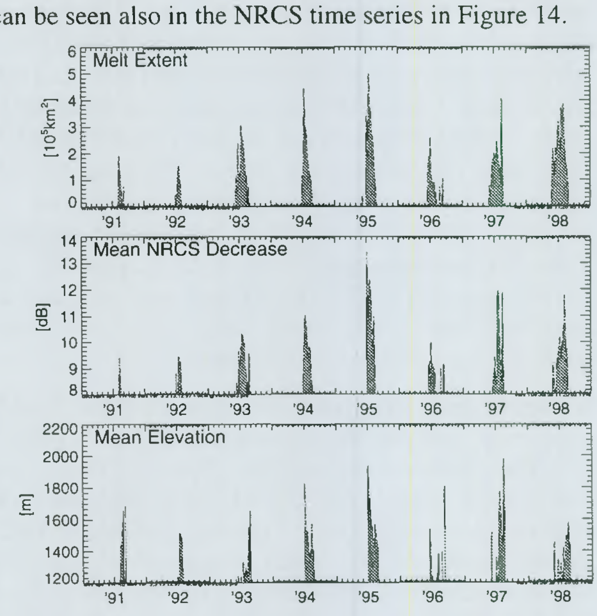


Figure 15 : Time series of the extent of the area affected by summer melting (top) and the average decrease in σ_{40} within the melt area (middle), and the mean elevation of the melt area (bottom).

ESTIMATING SNOW ACCUMULATION

A different temporal behaviour of the NRCS can be observed along the dry-snow line between the dry snow and percolation zone. Figure 16 shows a nearly linear decrease of σ_{40} for the ERS and the NSCAT data. For this location Ohmura and Reeh (1991) reported high precipitation and snow accumulation rates. The short interruption in this trend in summer 1995 will be discussed in the next section. The model describing the decreasing radar backscatter at the dry-snow/percolation zone boundary is based on the following physical picture: During 1991-1995, the boundary of the dry-snow zone evidently moved downslope from its previous location (Wismann et al., 1997). Thus ice layers and pipes in the percolation zone (which cause strong backscattering) were progressively buried under several years accumulation of dry-snow which did not melt, even during summers, until 1995.

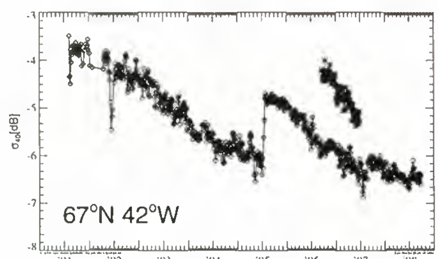


Figure 16 : Time series of σ_{40} at 67°N; 42°W, diamonds and stars represent ERS C-Band and ADEOS K_u-Band (NSCAT) data, respectively.

Dry-snow overlying the percolation zone structure diminishes backscattering by means of at least two mechanisms. First, scattering depends on the dielectric contrast between ice structures and the surrounding snow. The ice structures are incompressible but the snow densifies as being buried, reducing this contrast. Quantitative modeling shows, however, that this effect is likely to be small. Second, propagation through the dry-snow layer to and from ice structures of the firn layer is effectively attenuated by absorption in the snow, by scattering within the dry-snow layer, and by reflection away from the ice structures of the firn as well as from the radar at internal density layer interfaces (West et al., 1996). Interface reflection, in particular, leads to significantly greater effective attenuation than would otherwise be expected. Figure 17 depicts the results of a physical 2-layer model of radar backscatter from a dry-snow covered firn layer (Wismann et al., 1997). For a thin dry-snow layer the radar cross section is dominated by the scattering from the firn, but with increasing thickness of the dry-snow layer the backscatter from the buried firn is reduced significantly. Maps of the NRCS decrease at C-band (ERS) and K_u-band (NSCAT) are depicted in Figure 18 for the period September 1996 to June 1997 when ERS and ADEOS provided simultaneously data. Both data sets show similar spatial patterns, although, at K_u-band the sensitivity to snow accumulation is approximately twice as high as at C-band. This is in accordance with the NRCS trends depicted in Figure 16 and the results of the modelling shown in Figure 17.

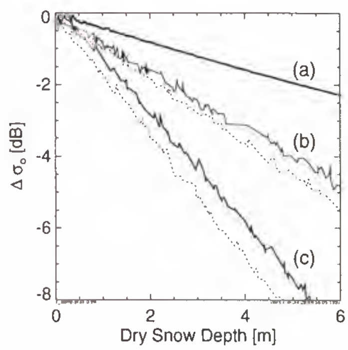


Figure 17 : Reduction of σ_{40} of firn as a function of the thickness of a covering dry-snow layer. For C-band and 1 mm grain size (a), and K_u-Band with grain sizes of 0.5 (b) and 1 mm (c). The solid and dotted lines are for VV and HH polarisation, respectively.

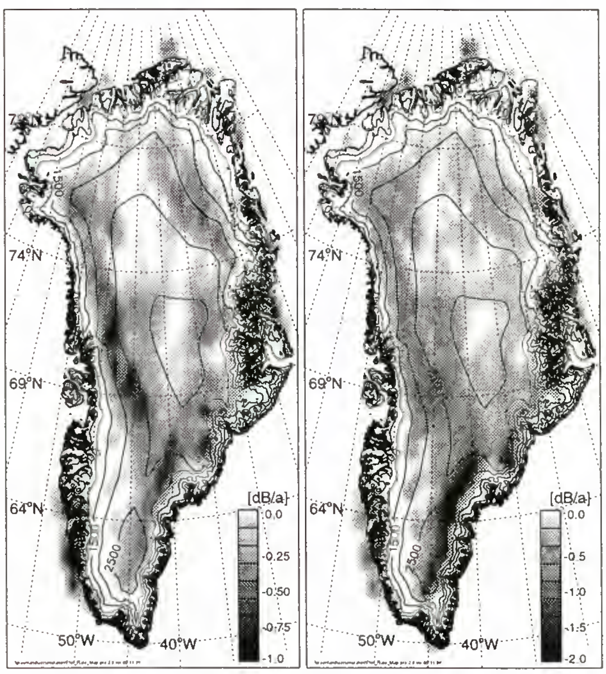


Figure 18 : Maps of NRCS decrease for C-band (left) and K_u-band (right) for the period Sep.'96 to June '97. The scales differ by a factor of 2.

These model curves can be used to transform the NRCS decrease to snow accumulation rates. The spatial pattern of these snow accumulation rates as well as their absolute values agree well with observations reported by Ohmura and Reeh (1991). There is a zone of high accumulation stretching along the westward slope having one maximum around 69°N and a second maximum far in the north around 76°N 60°W. Ohmura and Reeh gave precipitation rates of 500-600 mm/year for the maxima which corresponds approximately to 2 metres of dry-snow when assuming values for snow density of 0.35 g/cm³ to 0.4 g/cm³.

OBSERVING SNOW METAMORPHISM

The sudden and strong increases in radar cross section in summer 1995 and summer 1997 is striking as shown in Figure 19.

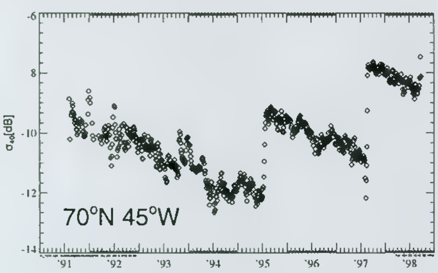


Figure 19 : Time series of σ_{40} at 70°N; 45°W.

The location and spatial distribution of the magnitude of these *"jumps"* is depicted the maps in Figure 20. In summer 1995, snow melting had its maximum extent and most of the percolation zone in western Greenland was affected, while in summer 1997 mainly the southern part of Greenland was influenced. Detailed data analysis revealed that σ_{40} is increasing directly after the event of snow melt (Wismann and Boehnke, 1997). This reflects the so-called melt-related metamorphism of dry snow into firn (Echelmeyer et al., 1992) resulting in enhanced NRCS is for these areas.

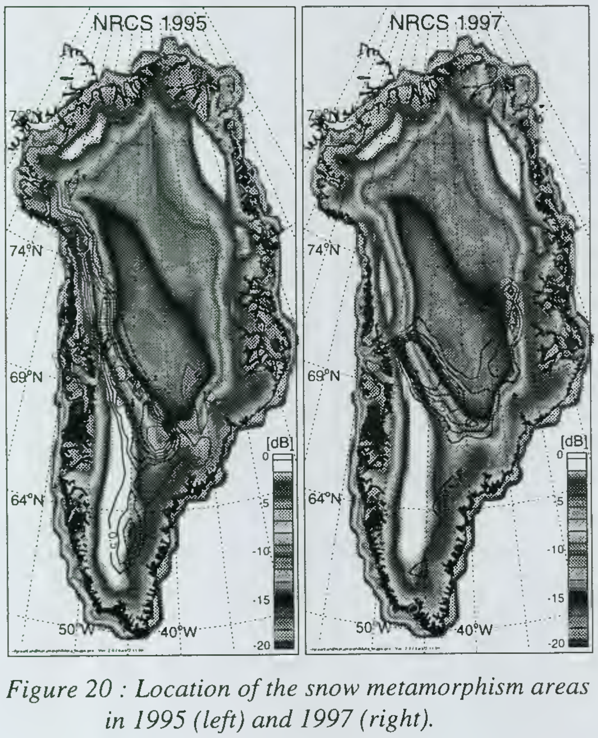


Figure 20 : Location of the snow metamorphism areas in 1995 (left) and 1997 (right).

MONITORING VEGETATION

The radar cross section of the Earth's surface decreases typically with incidence angle θ. The slope of the NRCS versus incidence angle curve is indicative for the responsible radar backscatter mechanism and/or for the properties of the backscattering surface. For pure volume scattering, this slope is zero, while for surface scattering this slope decreases with surface roughness. For vegetated areas the radar backscatter is governed by a combination of volume and surface scattering, whereby the amount of volume scattering depends on the canopy coverage and its density. Thus, the slope provides additional information to the NRCS measurement. We define the slope (α) as:

$$\sigma_0 = a\theta^{\alpha}$$

whereby α is computed by a linear regression between $\log(\sigma_0)$ and the incidence angle θ. Wismann et al. (1996b) introduced the Radar Backscatter Index (RBSI) in order to combine the complementary information of the slope and the NRCS:

$$RBSI = \frac{\overline{\sigma_0(40^O - 57^O)}}{|\alpha|}$$

This parameter improved the discrimination between different surfaces by taking into account not only the strength of the radar backscatter, but also information on the backscatter mechanism.

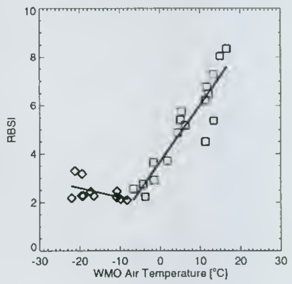


Figure 21 : RBSI as a function of air temperature at the WMO-station Salechard.

Siberia is well-suited to study the effect of air temperature on the development of vegetation covers since the temperature varies from -30°C to +20°C throughout the year. The growth of vegetation follows the increase of the temperature in spring and summer, giving rise to a gradual increase in vegetation density, with a maximum in late summer. This process is reversed starting in autumn with the decrease of temperature. With decreasing vegetation density, the ratio between the contributions of volume and surface scattering to the NRCS also changes. This phenomenon is well-reflected when looking to the RBSI dependence on temperature as shown in Figure 21. When the canopy is frozen (at least below -10°C) the RBSI stays constant. Although the dielectric properties of moist vegetation change by more than an order of magnitude due to freeze/thaw processes, this transition does not occur typically at 0°C, but at temperatures well below due to the presence of super-coolants such as sucrose solutions in the plant fluids. For temperatures above 0°C, the dielectric constant of vegetation is relatively insensitive to changes in temperature. Therefore, the increase in RBSI with temperature has its origin in the increase of vegetation density. Figure 22 illustrates the correlation between canopy density and RBSI for Siberia (Wismann et al. 1996b).

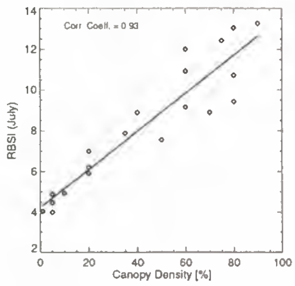


Figure 22 : Correlation between the RBSI and canopy density.

The high correlation between the RBSI and the canopy density was used to produce a map of canopy density for Siberia, which is shown in Figure 23. The boundaries between the forest and the Tundra in the north and between the forest and the wheat belt in the Kazakhstan steppe in the south are well depicted on the map. The latter are classified as areas of negligible canopy density due to extremely low RBSI values. When comparing this map with a forest map of this area, a lot of details are in good agreement, e.g., the dense forest along the Ural mountains, especially on the western slope, the variability of the vegetation along the river Ob in the West Siberian Lowland, and different forest types and densities in the Central Siberian Plateau. Also the small and isolated forests north of Semipalansk (51^{O}N; 80^{O}E) are detected.

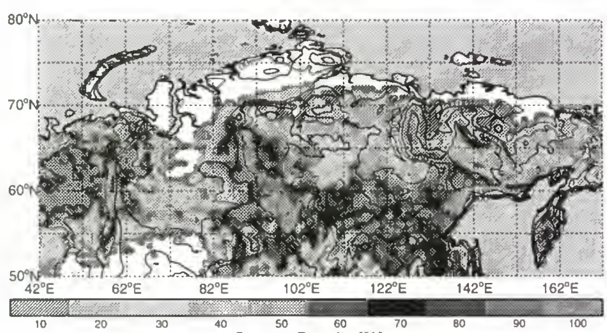


Figure 23 : Canopy density map of Siberia, derived from RBSI data.

The seasonal variability of the NRCS as depicted in Figure 24 accentuates the areas with strong seasonal variations in the vegetation cover, e.g., due to seasonal rainfall like in the Sahel region. Here the scatterometer data allow monitoring the interannual variability as shown in Figure 25. The comparatively high wetness in 1994 can be delineated from the enhanced NRCS as well as from the fact that the seasonal NRCS increase reached far more north than in the other years (Wismann et al., 1995).

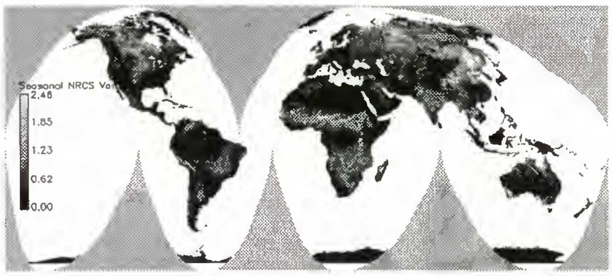


Figure 24: Seasonal NRCS Variability.

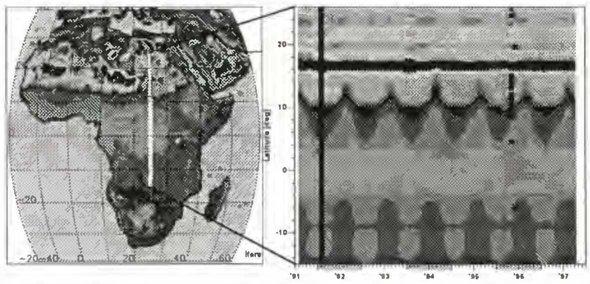


Figure 25 : Hovmoeller diagram of the NRCS along a cross section through Africa.

SOIL MOISTURE MEASUREMENTS

The dielectric properties of soil strongly depend on its moisture content. However, due to the limited penetration depth of microwaves into the soil, scatterometer can only provide moisture estimates in the soil surface layer (0.5 – 5 cm). This layer is the interface between the highly dynamic atmosphere and the deeper soil layers, and hence its moisture content is subject to short-term fluctuations on temporal scales of less than one day, depending on the type of soil. The resulting NRCS fluctuations are shown in Figure 26[5]. Increases in NRCS are well correlated with precipitation events while the seasonal variation due to vegetation effects (dashed curve in the NRCS times series) is much smaler. Wagner et al. (1998) used the ERS scatterometer data in order to estimate the degree of saturation of the soils. As a result maps of the Iberian Peninsula were produced which show the spatial and temporal variability of the soil moisture (see Figure 27) which reveal the seasonal rain variability (winter rain) as well as the interannual

[5] Figure 26 and Figure 27 were provided by W. Wagner, Space Application Institute, Joint Research Centre, Ispra, Italy.

variation. A drought was encountered in 1992 and 1993 and flooding in 1995/1996).

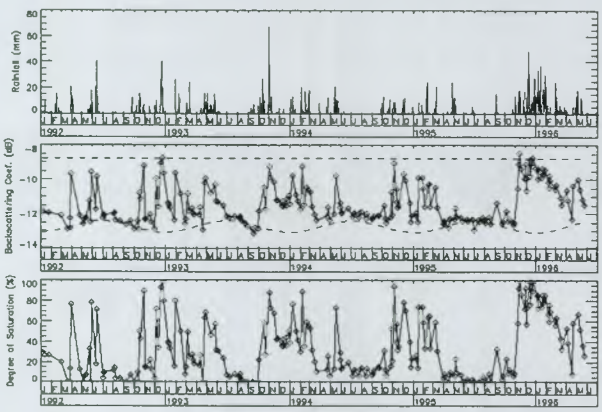


Figure 26 : Time series of rainfall (top), NRCS (middle), and degree of saturation for the station Beja, Portugal.

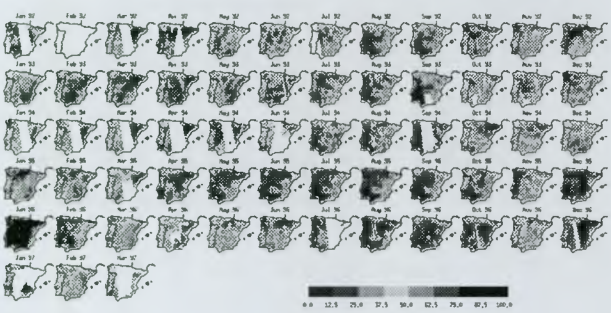


Figure 27 : Time series of maps of the degree of saturation for the Iberian Peninsula.

CONCLUSIONS

Despite its poor spatial resolution the scatterometer is well suited for global monitoring task due to the good temporal and spatial coverage and the relative small amount of data. Examples were presented demonstrating the sensitivity of radar cross section to changes in ice, snow, soil, and vegetation properties, allowing a wide range of applications of these data in the field of global monitoring.

The ERS data are of excellent quality in terms of calibration and temporal stability, they are provided in near real-time and thus well suited for any kind of operational data assimilation tasks, e.g., into numerical weather models.

So far the ERS scatterometers have collected a unique data set for climate monitoring which should be continued in the future.

ACKNOWLEDGEMENT

A part of this work was conducted under ESA study contract No. 11103/94/NL/CN. The ERS and NSCAT data were provided by: the Centre ERS d´Archivage et de Traitement (CERSAT), France.

REFERENCES

Benson, C. S., Stratigraphic studies in the snow and firn of the Greenland ice sheet, SIPRE Res. Rep. 70, 1962.

Boehnke, K. and V. Wismann, ERS Scatterometer Land Applications: Detecting the Thawing of Soils in Siberia, *Earth Observation Quarterly*, ESA Publication Division, 52, 4-7, 1996a

Boehnke, K. and V. Wismann, Thawing of Soils in Siberia Observed by the ERS-1 Scatterometer Between 1992 and 1995, *Proceedings of the International Geoscience and Remote Sensing Symposium – IGARSS´96*, Lincoln, USA, May 27-31, Volume IV, 2264-2266, 1996b.

Boehnke, K. and V. Wismann, Detecting soil thawing in Siberia with ERS Scatterometer and SAR, *Proceedings of the 3rd ERS Symposium*, Florence, Italy, March 17-21, ESA SP-414, Vol. I, 35-40, 1997a.

Boehnke, K. and V. R. Wismann, Thawing processes during Siberian spring observed by ERS Scatterometer and SAR, *Proceedings of the International Geoscience and Remote Sensing Symposium - IGARSS´97*, Singapore, August 3-8, Volume IV, 1826-1828, 1997b.

Echelmeyer, K., W. D. Harrison, T. S. Clarke and C. Benson, Surficial glaciology of Jakobshaven Isbræ, West Greenland: Part II. Ablation, accumulation and temperature, Journal of Glaciology, Vol. 38, No. 128, 169-181, 1992.

ESA, ERS-1 System, ESA Publications Division ESTEC Noordwijk, The Netherlands, ESA SP-1146, 87 pages, 1992.

Ezraty R., A. Cavanie and F. Gohin, Monitoring Arctic sea ice using C and Ku-band scatterometers. Proc. Conference on Polar Processes and Global Climate, pp. 66-67. WMO/ICSU/IOC WCRP, Orcas Island, USA, 3-6 Nov. 1997.

Fahnstock, M. R., Bindschadler, R. Kwok, and K. Jezek, Greenland ice sheet surface properties and ice dynamics from ERS-1 SAR imagery, Science, Vol. 262, 1530-1534, 1993.

Frison P.-L, Mougin E., Monitoring global dynamics with ERS-1 wind scatterometer data, *Int. J. of Remote Sensing,* 17, 3201-3218, 1996a.

Frison P.-L, Mougin E., Use of ERS-1 wind scatterometer data over land surfaces, *IEEE Trans. Geosci. Remote Sensing,* 34, 550-560, 1996b.

Gohin, F., A. Cavanié, and R. Ezraty, Evolution of the passive and active microwave signatures of a large sea ice feature during its 2.5-year drift through the Arctic Ocean, *J. Geophys.Res.*, 103, C4, 8177-8189, 1998.

Hamilton, S. K., S.J. Sippel, and J.M. Melack, Inundation patterns in the Pantanal wetland of South America determined from passive microwave remote sensing, submitted to Water Resources Research, 1995.

IEEE, "Electromagnetic properties of sea ice" IEEE Transactions on Geoscience and Remote Sensing, 36, 5,September 1998.

JGR, "Advances in oceanography and sea ice research using ERS observations", *J. Geophys. Res.*, 103, C4, April 15, 1998.

Kennett, R. G., and F. K. Li, Seasat over-land scatterometer data, part I : Global overview of the Ku-band backscatter coefficients, IEEE Trans. Geoscience and Remote Sensing, 27,5,592-605,1989a

Kennett, R. G., and F. K. Li, Seasat over-land scatterometer data, part II : Selection of extended area land-target sites for the calibration of spaceborne scatterometers, IEEE Trans. Geoscience and Remote Sensing, 27,6,779-788,1989b

Kerr, Y.H., and R.D. Magagi, Use of ERS-1 windscatterometer data over land surfaces, Space at the Service of our Environment, Proceedings of Second ERS-1 Symposium held in Hamburg, Germany, 11-14 October 1993, ESA SP-361 (European Space Agency), 381-388, 1994.

Lee, W. R. III and D. Hastings, TerrainBase Global DTM Version 1.0 (on CD-ROM), National Geophysical Data Center and World Data Center-A for Solid Earth Geophysics, Boulder, Colourado, USA, 1995.

Long, D. G., and Hardin, P. J., Vegetation studies of the Amazon Basin using enhanced resolution Seasat scatterometer data, IEEE Transactions on Geoscience and Remote Sensing, 32(2), 449-460, 1994.

Long, D. G., and M. R. Drinkwater, Comparison of enhanced resolution images of Greenland from ERS-1 and Seasat scatterometer, *Proceedings of the International Geoscience and Remote Sensing Symposium, IGARSS'94*, Pasadena, California, USA, 8-12 August, 2382-2384, 1994a.

Long, D. G., and M. R. Drinkwater, Greenland ice-sheet surface properties observed by the Seasat-A scatterometer at enhanced resolution, J. of Glaciology, Vol. 40, No. 135, 213-230, 1994b.

Magagi, R.D., and Y.H. Kerr, Temporal and Spatial variabilities of surface roughness response over Hapex-Sahel area : use of ERS-1 windscatterometer. *Proceedings of the International Geoscience and Remote Sensing Symposium, IGARSS'95*, Congress Centre Firenze, Italy, 10-14 July, 1287-1289, 1995.

Mougin, E., A. Lopes, C. Proisy, A. Warich, P. L. Frison, D. Lo Seen, and P. Lecomte, Analysis of ERS-1 scatterometer data over land surfaces – preliminary results, *Proceedings of Second ERS-1 Symposium*, Hamburg, Germany, 11-14 October 1993, ESA SP-361 (European Space Agency), 393-398, 1994.

Mougin E., Lo Seen D., Rambal S., Gaston A., Hiernaux P., A regional Sahelian grassland model to be coupled with multispectral satellite data. I. Description and validation. *Remote Sens. Environ., 52, 181-193,* 1995a.

Mougin, E., Lopes, A., Frison, P. L., and Proisy, C., Preliminary analysis of ERS-1 wind scatterometer data over land surfaces, *International Journal of Remote Sensing*, 16(2), 391-398, 1995b.

Ohmura, A. and N. Reeh, New precipitation and accumulation distribution maps for Greenland, J. of Glaciology, Vol. 37, No. 125, 140-148, 1991.

Rott, H. and W. Rack, Characterisation of Antarctic Firn by means of ERS-1 scatterometer measurements, *Proceedings of the International Geoscience and Remote Sensing Symposium, IGARSS'95*, Congress Centre Firenze, Italy, 10-14 July, 2041-2043, 1995.

Wagner, W., G. Lemoine, H. Rott, A method for estimating the profile soil moisture content from ERS scatterometer and soil data, submitted. To *Remote Sens. Environ.*, 1998.

West, R. D., D. P. Winebrenner, L. Tsang, and H. Rott, Microwave emission from density-stratified Antarctic firn at 6 cm wavelength, J. of Glaciology, 42(140), 63-76, 1996.

Wiesmann, A. and C. Mätzler, Monitoring the temporal behaviour of land surfaces with ERS-1 wind scatterometer data, *Proceedings of Second ERS-1 Symposium*, Hamburg, Germany, 11-14 October 1993, ESA SP-361 (European Space Agency), 399-404, 1994.

Winebrenner, D. P., E. D. Nelson, R. Colony, and R. D. West, Observation of melt onset on multiyear Arctic sea ice using the ERS-1 synthetic aperture radar. Journal of Geophys. Res., 99, 22425-22441, 1994.

Wismann, V. and K. Boehnke, Land surface monitoring using the ERS-1 scatterometer, *Earth Observation Quarterly*, ESA Publication Division, No. 44, 11-15, 1994.

Wismann, V. and K. Boehnke, Can the ERS-1 Scatterometer be used to forecast the inundation extent in the Pantanal Wetland ?, *Proceedings of the International Geoscience and Remote Sensing Symposium – IGARSS'96*, Lincoln, USA, May 27-31, Volume II, 839-841, 1996a.

Wismann, V. and K. Boehnke, Monitoring snow Properties on Greenland with ERS Scatterometer and SAR, *Proceedings of the 3rd ERS Symposium*, Florence, Italy, March 17-21, ESA SP-414, Vol. II, 857-862, 1997.

Wismann, V., K. Boehnke, and C. Schmullius, Large scale C-band radar signatures of land surfaces measured by the ERS-1 Scatterometer, *Proceedings of the International Geoscience and Remote Sensing Symposium 93, August 16-19, 1993, Tokyo, Japan,* 1392-1394, 1993.

Wismann, V., K. Boehnke and C. Schmullius, Radar signatures of land surfaces measured by the ERS-1 Scatterometer, *Proceedings of The Second ERS-1 Symposium : Space at the Service of Our Environment, October 11-14, 1993, Hamburg, Germany*, ESA SP-361,405-410, 1994a.

Wismann, V. R., K. Boehnke, and C. Schmullius, Global land surface monitoring using the ERS-1 scatterometer, *Proceedings of the International Geoscience and Remote Sensing Symposium, IGARSS'94*, California Institute of Technology, Pasadena, California, USA 8-12 August, 1488-1490, 1994b.

Wismann, V. R., K. Boehnke, and C. Schmullius, Monitoring ecological dynamics in Africa with the ERS-1 scatterometer, *Proceedings of the International Geoscience and Remote Sensing Symposium, IGARSS'95*, Congress Center Firenze, Italy, 10-14 July, 1532-1525, 1995.

Wismann, V., A. Cavanie, D. Hoekman, I. Woodhouse, K. Boehnke and C. Schmullius, Land surface observations using the ERS-1 scatterometer, *Final Report Part I*, ESTEC Contract 11103/94/NL/CN, 57 pages, 1996a.

Wismann, V., K. Boehnke A. Cavanie, R. Ezraty, F. Gohin, D. Hoekman, and I. Woodhouse, Land surface observations using the ERS-1 scatterometer, *Final Report Part II*, ESTEC Contract 11103/94/NL/CN, 65 pages, 1996b.

Wismann, V., D. P. Winebrenner, and K. Boehnke, Snow accumulation on Greenland estimated from ERS scatterometer data, *Proceedings of the International Geoscience and Remote Sensing Symposium – IGARSS'97*, Singapore, August 3-8, Volume IV, 1823-1825, 1997.

Processing Technologies

ON-BOARD PROCESSING: FASTER, BETTER, CHEAPER!

Dr SM Parkes

Applied Computing, University of Dundee, Dundee, DD1 4HN, Scotland, UK

Tel: 44 1382 345194, Fax: 44 1382 345509, Email: sparkes@computing.dundee.ac.uk

ABSTRACT

Important objectives of the ESLAB 32 symposium are to review the status of remote sensing instrument technology, to consider possible future developments and to address the issue of pre-development of key instrumentation. On-board processing is a technology which is used in one form or another in all instruments. It covers instrument control, data processing, data storage and data transfer.

This paper reviews on-board processing technology and considers in what ways the development of on-board processing systems can be made more efficient and can meet the demands of shorter time scales (faster).

Standardisation and re-use are examined as two possible means of achieving the necessary efficiency (cheaper). Space-based instruments are usually one-off developments. Seldom can reuse be applied to complete instrument sub-systems. Reuse must be applied at the lower sub-system, component and sub-component levels. Another approach to increasing efficiency is to use a sub-system for more than one purpose. An on-board processing sub-system at the heart of an instrument can be used to support instrument integration and testing. This can simplify and reduce the time taken for integration and test and lower the cost of EGSE (Electronic Ground Support Equipment).

A short development lifecycle increases risk. This risk can be managed and controlled by the use of simulation and the re-use of tried and tested components and designs, both hardware and software.

The paper begins with an overview of a typical on-board processing architecture and introduces the type of processing device that are used in space applications. A range of on-board processing applications is then examined providing the foundation for a more detailed look at on board processing technology. Efficiency gains that can be achieved through component and design reuse are then addressed. The use of the on-board processing system to support integration and testing of instrument and payload systems is also examined. Finally thought is given to the use of simulation techniques to help mitigate the risk inherent in short time-scale programmes.

INTRODUCTION

1.1 Typical On-Board Processing Architecture

A typical on-board processing architecture is illustrated in figure 1. The instrument comprises a sensor, sensor interface, instrument control unit, and a data processor. The instrument is controlled and monitored via telecommand and telemetry from/to the ground. Phenomena detected by the sensor are transformed to digital form for ease of information handling by the sensor interface unit. The raw data may be processed to extract information vital to the control of the instrument or to reduce, in some way, the volume of data being produced by the instrument. The data from the instrument is usually stored in a data storage unit prior to sending it to ground via an appropriate down-link telemetry system.

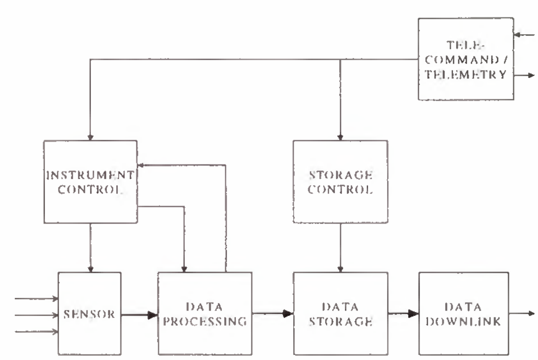


Figure 1 : Typical On-Board Processing Architecture

1.2 On-Board Processing Components

Several different types of digital electronic component are used for on-board processing.

Programmable processors like the ERC32 general purpose processor (GPP) or the TSC21020E digital signal processing (DSP) processor may be programmed to implement many on-board processing functions. Programmable processors are extremely flexible and can handle complex data processing and control tasks.

An Application Specific Integrated Circuit (ASIC) is a type of digital logic device which is designed to implement a specific task or related set of tasks. This specialisation requires significant design effort, so

Proc. 32nd ESLAB Symp., 'Remote Sensing Methodology for Earth Observation and Planetary Exploration', ESTEC, Noordwijk, The Netherlands, 15-18 September 1998 (ESA SP-423, December 1998)

ASICs have a high development cost. ASICs provide high performance for the task for which they have been designed but are usually inflexible.

Field Programmable Gate Arrays are a low cost alternative to ASICs. They are in effect a programmable ASIC device where the logic functions are configured or defined after the device has been manufactured and tested. FPGA devices are of lower performance than comparable ASIC devices.

Processing devices can be classified according to the performance they offer, usually expressed in million instructions per second (MIPS), million operations per second (MOPS), million floating-point operations per second (MFLOPS) or data samples per second. They can also be classified according to the level of flexibility offered by the device. Figure 2 shows the processing devices mentioned above positioned according to their performance and flexibility.

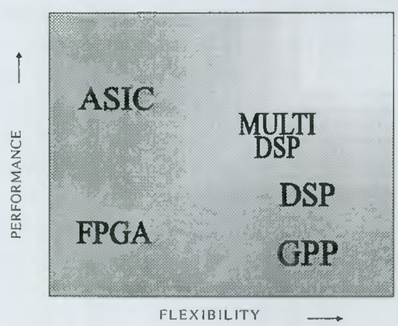


Figure 2: On-Board Processing Devices

1.3 On-Board Processing Applications

Digital processing is used in many areas on-board a satellite or spacecraft. Within an instrument, digital processing may be used for calibration, control and data processing. The data handling system responsible for managing the data gathered by sensors, storing it and transferring it to the down-link telemetry system, relies on digital processing for data compression, data formatting and control of data storage. The spacecraft's guidance and navigation control (GNC) system will use digital processing to implement the necessary control functions and also to extract relevant information from GNC sensors like star-trackers and global positioning system (GPS) sensors. Digital processing may also be used in the satellite communication system for encoding of the data to be transmitted and in some systems for digital formation of the antenna beam.

2 ON-BOARD PROCESSING APPLICATIONS

On-board processing applications may be classified according to the performance requirements and complexity of the processing task. Several applications of on-board processing are shown in figure 3 positioned according to the complexity and performance of the processing required for the application. As can be seen the set of applications illustrated has been chosen to span the performance /complexity domain. Each of these applications will be examined in turn and the type of processing device used for the applications considered.

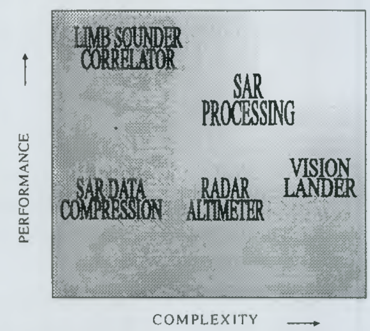


Figure 3: On-Board Processing Applications

2.1 Raw SAR Data Compression

A space-based SAR instrument produces a large amount of data which has to be telemetered directly to ground or stored on-board for later transmission to a ground station. The communication or data storage bottleneck is one factor currently limiting the coverage and/or resolution of SAR instruments. Data compression applied to the data from the SAR instrument would enable lower bandwidth down-link communications and lower on-board storage requirements. Conversely this equates to a capacity for higher data rates from the SAR instrument which would support higher resolution, wider coverage, multi-polarisation or multi-frequency operation. Because adjacent samples in raw SAR data are uncorrelated, compression techniques used for image data do not work on raw SAR data, so the high compression ratios that these techniques provide cannot be achieved for raw SAR data.

Raw SAR data compression is achieved by providing an optimum quantiser which gives an acceptable approximation to the received analogue signal with as few bits as possible. In practice an 8-bit analogue-to-digital converter (ADC) is used and the 8-bit samples are reduced to 2, 3 or 4-bit samples by a digital optimal quantiser. This technique was developed for use on the Magellan radar (Ref. 1) and is known as block adaptive quantisation or BAQ. Data is quantised in blocks because the expected power of SAR data is slowly varying. The expected power of one sample is very close to that of an adjacent sample even though the actual sample values may be completely different. The quantisation thresholds used in BAQ are adapted to the expected power of the block to perform optimum quantisation (Ref. 2,3).

Figure 4 shows a block diagram of a BAQ.

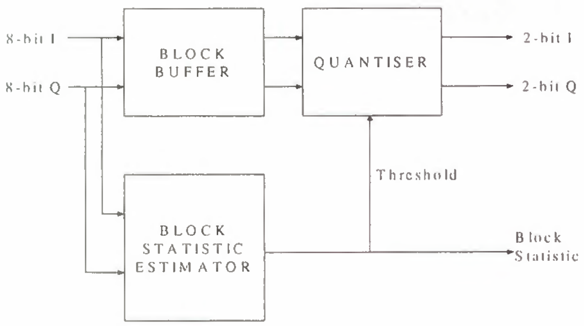


Figure 4: Block Adaptive Quantiser

A block of data is buffered in the block buffer while the block statistic estimator estimates its standard deviation. The block standard deviation defines the optimal quantisation thresholds to use for re-quantisation of the data block to 2, 3 or 4-bits. The standard deviation of the block is estimated by calculating the average magnitude (Σ |I| + |Q|) and using a look-up table to extract the standard deviation or quantiser threshold values. Quantisation is performed by comparing an input sample to a set of thresholds. The output of the comparators are then coded to 2, 3 or 4-bits.

This type of processing is ideally suited to an ASIC or FPGA. It is simple and requires a moderate data rate. A BAQ ASIC has been developed for the SAR on ENVISAT which operates at a sample rate of 20MHz (Ref. 4,5).

2.2 Radar Altimeter

A typical space-based radar altimeter can be used to determine information about the Earth's geoid, the average wave height of the ocean and the surface wind speed over the ocean. In order to do this with sufficient accuracy (a few centimetres over several hundred kilometres) a very narrow radar pulse (wide bandwidth) is required. The radar altimeter transmits a narrow pulse radar signal which travels to Earth. When it hits the ocean surface the radar signal is scattered and some of its energy travels back to the satellite where it is received by the radar altimeter. The radar altimeter measures the time from transmitting the signal to receiving the return pulse. A problem arises because the radar pulse has to travel so far, and is scattered at the ocean surface, so that the received signal is very faint. To make the signal detectable a high energy pulse has to be transmitted. For a narrow pulse signal this means that a very, very high power transmit amplifier has to be used which is not feasible.

The radar altimeter uses a long duration wide bandwidth chirp signal (chirp) of the same bandwidth as the required narrow pulse. The received signal has to be pulse compressed to transform the relatively long duration, wide bandwidth signal into a narrow pulse of the same effective bandwidth. The advantage is that a lower power amplifier can be used to put the required high energy in to the longer radar pulse.

Pulse compression in a radar altimeter is performed by the deramp technique where the radar return signal is multiplied by a copy of the transmitted chirp signal. The power spectrum of the deramped signal then provides the required pulse compressed signal. The advantage of the deramp technique is that after deramping the bandwidth of the received signal is much reduced. The problem is that the expected position of the radar return has to be known reasonably accurately so that deramping can be done at the right time. This requires tracking of the received pulse and appropriate adjustment of the position of the deramp chirp signal.

A block diagram of a radar altimeter is shown in figure 5 (Ref. 6).

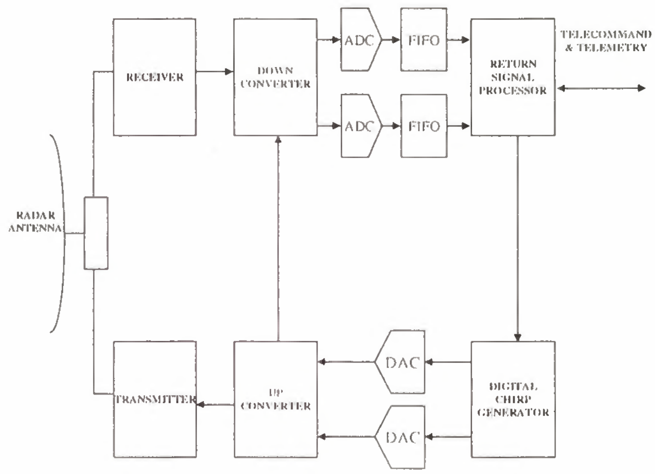


Figure 5: Radar Altimeter Block Diagram

The digital chirp generator generates the required transmit signal which is up converted to the required radar frequency, amplified by the transmitter and sent out of the antenna. The radar return signal is received by the antenna and amplified by the receiver some time later. It is down converted and deramped using a second chirp signal produced by the digital chirp generator. The resulting baseband signal is converted to digital form and passed to the return signal processor. The return signal processor performs the required signal processing on each return pulse, controls the digital chirp generator and other sub-systems, tracks the radar return signal and extracts and formats the required science data.

The radar altimeter processing requires some demanding DSP tasks to be performed together with general control and data handling tasks. A programmable DSP processor is ideally suited to this mix of DSP, control and data-handling.

2.3 Vision Guided Planetary Lander

A particularly complex application of on-board processing is the use of vision for the guidance of an autonomous planetary lander. Consider landing on the moon. Future missions are likely to require a lander to land close to a precise point on the lunar surface, for example next to the edge of a crater in mountainous terrain rather than in the middle of a flat plain (Ref. 7). When a lander is in orbit around the moon its position is likely to be known to about ± 1km. Inertial guidance is inadequate to take the lander to a required target landing spot because of this initial uncertainty in position. Computer vision can provide surface relative navigation able to direct the lander to a required spot. It can also be used to detect possible hazards (boulders, steep slopes, small craters) in the vicinity of the target landing spot (Ref. 8).

Figure 5 shows the various phases of descent of an autonomous vision guided lunar lander. The first step is to determine the position of the lander relative to the lunar surface. This can be done by taking an image of the terrain beneath the lander and matching this image to a larger scale ortho-image of the region around the target landing-site. Once the initial position of the lander has been determined its motion toward the target landing-site can be tracked by tracking the motion of features within images taken by the lander. When the lander is close to the target landing site hazards must be detected and the actual target landing-site determined.

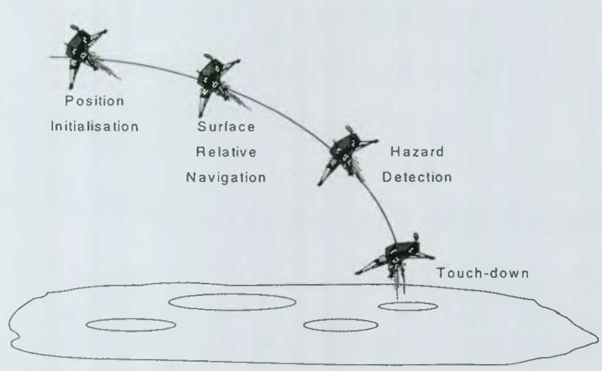


Figure 6: Vision Guided Lunar Lander

The complex image processing required is once again ideally suited to a DSP processor. A single DSP processor may not be able to cope with all the processing that needs to be done so a cluster of two or more DSP processors may be needed to perform the task at an appropriate speed.

2.4 On-Board SAR Processing

A block diagram of a typical SAR processor is given in figure 7. Each range line is range compressed and then placed in a corner turning memory. Data is read out of the corner turning memory in the azimuth direction and azimuth compressed. Detection to produce a real image from the complex data and multi-look processing to reduce speckle noise are performed after azimuth compression. Attitude and orbit knowledge and partially processed data are used to control the SAR image formation process.

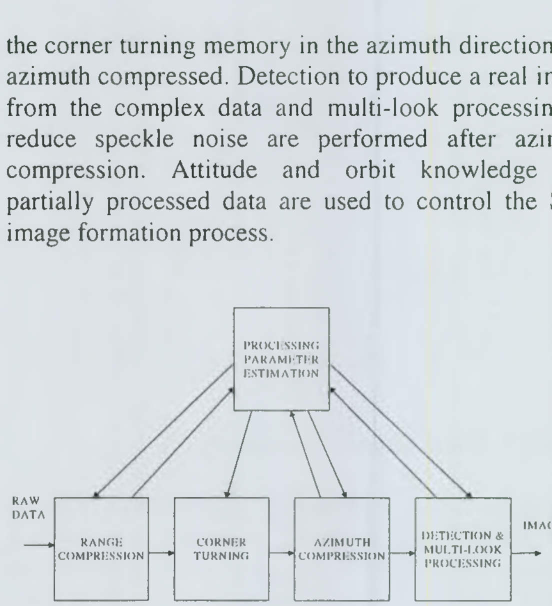


Figure 7: Block Diagram of SAR Processor

The range and azimuth compression operations can be performed efficiently in the frequency domain. For example, a range line may be compressed by transforming it to the frequency domain using an FFT, multiplying it by the complex chirp replica and then applying the inverse FFT to transform the compressed data back to the time domain. Although more efficient than time-domain compression frequency domain compression of SAR data is still very demanding because of the high data rate of the instrument. ASIC devices are ideal for implementing the frequency domain compression (Ref. 9).

The calculation of the processing parameters and the related manipulation of data within the corner-turning memory requires several complex processing tasks. The DSP processor is ideal at this complex processing. A combination of ASICs implementing FFT based pulse compression and DSP processors for parameter calculation and data manipulation is required for effective real-time on-board SAR processing.

The attraction of on-board SAR processing is that once a SAR image has been produced image compression techniques can be applied to reduce the data volume by a factor of 100 or more (Ref. 10). This significantly reduces the data down-link bandwidth required. The difficulty here is in defining what product is to be produced by the on-board SAR processor – since new applications for SAR data are still emerging which require different forms of processing. Raw SAR data, or fully focused, complex image data will be required to support the diverse and growing applications of space-based SAR. On-board SAR processing together with data compression could make a restricted data product available almost immediately to small receiving stations.

SAR (Synthetic Aperture Radar) processing transforms the raw radar data into an image of the terrain being illuminated. The high continuous data rate of a SAR instrument together with the complexity of full-resolution processing makes this difficult to achieve on-board with sensible power consumption. The level of processing power required can be reduced to a more realistic level by either not processing all the data or not processing the image to full resolution. In the first case only images of specified illuminated areas are produced. This could be, for example, one image per orbit. In the second case simplifications can be made to the SAR processing which reduces the processing power requirements at the cost of reduced image quality.

2.5 Microwave Limb Sounder Autocorrelator

Microwave limb sounders measure the spectrum of microwave radiation coming from the upper atmosphere in order to deduce its chemical composition. Wide bandwidths are required of the order of several hundred megahertz. The required signal is the emission spectra of various chemical constituents of the upper atmosphere. This signal is very weak, well below the noise level of the microwave receiver itself. To extract the required signal from the noise many measurements are made of the spectrum and averaged together to improve the signal to noise ratio. On-board processing is essential because of the wide bandwidth of the raw signal.

The on-board processing system has to calculate the spectrum of the received signal and then average many spectra together to reduce the noise level. The spectrum could be calculated using the FFT but this is difficult because of the very wide bandwidth. An alternative is to calculate the power spectral density by performing an auto-correlation, averaging the auto-correlation signals and transforming to the frequency domain. This is equivalent to the FFT and average approach and would normally be more expensive computationally except that the auto-correlation can be performed with only two or fewer bits. The additional noise introduced by quantisation to two or fewer bits is itself removed by the averaging process.

A block diagram of a 2-bit auto-correlator is given in figure 8 (Ref. 11). Each data sample is fed into a delay chain and to a set of 2-bit multiplier-accumulators. The 2-bit multiplier-accumulators form a tap of the auto-correlator and measure the correlation between the incoming signal and a delayed version of that signal where the delay is determined by the position of the tap in the delay chain. The outputs of the multiplier-accumulators are integrated and may be read out after integration using the chain of multiplexers (MUX) and holding registers illustrated in figure 8.

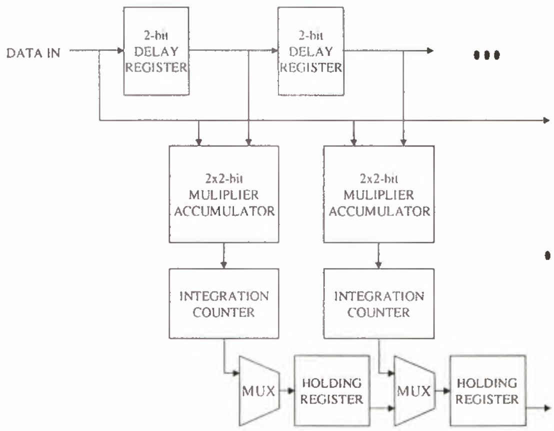


Figure 8: Digital Autocorrelator

The digital auto-correlator is ideally suited to implementation in a ASIC because of the high sample rate required and the simplicity of the processing algorithm.

3 ON-BOARD PROCESSING DEVICES

Having considered a range of applications for on-board processing this section provides a more detailed review of available on-board processing devices. In each case an introductory description of the type of device is given, followed by details of actual current state-of-the-art components suitable for use in space.

3.1 ASIC

Integrated circuits have to go through many manufacturing processes to diffuse transistor materials into the device substrate and lay down interconnecting polysilicon or metal tracks. An integrated circuit designed for a special application is expensive because all these process stages have to be custom designed for the particular device. To make the customisation process cheaper semiconductor manufacturers developed the gate array. This is a device that comprises an array of unconnected simple logic gates, NAND gates for example. Any logic design can be implemented by defining the interconnection between a set of simple gates. Gate arrays are mass produced ameliorating the expense of the custom design with only the final metal layer of the gate array left for customisation. A logic design is converted into the pattern for the metal layer which interconnects the gates to implement the required design. Since only the final few manufacturing process stages have to be customised the gate array is significantly cheaper than a full custom design. However, the gate array design will be slower and consume more power than the custom design.

The gate array is one form of ASIC another is the standard cell. A standard cell device is built from standard building blocks (e.g. RAM block, adder, multiplier, register) for which the transistor level design has been completed. The designer connects many standard cells together to form the required design. The different manufacturing stages are defined automatically from the standard cells, reducing design cost, but all the manufacturing stages are customised for the particular design. The standard cell is somewhere between the gate array and the full custom design in terms of cost and performance.

Commercial gate array devices now offer well over a million gates in a single device and some have in-built functions like RAM blocks designed and manufactured onto the gate array.

An example of a gate array suitable for space use is the TEMIC MG2RT radiation tolerant ASIC. This family of gate arrays comprises up to 700k gates and is manufactured on a 0.5-micron CMOS process. Operating at 5V the toggle frequency is specified as 625MHz allowing maximum system clock speeds of around 200MHz to be achieved for simple designs. These devices can also operate at 3.3V and 2.5V providing lower speed at the reduced voltages. The power consumption is 2uW/gate/MHz when operating at 5V and less than a third of this when operating at 3.3V. The MG2RT has a total dose tolerance of 100kRad (Ref. 12).

3.2 FPGA

An FPGA (Field Programmable Gate Array) is a type of gate array where the interconnection between gates can be defined by programming the complete, packaged device. The interconnections are defined using anti-fuses or transmission gates to connect gates to existing metal or polysilicon interconnection tracks. Anti-fuses are connections that are made between two tracks by applying a high voltage (the opposite of blowing a fuse). FPGAs made using anti-fuses are one-time programmable – once an anti-fuse connection has been made it cannot be broken. A different type of FPGA is made using transmission cells. Static RAM (SRAM) cells are used to control the transmission cell. If the SRAM cell holds a 1 then the transmission cell makes the connection, if it holds 0 then no connection is made. The SRAM is programmed with the required pattern to make the necessary interconnections between gates to implement a particular design. The SRAM can be reprogrammed at any time so that the design can be changed at will without removing the device from the circuit.

Since the FPGA is customised after manufacture they can be mass produced. This reduces the cost but this is offset against the cost of the additional circuitry required to form the interconnection and to program the device. FPGAs tend to be very cheap compared to ASICs for small quantities, but offer reduced performance.

Most FPGA devices provide a higher level of basic cell rather than the simple gate. A typical FPGA cell will comprise a combinatorial logic block (e.g. any function of four inputs) followed by a D-type flip-flop. This higher level basic cell eliminates much of the low level routing enabling higher cell density (less routing, more logic) and higher speeds to be achieved.

Commercial FPGA devices are approaching the equivalent of 500k gates and can operate at system speeds of 100MHz or so.

Radiation tolerant FPGAs are produced by Actel. The RH1020 and RH1280 provide 2k and 8k gates respectively. These devices use anti-fuse technology and are tolerant to 300kRad. Recently Actel announced the availability of the 16k-gate RT54SX16 and the 32k-gate RT54SX32 FPGA devices with a radiation tolerance of up to 100kRad (Ref. 13).

Another FPGA manufacturer, Xilinx, has just announced a range of radiation tolerant FPGA devices: XQR4036XL, XQR4062XL and XQR4013XL which contain the equivalent of 30k, 60k and 130k gates respectively. These devices use SRAM technology, so are reprogrammable, and are tolerant to 60kRad total dose (Ref. 14).

3.3 General-Purpose Processor

The Pentium processor at the heart of a PC or the single-chip microcomputer embedded in a washing machine are general purpose processors. Their inherent flexibility makes this type of processor suitable for a very wide range of general computing and control tasks. The general-purpose processor uses a von-Neuman architecture which has a single bus structure connecting the processor to memory. The memory is used to store both program instructions and data. The processor operates by fetching an instruction from memory, decoding it, getting a data operand from memory, performing some calculation on it and storing the result back in memory. The frequent accessing of memory limits the performance of the processor to the speed of the memory. As a consequence, general-purpose processors have developed a layered memory architecture. Close to the processor core is a bank of registers where data is held while it is being processed. A high-speed cache memory (of limited size) holds the most recently accessed instructions and data. Since a typical program spends a lot of time executing loops, an instruction or data item that has been used recently, and hence stored in cache, is likely to needed again soon. Retrieving this instruction or data item from cache is much more efficient than having to get it from main memory. The next level of memory is the main memory for the processor usually comprising several Mbytes of dynamic RAM. Main memory is large and relatively slow. The processor will need to insert one or more wait states when accessing DRAM.

As computers and microcomputers were developed their designers thought of more and more advanced instructions to design into the processor. Complex instructions were developed which took many machine cycles to perform. These instructions added complexity to the microcode and functional units of the processor core and tended not to be often used by high-level language compilers. This led to the development of the reduced instruction set computer (RISC) where the number of instructions was reduced to a relatively small set, all of which could be implemented to execute efficiently. Complex operations were implemented using several RISC instructions. The advantage of RISC is the small size of the processor core, the speed of execution, the simplicity of the design and the simplicity of the instructions which made compiler writing and code optimisation easier. After the term RISC had been coined, the previous complex instruction set computers became known as CISC.

General purpose processors are generally supported by a range of high level programming languages including C, C++ and Ada.

An example of a radiation tolerant general purpose processor developed for space use is the TEMIC ERC 32 processor. This processor is based on the SPARC architecture (V7) which is a open standard RISC architecture developed by SUN Microsystems and used in their range of SUN workstations.

A block diagram of a processor constructed using the TEMIC ERC32 is given in figure 9. The ERC32 processor comprises three devices the TSC691E integrer unit, the TSC692E floating-point unit and the TSC693E memory controller. These devices are implemented on a 0.8 micron radiation tolerant CMOS process. The performance of the ERC32 is 10 MIPS and 2 MFLOPS when operating at 14MHz. Its power consumption is around 2.5 Watts (Ref. 15).

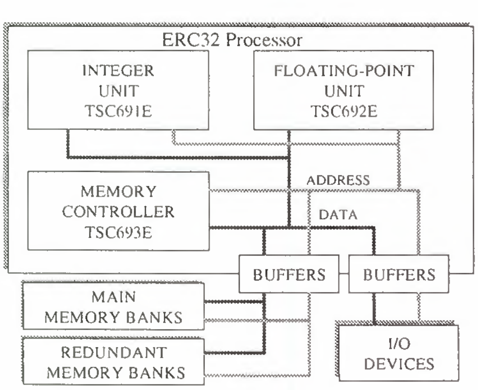


Figure 9: ERC32 Processor

A single chip version of the ERC-32 is currently being developed by TEMIC on 0.5 micron process. This device will offer much improved performance and will operate at both 5V and 3.3V. Operating at 5V and 35 MHz it will provide a performance of 35 MIPS and 5 MFLOPS while consuming 1.5 W (typical). At 3.3V the device is expected to provide a performance of 14 MIPS and 3 MFLOPS and have a power consumption of just 0.4 W (typical).

3.4 Digital Signal Processor

A digital signal processing (DSP) processor is a programmable processor that is particularly efficient at implementing DSP algorithms.

At the heart of most digital signal processing (DSP) algorithms is the multiply-accumulate operation. For example the FIR (Finite Impulse Response) filter which may be expressed as

$$y(n) = \sum_{m=0}^{M} a(m) \cdot x(n-m)$$

A multiplier-accumulator performs a multiplication of two operands, a(m).x(n-m), and accumulates the result with a previous result, Σ. The multiplier-accumulator is a key part of a DSP processor.

An efficient DSP processor will have a multiplier-accumulator that operates in a single processor cycle and will ensure that the multiplier-accumulator is kept busy by providing it with new data to process each cycle. The architecture of a DSP processor is designed with this firmly in mind.

The structure of a DSP processor is illustrated in figure 10. Note that this block diagram and the subsequent description of operation is for the ADSP21020 DSP processor (Ref. 16). Other processors have similar features but differ in detail.

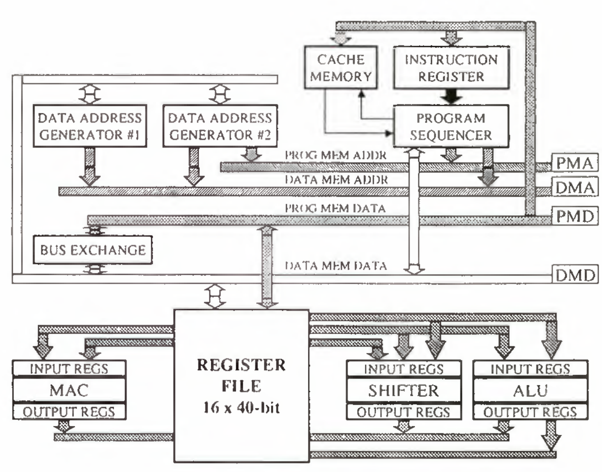


Figure 10 : Block Diagram of TSC21020E DSP Processor

There are three basic elements:-

- Program Control Unit – made up of the instruction register, cache memory and program sequencer.

The program controller is responsible for getting the next instruction to be executed from program memory, decoding that instruction and sending out appropriate control signals to the rest of the processor.

- Computation Unit – made up of the register file, arithmetic-logic unit (ALU), multiplier-accumulator (MAC) and the shifter. The computation unit is responsible for performing the required calculations on the data. The MAC is used for most signal processing functions, the ALU provides general arithmetic capability and the shifter is used for scaling. The ALU, MAC and shifter each operate in a single processor cycle. The register file is used to supply the arithmetic units with data. It de-couples the transfer of data to/from the arithmetic units from the transfer of data to/from memory.
- Data Control Unit – made up of the two data address generators. The data address generators are responsible for calculating the address of the data item required by the computation unit, getting that data from memory and passing it to the register file ready for access by the computation unit. They are also responsible for transferring any computation results from the register file to memory.

The TEMIC TSC21020E is a radiation tolerant version of the Analog Devices ADSP21020 32-bit floating-point processor. It operates at a clock speed of 20 MHz providing a sustained performance of 40 MFLOPS. It is able to implement a 1024-point complex FFT in under 1msec. The power consumption of the TSC21020E is 2.5W maximum (i.e. 62mW per MFLOP) with a typical power of around half this value (Ref. 17)

A block diagram of a complete DSP processing node including program and data memory banks is illustrated in figure 11 (Ref. 18). This design was used for the Matra Marconi Space (MMS) Star-DSP board which flew on the TEAMSAT as part of the Visual Telemetry System (Ref. 19, 20). A photograph of the Star-DSP board is given in figure 12.

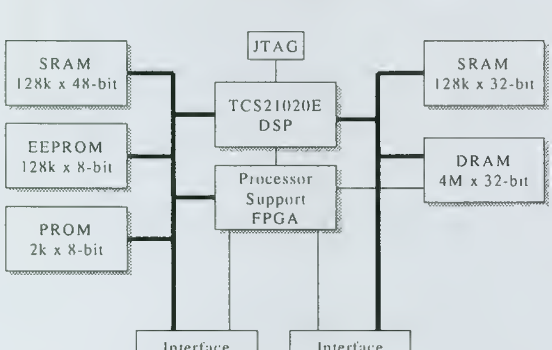


Figure 11: Block Diagram of MMS Star-DSP

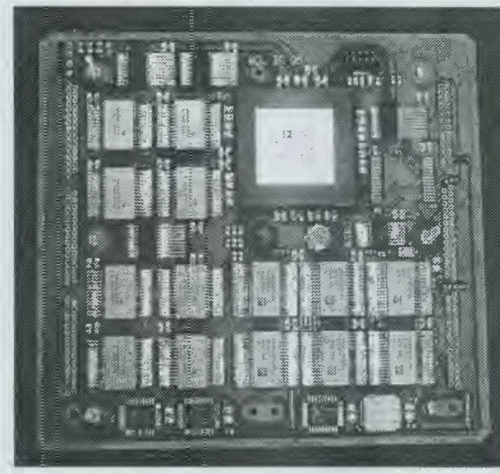


Figure 12: Photograph of MMS Star-DSP Board

3.5 Multi-Chip Modules

Mass and size are important constraints on board a spacecraft. Multi-chip module technology where several integrated circuit die (chips) are packaged together. The die are placed, unpackaged, onto a substrate typically made of ceramic. The die are gold wire bonded to tracks on the substrate which provide the required interconnection between the integrated circuits.

An example of a multi-chip module is illustrated in figure 13. This is the MMS Quasar-DSP multi-chip module which fits the ADSP21020 DSP processor, 11 memory chips and an FPGA for input/output purposes, into a single package of size 60 x 60 x 5 mm, weighing 50g (ref xx). Multi-chip modules have also been developed for the ERC32 device (Ref. 21).

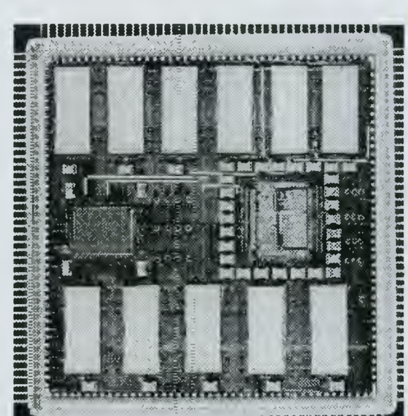


Figure 13: Quasar-DSP Multi-Chip Module

4 FASTER, CHEAPER, BETTER

The key to "faster, cheaper, better" is the re-use of existing proven designs, components and sub-systems.

Using existing technology eliminates much of the risk in developing space-based instrumentation.

This section considers the re-use of technology from the logic design level to the system level.

4.1 Logic Design Re-Use

The development and widespread acceptance of VHDL, VHSIC (Very High Speed Integrated Circuit) Hardware Description Language, has enabled logic designs to be transferred from one component to another. VHDL is a programming language for specifying, designing, simulating and testing digital logic designs. It can define hardware at several different levels: behavioural level, register-transfer-level (RTL) and gate-level. The behavioural-level models how a block of logic responds to inputs without defining how the block of logic will implement that behaviour. Behavioural level modelling is ideal for system specification and for providing test-benches to verify the correct functioning of implemented logic blocks. The RTL-level implements a design as a set of combinatorial logic functions separated by registers. Data is loaded into a register, fed through a combinatorial logic function and the result transferred to another register. The gate-level is the detailed design level which specifies the interconnection of simple gates to make up the primitive logic functions like flip-flops, multiple input NAND gates etc. Synthesis tools are available which are able to generate gate-level designs for ASICs from an RTL level description.

A design for a specific logic function, once developed to RTL level can be used in many different ASIC designs by simply copying the VHDL code from one design to another. VHDL components are now available commercially for many common functions. The investment in design can be recouped over many projects thanks to the availability of VHDL. Faster, cheaper, better is a hallmark of VHDL.

4.2 Hardware Component Re-Use

Programmable processors and programmable logic devices, by their very nature, can be used in many different applications. The problem for space applications is sourcing such devices in a radiation tolerant, high-reliability form. The development of a device like the ADSP21020 processor takes an enormous investment. Not only does the device have to be developed but the supporting software (assembler, compilers, real-time operating system, function libraries) and hardware (development boards, in-circuit emulators) also has to be provided. Significant benefit can be drawn from building on commercial technology, transferring it where necessary to a radiation tolerance process.

The advantages of re-using commercial technology as the basis for a radiation tolerant component are

1. The reduction in risk because a commercially proven design is being used and because test vectors are available for validating the design.
2. Shorter design cycle since effort is put into radiation proofing the design rather that in the logical design.
3. Availability of support both hardware and software.
4. Availability of commercial parts for developing applications – work can start on applications before the radiation tolerant part is available.

The outcome of this is

- a more rapid application development (faster) especially since the application design can start with commercial parts well before space-qualified parts are available,
- a more reliable design (better) since the logic design has been proven on many commercial applications, and
- a lower development cost (cheaper) even though license fees may have to be paid for use of the commercial technology.

The TEMIC TSC21020E DSP processor was developed in this manner from the commercial ADSP21020 device from Analog Devices (Ref. 22, 23).

4.3 Sub-System Level Re-Use

4.3.1 Board Level Re-use

Careful design of a processing board considering generic applications can enable the processing board to be re-used in many applications. The trap of making an “all-singing, all-dancing” design, which is complex and expensive, must, however be avoided. Simple and generic design is the goal.

An example of this is the MMS Star-DSP board. When the Star-DSP board was being developed it was for one specific purpose – as the processing unit of the TEAMSAT Visual Telemetry System (VTS). It had to interface to the OBDH (On Board Data Handling) bus and to one or more cameras. Other applications for the board were considered during the system design phase and a generic design developed. The core processing components were placed on one PCB, the camera interface on a second and the interface to the OBDH interface on a third. If a VTS system was required in future for a satellite with a different form of on-board bus (e.g. 1553) then another spacecraft interface board could be developed without modifying the DSP processor board. If another application was to be supported, for example a radar altimeter, the camera interface board could be replaced by a board containing the required sensor interface (e.g. digital chirp generator, DACs and ADCs). Again there is no need to modify the DSP processor board. This modular approach had the side effect of simplifying the design of the Star-DSP board itself (see figures 11 and 12). Figure 14 shows the VTS processing unit with the camera interface board on the top, the Star-DSP board in the middle and the OBDH interface board at the bottom.

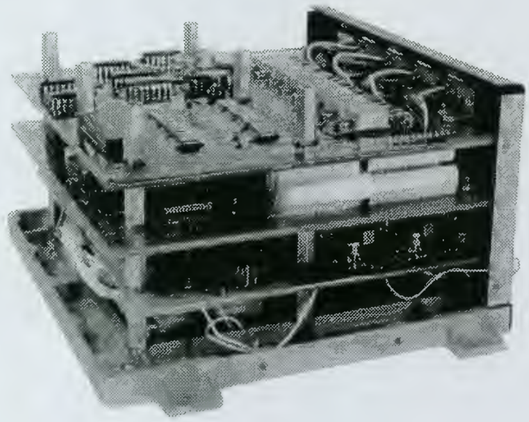


Figure 14: Visual Telemetry System Processing Unit

4.3.2 System Architecture for Re-Use

The ability to re-use sub-systems for different applications can be enhanced if the sub-systems are designed with a common, standard interface. Work on a high-speed serial interface standard for interconnecting processing sub-systems, sensors, solid-state recorders and telemetry units has been sponsored by ESA (Ref. 24, 25, 26, 27, 28). The data link being developed will operate at over 100Mbits/sec, at distances of up to 10m and each link will consume less that 0.5 W. The data link is based on the IEEE 1355 standard with the physical layer replaced by low voltage differential signalling (LVDS). A draft standard for this data link will be produced early in 1999. Sub-systems using this "space-wire" data link can be interconnected readily.

One processing board that has been designed using the "space-wire" interfaces is the Dornier Satellitensysteme (DSS) Mosaic020 board, a photograph of which is given in figure 15 (Ref. 29). The Mosaic020 board is a DSP board based on the TSC21020 DSP processor. The two large components to the right of this board are DSS SMCS332 devices, which each implement three "space-wire" interfaces. The Mosaic020 board can be easily connected to sensors, solid-state recorder units or other processing boards to form a complete on-board data processing system.

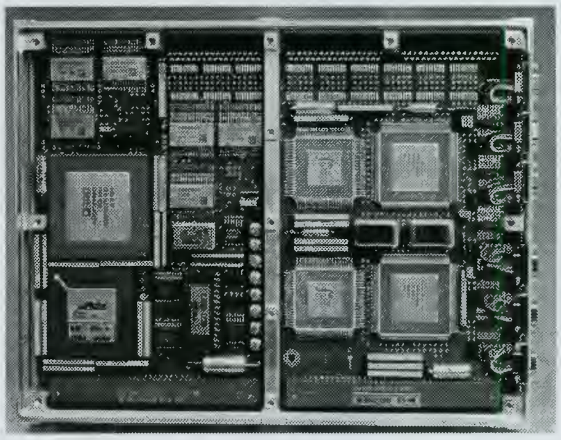


Figure 15: DSS Mosaic020 DSP Board

4.4 Software Re-Use

Software should be inherently re-usable, but it rarely is!

Real-time operating systems (e.g. Eonic Virtuoso Ref. 30) provide a multi-tasking platform for building software systems. The real-time operating system provides the necessary hooks for developing device drivers for new hardware peripherals and enables these device drivers to be used across several applications. Function libraries provide sets of optimised functions which can be utilised within an application program. Function libraries are particularly important for DSP processors where they provide all the common DSP functions hand-coded in assembly language to achieve the best possible performance by making full use of the available processor facilities.

The closer one gets to the application level the more difficult re-use of existing software becomes. This is largely because existing software is poorly understood or does not provide exactly the required functionality.

Object oriented design techniques can help by providing encapsulation of data and functionality. This makes software easier to re-use because each object is self-contained and hence easier to understand. Object oriented analysis and design has recently received a major boost by the development of the Unified Modelling Language (UML) which provides a common graphical language for object oriented modelling. UML embodies use case analysis, class diagrams and object interaction/collaboration diagrams enabling systems to be analysed, modelled and understood by other software engineers (Ref. 31).

5 RISK REDUCTION WITH SIMULATION

Simulation can test designs, components, processing boards and sub-systems before they are manufactured. The structural and behavioural simulation capabilities offered by VHDL support simulation at all these levels (Ref. 32). Simulation to reduce risk (testing before manufacture) is an essential practice to achieve efficient and timely development of on-board processing systems.

Simulation can also be used to validate complex processing algorithms for applications like the vision guided lunar lander. Sensors and the environment that they are sensing can be simulated to provide data to evaluate and test different processing algorithms. This can greatly reduce the risk involved in some of the more exotic missions.

A lunar surface simulation system is being developed at the University of Dundee to support the development of computer vision techniques for autonomous, semi-autonomous and operator-guided planetary landing systems. The required output from a computer simulation of the lunar surface is an image of the surface taken from a camera on a lander at a particular position and orientation relative to the surface. Ray tracing is used to produce the image from the simulated surface and to model Sun and Earth illumination. The lunar surface model is built by simulating impact cratering on an initial terrain model. An image produced by this simulation system is given in figure 16.

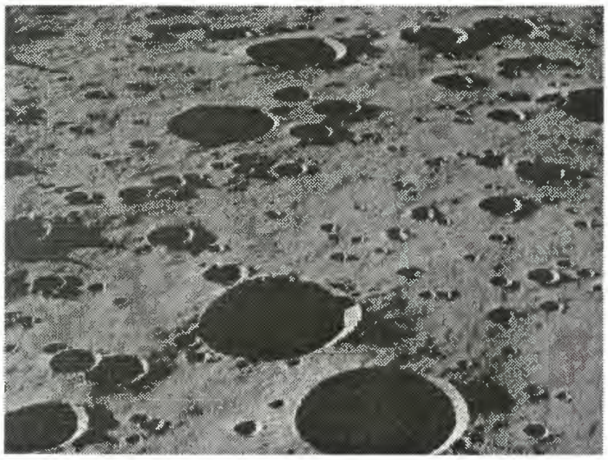


Figure 16: Simulation of Lunar Surface

This type of sensor simulation, running on a PC or workstation can simulate a sensor gathering information about a planetary surface. A "space-wire" interface fitted to the PC enables the sensor to be replaced in the engineering model of the flight hardware by the PC running the simulation. If the sensor is connected to the processing unit by a "space-wire" data link then it can be simply unplugged and replaced by the sensor simulation system. The on-board processing unit can then be extensively tested to confirm its robust operation. This can dramatically reduce the risk of advanced on-board processing developments.

6 INTEGRATION SUPPORT

In-circuit emulation has been widely used to support the development of processing hardware and for hardware/software integration. An in-circuit emulator is a pod containing a special "bonded-out" version of a processor which replaces the processor in a circuit under development. The emulator is controlled from a PC and enables code to be loaded into memory, programs to be stepped through one instruction at a time, values held in processor registers and memory to be viewed, break-points to be set to halt processor execution, etc.

Recent processors, including the TSC21020E, have included much of the emulation circuitry on the processor chip itself to avoid having to replace the processor by a special pod. The on-chip emulation circuitry is accessed through a serial port (usually a JTAG boundary scan test port) and controlled from a host PC. The on-chip emulation facility is extremely useful for hardware and software development. It can also be used to test other parts of the system under development e.g. sensor interfaces. Access to the on-chip emulation facility via a JTAG port or another type of data link driving the JTAG port can provide a valuable form of support for system integration and test.

7 CONCLUSIONS

This paper has examined a range of on-board processing applications and reviewed the technology that is used to implement the range of applications. Processing technology is essential for most remote sensing and planetary exploration missions. Techniques that can help to make the development of future processing systems "faster, better, cheaper" have been considered. Re-use of existing technology at the logic design, component, board and sub-system levels is important. The use of simulation to test components and boards before they are manufactured, and processing sub-systems before they are launched can help with the management of risk in projects with short development schedules. On-board processors which have some form of on-chip emulation capability can be used to support system integration and test, again alleviating risk and reducing costs.

8 ACKNOWLEDGEMENTS

I would like to thank Philippe Armbruster, Patrick Plancke and Josep Rosello, ESTEC study managers for many of the ESA studies that I have been involved in, for their vision and support. I would also like to thank the DSP teams at Matra Marconi Space UK and Dornier Satellitensysteme for the stimulating and enjoyable time spent working with them.

The photographs of the Star-DSP and Quasar-DSP are from Matra Marconi Space UK and that of the Mosaic020 from Dornier Satellitensysteme.

9 REFERENCES

1. Kwok R. & Johnson WTK, "Block Adaptive Quantization of Magellan SAR Data", IEEE Transactions on Geoscience and Remote Sensing, Vol. 27, No. 4, pp 375-383, July 1989.
2. Parkes SM and Clifton HL, " ShrinkSAR Raw SAR Data Compression Study Final Report", Matra Marconi Space Report TR0115, DRA Contract No. CSM2/002, April 1995.
3. Parkes SM and Clifton HL, "The Compression of Raw SAR and SAR Image Data", International Journal of Remote Sensing, to be published.
4. Dutkiewicz, M. and Kuduvalli, G. "SAR Pre-Processing On-Board Study Final Report", MacDonald Dettwiler, ESA Contract No: 10135/92/NL/FM(SC).
5. Saab Ericsson Space, "On-Board SAR Signal Conditioner Development (OBSP), Work Order 4 to ESTEC Contract No. 10737/94/NL/JG, Final Report (Phase 2), OBSP/TNT/011/SE, Saab Ericsson Space, October 1997.
6. Durrant RJ & Parkes SM, "The Implementation of High Time-Bandwidth Chirp Pulses using Digital Techniques", IEE Colloquium on High Time-Bandwidth Product Waveforms in Radar and Sonar, 1st May 1991.
7. Ockels WJ, "A Plan for a European Lunar South Pole Expedition, EuroMoon 2000" ESA Publication No. BR-122, ISBN 92-9092-428-4, ESA, 1996.
8. Paar G, Parkes SM, Classen HJ, "3D Planetary Modelling Study Final Report" ESTEC contract 9195/90/NL/SF, Joanneum Reseach 1996.
9. Parkes SM and Richards B, "Synthetic Aperture Radar (SAR) Processing Requirements, WP1200 Report for ESA Contract No. 8741/90/NL/JG(SC), British Aerospace Space Systems Ltd., May 1990.
10. Bierens L, Van Heijster R & van Bezouwen H, "Data Reduction Using Real-Time On-Board Satellite SAR Processing", Proceedings Fourth International Workshop on Digital Signal Processing Techniques Applied to Space Communications, ESA WPP-086, pp2.21-2.30, September 1994.
11. Parkes SM and Gillions AR, "Tenor-ACS Digital Autocorrelator", MMS Applied Resarch Report No. TR0130, Matra Marconi Space, November 1995.
12. TEMIC, "MG2RT Radiation Tolerant 0.5-um CMOS Sea-of-Gates", Data Sheet, Rev.B, TEMIC Semiconductors, Feb 1998.
13. Actel web site, http://www.actel.com/
14. Xilinx, "QPRO XQR4000XL QML Radiation Hardened Field Programmable Gate Arrays", Data sheet, Xilinx September 1998, (web site, http://www.xilinx.com/).
15. TEMIC, "SPARC RT Processor", http://www.temic-semi.de/nt/ams/sparc/sparc.html
16. Analog Devices, "ADSP21020 User's Manual", Analog Devices Inc., 1995.
17. Temic, "TSC21020E Radiation Tolerant 32/40-Bit IEEE Floating-Point DSP Microprocessor", Data Sheet, Rev. D, May 1998.
18. Parkes SM, "Star-DSP: Specification and Design Trade-Offs", Applied Research Report, Matra Marconi Space, November 1995.
19. Durrant RJ, "VTS Demonstrator, Image Compression and ADSP21020 Implementation", ESA Round Table on Data Compression Techniques for Space Applications, ESTEC, October 1996.
20. Teamsat web site, http://www.estec.esa.nl/teamsat/page_menu_team.html
21. Parkes SM et al, "Programmable DSP Developments at MMS", ESA Open Interface Data Link Round Table Session 3 Handouts – Applications and Architectures, ESA-ESTEC, 22 June 1995.
22. Beghin F and Stachetti V, " DSP Development", ESA Open Interface Data Link Round Table Session 3 Handouts – Applications and Architectures, ESA-ESTEC, 22 June 1995.
23. Pike T et al, "The Development Programme for European-Sourced 32-bit DSP for Space Applications", ESA Open Interface Data Link Round Table Session 3 Handouts – Applications and Architectures, ESA-ESTEC, 22 June 1995.
24. IEEE, "IEEE Standard for Heterogeneous InterConnect (HIC) (Low-Cost, Low-Latency Scalable Serial Interconnect for Parallel System Construction)", IEEE Std 1355-1995, IEEE, June 1996.
25. Christen AZ, "SMCS User's Guide," DIPSAP-II-DAS-31-06, Draft D DIPSAP-II Consortium, February 1998. http://www.omimo.be/new/companies/dasa_001.htm
26. Parkes SM and Price SD, "DSP-Link Study Final Report", ESTEC contract no. 8741/90/NL/JG(SC), Matra Marconi Space TR 0104, October 1994.
27. Parkes SM et al, "Review of Standard and Status", Digital Interface Circuit Evaluation Study WP1000 Technical Report, Document No. UoD-DICE-TN-1000, ESA Contract No. 12693/97/NL/FM, University of Dundee, July 1998.
28. Parkes SM, "High-Speed, Low-Power, Excellent EMC: LVDS for On-Board Data Handling", Proceedings of the 6th International Workshop on Digital Signal Processing Techniques for Space Applications, ESTEC, Sept. 1998.
29. DSS, "Mosaic020 Digital Signal Processor Board", http://www.omimo.be/new/companies/dasa_004.htm
30. Eonic, "Virtuoso the SoftStealth RTOS for DSP and ASIC Cores", Technical Brief, TB201-0398, Eonic Systems, 1998.
31. Fowler M, "UML Distilled", Addison-Wesley 1997, ISBN 0-201-32563-2.
32. Habinc S and Sinander P, "Using VHDL for Board Level Simulation", IEEE Design and Test of Computers, pp 66-78, 1996.

MARS RIOMETER SYSTEM

E. Nielsen

Max-Planck-Institut für Aeronomie (MPAE),
D-37191 Katlenburg-Lindau, Germany,
Telf.: +49 5556 979 450; Fax: +49 5556 979 240; E-mail: nielsen@linmpi.mpg.de

ABSTRACT

'Mars Riometer System' is an experiment designed to operate on the surface of Mars. It will measure the intensity, variations with time, and occurrence of precipitation of energetic particles and x-rays into the Martian ionosphere, i.e. provide information on the space weather on Mars. This is done by recording the variations in the cosmic noise intensity at the surface of the planet. The riometer technique is closely connected with the response of the planetary atmosphere to an ionizing agent (particles or x-rays). Since the atmosphere of Mars is mainly composed of CO_2, and so very different from that of the Earth (where N_2 is the major molecule of interest to riometry), a close analysis of the Martian atmosphere for riometry has been made. The analysis shows a Martian riometer to be more sensitive than the equivalent Earth riometer. and, thus, Mars has actually a better atmosphere for riometry than the Earth. The observations will show if predicted micro-lightening occurs during Martian dust storms, and be used to determine the associated increase in the electro-magnetic noise level. The experiment will provide real time information about radio wave absorption in the ionosphere, and possibly about energetic particle bombardment of the surface, both aspects of interest in connection with manned landings on Mars. The experiment will probably be characterized by low power (<100 mW), low mass (<100 g), low volume (<100 cm^3), and low data rate.

Key words: Space weather, ionosphere, space plasma physics, energetic particles, x-rays, solar flare

1. INTRODUCTION

The name 'riometer' stands for relative-ionospheric-opacity-meter. The experiment is a passive (high frequency) radio wave receiver which measures the intensity of the cosmic noise at ground level, i.e. after the radio wave signal has passed through the ionosphere. The intensity of the cosmic noise imprinning on the ionosphere, from any given direction, is very stable in time. The observed cosmic noise intensity is therefore a measure of the absorption of high frequency radio waves in the ionosphere. This is basically a measure of the height integrated ionospheric electron density over the riometer site. A decrease in the cosmic noise power, not related to the variations in the cosmic noise power caused by the rotation of the planet, is therefore a measure of increased radio wave absorption in the ionosphere. This increase in absorption is directly related to an increase in the height integrated electron density over the riometer site.

The ionospheric electron density is controlled and influenced by several agents. It is fundamentally controlled by the kind of atmosphere the planet has, and by its response to ionizing radiation. The diurnal variations of the electron density is controlled by the solar ultra violet radiation. Also solar X-rays emitted during solar flares ionize the planetary atmosphere. The ionospheric density increases also owing to precipitation of energetic charged particles, which loose energy through collisions with and ionization of atmospheric molecules. Thus, a riometer measures the occurrence, intensity and time variations of ionospheric absorption caused by solar ultra violet radiation, X-ray radiation, and particle precipitation. The riometer technique has also been used to detect horizontal drifts of large irregularities of enhanced electron density across the riometer site.

Precipitating energetic charged particles can originate in solar flares or have their origin in acceleration processes near the planet. The different acceleration processes are producing different particle energy spectra and have different time histories, and they will therefore be associated with different kinds of absorption events, each with its distinct signature in occurrence, intensity and time variation. The riometer experiment therefore allows an identification of the kind of event causing the enhancement in ionospheric absorption.

Since the riometer technique is closely connected with the response of the atmosphere to an ionizing agent

Proc. 32nd ESLAB Symp., 'Remote Sensing Methodology for Earth Observation and Planetary Exploration', ESTEC, Noordwijk, The Netherlands, 15-18 September 1998 (ESA SP-423, December 1998)

(ultra-violet -, x-ray radiation, or particles precipitation), it is clear that the particular atmosphere of the planet on which a riometer is operating must be considered. So far riometry has only been carried out at the Earth's, where the nitrogen molecule is of dominating interest for riometry. Since the atmosphere of Mars is dominated by CO_2, and so very different from that of the Earth, a close analysis of the Martian atmosphere for riometry is required. In our analysis (see below), which we want to test against experiment, we find a Martian riometer to be more sensitive than the equivalent Earth riometer, and, thus, Mars has actually a better atmosphere for riometry than the Earth.

If there are increases in the observed radio wave signal intensity occurring on the surface of Mars the riometer will of course also record these increases. Since the cosmic noise signal imprinning on the ionosphere is very constant in time, such an increase in intensity most likely has a source in the atmosphere. It has been speculated that micro-lightening may occur during dust storms. The friction between dust grains causes electric charges, which discharge in lightening. Lightening is known to produce radio waves, and therefore lightening may be a source of signal which the riometer will record. A riometer measurements with a good time resolution could provide information on the intensity of lightening electric fields, and on the increase in noise level.

The ionospheric radio wave absorption measured in deciBell is a parameter that is readily accessible from the riometer observations. The experiment operates continuously, and is to a large extend independent on the environment (at the Earth it is independent of sunshine, moonshine, rain, clouds, fog, wind etc). The simplicity and robustness of the experiment taken together with the easy interpretation of the data in a physically meaningful parameter, makes the riometer a valued experiment at all ionospheric stations for a first reliable determination of the state of the ionosphere.

The measurement of the ionospheric absorption can of course be used directly to evaluate the attenuation of a communication signal. This would be useful in connection with future manned landings on Mars. At the Earth particle precipitation can give rise to 'black out', i.e. total loss of communication. A riometer experiment would give experience with the possible absorption levels that can occur in the Martian ionosphere, for use in design of future communication equipment. In a manned mission on the surface of Mars a riometer would give the astronaut ready access to a reliable and robust real time indicator of the level of the ionospheric absorption.

With some experience the riometer data may be used to estimate the spectrum of the precipitating particles. Since the Martian atmosphere is much thinner than the Earth's some of the more energetic particles is expected to penetrate the atmosphere and to hit the surface of the planet. The associated radiation hazard for astronauts could be estimated from the riometer measurements, and used in real time warning.

The scientific goals of a riometer experiment are:

1) Study and experimental application of riometry in an atmosphere totally different from that of the Earth's;
2) Geophysical studies of precipitation of energetic charged particles and solar X-rays; and the atmospheric response to ionizing agents;
3) Detections of and studies of micro-lightening;
4) The radio wave absorption measurements can be used in evaluating communication hazards;
5) Evaluated particle spectra may be used to estimate radiation hazard from energetic particles at the surface of planet;
6) Evaluation of the micro-lightening electric fields as a hazard to electronics equipment, and of increase of electro magnetic noise level owing to lightening.

Little is known about Mars's ionosphere, and less is known about local processes that accelerate charged particles. In the following our knowledge is outlined with a view to areas where a riometer can yield new insight.

2. SCIENCE AND PRACTICAL APPLICATIONS

A fundamental aspect of planetary research is the study of the interaction of a planet with the solar wind. At the Earth this interaction is mainly between the geomagnetic field and the solar wind. The geomagnetic field deflects the solar wind around the planet, and so it shields most of the atmosphere/ ionosphere from direct contact with the solar wind. At Mars the solar wind interacts directly with the Martian atmosphere/ionosphere, so the interaction is very different from the Earth's and therefore of also of great interest. What are the processes and forces at work to deflect the solar wind around a weakly magnetized planet? Theories predict that some of these processes results in acceleration of charged particles, and when these precipitate into the atmosphere they cause electron density enhancements in the ionosphere, which are detectable with a riometer experiment.

2.1. Ionosphere

Mars's top-side ionosphere is known from radio occultation measurements (Kliore et al., 1972), from the landings of Viking 1 and 2 (Hanson et al., 1977). Because Mars is an external planet the occultation technique only allows measurements for solar zenith angles larger than about 45 degrees. The variations in the vertical profile of electron density and electron and ion temperatures have been taken to indicate that Mars has a (possibly very weak) intrinsic magnetic field (Krymskii et al., 1995). There are no observations - direct or indirect - of the bottom side ionosphere below the ionization maximum, and there are no observations at all of the midday ionosphere (zenith angle less than 45 degrees). It has been suggested that merging of the interplanetary magnetic field and a Martian magnetic field may occur on the dayside. That would give rise to acceleration of charged particles, some of which are expected to precipitate into the atmosphere. As the energetic particles penetrate into the denser atmosphere they and cause ionization owing to collisions with atmospheric molecules. This ionization in the bottom side ionosphere is detectable with a riometer. Even if the interaction of Mars with the solar wind is cometary-like acceleration of particles can occur in the interplanetary magnetic field as it may become folded upon itself when passing the planet (Verigin et al., 1987). The nightside Martian ionosphere is weak and highly variable (Lindal et al., 1979). It has been suggested that the nightside ionosphere at Mars is a result of energetic particles impacting on the atmosphere (Zhang et al., 1990), either solar wind electrons and/or photoelectron from the dayside. Thus, it appears that also the nightside ionosphere will be a fertile region for riometer observations.

2.2. Energetic Charged Particles

The energetic particle populations at the Earth and at Mars have very different effects on the respective ionospheres. The absence of, or very weak, magnetic field at Mars, will not channel the charged particles into the high magnetic latitude region of Mars, as is the case at Earth, but there will be more direct access to the ionosphere from the source or acceleration regions. The atmosphere of Mars will be penetrated by $\sim$ 85 MeV protons and by $\sim$ 12.5 MeV electrons. Charged particles of lower energy will loose all energy through collisions in the atmosphere and will not reach the surface.

Solar cosmic rays, i.e. particles energized and injected into interplanetary space during solar flare events, are detected at Mars. But also populations of particles which appear to be energized by local processes are observed. Furthermore, galactic cosmic rays may have detectable ionospheric effects on Mars.

At the Earth solar flare particles are magnetically channeled into the polar caps and high latitude regions. At Mars we expect the planet to be fully exposed to the oncoming solar particles. The particles travelling roughly along the Archimedean spiral of the interplanetary magnetic field impact on the atmosphere to produce ionospheric ionization (and riometer absorption) over one-half of the planet. The particle signature of this impacting is a 'shadow' in the energetic particle fluxes in the region above the planet opposite to that threaded by the interplanetary field. Co-rotating interaction regions, or interplanetary shocks, also produce energetic particle fluxes that may be detectable by their ionospheric ionization effects.

Close to the planet there exist three distinct populations of energetic particles, which appear to be locally accelerated. Actual particle measurements near Mars were obtained during the Phobos mission in January-March 1989. One population travel down the tail with energies > 55 KeV. It has been suggested that these particles are the result of magnetic field line reconnection in the dayside and/or nightside Martian regions. Merging of interplanetary magnetic fields with intrinsic magnetic fields on the dayside, or merging of fields around the neutral sheet on the nightside may be invoked. Another population occur just inside the planetopause with energies < 225 keV. Such particle fluxes were observed uninterrupted over an 8 day period. This enduring signature has lead to the suggestion that the particles were quasi-trapped. In view of recent indications in Mars Surveyor data of the presence of a magnetic field (albeit contained in magnetic anomalies) makes the suggestion of quasi-trapped particles especially interesting. A third population, with energies up to 600 keV is found close to the bow shock at the terminator. It is thought that acceleration in the bow shock may account for this population.

It is also to be determined if galactic cosmic rays could possibly have ionospheric effects. If they have then solar wind modulations of the fluxes, as occurs in a Forbush decrease, may be observable with a riometer.

2.3. Micro-Lightening

It is also proposed to study atmospheric electricity through observations of electrostatic discharges predicted to occur in a dust storm. The friction between dust grains blowing about in a dust storm may lead to electrical charging of the grains, and discharges can take place between grains or clouds of charged grains. It has been estimated that electrical charges of 20 to 25 kV/m can occur in a dust storm. The lightening will radiate electro magnetic energy in a wide bandwidth. It is to be expected that radiated energy at the riometer frequency will be detected by the riometer experiment provided that the sampling rate is sufficiently high or that the rate of lightening is sufficiently high. Thus, a riometer could be used as a detector of the intensity and occurrence of lightening.

2.4. Communication, Radiation, and Lightening Hazards

With a view to future landings of astronauts on Mars, it is essential to consider and test instrumentation for use by the astronauts to asses the environment. Also the environment as far as radio wave absorption and energetic particle radiation is concerned. The riometer is uniquely suited to test these two aspects of the Martian environment. At ionospheric stations at the Earth there is also a need to quickly asses the state of the ionosphere. Here the riometer is the essential instrument. It is operated continuously, is independent of weather and light level, and has an output that can easily be converted to a meaningful physical parameter: the ionospheric absorption in decibel. Data from other instruments are either not readily available or need extensive analysis before a physical parameter can be extracted (cameras: need darkness, or need to develop film; scintillation: have to wait for satellite to pass; magnetometer: easily accessible, but what does it mean? the equivalent current may have no relationship to the actual current overhead; pulsation magnetometer: need fourie analysis; VLF propagation experiment: sensitive to far away regions; etc.etc.). Other ionospheric experiments which can yield more information than a riometer, as for example an incoherent scatter radar, are much too complicated to even be considered. The simplicity and reliability of a riometer makes it a strong candidate for monitoring of the state of the Martian ionosphere.

Detrick et al. (1997) has pointed out that for the same spectrum of precipitating particles the radio wave absorption in the Martian ionosphere is 2 to 4 orders of magnitude larger than at the Earth (see below). At the Earth absorption values of up to 10 dB at 30 MHz are not unusual. The comparable absorption values at Mars will depend on the Martian particle fluxes compared with those at the Earth. Since fluxes during solar cosmic ray events will be comparable at 1 AU and 1.5 AU, we expect that Martian absorption values from 20 to 30 dB at 30 MHz may occur, or 40 to 60 dB for two-way communications. Thus, it is clear that attenuation of radio waves at Mars can be considerable, and it cannot be ignored in cases where communications are essential. We propose to measure the level of attenuation, its variation with time, and its occurrence, and to demonstrate that a riometer yields real time information on the attenuation level.

At the Earth the thick atmosphere (and the geomagnetic field) shields the planet surface from bombardment of cosmic rays. Owing to the low atmospheric density at Mars, the Martian surface is much more exposed to harmful radiation from space. The atmosphere is dense enough to absorb all except the more energetic particles; but during the early phase of solar flare events there may be dangerously large fluxes of particles with sufficient energy to penetrate to the surface. The riometer detects these fluxes when they occur, i.e. when the particles arrive at Mars. However, when a solar flare occurs the sun emits X-rays, which propagate with light velocity, and gives rise to SCNA (= Sudden Cosmic Noise Absorption) on the whole sunlit hemisphere. This kind of absorption event has a distinct signature, which can be taken an early warning of a coming radiation hazard.

The internal friction in a dust storm may produce large voltages which are released in (micro-) lightening that produce electro magnetic waves in a wide spectrum. We propose to detect the electric fields associated with the lightening if it occurs. Electric fields emitted by lightening may, if they are large enough, be a threat to electronics equipment. Even though lightening may be observed from orbit, it is therefore essential to make in situ observations from the surface inside a dust storm.

3. MARTIAN RIOMETRY

The Martian atmosphere is mainly composed of CO_2, in contrast to the Earth atmosphere which is mainly N_2 at the altitudes of interest for riometry. This is the fundamental difference between the two atmospheres. The expected behavior of a CO_2 atmosphere for riometry is evaluated below (see Detrick et al., 1997).

To determine the riometer absorption caused by a given precipitating electron flux, it is necessary to calculate the production rate of electrons that are produced by collisional ionization of the precipitating particles and the air molecules. One must then calculate the equilibrium electron density that results from the electron-ion production and the recombination that will take place. It is the equilibrium density that determines the radio wave absorption.

3.1. Production Rate

The energy loss of the energetic electrons is determined by the atmospheric mass traversed (g/cm^2). The atmospheric depth (g/cm^2) at Mars can be derived dividing the Viking atmospheric pressure data reported by Seiff and Kirk (1977) by the gravitational acceleration, taken to be 392 cm/s^2. The production rate of ions for an isotropic electron flux can then be written

$$q(T_o, h) = \frac{T_o}{r_o} J_{NI}(x) \frac{\rho(h)}{\rho(R)} \cdot \frac{1}{33} \qquad (1)$$

where
T_o = electron kinetic energy in keV;
h = altitude;
R = range in g/cm^2 of an electron with initial energy T_o;
$\rho(R)$ = atmospheric density at the altitude electron range;
r_o = R*1000/ρ(R);

x = residual range;
J_{NI} = isotropic normalized energy dissipation function;

The energy required to produce an electron-ion pair in a carbon dioxide atmosphere is 33 eV (Klots, 1968).

For a particle flux $F(T_o)$[electrons/cm^2-str-sec-keV] the production rate at altitude h is

$$q(h) = \int q(T_o, h) 2\,\pi F(T_o) dT_o \qquad (2)$$

3.2. Effective Recombination Coefficient

The effective recombination in the Earth's atmosphere is controlled by three-body recombination (involving neutral, ion, electron) and by dissociative recombination. At Mars the recombination is controlled by dissociative recombination of O_2^+ and NO^+, the main species of ions below 120 km (Schunk and Nagy (1980)), and by three-body interactions. Considering only dissociative recombination Detrick et al (1997) finds

$$\alpha_e(h) = 1.9 * 10^{-7} (\frac{300}{T})^{0.5} (\frac{N(O_2)}{N}) \qquad (3)$$

$$+3.8 * 10^{-7} (\frac{300}{T}) (\frac{N(CO_2)}{N}) \qquad (4)$$

where
T = temperature of neutral atmosphere;
$N(O_2)$ = density of O_2; (from Schunk and Nagy, 1980);
$N(CO_2)$ = density of CO_2; (from Schunk and Nagy, 1980);
N = $N(O_2) + N(CO_2)$.
Three-body interactions have not been considered in this equation.

3.3. Equilibrium Electron Density

The steady state electron density is determined by the equality of production rate and recombination at altitude h,

$$n_e(h) = \sqrt{(\frac{q(h)}{\alpha_e(h)})} \qquad (5)$$

3.4. Radio Wave Absorption

Sen and Wyller (1960) calculated the complex refractive index of weakly ionized gas, and determined the radio wave absorption, which in the case of no magnetic field, is

$$A(dB/km) = 1.5 * 10^5 \frac{n_e(h)}{v_m(h)} C_{5/2}(\frac{\omega}{v_m}) \qquad (6)$$

and since to a good approximation

$$C_{5/2}(\frac{\omega}{v_m}) = 1/2.5 * 1/(1 + (\omega/v_m)^2) \qquad (7)$$

we find

$$A(dB)/km) = 4.61 * 10^4 \frac{n_e(h) v(h)}{v^2(h) + \omega^2} \qquad (8)$$

where

v = 1.5 v_m, the momentum-transfer collision frequency; shown in Figure 1.

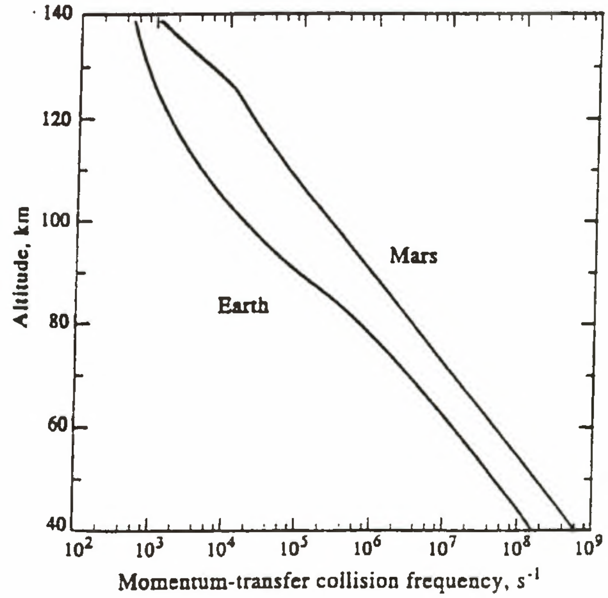


Figure 1. Electron-neutral collision frequency for the terrestrial and martian atmospheres.

To illustrate the difference between the Earth and Mars atmosphere with respect to riometry, the altitude profile of absorption for the two planets is calculated for the same weak flux of precipitating electrons, in Figure 2. The height integrated absorption is at the Earth 0.04 dB, and at Mars 2.6 dB, showing the larger sensitivity of the Martian atmosphere.

4. TECHNICAL

A riometer is a passive radio wave experiment, and consists of an antenna, a receiver, and a data recording device that records the receiver output voltage as a function of time.

The riometer technique is fully dependent on the electron density enhancement resulting in the ionosphere from a precipitating particle flux. The atmosphere

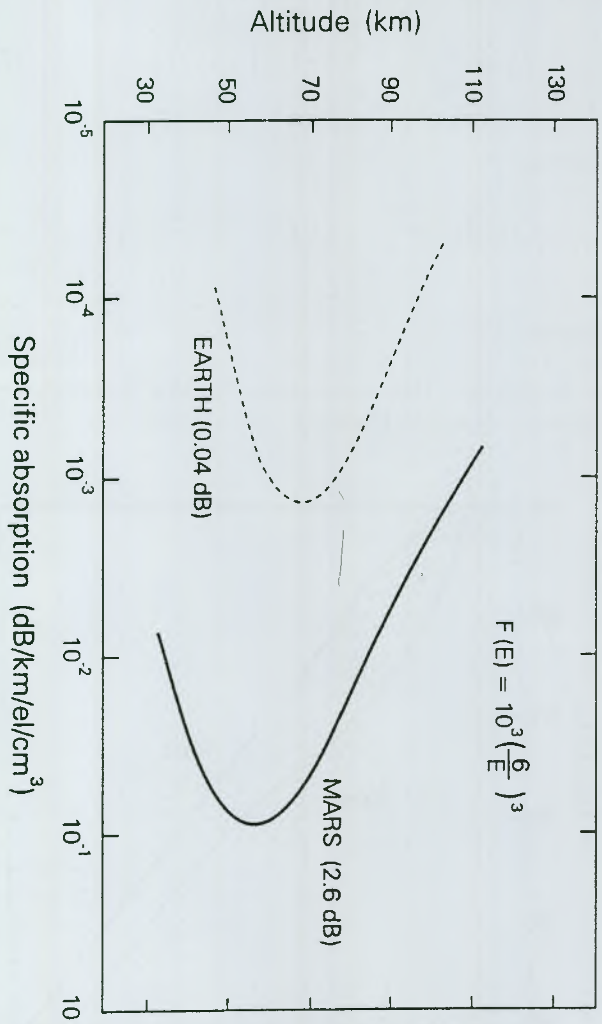


Figure 2. Altitude profiles of the specific absorption in the terrestrial and martian atmospheres for the same precipitating electron flux; the numbers in parenthesis is the total absorption values in dB.

acts in a sense as a particle detector. Thus, as part of the instrument an analysis of the Martian atmosphere with regard to riometry has been outlined above.

It has already been pointed out that the Martian atmosphere is much more sensitive to particle precipitation than the Earth's atmosphere. Since the range of precipitating particle spectra may be as large or larger than at the Earth, it is to be expected that the range of absorption values is much larger than at Earth. Thus, if for example a low frequency is used in order to be very sensitive to absorption variations, the instrument may go in saturation. If a large frequency is used, then the instrument may not be sensitive enough for observation of weak particle fluxes. It is therefore to be considered that the riometer should be operated at least two frequencies, one low and one high, for example at 10 and 100 MHz.

4.1. Antenna

The radiation pattern of a riometer antenna is typically wide beam, with a 3 dB opening angle of, say, 60 degrees, and with maximum gain in the vertical direction.

In order to ensure good gain during communications between the Lander and orbiter, the communication antenna on the Lander may be a horizontal dipole (or dipoles). If that is the case it should be examined if the communication antenna could be used also in the riometer experiment.

If the communication antenna cannot be used, then the riometer antenna will be either a horizontal short dipole mounted on the Lander, or placed on the Martian surface, or a horizontal monopole extending from the side of the Lander. In this case a further requirement for a Martian riometer antenna is that it must be deployable from a Lander and it must preferably have a very low weight, of the order of 100 g in order to be accommodated on the Lander.

The antenna is placed close to the Martian surface, at a height of less than 0.1 wave length. Owing to the very low conductivity of the Martian ground, this proximity of the antenna to the ground is not a problem. The antenna pattern in the vertical direction is modified and the gain reduced relative to its free space value, but the gain is still sufficient for the antenna to function effectively. Figure 3 shows the calculated antenna diagram for a horizontal dipole placed a height h=0 and h=0.1 wave length over the Martian surface (Smith, 1984). For h=0.1 (0) the gain in the zenith is 1 (0) dB, and the beam width is about 60 degrees.

To keep the antenna light weight but still with sufficient strength to withstand wind pressures, it could be constructed of thin-walled metal tubes, and the antenna feed point connected to the receiver by a thin coaxial cable.

If the communication antenna is used a switch must be included (controlled by the Lander system) to connect and disconnect the riometer to the antenna. If more than one frequency is used then an impedance tuning network for each frequency is required, and a switch must be included (controlled by the experiment) to connect and disconnect the impedances as required. Each impedance network should be wide band enough to allow fine tuning of the frequency to avoid interference (see below).

The location, storage and deployment of the antenna will depend on the physical appearance of the Lander.

Since lightening is predicted to occur in the dusty Martian atmosphere it may be necessary to protect the receiver input against large voltages induced in the antenna. A 'gas tube' that short circuits for large voltages on the antenna output could provide such protection.

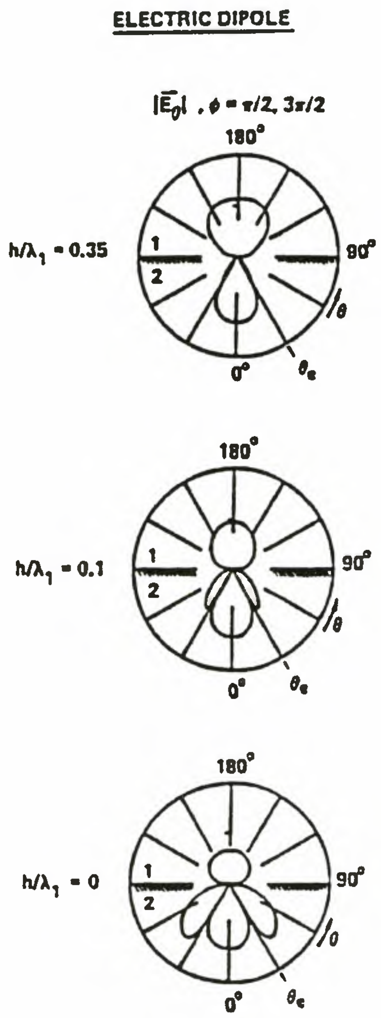


Figure 3. Electric field patterns for a dipole at various heights above an interface between lossless dielectric media, with permittivity 1 and 4.

4.2. Receiver

A very important key property of a riometer experiment is, that its response to a given signal intensity must be as constant and unchanged in time as the cosmic ray intensity itself at the top-side of the ionosphere. This property is essentially a key requirement to the receiver. For a given input radio wave signal intensity the riometer receiver must produce the same output signal, unchanged with time. The reason for this requirement lies in the procedure for calculating the ionospheric absorption. Since at any given time (t) the actual intensity at the planet surface is measured, the absorption at hat given time is calculated by comparing the intensity measured at the time (t) during disturbed times, with the intensity (I) measured at the same time (t) during quiet times. Thus, the absorption (A) in deciBell is,

$$A(dB) = 10 * \log(I(\text{'Quiettime'})/I(\text{'Disturbedtime'})) \quad (9)$$

Since the 'Quiet time' and the 'Disturbed time' may be one sol (at the Earth 24 hours) or, more realistically, weeks apart, it follows that the receiver must have constant performance over such time intervals.

Calibration of the receiver is normally ensured by the use of a time constant reference noise source. The receiver input is switched back and forth between the antenna and the noise source (for example every 10 minutes). Use of a constant reference noise source (maybe temperature calibrated) makes it possible to correct the measured intensities for time variations in the receiver performance.

Since the experiment will be used in space applications it is essential in the receiver design to minimize the instrument mass, and to minimize power consumption. It is also to be wished that the receiver is flexible so that operations can be automatically optimized or controlled remotely:

1) The absorption values that ocurrs in the Martian atmosphere may vary over an order of magnitude for a frequency of about 30 MHz. We expect a solar flare event to give rise to large absorptions, and such events should therefore be observed at a relative high frequency to ensure good accuracy of the absorption measurements. On the other hand events associated with local acceleration events, in particular night time events, may be very weak, and should therefore be observed at a relative high frequency. The frequency range to be considered may be from 10 to 100 MHz. Thus, it would be advantageous if the riometer could,
   a) operated at a selected fixed frequency appropriate to the kind of event to be studied,
   b) step through several discreete frequencies, or
   c) select the frequency to keep the absorption within a pre-set absorption range to ensure good accuracy of the absorption measurements.
2) The riometer is sensitive to all signals within the bandwidth that excite the antenna. The Station (to which the riometer is a part) may produce noise, owing to computer activity, to operation of solar cells, and to activity of other experiments. The riometer operating center frequency must therefore be carefully chosen to lie in a quiet frequency range. Fortunately, the choice of riometer frequency is very flexible; it is for example not crucial if we measure at for example 20, 22, or 19 MHz. If the riometer automatically can determine a quiet band, then it will probably also be possible to find such a quiet band to use for riometry.

3) It would probably be an advantage if the time resolution of the experiment (i.e. the time over which the measurements are averaged before bing processed) could be changed to on the one hand fit the requirements to the event studied, and to limit the amount of data to be saved.

To realise such flexibility a programmable Digital Signal Processor would probably have to be part of the riometer. This unit would also serve to digitize the analog output of the receiver.

Design goals for the receiver are,

Frequency range	10 - 100 MHz
Power	< 100 mW
Mass	< 100 g
Dimensions	< 100 cm^3
Programmable DSP	

4.3. Receiver Operations

The receiver measures the intensity of cosmic radio noise at the antenna input. It is calibrated by periodically switching between the antenna input and a known stable noise source. Sky noise (cosmic noise) is determined by comparing the signal received at the antenna to the signal from the noise source. The receiver operates at a frequency, or frequencies, over the range from 10 MHz to 100 MHz. Receiver parameters such as power bandwidth (nominally 100 kHz), calibration options, and operating mode can be programmed via the DSP. In addition, other receiver characteristics are programmable, such as data sample rate and integration time.

There are several selectable operating modes: fixed frequency, stepped frequency, or adjusting the frequency to yield absorption in pre-selected range. A default frequency is selected when the receiver is switched on, or when it is reset. In the stepped frequency mode, a limited number of specific frequency channels are selected sequentially (for example, 10, 25, 50 and 100 MHz). In the adjust-frequency mode, the absorption is calculated (using stored information about the 'Quiet time' intensities, and an appropriate frequency selected to ensure that the uncertainty on the absorption is at an acceptable level.

To allow autonomous operation, the receiver can be programmed to search several adjacent frequencies around the desired one and to select the channel with the lowest noise.

5. DATA ANALYSIS

During a day/night sequence (hereafter called a 'sol'; one sol at Mars is equal to 24.6 hours) the riometer antenna sweeps the sky and encounters a steadily varying sky temperature, i.e. cosmic noise intensity. If the time of measurement is recorded in star time, then the cosmic noise intensity (above the ionosphere, or equivalently for a quiet time ionosphere) is always the same for the same time. The associated variation of the riometer output, V1 (see Section I.6.2), is referred to as 'the quiet day curve', with the cosmic noise power Q. During an absorption event the cosmic noise power is equal to D, and with Q measured at the same star time as D, the absorption is

$$A[dB] = 10 * \log(\frac{Q}{D}) \tag{10}$$

Thus, to determine the absorption one must know the quiet day curve. In principle this requires only measurements during one sol, provided there are no absorption events during that time. Since one must assume some events occurring every sol, in practice determination of the quiet day curve would require observations over 5 to 10 sols.

The accuracy of the absorption is given by

$$dA[dB] = 10 * \log(1 + (1/\tau * B) \tag{11}$$

where

τ	=	integration time in second;
B	=	bandwidth in Hz;
for τ	=	5 and B = 10^5, dA = 0.01 dB
for τ	=	0.1 and B = 10^5, dA = 0.05 dB

The riometer measures continuously the radio wave intensity with a time resolutions of a few seconds, allowing studies of absorption with time scales from 10ths of seconds to hours. The time profile of the absorption allows recognition of different types of events, or makes it possible to separate the events into different types. The measurements allows the pattern of occurrence of these different event types during a sol to be determined, and this will assist in identifying the origin of the events. With a few assumptions it is possible to calculate the expected absorption from a given particle spectrum. Varying the given spectrum until a fit to the observed one is obtained, allows to determine limits of the actual spectrum of the precipitating particle fluxes.

When the riometer is operated with a time resolution of a tenth of a second increases in the signal intensity which may occur owing to lightening will be observed with a time resolution sufficient to determine the occurrence and intensity of lightening.

In order to possibly reduce the amount of data to be transmitted to Earth, it will be investigated if the riometer could include a data compression procedure in the form of automatic event selection software. The software will be responsible for deciding which data is of interest and should be transmitted back to Earth. This software will also be responsible for the generation of the Martian quiet-day curve (QDC) used for the absorption calculations.

REFERENCES

Detrick, D.L., Rosenberg, T.J., and Fry, C.D. (1997), Analysis of the Martian atmosphere for riometry, *Planet. Space Sci. [*45, 289.

Hanson, W.B., Sanatani, S., and Zuccaro, D.R. (1977), The Martian ionosphere as observed by the Viking retarding potential analyzers, *J. Geophys. Res.* **82**, 4351.

Kliore, A.J., Cairn, D.L., Fjeldbo, G., Seidel, B.L., Sykes, M.J., and Rasool, S.I. (1972), The atmosphere of Mars from mariner 9 radio occultation measurements, *Icarus* **17**, 484.

Klots, C.E. (1968), Energy deposition processes, In Fundamental processes in Radiation Chemistry, ed. P.J. Ausloos, Wiley, New York.

Krymskii, A.M., Breus, T., and Nielsen, E. (1995), On possible observational evidence in electron density profiles of a magnetic field in the Martian ionosphere, *J. Geophys. Res.* **100**, 3721.

Lindal, G.F., Hotz, H.B., Sweetnam, D.N., Shiponny, Z., Brenkle, J.P., Hartsell, G.V., Spear, R.T., and Michael,Jr., W.H. (1979), Viking radio occultation measurements of the atmosphere and topography of Mars: Data acquired during 1 Martian year of tracking, *J. Geophys. Res.* **84**, 8443.

Ranta, A., Ranta, H., Turunen, T., Silen, J., and Stauning, P. (1985), High resolution observations of D-region by EISCAT and their comparison to riometer measurements, *Planet. Space Sci.* **33**, 583-589.

Schunk, R.W., and Nagy, A.F. (1980), Ionospheres at the terrestrial planets, *Review of Geophys. Space Phys.* **18**, 813.

Seiff, A., and Kirk, D.B. (1977), Structure of the atmosphere of Mars in summer at mid-latitudes, *J. Geophys. Res.* **82**, 4364.

Smith, G.L. (1984), Directive properties of antennas for transmission into a material half-space, IEEE Transactions on Antennas and propagation, AP-32, 3.

Verigin, M.I., Axford, W.I., Gringauz, K.I., and Richter, A.K. (1987), Acceleration of cometary plasma in the vicinity of comet Halley associated with an interplanetary magnetic field polarity change, *Geophys. Res. Lett.* **14**, 987.

Zhang, M.H.G., Luhman, J.G., and Kliore, A.J. (1990), An observational study of the nightside ionosphere of Mars and Venus with radio occultation methods, *J. Geophys. Res.* **95**, 17095.

List of Participants

Aben, I.
SRON
Sorbonnelaan 2
3584 CA Utrecht
The Netherlands

Tel: +31 30 2538579
Fax: +31 30 2540860
e-mail: i.aben@sron.nl

Algra, T.
NLR - National Aerospace Laboratory
Postbus 90502
1006 BM Amsterdam
The Netherlands

Tel: +31 527 248464
Fax: +31 527 248210
e-mail: algra@nrl.nl

Amore, O.A.
Thomson CSF
Avenue de Rochepleine
BP 123
F-38521 Saint-Egreve Cedex
France

Tel: +33 2 76 58 3440

Attema, E.
ESTEC, SCI-VRS
Postbus 299
2200 AG Noordwijk
The Netherlands

Tel: +31 71 565 4461
Fax: +31 71 565 5675
e-mail: eattema@estec.esa.nl

Atzei, A.
ESTEC, IMT-TS
Postbus 299
2200 AG Noordwijk
The Netherlands

Tel: +31 71 565 3418
e-mail: aatzei@estec.esa.nl

Baumert, T.H.
DLR
Postfach 1116
D-82230 Wessling
Germany

Tel: +49 8153 28 1103
Fax: +49 8153 28 1271
e-mail: thomas.baumert@dlr.de

Benz, R.
Dornier Satellitensysteme
P.O. Box 1420
D-88039 Friedrichshafen
Germany

Tel: +49 7545 83250
Fax: +49 7545 84105
e-mail:
rudolf.benz@dss.dornier.dasa.de

Berndt, K.
DASA-DJO
Jena-Optronik GmbH
Pruessingstrasse 41
D-07745 Jena
Germany

Tel: +49 3641 200 132
Fax: +49 3641 200 222
e-mail: kabe@djo-jena.de

Billebaud, F.
Observatoire de Bordeaux
Boite Postale 89
F-33270 Floirac
France

Tel: +33 5 57 77 61 00
Fax: +33 5 5777 6110
e-mail: billebaud@observ.u-bordeaux.fr

Bloemink, H.I.
Royal Meteorological Institute
(KNMI)
Wilhelminalaan 10
P.O. Box 201
3730 AE De Bllt
The Netherlands

Tel: +31 30 2206461
Fax: +31 30 2210407
e-mail: bloemink@knmi.nl

Borgeaud, M.
ESTEC, EEP
Postbus 299
2200 AG Noordwijk
The Netherlands

Tel: +31 71 565 4830
Fax: +31 71 565 4999
e-mail: Maurice@xe.estec.esa.nl

Boucharlat, G.
Thomson-CSF Semiconducteurs Spec.
BP 123
F-38521 Saint Egreve
France

Tel: +33 4 7658 3000
Fax: +33 4 7658 3497

Brandt, G.
Delft Sensor Systems
Westerring 21
B-9700 Oudenaarde
Belgium

Tel: +32 55 333 811
Fax: +32 55 316 895
e-mail: gbr@oip.be

Bruijn, M.P.
SRON
Sorbonnelaan 2
3584 CA Utrecht
The Netherlands

Tel: +31 30 2535 600
Fax: +31 30 254 0860
e-mail: m.p.bruijn@sron.nl

Burrows, J.P.
University of Bremen
Inst. of Environmental Physics
Kufsteiner Str NW1
P.O. Box 330440
D-28334 Bremen
Germany

Tel: +49 421 218 4548
Fax: +49 421 218 4555
e-mail: burrows@iup.physik.uni-bremen.de

Bussey, B.
ESTEC, SCI-SO
Postbus 299
2200 AG Noordwijk
The Netherlands

Tel: +31 71 565 4720
e-mail: bbussey@estec.esa.nl

Callies, J.
ESTEC, APP-LTO
Postbus 299
2200 AG Noordwijk
The Netherlands

Tel: +31 71 5655179
e-mail: jcallies@estec.esa.nl

Carpay, J.G.H.
NIVR
P.O. Box 35
2600 AA Delft
The Netherlands

Tel: +31 15 2782015
Fax: +31 15 2623096
e-mail: j.carpay@nivr.nl

Chautemps, J.
Thomson CSF
Avenue de Rochepleine
BP 123
F-38521 Saint-Egreve Cedex
France

Tel: +33 4 76 58 3112
Fax: +33 4 76 58 3125
e-mail: chautemps@tcs.thomson.fr

Chicarro, A.
ESTEC, SCI-SO
Postbus 299
2200 AG Noordwijk
The Netherlands

Tel: +31 71 565 3613
Fax: +31 71 565 4697
e-mail: achicarr@estec.esa.nl

Court, A.
TNO Institute of Applied Physics
P.O. Box 155
2600 AD Delft
The Netherlands

Tel: +31 15 269 2823
Fax: +31 15 269 2111
e-mail: court@tpd.tno.nl

Cox, G.
Matra Marconi Space UK
Antenna Systems
Anchorage Road
Portsmouth PO3 5PU
United Kingdom

Crone, G
ESTEC, XEA
Postbus 299
2200 AG Noordwijk
The Netherlands

Tel: +31 71 565 3934
Fax: +31 71 565 4999
e-mail: gcrone@estec.esa.nl

Cutter, M.A.
SIRA Electro-optics Ltd
South Hill
Chislehurst
Kent BR7 5EH
United Kingdom

Tel: +44 181 467 2636
Fax: +44 181 467 6515
e-mail:Mike_Cutter@siraeo.co.uk

De Nivelle, M.
SRON
Sorbonnelaan 2
3584 CA Utrecht
The Netherlands

Tel: 030 2535600
Fax: 030 2540860
e-mail: m.denivelle@sron.ruu.nl

De Vos, L.
Delft Sensor Systems
Westerring 21
B-9700 Oudenaarde
Belgium

Tel: +32 55 333 811
Fax: +32 55 316 895
e-mail: ldv@oip.be

Detaille, M.
Societe Europeenne de Systemes
Optiques SESO
BP 55000
F-13792 Aix en Provence Cedex 3
France

Tel: +33 4 4216 8500
Fax: +33 4 4216 8585

Donovan, D.P.
Royal Meteorological Institute
(KNMI)
Wilhelminalaan 10
P.O. Box 201
3730 AE De BIlt
The Netherlands

Tel: +31 30 2206 770
Fax: +31 30 221 0407
e-mail: donovan@knmi.nl

Endemann, M.
ESTEC, APP-NWI
Postbus 299
2200 AG Noordwijk
The Netherlands

Tel: +31 71 565 5129
Fax: +31 71 565 5153
e-mail: mendeman@estec.esa.nl

Foing, B.H.
ESTEC, SCI-SO
Postbus 299
2200 AG Noordwijk
The Netherlands

Tel: +31 71 565 5647
Fax: +31 71 565 4698
e-mail: bfoing@estec.esa.nl

Giunta, D.
ESTEC, TOS-ETD
Postbus 299
2200 AG Noordwijk
The Netherlands

Tel: +31 71 565 3863
Fax: +31 71 565 4596
e-mail: domenico@ws.estec.esa.nl

Grard, R.J.L.
ESTEC, SCI-SO
Postbus 299
2200 AG Noordwijk
The Netherlands

Tel: +31 71 565 3596
Fax: +31 71 565 4697
e-mail: rgrard@estec.esa.nl

Gudmandsen, P.E.
Technical Univ of Denmark
Building 348
DK-2800 Lyngby
Denmark

Tel: 45 45 25 37 88
Fax: 45 45 93 16 34
e-mail: pg@emi.dtu.dk

Guijarro, J.
ESTEC
Postbus 299
2200 AG Noordwijk
The Netherlands

Tel: +31 71 565 4435
e-mail: jquijarro@estec.esa.nl

Hagfors, T
Max-Planck-Institut f. Aeronomie
Postfach 20
D-37191 Katlenburg-Lindau
Germany

Tel: +49 5556 979 415
Fax: +49 5556 979 410
e-mail:
hagfors@linghag.mpae.gwdg.de

Hartogh, P.
Max Planck Institute for Aeronomie
Max Planck str. 2
D-37191 Katlenburg-Lindau
Germany

Tel: +49 5556 979342
Fax: +49 5556 979240
e-mail: paul@chirp.mpae.gwdg.de

Heilimo, J.
Space Systems Finland
Kappelitie 6
FIN-02200 Espoo
Finland

Tel: +358 9 613 28 641
Fax: +358 9 613 28 699
e-mail: jyri.heilimo@ssf.fi

Hoogeveen, R.
SRON
Sorbonnelaan 2
3584 CA Utrecht
The Netherlands

Tel: +31 30 2535703
Fax: +31 30 2540860
e-mail: r.hoogeveen@sron.nl

Hugon, X.
Thomson CSF
Avenue de Rochepleine
BP 123
F-38521 Saint-Egreve Cedex
France

Tel: +33 4 76 58 31 93
Fax: +33 476 58 31 05

Josset, J.L.
CSEM
1, rue Jaquet Droz
BP 41
CH-2000 Neuchatel
Switzerland

Tel: 41 32 7205 534
Fax: 41 32 7205 720
e-mail: jean-luc.josset@csemne.ch

Kattnig, Mr., A.
ONERA
BP 72
F-92322 Chatillon Cedex
France

e-mail: kattnig@onera.fr

Klooster Van 'T, C.G.
ESTEC, TOS-EEA
Postbus 299
2200 AG Noordwijk
The Netherlands

Tel: +31 71 565 3940
e-mail: kvtkloos@estec.esa.nl

Koopman, R.M.
ESTEC
Postbus 299
2200 AG Noordwijk
The Netherlands

Tel: +31 71 5655240
Fax: +31 71 565 5617
e-mail: rkoopman@jw.estec.esa.nl

Kuenzi, K.
University of Bremen
Inst. of Environmental Phys.
P.O. Box 330440
D-28334 Bremen
Germany

Tel: +49 421 218 3909
Fax: +49 421 218 4555
e-mail: kunzi@physik.uni-bremen.de

Kummer, U.
Dornier Satellitensysteme
P.O. Box 1420
D-88039 Friedrichshafen
Germany

Tel: +49 7545 8 4147
Fax: +49 7545 8 4105
e-mail:
Uwe.Kummer@dss.dornier.dasa.de

Labate, D.
Officine Galileo
Via. A. Einstein, 35
50013 Campi Bisenzio (Firenze)
Italy

Tel: +39 55 89 50 313
Fax: +39 55 89 50 613
e-mail: dis_optics@galileo-space.it

Landi, A.
Officine Galileo
Via. A. Einstein, 35
50013 Campi Bisenzio (Firenze)
Italy

Tel: +39 55 8950700
Fax: +39 55 8950613
e-mail: dis_optics@galileo-space.it

Lebreton, J.-P.
ESTEC, SCI-SO
Postbus 299
2200 AG Noordwijk
The Netherlands

Tel: +31 71 565 3600
Fax: +31 71 565 4697
e-mail: jlebreto@estec.esa.nl

Lin, C.
ESTEC, JPE
Postbus 299
2200 AG Noordwijk
The Netherlands

Tel: +31 71 565 5813
Fax: +31 71 565 4696
e-mail: clin@estec.esa.nl

Livi, S.
Max Planck Institute for Aeronomie
Max Planck str. 2
D-37191 Katlenburg-Lindau
Germany

Tel: +49 5556 979429
Fax: +49 5556 979 139
e-mail: livi@linmpi.mpg.de

Lobb, D.R.
SIRA Limited
South Hill
Chislehurst
Kent BR7 5EH
United Kingdom

Tel: +44 181 467 2636
Fax: +44 181 467 6515
e-mail: Dan.Lobb@sicaeo.co.uk

Mahdi, S.
Dornier Satellitensysteme
P.O. Box 1420
D-88039 Friedrichshafen
Germany

Tel: +49 7545 8 4887
Fax: +49 7545 8 4177
e-mail:
semir.mahdi@dss.dornier.dasa.de

Martimort, PH.
Alcatel Space Industrie
100, Boulevard du Midi
BP 99
F-06156 Cannes La Bocca Cedex
France

Tel: +33 4 9292 7993
Fax: +33 4 9292 7660

Martin-Neira, M.
ESTEC, TOS-ETP
Postbus 299
2200 AG Noordwijk
The Netherlands

Meynart, R.
ESTEC, JPE
Postbus 299
2200 AG Noordwijk
The Netherlands

Tel: +31 71 565 4606
e-mail:
RMEYNART@ESTEC.ESA.NL

Michaelis, H.
DLR Berlin
Rudower Chaussee 5
D-12489 Berlin-Adlershof
Germany

Tel: 49 30 6 70 55-0
e-mail: Harald.Michaelis@dlr.de

Mutlow, C.
Rutherford Appleton Laboratory
Space Science Department
Chilton, Didcot
Oxfordshire OX11 OQX
United Kingdom

Tel: +44 1235 446225
Fax: +44 1235 445848
e-mail: c.t.mutlow@rl.ac.uk

Nemeth, S.
IMEC
Kapeldreef 75
Heverlee
B-3001 Leuven
Belgium

Tel: +32 16 281 684
Fax: +32 16 281 501
e-mail: nemeth@imec.be

Nielsen, E.
Max Planck Institute for Aeronomie
Max Planck str. 2
D-37191 Katlenburg-Lindau
Germany

Tel: +49 5556 979 450
Fax: +49 5556 979 240
e-mail: nielsen@linmpi.mpg.de

Olij, C.
TNO-TPD
Stieltjesweg 1
P.O. Box 155
2600 AD Delft
The Netherlands

Tel: +31 15 269 2000
Fax: +31 15 269 2111
e-mail: olij@tpd.tno.nl

Parkes, S.M.
University of Dundee
Applied Computing
Dundee DD1 4HN
United Kingdom

Tel: +44 1382 345194
Fax: +44 1382 345509
e-mail:
sparkes@computing.dundee.ac.uk

Peacock, A.
ESTEC, SCI-SA
Postbus 299
2200 AG Noordwijk
The Netherlands

Tel: +31 71 565 3563
e-mail:
APEACOCK@ESTEC.ESA.NL

Pimm, I.
ESTEC, QCT
Postbus 299
2200 AG Noordwijk
The Netherlands

Tel: +31 71 565 4986
e-mail: IPIMM@ESTEC.ESA.NL

Proetel, K.
German Aerospace Center, DLR
Koenigswintererstr. 522-524
PO Box 300364
D-53227 Bonn
Germany

Tel: +49 228 447 351
Fax: +49 228 447 745
e-mail: klaus.proetel@dlr.de

Readings, C.J.
ESTEC, SCI-VR
Postbus 299
2200 AG Noordwijk
The Netherlands

Tel: +31 71 565 5673
e-mail:
CREADINGS@ESTEC.ESA.NL

Riemann, R.M.
ESTEC, TOS-EUM
Postbus 299
2200 AG Noordwijk
The Netherlands

Tel: +31 71 565 5523
Fax: +31 71 565 5731
e-mail: rriemann@estec.esa.nl

Rivada Antich, A.L.
ESTEC, TOS-EEO
Postbus 299
2200 AG Noordwijk
The Netherlands

Tel: +31 71 565 3985
e-mail: alrivada@estec.esa.nl

Saari, H.K.
VTT Automation
P.O. Box 1304
FIN-02044 ESPOO
Finland

Tel: 358 04564360
Fax: 358 0 4567029

Saint-Pe, O.
Matra Marconi Space
31, Av. des Cosmonautes
Z.I. du Palays
F-31402 Toulouse CEDEX 4, France

Tel: +33 562196716
Fax: +33 5 62197785
e-mail: olivier.saintpe@mms.tls.fr

Sawyer, G.
European Commission
200, Rue de la Loi
B-1049 Brussels
Belgium

Tel: +32 2 295 9857
Fax: +32 2 296 0588
e-mail: geoff.sawyer@dg12.crc.be

Schwehm, G.H.
ESTEC, SCI-SO
Postbus 299
2200 AG Noordwijk
The Netherlands

Tel: +31 71 565 3539
e-mail:
GSCHWEHM@ESTEC.ESA.NL

Silvestrin, P.
ESTEC, APP-JPP
Postbus 299
2200 AG Noordwijk
The Netherlands

Tel: +31 71 565 3689
Fax: +31 71 565 4696
e-mail: psilvest@estec.esa.nl

Sipkes, J.H.A.
Earth Monitoring Consultancy
P.O. Box 2960
Florida, 1710
South Africa

Tel: +27 114728012
Fax: +27 11 472 8012
e-mail: emc@pixie.co.za

Southwood, D.
ESA Headquarters, SCI-E
8-10 Rue Mario Nikis
F-75738 Paris Cedex 15
France

Tel: +33 1 53697172
Fax: +33 1 53697674

Svedhem, H.
ESTEC, SSD
Postbus 299
2200 AG Noordwijk
The Netherlands

Tel: +31 71 5653370
Fax: +31 71 5654697
e-mail: hsvedham@estec.esa.nl

Tanzi, C.P.
SRON
Sorbonnelaan 2
3584 CA Utrecht
The Netherlands

Tel: +31 30 253 8585
Fax: +31 30 254 0860
e-mail: c.tanzi@sron.nl

Testud, J.
CETP/CNRS
10-12 Av. de l'Europe
F-78140 Velizy
France

Tel: +33 1 39 25 4776
Fax: +33 1 39 25 4778
e-mail: testud@cetp.ipsl.fr

Tribolet, P.
Sofradir
43-47 Rue Camille Pelletan
F-92298 Chatenay-Malabry
France

Tel: +33 1 41 13 4530
Fax: +33 1 46 61 5884
e-mail:
sofradirchatenay@compuserve.com

Tuominen, T.
Patria Finavitec
OY Systems
Naulakatu 3
FIN-33100 Tre
Finland

Tel: +358 3 2450 111
Fax: +358 3 2450 243
e-mail: taisto.tuominen@patria.fi

Ulaby, F.T.
University of Michigan
Electrical Eng.& Comp. Science
3228 EECS Building
Ann Arbor, MI 48109-2122
USA

Tel: +1 734 764 0501
Fax: +1 734 547 2106
e-mail: ulaby@eecs.unmich.edu

Van Hoof, C.
IMEC
Kapeldreef 75
Heverlee
B-3001 Leuven
Belgium

Tel: +32 16 281518
Fax: +32 16 281 501
e-mail: vanhoofc@imec.be

Velluet, M.T.
ONERA
BP 72
F-92322 Chatillon Cedex
France

Tel: +33 1 4673 4749
Fax: +33 1 4673 3772
e-mail: velluet@onera.fr

Ward, R.
GEC-Marconi Infra-Red Limited
PO Box 217
Millbrook Industrial Estate
Southampton S09 7QG
United Kingdom

Tel: +44 1703 316805
Fax: +44 1703 316777
e-mail: rodney.ward@gecm.com

Wenzel, K.P.
ESTEC
Space Science Department
Postbus 299
2200 AG Noordwijk
The Netherlands

Tel: +31 71 565 3573
Fax: +31 71 565 4697
e-mail:
KWENZEL@ESTEC.ESA.NL

Werner, CH.
DLR
Inst. of Optoelectronics
Postfach 1116
D-82230 Wessling
Germany

Tel: +49 8153 282573
Fax: +49 8153 281608
e-mail: Christian.Werner@dlr.de

Wijnberger, J.J.
SRON-Groningen
Universiteitscomplex 'Paddepoel'
Landleven 12, P.O. Box 800
9700 AV Groningen
The Netherlands

Tel: +31 50 3634074
Fax: +31 50 3634033
e-mail: j.j.wijnberger@sron.nl

Wingham, D.J.
Mullard Space Science Laboratory
University College London
Holmbury St. Mary
Dorking, Surrey RH5 6NT
United Kingdom

Tel: +44 171 419 3677
Fax: +44 171 419 3418
e-mail: DJW@mssl.ucl.ac.uk

Wismann, V.
Space Applications Institute
via E. Fermi
I-21020 Ispra
Italy

Tel: +39 0332 78 6245
Fax: +39 0332 78 5461
e-mail: volkmar.wismann@jrc.it

Zender, J.
ESTEC, SCI-SO
Postbus 299
2200 AG Noordwijk
The Netherlands

Tel: +31 71 565 4919

Zheng, Y.
Surrey Space Centre
University of Surrey
Guildford
Surrey GU2 5XH
United Kingdom

Tel: +44 1483 259078
Fax: +44 1483 259503
e-mail: y.zheng@ee.surrey.ac.uk

European Space Agency
Agence spatiale européenne

Contact: ESA Publications Division
c/o ESTEC, PO Box 299, 2200 AG Noordwijk, The Netherlands
Tel (31) 71 565 3400 - Fax (31) 71 565 5433