The High Resolution Spectrometer for SOFIA-GREAT
Instrumentation, Atmospheric Modeling and Observations

Dissertation
zur Erlangung des Doktorgrades
der Fakultät für Angewandte Wissenschaften
der Albert-Ludwigs-Universität zu Freiburg

vorgelegt von
Gerónimo Luis Villanueva Sozzi
aus Mendoza / Argentinien

Freiburg, 2004

Bibliografische Information Der Deutschen Bibliothek
Die Deutsche Bibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet `http://dnb.dbb.de` abrufbar.

D7
Referent: Prof. Dr. Leonhard Reindl
Korreferent: Prof. Dr. Oskar von der Lühe
Tag der mündlichen Prüfung: 30. November 2004

ISBN 3-936586-34-9

Copernicus GmbH, Katlenburg-Lindau
Druck: Schaltungdienst Lange, Berlin

Printed in Germany

The time will come when diligent research over long periods will bring to light things which now lie hidden. A single lifetime, even though entirely devoted to the sky, would not be enough for the investigation of so vast a subject … And so this knowledge will be unfolded only through long successive ages. There will come a time when our descendants will be amazed that we did not know things that are so plain to them … Many discoveries are reserved for ages still to come, when memory of us will have been effaced. Our universe is a sorry little affair unless it has in it something for every age to investigate … Nature does not reveal her mysteries once and for all.

Seneca, Natural Questions, Book 7, first century.

DEDICATED TO
MY PARENTS

Abstract
(Deutsch)

Die vorliegende Dissertation beschreibt die Entwicklung eines hochauflösenden Spektrometers und seine Anwendung in der Erforschung planetarer Atmosphären anhand eines Zirkulationsmodells und Observationen der Marsatmosphäre. Das neue Spektrometer ist das "Chirp-Transform-Spektrometer" (CTS). Es ist ein Teil des "German REceiver for Astronomy at Terahertz frequencies" (GREAT) am "Stratospheric Observatory For Infrared Astronomy" (SOFIA). Das CTS wird über ein bisher unerreichtes spektrales Auflösungsvermögen bei linearer Charakteristik verfügen. Die analoge Fourier-Transformation des CTS wurde entscheidend verbessert durch ein neues Design, im Folgenden als "Adaptive Digital Chirp Processor" (ADCP) bezeichnet wird. Die Funktionsweise des ADCP beruht auf der digitalen Erzeugung eines Signals, dessen dispersive Charakteristika den Eigenschaften der Verzögerungsleiter der SAW-Filter optimal angepasst werden können, wodurch die höhere spektrale Auflösung und eine größere Dynamik erreicht werden. Die Entwicklung des ADCP erforderte die optimale Kombination verschiedener Technologien, beispielsweise von Quadraturmodulation, Einsatz von Hochleistungsfrequenzfiltern und ultrahochstabilen Oszillatoren, sowie die grundlegende Auseinandersetzung mit der Problematik durch numerische Simulationen mit Hilfe eines ad hoc- Modells.
Das CTS stellt eine neue leistungsstarke Methode zur Erforschung planetarer Atmosphären dar, besonders im Rahmen des SOFIA-Projekts. Es ermöglicht die Messung hochaufgelöster Höhenprofile von Temperatur, Windgeschwindigkeit, Wasserdampf und Spurenstoffen. Die Verbindung zwischen den spektroskopischen Untersuchungen und den zugrundeliegenden physikalischen Phänomenen wurde im Rahmen der Arbeit durch die Entwicklung eines allgemeinen Zirkulationsmodells ("general circulation modell", GCM) für die Marsatmosphäre vollzogen. Ein funktionsfähiger Prototyp des SOFIA-GREAT-CTS wurde am Submillimeter-Teleskop in Arizona in den USA zur Messung der Kohlenmonoxidlinie (CO) bei 345 GHz verwendet. Aus diesen Daten wurden Temperaturprofile bestimmt, die als Eingangsdaten für die Simulationsrechnungen mit dem GCM dienten. Des Weiteren ist das CTS ein leistungsfähiges Instrument für Analyse von Linien geringer Breite, wie sie typischerweise im Spektrum von Kometen auftreten. So ist es beispielsweise möglich, durch die Detektion der Linie von Blausäure (HCN) mit Hilfe des CTS Asymmetrien beim Ausgasen des Kometen Encke nachzuweisen.

Abstract
(English)

The purpose of the thesis is the development of a high resolution spectrometer, in association with atmospheric modeling and observations. The new spectrometer, the “Chirp-Transform-Spectrometer (CTS)” as part of the “German REceiver for Astronomy at Terahertz frequencies (GREAT)” on the “Stratospheric Observatory For Infrared Astronomy (SOFIA)”, will provide unprecedented spectral resolving power and linearity response. The analog Fourier transform performed by the CTS spectrometer was significantly improved through a new design called “Adaptive Digital Chirp Processor (ADCP)”. The principle behind the ADCP consists of digitally generating the dispersive signal which adapts to the convolver dispersive properties, achieving higher spectral resolution and higher dynamic range. This development demanded a rational and optimum combination of different technologies, such as quadrature-modulation, high performance spectral filtering, ultra-stable frequency sources and a deep study of the problem through numerical simulations with an ad hoc model.

The newly developed CTS is a prevailing instrument for the study of planetary atmospheres, especially as part of the SOFIA mission providing highly resolved altitude profiles of temperature, wind, water vapour and minor species from ground level up to an altitude of 80 km. Consequently, the link between the spectroscopic observations and the physical phenomena in study was addressed via the development of a general circulation model (GCM) for the Martian atmosphere. A running prototype of SOFIA-GREAT-CTS mounted at the Submillimeter Telescope in Arizona was used for remote sensing of Mars through the broadened line of CO at 345 GHz, providing temperature measurements which were then used as data assimilation inputs for the GCM. The new CTS also demonstrated itself to be a powerful tool for the study of narrow line features typical of cometary bodies, through the detection of hydrogen cyanide and outgassing asymmetries on comet Encke.

Abstracto
(Castellano)

Esta tesis describe el desarrollo de un espectrómetro de alta resolución, que es combinado con simulaciones de circulación atmosféricas y observaciones. El nuevo espectrómetro, el "Chirp Transform Spectrometer" (CTS) como parte del "German REceiver for Astronomy at Terahertz frequencies (GREAT)" en el "Stratospheric Observatory For Infrared Astronomy (SOFIA)", proveerá de una resolución espectral sin precedentes. La transformada analógica de Fourier como base del espectrómetro CTS fue mejorada sustancialmente a través de un nuevo diseño llamada "Adaptive Digital Chirp Processor (ADCP)", o procesador adaptivo dispersivo. El principio del ADCP consiste en generar digitalmente la señal dispersiva que se adapta a la característica dispersiva del elemento que funciona como operador de convolución. Ésto conlleva a un incremento de la resolución espectral y del rango dinámico del instrumento. El nuevo desarrollo demandó una combinación óptima y racional de diferentes tecnologías, como modulación por cuadratura, filtros espectrales de alto rendimiento, osciladores ultra estables y un profundo estudio del problema a través de simulaciones numéricas con un modelo ad-hoc.

El nuevo CTS es un instrumento único para el estudio de atmósferas planetarias, especialmente como parte de la misión SOFIA, permitiendo la obtención de perfiles de temperatura, vientos, vapor de agua y otras especias menores hasta 80 km de altura. Como consecuencia, se desarrolló un modelo de circulación general (GCM) para la atmósfera de Marte como vínculo entre las mediciones espectroscópicas y el fenómeno físico de estudio. Un prototipo de SOFIA-GREAT-CTS integrado al "Heinrich Hertz Submillimeter Telescope" (HHSMT) en Arizona, fue usado para el sensado remoto de Marte a través de la transición rotacional de monóxido de carbono (CO) a 345 GHz. De las mediciones se obtuvo un mapa global de temperaturas en Marte que actualmente está siendo usado como parámetro de entrada para el GCM. El nuevo espectrómetro demostró también ser una herramienta poderosa para el estudio de la estructura hiperfina de líneas espectrales en cometas, con la detección de cianuro de hidrógeno (HCN) y asimetrías en la desgasificación del cometa Encke.

Contents

List of Tables

Chapter 1

Introduction

The recent technological developments in the receiver and spectrometer fields permit at present, observations of extremely weak spectral line features at the far-IR and submillimeter frequency range. Forty years ago, this frequency range started to become accessible by coherent and incoherent detectors, opening a new window for the study of unrevealed molecular transitions. The undesirable strong atmospheric absorption ultimately led to the development of SOFIA, a flying observatory that will avoid most of this absorption. Its high frequency dual heterodyne receiver (GREAT) combined with the high resolution spectrometer (CTS) will provide spectrometry of an extended region into the far-IR and submillimeter region (1.6 - 5 THz) with a spectral resolution of 50 kHz. This represents a unique tool for addressing a wide range of topics of modern astrophysics, especially in planetary research on the retrieval of atmospheric information through the inversion of highly resolved spectral lines and the study of outgassing asymmetries on emission lines from cometary bodies.

1.1 Far-IR and submillimeter remote sensing

Remote sensing through the far-IR and submillimeter region (300GHz - 5THz) was for a long time unfeasible due to the reduced sensitivity when compared to incoherent detectors in the infrared region and coherent detectors in the microwave range. Ignoring the fact that extensive regions of the submillimeter and far-IR radiation are opaque from sea level, an important limitation in the development of submillimeter astronomy was the complexity of the techniques necessary in the generation and detection of submillimeter radiation. The instrumental techniques were constrained by the absence of reliable oscillators and sensitive detectors together with difficulties of guidance and transformation of these high frequency signals. These difficulties however, were initially addressed in the early 1960's with the advent of semiconductor physics and quantum electronics. In addition, the rapid development and improvements of techniques of space and airborne astronomy techniques helped eliminate the influence of the absorbing terrestrial atmosphere, vastly improving the sensitivity.

Initially the submillimeter region was explored by extending the capabilities of infrared photoconductive detectors to longer wavelengths. At wavelengths shorter than 100 μm, detectors

were made by doping Ge with suitable impurities whose ionization energies determined the long-wave threshold of the detector [Bratt et. al, 1961]. With the new developments in the production of semiconductors with other base compounds having lower ionization energy, such as In-Sb, led to providing a complete coverage of the submillimeter range [Putley, 1963]. Even though this technology has reduced bandwidth at these wavelengths, it rapidly became a standard tool for astronomical ground-based observations together with bolometric detectors using semiconductive sensitive elements (In-Sb and Ge).

From the side of long wavelength heterodyne detection, the incursion into the submillimeter range took longer mainly due to the low efficiency of harmonic multiplication and also to the complexity of the radiation spectrum produced by the harmonic generators. There were also other difficulties, such as the problem of the broad-band tuning, the increasing problem of mode interference in the oversized guides, the worsening of the receiver characteristic with increasing frequency as well as the inefficiency of the conventional methods of signal recovery. The appearance of backward-wave oscillators (BWOs) [see appendix A] in the late 1960's with considerably large output power and relatively broad-band coverage stimulated the development of new detectors. The first approach consisted of extending the operational microwave point contact detectors, which would produce a substantial increase in the resolving power, the accuracy of the measurement of the line frequency and the sensitivity of the observations when compared to infrared-type incoherent detectors. Nevertheless, due to limitations in the fabrication of the point contact diodes, in which the performance of the crystal detector falls with increasing frequency because of the shunting effect of the contact capacity, the first generation submillimeter detectors had extremely poor sensitivity and were not able to scan a broad range of frequencies. Many new developments have led to the present optimum performance up to the THz regime, for instance the use of diode mixers which employ whisker contacted Schottky diodes in a quasi-optical corner cube mount [Schwaab et al., 1998]. However, the two most successful coherent detector technologies for mm/submm/IR astronomy are based on superconducting structures, and these are the SIS (Superconductor-Isolator-Superconductor) and the HEB (Hot-Electron-Bolometer) detectors [see appendix B]. At frequencies below 1 THz, the SIS technology provides high sensitivity and bandwidth combined with low power requirements from the mixing oscillator. At higher frequencies, the tuning circuit between the junction and the coupling antenna becomes very difficult to develop, because the RF impedance of the junction capacitance scales inversely with frequency and the superconductors or metals used for the tuning inductor become increasingly lossy. Recently, several improvements in the field of SIS mixers have achieved relatively good performance up to 1.2 THz [Jackson et al., 2002; Klapwijk et al., 1997]. Nevertheless with actual material technologies, above 1.5 THz the SIS detectors are too unstable and noisy [Zmuidzinas, 2002] and that is why for current developments on THz coherent detectors the focus is on HEBs. On SOFIA, the heterodyne detection at Terahertz frequencies performed by GREAT is based on diffusion-cooled HEB for its detector at 1.6-1.9 THz, and on phonon-cooled HEB for the detectors at 2.6 THz and 4.7 THz.

Alongside the evolution of submillimeter detectors of higher sensitivity and bandwidth was a great growth in different spectrometric techniques for the detected signal. The first steps on the retrieval of the spectral information of the detected signal were based on filterbanks. However, their high demands in mass and space stimulated the development of new types of spectrometers. Broadband coverage was initially achieved by "Acousto Optical Spectrometer (AOS)" which makes use of a diffractive element (Bragg-cell), and later by autocorrelators. Also in the early sixties, while developing new tools for radar systems, Klauder et al. [1960] demonstrated that the

pulse compression can be used to perform Fourier Transforms. Subsequently, Darlington [1964] showed in detail how the method could be applied to frequency measurements and in present the Darlington's system is called chirp transform spectrometer. This technique was later extended to radiometry of the atmosphere by Hartogh [1989]. Furthermore with its contribution to the space mission Rosetta and at present to SOFIA, it can be readily used for the study of a wider range of topics in modern astrophysics, from questions about comets, planetary atmospheres and the interstellar medium in the galaxy to investigations related to the early Universe.
One important constraint in the development of far-IR and submillimeter astronomy arises from the strong telluric absorption at these wavelengths, which demands that the receiving apparatus is above much of the Earth's atmosphere. Consequently, many ground based submillimeter telescopes are located at high altitudes, for instance the "Heinrich Hertz Submillimeter Telescope" (HHSMT) at Mt. Graham (3200 m), the "James Clerck Maxwell Telescope" (JCMT) at Manua Kea (4092 m), and the current development of the 64- antennae array at the Atacama desert (5400 m), "The Atacama Large Millimeter Array" (ALMA). Above an altitude of 14 km, the atmospheric transmission is significantly enhanced due to the fact that the main submillimeter absorber (water vapor) decreases abruptly above the tropopause. This encouraged the development of airborne and space observations, especially the former which in addition, contains some of the advantages of ground based observatories such as long lifetime, high reusability and easy maintenance. In the early seventies, NASA initiated the development of airborne observatories and then in 1974 it started the KAO program that ended in 1996. Many discoveries were made during this mission, including the detection of interstellar water in comets and the rings of Uranus [Elliot, 1998]. The idea of a new, bigger telescope mounted on a Boeing 747 SP surfaced in 1984, under the name SOFIA [see appendix C]. This new approach would provide a higher spatial resolution with its 2.5 m diameter main dish (compared to 0.91m of KAO) and the possibility of including the latest IR and submillimeter detectors and spectrometers. Together with the evolution of airborne observatories, space missions as ODIN and SWAS (Submillimeter Wave Astronomy Satellite) provided limited coverage of the lowest region of the submillimeter range but with unprecedented sensitivity. Also, currently under development is the promising space telescope "Heterodyne Instrument for the Far-Infrared" (HIFI), which will provide from 2007 to 2009 a continuous coverage over the range of 0.5-1.2 THz and 1.4-1.9 THz.

1.2 Scientific motivation

1.2.1 Planetary research

Planetary and cometary research are the main motivations for the development of SOFIA-GREAT-CTS, since its spectral resolution allows high definition retrieval of atmospheric information through the inversion of spectral lines. The first microwave detection of carbon monoxide on Mars by Kakar et al. [1977] at 115 GHz demonstrated the potential of the study of broadened lines from our neighbour planets on the retrieval of their atmospheric structure and more importantly of their chemical diversity. However, most of our extra-terrestrial remote sensing knowledge of atmospheres started from the radiometry of our own atmosphere. Several studies of the stratosphere, mesosphere and ionosphere were realized by different groups around the world, providing important information about the physicochemical processes in the

atmosphere and their interrelation with the general circulation. This reproduced a great boom in the study of extra-terrestrial planetary atmospheres through ground-based, airborne and space borne observations. The capability of SOFIA to avoid most of the Earth's atmospheric absorption will lead to an important increase in sensitivity, representing a great advantage for the study of weak trace gases and their influence to the overall atmospheric energy balance.

The characterized spectrum of the molecular transition allows defining our target, while the correlation between pressure and therefore altitude with the line broadening allows retrieving atmospheric composition profiles from the line shape. Atmospheric research with a CTS spectrometer started with the study of halogens-bearing species, ClO and HCl, during the European Artic Stratospheric Ozone Experiment [Nett et al. 1992; Crewell et al., 1994; Wehr et al., 1995], and during the intercomparison campaign with the MAS satellite experiment [Aellig et al., 1996; Feist et al., 2000]. Furthermore, the CTS has been applied to ozone research in the stratosphere and mesosphere [Hartogh & Jarchow, 1994/5]. Recent scientific results based on the CTS ozone data helped to explain and quantify the building mechanism of the tertiary ozone maximum [Hartogh et al., 2004], and the comparison between the groundbase experiments CTS and LIDAR has shown very good agreement between 15 and 35 km altitude [Song et al., 1998]. Microwave CTS data, in addition, significantly improved our understanding of the annual and inter-annual variation of middle atmospheric water vapour [Hartogh & Jarchow, 1995; Seele & Hartogh, 1999] in high and mid latitudes [de La Noe et al., 2000; Hartogh 2000; Nedoluha et al., 2000], and the building mechanism of noctilucent clouds [von Zahn et al, 2004]. Also it has been demonstrated that the CTS water vapour data can be used as a tracer of air movement in the middle [Seele & Hartogh, 2000] and upper atmosphere [Stevens et al., 2003]. The instrumental and retrieval knowledge was then evolved and introduced into remote sensing of extra-terrestrial sources through the integration of a CTS spectrometer into the HHSMT Telescope at Mt. Graham. The successful observations performed with the SOFIA-GREAT-CTS by Villanueva et al. [2003b] at HHSMT on Mars, and those performed by Hartogh et al. [1997], demonstrated the potential of such an instrument for addressing planetary research.

In comparison with space missions, where weight and power consumption constrain the scientific research, the airborne observatory SOFIA will let us observe Mars through coherent detection up to 5 THz with a spectral resolution of 50 kHz. This means the capability of performing different studies, such as a chemical survey of Mars of weak transitions in the far-IR and submillimeter ranges, and the study of spectral lines currently hidden behind telluric lines. Mars continues to be of great interest since there is evidence that it once had liquid water running across the surface and has undergone recent climate changes. The driving processes of these changes are still not clearly understood, because of our still restricted knowledge of the Martian general circulation and atmospheric dynamics and the effects of minor atmospheric constituents to the energetic balance (such it is the case of ozone in earth's atmosphere). For this purpose I am working in the development of a GCM (General Circulation Model) for the Martian atmosphere which will give us an important tool for understanding the implications of different physical phenomena to the overall Martian circulation. The model will be able to assimilate information from observations performed with SOFIA and other missions, defining the boundaries for the simulations and from them retrieve atmospheric information of the related physical effects. The question that immediately arises is "What type of information could SOFIA provide to the GCM?" It will mainly be focused on the study of minor chemical components and their implications to the overall energetic and chemical budgets. The study of the water cycle could be of great interest, since its interaction with the dust particles would lead to the

precipitation of the condensed particles, reproducing a large thermodynamic effect as the atmospheric opacity, and therefore sunlight absorption is strongly affected. The observations of different water vapour lines in the infrared range would lead to a better understanding of the seasonal cycle and would help improve and enhance the rich spatial but limited spectral resolution of the water vapour data-sets actually available. This data was obtained by the successful Martian mission MGS (Martian Global Surveyor) and its atmospheric instrument TES (Thermal Emission Spectrometer) [Smith, 2002]. SOFIA observations of the water vapour cycle correlated with the dust cycle (as derived from TES measurements) could be introduced in the model [Richardson et al., 2002; Newman et al., 2002], adding an important effect into the radiative transfer module, and hence leading to a more realistic status for our GCM. From the comparison between simulations and observations important parameters could be retrieved, such as the dust agglomerates size and shape, and also the atmospheric region where this water-dust interaction plays a stronger role [Rodin and Wilson, 2003]. Furthermore, the seasonal differences between north and south pole solar radiation (related to the orbit eccentricity and planet obliquity) could ultimately explain the present hemispheric asymmetry in water vapour abundance present in Mars [Hartogh, 1998; Tokano, 2003]. This linkage could be of relevant use for the study of the Martian past climate changes through the link with the evolution of the orbital parameters, which at the moment can be estimated with high accuracy [Laskar, 2004].

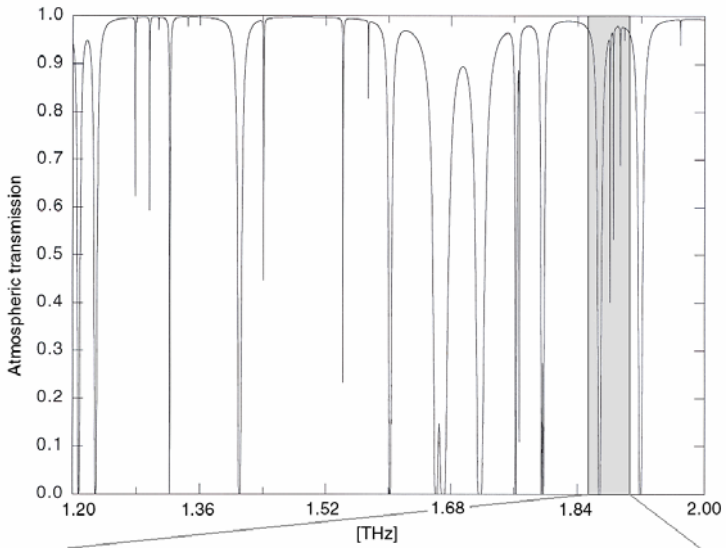

Figure 1.1: Water vapour lines and their contribution to the atmospheric transmission, considering a flying altitude of 14 km and no tropospheric water vapour content. The selected range is considers the low frequency detector on GREAT [Jarchow, 2003, priv. comm.].

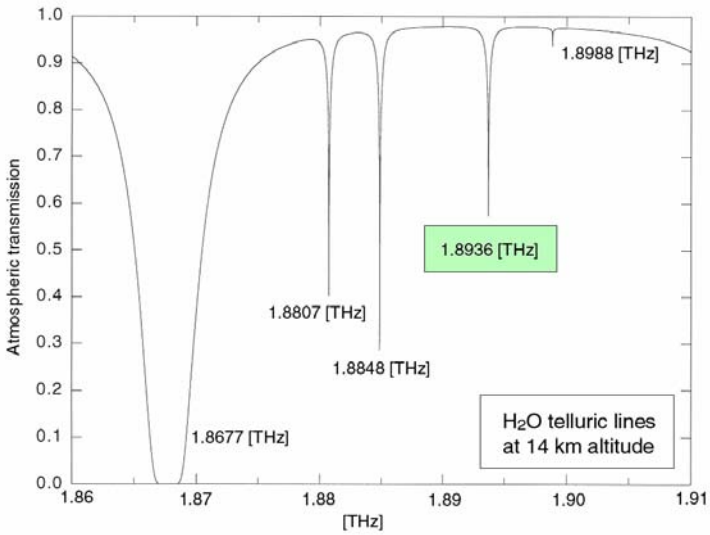

Figure 1.2: The line 1.8936 [THz] is selected since this spectral line is relatively strong and at 14 km the atmospheric transmission is relatively high (60%). Even at 14 km, only 1% total water vapour column totally prevents transmission at 1.8677 [THz]. [Jarchow, 2003, priv. comm.].

The rotational line of water vapour at 1.893 THz (158.3 μm) is a good choice for our extra-terrestrial water vapour observations, since this spectral line is relatively strong and at 14 km the absorption of the telluric water vapour is only 40%. Figure 1.1 shows a selected range of water lines for the lower frequency receiver at GREAT. Observations in this spectral region provide low receiver temperatures (~1500 K) and reasonable spatial resolution (15 arcsec), that in optimum Earth-Mars orbital situations would allow us to obtain a 7 point Martian map of water vapour. From the observed line, the water vapour content profile can be retrieved with high altitude resolution up to 80 km. The inversion of the spectra is based on a forward planetary spectral lines synthesis module and a minimization method [Jarchow, 1999]. The minimization method (Optimal Estimation Method) needs a radiative transfer model which relates the temperature profile with the corresponding measured spectrum. The relation between the z-coordinate (profile) and the f-coordinate (spectrum) is available due to the line-broadening effect; associating a region of the measured spectra with a particular height region. These relating functions are normally called kernel-functions of the inversion method.

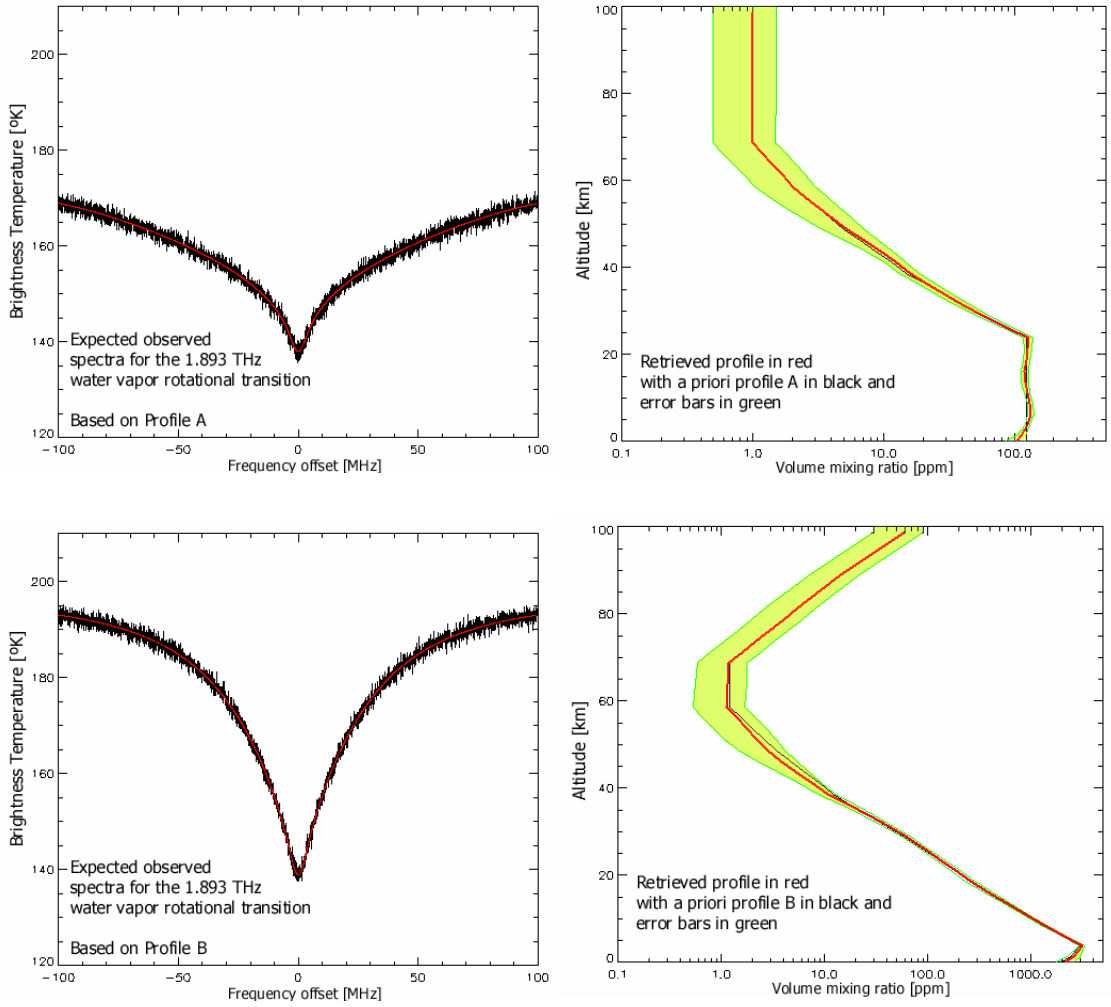

Figure 1.3: [Left] Simulated spectral lines for the rotational transition of the water vapour molecule at 1.8936 THz, based on a modelled Martian atmosphere. Fitted line in red. [Right] Possible retrieved profiles of water vapour. The retrieved profile in red, in black the a priori profile and the error bars in green [Jarchow, 2003, private communication].

The rotational water vapour line at 1.8936 THz shown in Figure 1.2 is then used to analyze what we expect from a SOFIA observation of this spectral line on Mars. As there is not a deep knowledge of the water vapour on Mars, the spectral lines were synthesized assuming two a-priori water vapour volume mixing ratio profiles. These a-priori concentration profiles were obtained from a photochemical model of the Martian atmosphere [Nair et al., 1994]. Gaussian noise was later added to the synthesized spectral lines, considering the transmission and sensitivity achievable with SOFIA-GREAT-CTS at this frequency and considering a zenith observation. The simulated CTS spectral lines are then inverted and the volume mixing ratio is retrieved. In Figure 1.3, it can be seen the different a-priori profiles for the water vapour column and the comparison between the a-priori and the retrieved profiles. These results are encouraging; and furthermore we can extrapolate this analysis and interpretation to other important atmospheric constituents, such as is the case of HDO constituents. The HDO concentration is a key element in order to understand the possible loss of water that could have escaped from Mars to space in the past, and therefore have a deeper knowledge of Mars' past climate [Hartogh, 1998; Bjorker, 1989; Mumma et al., 2003].
In order to address such motivating research topics, we studied optimum observations methods for planetary sciences, taking into account the characteristics of the CTS spectrometer as part of the SOFIA observatory. These methods are later described in this thesis, including beam filling factor and mapping methods, calculation of the spectral line strength, the expected signal to noise ratio, calibration methods, optimum observation modes, etc.

1.2.2 Cometary research

The study of comets through infrared observations is of great value since comets are relatively cold bodies, and therefore exhibit an emission maximum at long wavelengths. At a distance of 1 AU from the Sun, their equilibrium temperature is around 300 K (Planck's radiation law maximum at 10 μm) and 150 K at 4 AU (20 μm). The strong interest in comets lies in the principle that they are travellers and collectors of our ancient times, and can provide information about the origin of our Solar System. Their cold cometary nuclei have been stable for most of the last 4.6 Gyr [Crovisier & Encrenaz, 2000], where each layer of its constitution describes a unique stage of its history and its path. During the sublimation process that takes place when the comets approach the sun, the sublimed volatile species provide evidences about the physicochemical conditions at the formation of the ice layer, and from this we can infer information of past times and distant places.
Figure 1.4 describes the main physicochemical phenomena taking place in a comet, such as the formation of the cometary tail and its chemical composition. The mother products initially in the cometary nuclei are degraded into daughter products, due to the photodissociation process driven by high energetic photons coming from the sun. The sublimation process taking place in the nuclei forms a layer of gas, that together with lifted off dust particles covers the comet's nuclei, forming the coma, and consequently hiding to the external viewer the observation of the deep nuclei material. With SOFIA-GREAT-CTS a complex spectral study can be carried out, meaning not only the detection of parent and daughter products (and therefore the non-volatile and volatile abundance ratios), but also a precise measurement of the line shapes, and from them the retrieval of out-gassing asymmetries, gas temperature, collision rates, structure and size of the dust agglomerates and determination of the cometary nuclei shape and size [Hartogh et al., 2001; Rauer et al., 1998]. Infrared spectroscopy on comets has a successful heritage on airborne

observatories, as the first water vapour detection on a comet was obtained with the SOFIA predecessor, the KAO observatory (on comet Halley by Weaver et al. [1986] and on comet Wilson by Larson et al. [1989]). Later in this thesis, observations of HCN and H_2CO performed on comet Encke with the SOFIA-GREAT-CTS prototype on HHSMT are discussed and shown.

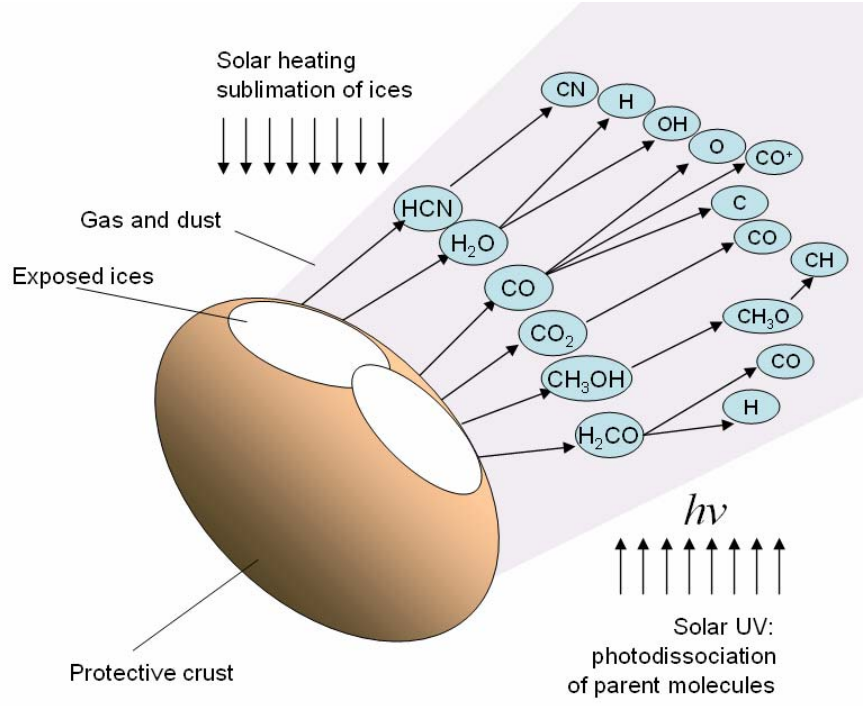

Figure 1.4: Description of the physicochemical phenomena taking place in a comet. The solar heating sublimes the ice, and from this gas, the tail of the comet is created. The low density in the tail enhances the photodissociation of parent molecules, forming secondary products (radicals, ions and atoms). Sketch based on [Bockelée-Morvan and Crovisier, 1994].

1.3 The measuring principle

There are fundamentally two different methods of detection: coherent detection (or heterodyne, provided with GREAT) and incoherent detection. With incoherent detectors spectroscopy can be obtained through the use of pre-detection spectral filtering, for instance via a Fabry-Perot interferometer or a diffraction grating. In the case of coherent detectors, the complete frequency at the observed wavelength is converted by mixing the signal with a local oscillator single frequency signal to a much lower center frequency. This signal can be amplified and spectrally analyzed without further degradation in S/N. The spectral resolution is constrained by the stability and purity of the local oscillator and the backend. In the case of SOFIA-GREAT-CTS, the spectral resolving power is approximately 10^8 at 4.7 THz.

In comparison to incoherent detection, coherent instruments offer important advantages such as high resolution spectroscopy. However, coherent instruments are limited in sensitivity as they preserve information about the amplitude and the phase. Since both quantities are non-commuting according to quantum mechanics, there is a corresponding uncertainty and therefore a "quantum limit" [Zmuidzinas, 2002]. In thermally dominated regimes kT >> hν, where the thermal background noise is comparably bigger than the quantum noise, the "quantum limit" is less significant and both detection methods are comparable in sensitivity. In addition, since the spectral differentiation is performed prior detection in incoherent detectors, the spectrometer must be cold in order to avoid additional thermal noise, and therefore this severely limits this type of detection.

From the technological point of view, the incoherent detectors are much simpler to build and thus have been used already for quite a number of years in 2-dimensional array configurations, giving a spatial multiplexing advantage for incoherent detectors. On the other hand, spectral multiplexing is much more simple to achieve, and is traditionally quoted as an advantage for this type of receiver.

The coherent detection principle can be explained as the multiplication of the incoming signal with a source signal at a comparable frequency range, where the detection process relies in their difference. The speed of the detector in being able to detect differences will define the output bandwidth. Consider the case of a monochromatic signal arriving at the SOFIA telescope, represented as a potential U:

$$U(t) = E\cos(2\pi\nu t + \phi) \qquad (1.1)$$

where E is the amplitude of the incoming signal, ν the frequency and ϕ is the phase at time t=0. In order to perform the detection, source signal with a "Local Oscillator" (LO) is created, represented by:

$$U_{LO}(t) = Q\cos(2\pi\nu_{LO} t + \phi_{LO}) \qquad (1.2)$$

where Q is the amplitude of the LO's output, ν_{LO} is the LO's frequency and ϕ_{LO} is the initial phase. This signal is generated in GREAT by the use of a BWO, a solid-state LO, or a FIR-Laser LO (see appendix A). These two signals are then sent into a non-linear device, called "mixer", that in the case of GREAT is a HEB bolometer (see apendix B).

$$I(t) = a_0 + a_1(U(t)+U_{LO}(t)) + a_2(U(t)+U_{LO}(t))^2 + a_3(U(t)+U_{LO}(t))^3 + \ldots \qquad (1.3)$$

The output of the mixing process can be represented then by a series (1.3). In the case of heterodyne detectors, the mixers are developed in order to enhance the second order term of its response; and thus developing the second order component of the (1.3) series, we would obtain the following

$$\begin{aligned} I(t) = {} & a_2E^2\cos^2(2\pi\nu t + \phi) \\ & + 2a_2EQ\cos(2\pi\nu t + \phi)\cos(2\pi\nu_{LO} t + \phi_{LO}) \\ & + a_2Q^2\cos^2(2\pi\nu_{LO} t + \phi_{LO}) \end{aligned} \qquad \begin{aligned} I(t) = {} & a_2E^2\cos^2(2\pi\nu t + \phi) \\ & + a_2EQ\cos(2\pi(\nu+\nu_{LO})t + \phi + \phi_{LO}) \\ & + a_2EQ\cos(2\pi(\nu-\nu_{LO})t + \phi + \phi_{LO}) \\ & + a_2Q^2\cos^2(2\pi\nu_{LO} t + \phi_{LO}) \end{aligned} \qquad (1.4)$$

It can be seen from the 2nd and 3rd terms in equation (1.4) that the incoming radiation is shifted in the frequency domain ($\nu - \nu_{LO}$, $\nu + \nu_{LO}$). By inserting a filter at the output of the mixer, only the difference term (3rd) is selected:

$$\nu_{IF} - \frac{\Delta\nu}{2} \le \left|\nu - \nu_{LO}\right| \le \nu_{IF} + \frac{\Delta\nu}{2} \qquad (1.5)$$

and ν_{IF} is the so-called "Intermediate Frequency" (IF) where the detected radiation is now spectrally located. Hence, after mixing and filtering, the output of the receiver is

$$U_{IF}(t) = a_2EQ\cos(2\pi(\nu - \nu_{LO})t + \phi - \phi_{LO}) \qquad (1.6)$$

The frequency change, usually towards a lower frequency, allows the backend instruments to analyze the incoming radiation at lower frequencies, where the signal processing, such as amplification and filtering is simpler.
The three most common backend technologies besides the CTS are: the filterbanks, the AOS and the autocorrelators. The filterbanks separate the different spectral components by a set of filters tuned at the corresponding frequency. The spectral power is measured by a quadratic detector and is furthermore integrated in order to reduce the data rate.
On an Acousto-Optical Spectrometer (AOS), the diffractive properties of a crystal (the so called Bragg-cell) are used to separate in space the different spectral components. This type of spectrometers are normally defined as wideband spectrometers, covering up to 3 GHz of spectral bandwidth and obtaining reasonable spectral resolution, being 1 MHz for the AOS instruments as part of SOFIA-GREAT [Sievertz, 2003]. Because of its relatively simple design and remarkable performance, it has become a standard tool in radioastronomy for spectroscopic observations. A piezoelectric transducer, driven by the RF-signal (the input signal coming from the receiver), generates an acoustic wave in the Bragg-cell, which then modulates the refractive index and induces a phase grating. The Bragg-cell is illuminated by a collimated laser beam. The angular dispersion of the diffracted light represents a spatial image of the RF-spectrum according to the amplitude and wavelengths of the acoustic waves in the crystal. The spectrum is detected by using a single linear CCD (charged-coupled device) array, which is placed in the focal plane of the imaging optics. This one dimensional array provides the spectral information of the incoming radiation, as the spatial deviation performed by the Bragg-cell into the laser ray path can be extrapolated into spectral information.
In the case of the autocorrelation spectrometers, their principle lies on the Wiener-Khinchin theorem: the Fourier transform of the autocorrelation of a stationary random variable gives an estimator of its power spectrum. The delay and multiplication of the signal can be done by analog methods using acousto-optical devices, such as the Analog AutoCorrelator Spectrometer (AACS) part of SOFIA, or by digital operations being the case of the Digital AutoCorrelator Spectrometer (DACS) part of HIFI. One major advantage of this type of spectrometer is their flexibility in the definition of the operational bandwidth and resolution. Nevertheless, they require of comparable high power consumption when processing very wide bandwidths [Emrich, 1996]. Furthermore, limitations in the computational power produce a data sampling with very coarse accuracy, only 1 or two bits. This implies a degradation of the signal to noise ratio which for a 2-bit Autocorrelator, represents a factor of 0.88. With the up coming of faster correlator processors, this data sampling limit can be taken to much higher values, which already for a 3 bit quantization scheme would produce a 0.95 signal to noise quality factor.

1.4 The SOFIA-GREAT Instrument

1.4.1 Instrument configuration

The heterodyne instrument GREAT will provide high resolution spectroscopy at THz frequencies based on a modular dual-channel heterodyne instrument. A consortium of German research laboratories between the Max-Planck-Institut für Radioastronomie (MPIfR), the Max-Planck-Institut für Sonnensystemforschung (MPS), the Physikalisches Institut der Universität zu Köln (KOSMA) and the Deutsches Zentrum für Luft- und Raumfahrt (DLR) was established for

the development of the instrument. GREAT is planned to take place in the first astronomical mission, scheduled for 2005, with a first-generation set of instruments. The instrument is integrated by different modules, separated between the frontend modules (receivers and LOs) and the backend modules (Chirp-Transform-Spectrometers and Acousto-Optic-Spectrometers). The modularity of the instrument design will allow an increasingly more complete coverage of the far-IR spectrum [Güsten et al., 2003]. GREAT in the first flight version will focus on three scientifically-selected spectral ranges:

A low-frequency band: 1.6–1.9 THz (188-158 μm), covering – among other lines – the rotational water line at 1.893 THz and the important atomic fine-structure transition of ionized carbon at 1.9 THz. The KOSMA institute is in charge of the development of this module, which uses a diffusion-cooled HEB as mixer and a BWO oscillator in order to perform the heterodyne detection.

A mid-frequency band: centered on the cosmologically relevant 1-0 transition of deuterated molecular hydrogen (HD) at 2.6 THz (115 μm) and the rotational ground-state transition of OH. The MPIfR Institute, also responsible for the overall GREAT project management, is developing this module based on a phonon-cooled HEB as a mixer. A Solid-State Local Oscillator Unit is used as heterodyne reference.

A high-frequency band: located in the boundary of the far-IR region, next to the middle-IR, targeting the 4.7 THz (63 μm) fine-structure transition of atomic oxygen. The German Aerospace agency (DLR) is developing a phonon-cooled HEB at this short wavelength, breaking many technological constraints, becoming the first operational heterodyne detector in this spectral range. The local oscillator is based on an optically pumped FIR ring laser.

Elliptical mirrors and lenses in front of the mixer elements match the antenna-characteristics with those of the telescope. For the injection of the LO-signal, a Martin-Pupplet interferometer or a semitransparent foil is used as the diplexer [van der Wal, 2003]. In the first flight configuration, as shown in Figure 1.5, the receivers are only sensitive to one polarization, which is why the separation between the two receivers is performed via a polarization splitter. However, it is planned in the future to provide a double polarization detection in each receiver unit and hence the splitter will be replaced by an optical filter system. This will double the detected information with 4 channels, meaning two per polarization and two per receiver. The modularity of the system allows a rapid change between frequency bands (receivers and LO units) without realigning the optical setup, which gives the flexibility of having different frequency bands in various combinations available for observations. In addition, based on new developments on HEB detectors, it is expected to reach up to 10 THz (30 μm) in the near future and this could be easily mounted on GREAT.
The sensitivity of the receivers at these high frequencies is importantly constrained by the quantum noise (hν) and the low efficiency in the antenna-mixer coupling. The overall receiver temperature (T_{rec}) can be then expressed by:

$$T_{rec} = T_{Back} + \frac{h\nu}{\eta k} \qquad (1.7)$$

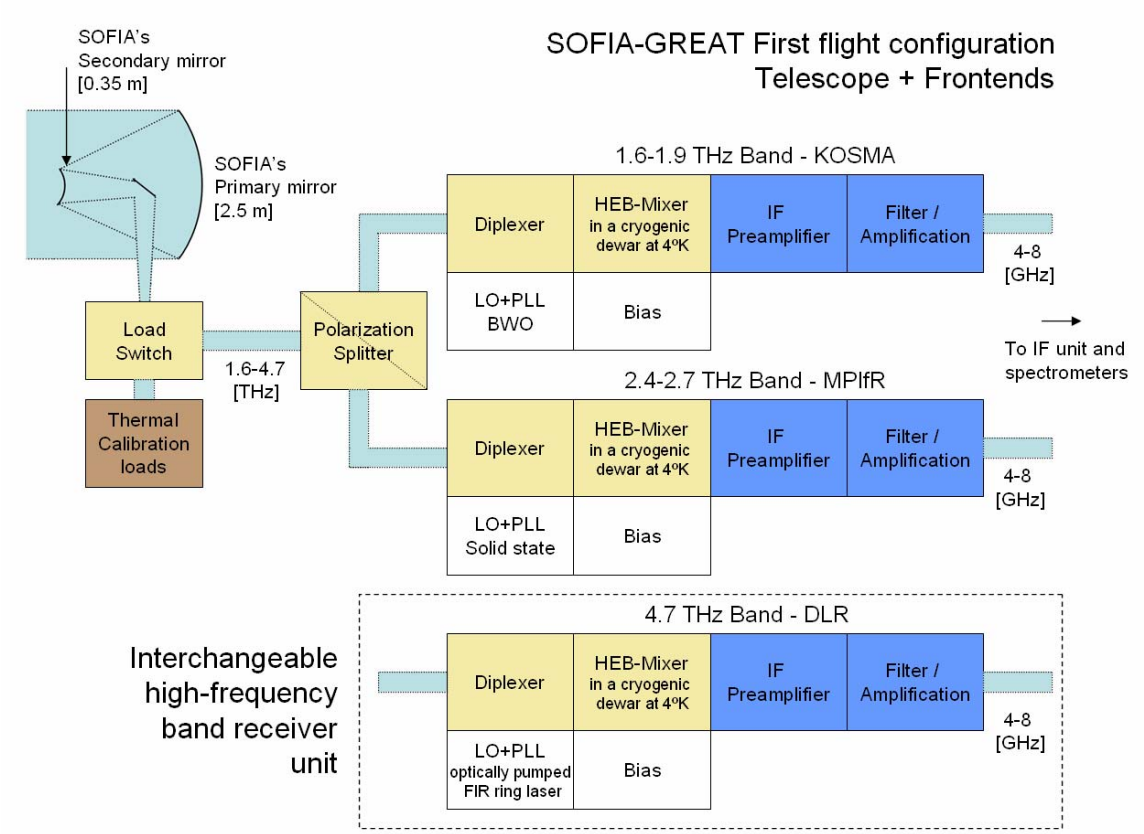

Figure 1.5 This block diagram shows the layout of the frontend components. High frequency signals from the telescope and the signal references are fed via pre-optics, such as the polarization splitter and the diplexers for local oscillator injection, to the mixers. The mixers together with the first low noise amplifiers are mounted inside a cryostat.

where T_{back} represents the induced background noise temperature, associated to electronics, operation temperature, etc; and the detectors efficiency η, describing the amount of photons needed in order to perform one detection, which for THz frequencies ranges from 0.05 to 0.4 [Schubert et al., 1999]. Based on preliminary measurements, a double-sideband receiver temperature of 1500 ºK is expected for the low-frequency band receiver [Jacobs, 2003, priv. comm.], 1600ºK for the mid-frequency band receiver [Heyminck, 2003, priv. comm.] and 5800º for the high-frequency band receiver [Hübers, 2003, priv. comm.].

One interesting aspect of the configuration of the backend spectrometers in SOFIA is that they will share different frontend receivers. An agreement was obtained for the common use of spectrometers between GREAT and the American receiver CASIMIR (CAltech Submillimeter Interstellar Medium Investigations Receiver). CASIMIR is a submillimeter heterodyne receiver covering the spectral range between 500 GHz (0.6 mm) and 2000 GHz (150 μm) [Zmuidzinas & Edgar, 1999]. As part of the CASIMIR facility instrument, a wideband Analog Autocorrelator Spectrometer (AACS) will provide wideband but lower resolution observations. GREAT will provide high-resolution spectroscopy with the CTS backend instruments and wideband spectral coverage with the AOS spectrometers.

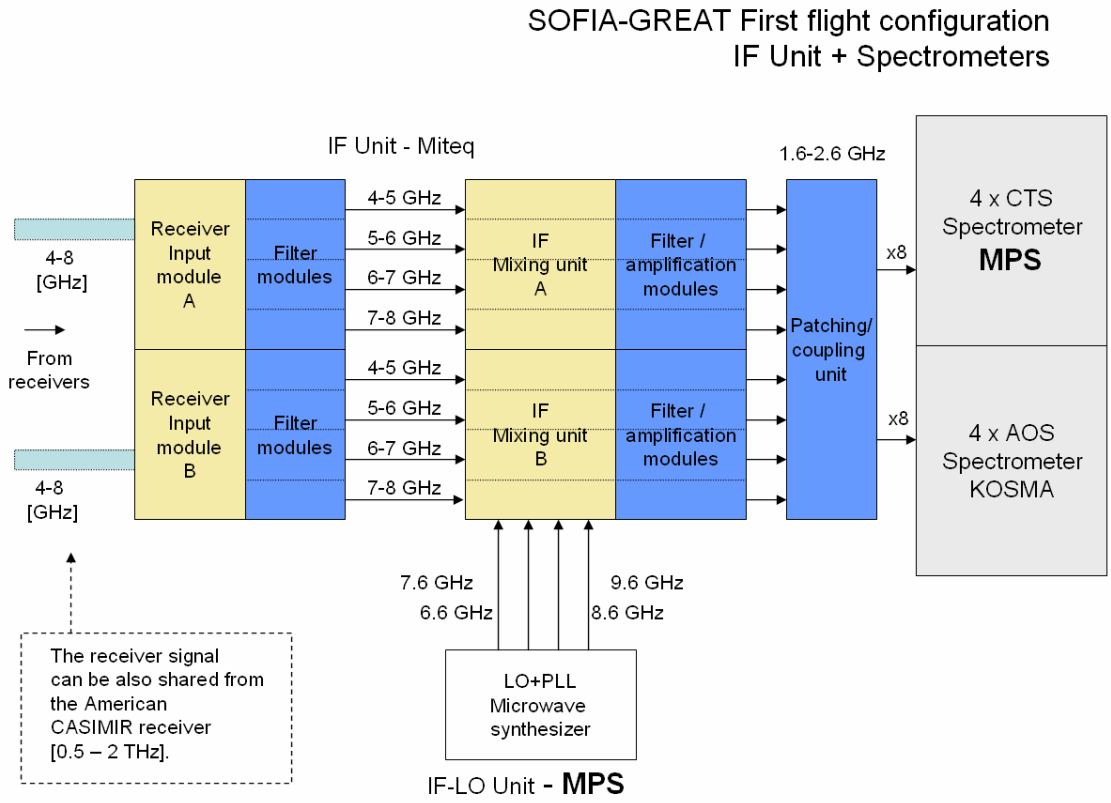

Figure 1.6: Scheme showing the IF-unit processing unit and the backend instruments layout. The detected signal coming from the receivers is pre-processed and down-converted to the center frequency of the spectrometers via the IF-LO unit. This signal is then distributed by the patching / coupling unit to the backend instruments.

The IF-unit which serves the spectrometers splits the 4-8 GHz mixer frequency band into four sub-bands, which are later down-converted to the input AOS/CTS band centered at 2.1 GHz. An interconnection scheme allows the distribution of the detected signals from the mixer sub-bands between a maximum of 8xAOS units and 8xCTS units. In the first configuration flight, the patching and coupling unit will allow one to dynamically select the connection between the frontend and backend instruments, while in the full configuration, each receiver of 2x[4-8] GHz will be completely covered by the backend units.

Backend-Spectrometer SOFIA-GREAT/CASIMIR	**Spectral Resolution [kHz]**	**Instantaneous Bandwidth [MHz]**
High-resolution spectroscopy CTS – Chirp Transform Spectrometer (MPS)	50	4 x 210 = 840
Wideband spectroscopy AOS – Acousto Optical Spectrometer (KOSMA)	1000	4 x 1100 = 4400
Wideband spectroscopy AACS – Analog Autocorrelator Spectrometer (CalTech)	16000	4 x 4000 = 16000

There will also exist an overlap between the spectral coverage of the different AOS backend instruments, since the instantaneous bandwidth is 1.1 GHz (1.55-2.65 GHz) and the band spacing is 1 GHz at 1.6-2.6 GHz. Furthermore, since the spectral coverage of the CTS is 200 MHz compared to 1 GHz of the down-converted signal, the IF-LO unit [Römer, Villanueva & Clement, 2003] allows a tunable selection of the mixing frequency and hence a displacement of the 200 MHz window across the down-converted signal. The IF-LO unit serves 2 times four LO channels: 6.6 ± 0.4 GHz for the 4-5 GHz band, 7.6 ± 0.4 GHz for the 5-6 GHz band, 8.6 ± 0.4 GHz for the 6-7 GHz band and 9.6 ± 0.4 GHz for the 7-8 GHz band. The oscillators are based on the phased-locked-loop (PLL) principle, obtaining high frequency stability (of the order 10^{-9}). This high frequency stability is needed because of the high resolution achieved by the GREAT backends (specially the CTS).

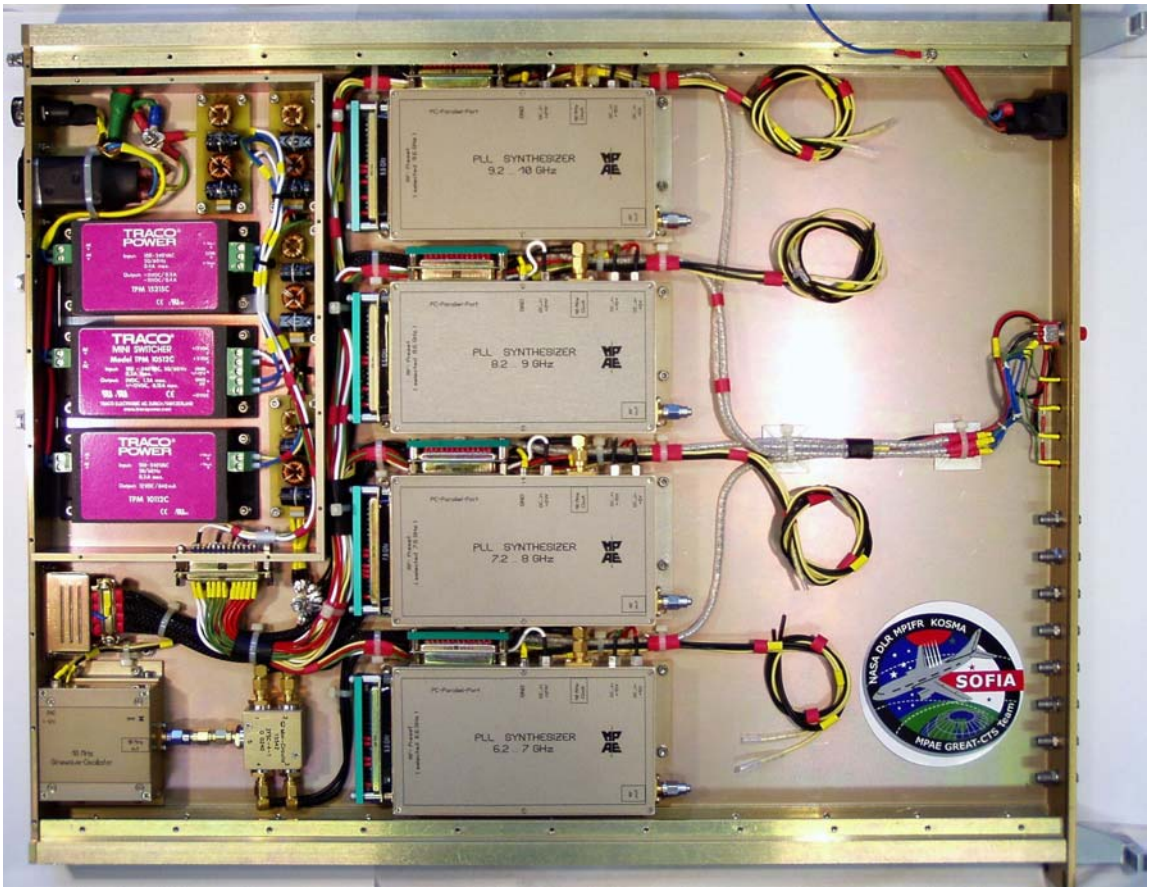

Figure 1.7: IF-LO unit developed at the Max-Planck-Insitut für Sonnensystemforschung. The PLL oscillators and the 10 MHz sinewave source are fed with a filtered energy supply from the power units. The output of the sinewave source is split 4 times and delivered to the independent LO units which are remotely programmed through a parallel interface.

Chapter 2

CTS – Chirp Transform Spectrometer

The unique opportunity provided by the SOFIA mission was an important driver for our study and development of new technologies. The culminating result was the development of a new CTS with notable performance, which is based on a proper combination of modeling, simulation and laboratory experimentation. The new instrument is improved through a dispersive matching system, represented by the "Adaptive Digital Chirp Processor (ADCP)", producing an increase of the spectrometer's dynamic range and frequency resolution. In this chapter, I will initially discuss the different blocks involved in the spectrometer and provide an overall view of their interrelation. I focus mainly on the ADCP technology, and its realization through digital signal processing, quadrature-modulation and compact spectral filters. Furthermore, I discuss the study of the optimum dispersive matching and its implementation combined with frequency multiplication and the RF analysis. The instrument's structural and thermal stability was studied in an attempt to increase the isolation of the instrument to external fluctuations (high accelerations and fluctuating environmental temperature), and thus improve its overall stability.

2.1 Principle of function

The technique of expanding and compressing signals started with the need to enhance the performance of radar, which needs a very narrow signal in time with high energy in order to obtain high spatial resolution and long distance coverage. Nevertheless, high energy in very short times requires enormous amount of instantaneous power ($P = E / t$). For that aim, dispersive devices were used to provide expanded and compressed signals, obtaining high power in short time. In order to achieve this instantaneous power, a longer signal with time t_{long} ($P_{long} = E / t_{long}$) is required. This is then compressed in time through a time dispersive element, obtaining a new compressed signal with time t_{short} ($P_{short} = E / t_{short}$). The energy delivered is the same, the difference lies in the instantaneous power.

The dispersive properties of a SAW filter are extremely convenient in performing this compression or expansion, since they delay the signal in time according to its frequency. Surface acoustic waves are modes of propagation of elastic energy along the surface of a solid. These

waves can be tapped and controlled along the propagation path [Slobodnik, 1976]. In addition, these dispersive properties can be arranged and coded, and can be used as an identification technique [e.g. Reindl et al., 2003; Reindl, 2002]. By means of a piezoelectric transducer, the incoming electrical energy is transformed into an ultrasonic surface traveling wave. The transducers are represented by an interdigital comb structure on the surface of a piezoelectric substrate (for example Lithium-niobate), that when an electrical voltage is applied to the comb terminations, the electrical field between the fingers causes the piezoelectric material to deform. This induced mechanical distortion is then propagated through the material by surface waves at a speed of about 3000-3500 m/s ≈ $10^{-5}v_{light}$ [Maines & Paige, 1976]. This low speed compared to the electromagnetic speed of light allows the performing of controlled phase delays of the incoming electrical signal. So this means $t_{out} = t_{in} + \mu^{-1}(f_{in}-f_o)$, being μ [Hz/s] the dispersive constant, t_{in} the input time, t_{out} the output time, f_{in} the input frequency and f_0 the frequency at $t_{in}=t_{out}$. It can be seen that if one provides an impulse (all frequencies in one instant) to a SAW filter, a signal that evolves in time increasing its frequency is obtained. This means that the impulsive response of a SAW device is a chirp signal, or a linear frequency modulated signal ($e^{j\pi\mu t^2}$). However, if a linear frequency modulated signal generated with –μ dispersive constant (complex conjugate transfer function) is provided, one obtains the autocorrelation of the chirp signal which is a sinc function or a compressed signal (all the energy of the component frequencies in one instantaneous moment).

The structure of the sinc function, that is, the location of the zero-crossings, the position of the main peak, etc., are defined by the properties of the initial chirp signal and the device with the complex conjugate transfer function. These two elements involved in the autocorrelation are the core of the analog Fourier Transform. The device in charge of producing the chirp signal is defined as the "expander", because it produces from a short impulse signal a time expanded signal. The "compressor", the device with the chirp complex conjugate transfer function, compresses the incoming signal as its dispersive characteristic is inverted. The resulting signal can be analyzed performing the inverse Fourier transform of a flat spectral signal with a bandwidth B:

$$f(t) = \frac{1}{B}\int_{-B/2}^{B/2} 1\cdot e^{j2\pi ft}df = \frac{e^{j\pi Bt} - e^{-j\pi Bt}}{j2\pi Bt} = Sinc(\pi Bt) \qquad (2.1)$$

where B represents the common bandwidth between the expander and the compressor. The formula above allows us to understand in more detail the structure of the response of the interaction between the "compressor" and the "expander". One important parameter is the first crossing by zero of this response.

$$f(t_0) = Sinc(\pi Bt_o) = 0 \;\Rightarrow\; \pi Bt_o = \pi \;\Rightarrow t_o = \frac{1}{B} \qquad (2.2)$$

Equation (2.2) dictates that the first crossing by zero depends on the amount of spectral components (B) compressed by the "compressor", and the width of the main lobe of the sinc response tells us about the "time-resolution" of the autocorrelation.

As sketched in Figure 2.1, by the proper arrangement of the two dispersive elements, the spectral information of the input signal is translated into the time domain, where the parameter Δt describes the relative delay of an spectral component (Δf) after the transformation. Considering the case of the expander and the compressor signals having matching dispersive properties, it can be demonstrated [see Hartogh, 1990] that a displacement Δf in the expander signal will produce a delay Δt in the time domain:

$$\Delta t = k + \mu\Delta f \qquad (2.3)$$

where k=B/μ and is a constant delay related to the dispersive elements and not to the input signal. The possibility of measuring a signal in time will ultimately give frequency information of the external source, and therefore this is the principle behind the chirp transform spectrometer. In order to perform this spectral shift of the expanded signal a mixing process is used. The multiplication of the external source with the chirp signal will produce two sidebands (as was described in the coherent detection), meaning that two chirp signals are obtained, one following the initial spectral order, the upper sideband (USB, $f_{USB} = f_{ext} + f_{chirp}$), and the other with the inverted spectral order, the lower sideband (LSB, $f_{LSB} = f_{ext} - f_{chirp}$). This allows the use of two dispersive elements with the same dispersive properties, but the LSB is provided to the compressor, the LSB having the inverted spectral components.

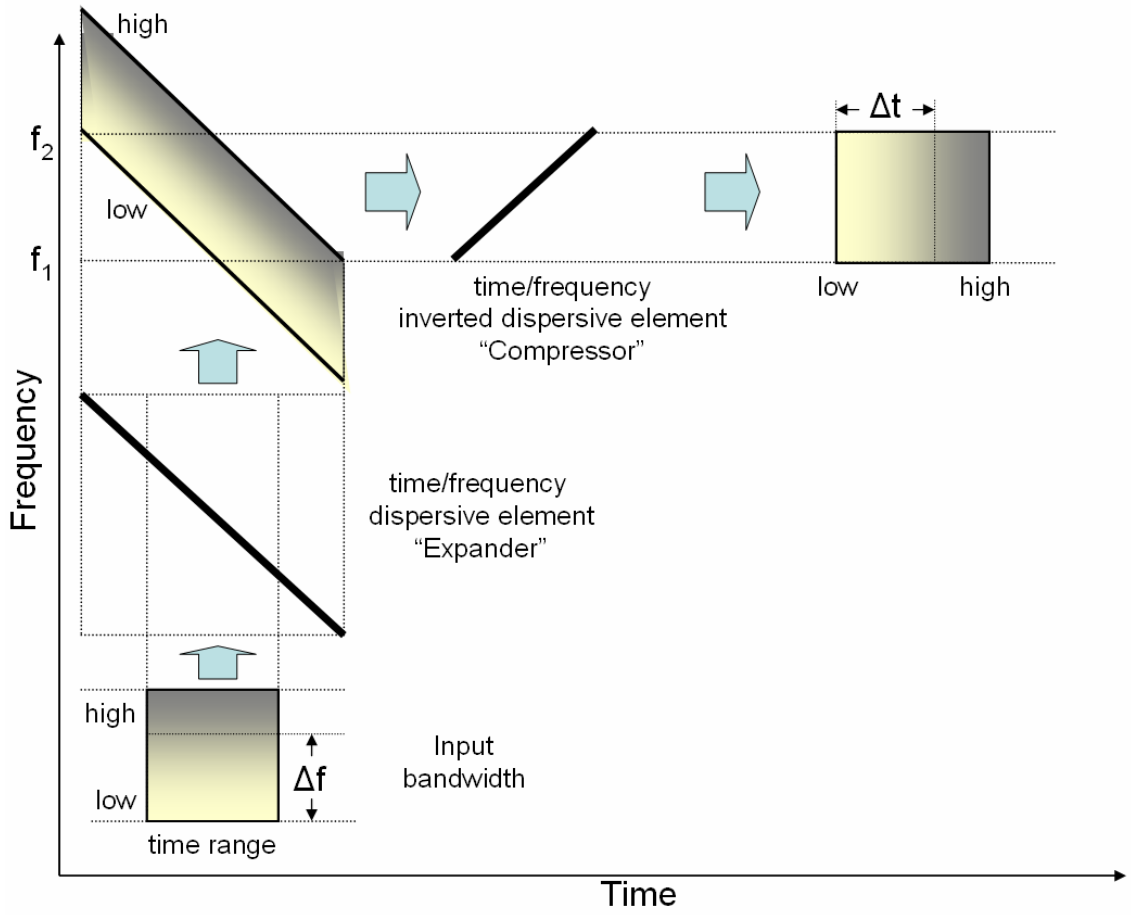

Figure 2.1: Sketch describing the principle of the chirp transform spectrometer. The input signal shown in the lower-left corner is mixed with a time/frequency dispersive element (the expander) and is later compressed by a second dispersive element (the compressor). The output of the setup is a signal where its time distribution is related to the frequency distribution of the input signal.

2.1.1 The Chirp Transform

The transform that takes place in a Chirp Transform Spectrometer is based on an analog frequency-time Fourier Transform. The chirp signal, a linear frequency modulated signal, is expressed also as a quadratic phase modulated signal:

$$f_{chirp}(t) = e^{-j\pi\mu t^2} \qquad (2.4)$$

In order to simplify the mathematical analysis, the chirp signal is centered to the null frequency and therefore the carrier component is not described. If this chirp signal, the impulsive response of a SAW filter, is then mixed with an arbitrary external signal f(t), the mixing product $f_{mix}(t)$ would be obtained:

$$f_{mix}(t) = f(t)e^{-j\pi\mu t^2} \quad (2.5)$$

It is known that the interaction of a signal with a linear system can be represented by a convolution of the impulsive response h(t) of the system with the signal:

$$g(\tau) = \int_{-\infty}^{\infty} f(t)h(\tau - t)dt \quad (2.6)$$

The components of equation (2.6) are analyzed considering that f(t) is the mixing product between the external source and the chirp signal coming from the expander ($f_{mix}(t)$), while h(t) is the impulsive response of the "compressor" (actually a chirp signal but with inverted dispersive properties), which means:

$$f_{comp}(\tau) = \int_{-\infty}^{\infty} f_{mix}(t)e^{-j\pi\mu_c(\tau-t)^2}dt \quad (2.7)$$

$$f_{comp}(\tau) = \int_{-\infty}^{\infty} f(t)e^{-j\pi\mu_e t^2}e^{-j\pi\mu_c(\tau-t)^2}dt \quad (2.8)$$

Equation (2.8) describes the output of the compressor and it provides the power spectral information of the signal f(t). In order to retrieve the complete information of f(t), i.e. spectral power and phase information, a phase factor $e^{-j\pi\mu\tau^2}$ should be added to the transform which is also a chirp signal, and is multiplied to the signal from the compressor. Considering a perfect matching between the expander and compressor $\mu=\mu_e=-\mu_c$, then:

$$f_{CTS}(\tau) = e^{-j\pi\mu\tau^2}\int_{-\infty}^{\infty} f(t)\cdot e^{-j\pi\mu t^2}\cdot e^{j\pi\mu(\tau-t)^2}dt \quad (2.9)$$

$$f_{CTS}(\tau) = \int_{-\infty}^{\infty} f(t)\cdot e^{j\pi\mu\left[-\tau^2-t^2+\tau^2-2t\tau+t^2\right]}dt \quad (2.10)$$

$$f_{CTS}(\tau) = \int_{-\infty}^{\infty} f(t)\cdot e^{j\pi\mu[-2t\tau]}dt \quad (2.11)$$

Based on equation (2.3) which linearly relates time and frequency, and comparing to the result of equation (2.11), it can be seen that the CTS output time domain τ is associated to the spectral domain of the external source. This would validate that the chirp transform spectrometer transfer function is a Fourier transform.

$$f_{CTS}(\tau) = \int_{-\infty}^{\infty} f(t)\cdot e^{-j2\pi ft}dt \quad (2.12)$$

$$F(f) \equiv \int_{-\infty}^{\infty} f(t)\cdot e^{-j2\pi ft}dt \quad (2.13)$$

2.1.2 The expander and compressor arrangements

The specifications of the chirp-transform-spectrometer, such as the spectral resolution and the spectral coverage, are defined by the properties of the involved dispersive elements and the expander-compressor arrangement. Two alternative arrangements of expander-compressor scheme must be distinguished; one in which the time duration of the premultiplier chirp signal (the expander chirp) is short compared to the impulsive response of the convolution filter (the compressor): the defined M(s)-C(l) arrangement, and the one in which the premultiplying chirp duration is long: the M(l)-C(s) arrangement [Hartogh, 1990; Jack et al., 1980]. The notation used is M denoting multiplication, C convolution, (l) long-duration chirp and (s) short duration chirp.
The expander SAW device would have a bandwidth B_e and a chirp-length T_e while the compressor B_c and T_c. However, the chirp rate is defined to be matched and therefore $\mu=B_e/T_e=-B_c/T_c$. As defined in equation (2.2), the width of the out coming sinc function is related to the common bandwidth between the compressor and the expander.
In the case of M(s)-C(l) as shown in Figure 2.2, it is clear that the common bandwidth is B_e, and therefore the first crossing by zero of the sinc function is at $1/B_e$. This gives the spectral resolution of the arrangement: $\Delta f = \mu/B_e \rightarrow \Delta f = 1/T_e$, while the number of independent spectral channels is $n = B / \Delta f = (B_e - B_c)\cdot T_e$, which in turn is related to the time bandwidth product of the expander SAW device. In this scheme the properties of the expander define the resolution of the setup, while the properties of the compressor are constrained by the external source bandwidth.

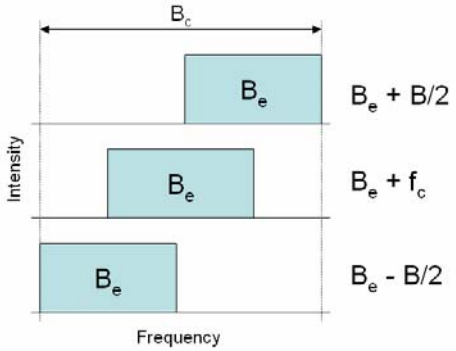

Figure 2.2: In the M(s)-C(l) setup, the bandwidth of the compressor B_c is bigger than the premultiplied chirp signal represented by B_e. The external source, represented by a center frequency fc and a bandwidth BW, spectrally shifts the expander chirp signal within the bandwidth of the compressor. The modification of the center frequency of the chirp-signal through the external source produces changes in the convolved output of the compressor filter.

In the M(l)-C(s) scheme, the expander chirp is longer than the compressor chirp signal. This setup is the one used by the Chirp Transform Spectrometer developed for SOFIA. From the theoretical point of view both arrangements are comparable in performance, but with the M(s)-C(l) the largest time bandwidth product chirp waveform (expander) is used only as a multiplying signal and, in consequence, techniques such as frequency multiplication and longer chirp expansion, can be used to achieve the required waveform parameters from SAW chirp filters of modest time-bandwidth products. The spectral resolution of the M(l)-C(s) arrangement can be defined as $\Delta f = \mu/B_c \rightarrow \Delta f = 1/T_c$.

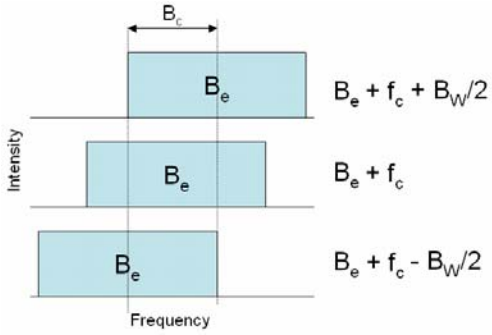

Figure 2.3: In the M(l)-C(s) the compressor bandwidth is smaller than the expander chirp signal. The spectral components affected by the compressor (Bc) are compressed, while the rest is filtered out. The image shows that the common bandwidth between the expander and the compressor (used to retrieve the spectral resolution) is Bc. In this arrangement the expander bandwidth is constrained by the desired spectral coverage.

The fact that both arrangements do not use the entire spectral bandwidth of the expander and the compressor means a reduction in the duty cycle. The duty cycle defines the ratio between the effective time and the total time required for the transformation. As the chirp transform provides a time signal, the duty cycle is defined by a ratio of time values. The effective time defines the duration of the outcoming signal that provides spectral information and it can be written as $t_{eff} = B/\mu$, while the total time is the time of the maximum chirp signal (T_c for M(s)-C(l) and T_e for M(l)-C(s)).

$$\eta_{M(l)-C(s)} = \frac{B}{T_e \mu} = \frac{B}{B_e} \qquad (2.14)$$

$$\eta_{M(s)-C(l)} = \frac{B}{T_c \mu} = \frac{B}{B_c} \qquad (2.15)$$

This ratio affects the real integration time during a spectral measurement, meaning that instead of measuring during time t_{obs} on source, we are measuring $\eta \cdot t_{obs}$. As it is known from the radiometric formula, this would lead to a reduction in the sensitivity.

$$S_{CTS} = \frac{S_{\max}}{\sqrt{\eta}} \qquad (2.16)$$

All possible M(l)-C(s) arrangements represent $\eta < 1$, and therefore the only possibility of improving the sensitivity is to use multiple expander-compressor arrangements running in tandem. In the case of SOFIA-GREAT-CTS with B_e=400 MHz and B=200 MHz → η=0.5, a push-pull configuration of two expander-compressor branches is used, obtaining a duty cycle of η equal to unity.

$$\eta_{CTS} = N \cdot \eta \quad \Rightarrow \quad S_{CTS} = S_{\max} \;,\; N = \frac{1}{\eta} \qquad (2.17)$$

2.2 The digital chirp transform spectrometer

In the arrangement used in SOFIA-GREAT-CTS, the largest time-bandwidth product chirp waveform is a well known signal represented by the complex-conjugate of the impulsive response of the compressor. If we are able to provide a chirp waveform that matches exactly to the compressor properties, we will be able to obtain the maximum from the arrangement, meaning spectral resolution, spectral coverage and linearity over the complete spectral range. In order to approach the theoretical limit, the properties of the SAW filters should be very precise and constant; nevertheless they are strongly sensitive to temperature and micro-structural differences. For that, I have studied and developed a novel idea called “Adaptive Digital Chirp Processor (ADCP)”. The principle behind the ADCP is to incorporate the dispersive properties of the compressor into a digital system that produces the perfect matching expander waveform. The core of the digital system is a direct digital synthesizer (DDS), that controlled through precise commands from a microcontroller generates a digital signal which is later converted to an analog signal. The benefits of the arrangement are various, the main ones being: optimum matching between the expander and the compressor, and a higher dynamic range since the digital expander has a higher signal to noise ratio than the one generated through SAW devices (which suffer high insertion loss).

A DDS system can be simplified as a controlled phase accumulator, which combined with digital to analog converters (DAC), allows production of digitally synthesized analog signals. As is known from Fourier theory, every signal can be represented by an integral if it is not periodic, or by a series if it is periodic. This means that providing a defined phase information of sine-wave harmonics to the DDS will actually produce the creation of arbitrary signals. The controlled output phase is regulated by the phase increment of the phase accumulator and the periodicity of the accumulation, providing a numerically-controlled oscillator which is the core of a highly-flexible DDS device.

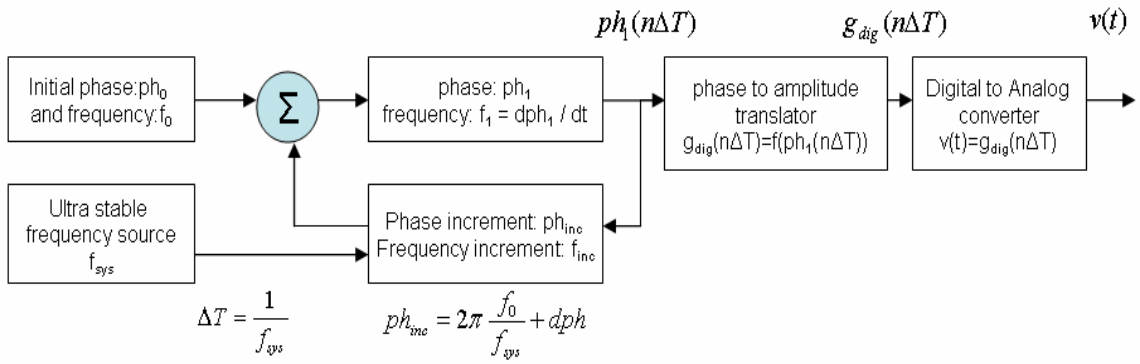

Figure 2.4: Description of the main blocks involved in a DDS system. The system accumulates phase with a defined phase increment parameter (dph), at a rate set by fsys. This phase accumulation process also produces a frequency modification, defined by $f_1=dph_1 / dt$. The resulting phase numerical values are converted to amplitudes and later converted to analog signals with the use of a digital to analog converters.

In Figure 2.4 the description of the main blocks involved are analyzed, where the kernel of the process is actually the phase processing blocks, represented by the sum operator and the numerical phase to analog signal translators. In each system clock, represented by ΔT, the resulting phase ph_1 is incremented with ph_{inc}. The phase incremental is defined by the initial conditions, as f_0, and the differential element dph used to produce arbitrary signals. As we are able to modify the differential element, represented internally in the device as a digital register, we would be able to advance or reward the phase of the resulting signal arbitrarily. This means the capability of producing any type of signal, most importantly an increasing phase signal. As described before, the chirp waveform is a linear quadratic phase modulated signal, meaning that if we monotonically increment the phase in a quadratic matter, we would be able to generate a chirp signal. The phase evolution of a chirp signal ($f_1=f_0+\mu t$) considering the system described in Figure 2.4 can be written as

$$ph_1 = 2\pi \int f_1 dt + ph_0 \qquad (2.18)$$

$$ph_1 = 2\pi \left(f_0 t + \frac{1}{2} \mu t^2 \right) + ph_0 \qquad (2.19)$$

The phase equation in (2.19) is considered for a continuous time frame, nevertheless in a digital system the time is discretized and each time interval is defined by the system clock ($\Delta T = 1/f_s$), where f_s is the clock frequency of the DDS. The phase information is translated through a correspondence table (discrete sinewave function) into amplitudes. This means:

$$g_{dig}(n) = \sin\left(ph_1\right) \quad (2.20)$$

$$g_{dig}(n) = \sin\left(ph_0 + 2\pi\left(f_0 n\Delta T + \frac{1}{2}\mu\left(n\Delta T\right)^2 \right)\right) \quad (2.21)$$

In this process of transforming discrete values of phase into amplitude values, errors appear in the transformation and depending on the ratio f_1/f_s, they will become more significant. This is called phase-truncation, and later in the next chapter I will describe in deeper detail this phenomena and its implication to the resulting signal.

According to equation (2.20), the signal will evolve in frequency infinitely, but it was discussed before that the chirp waveform should have a defined length, called chirp length (T_e). For that purpose, the signal is constrained in time and in frequency through a controlled re-initialization of the phase register at a defined time. The chirp length can also be represented as system cycles $N = T_e \cdot f_s$, obtaining a constrained time which can be analyzed in the complete chirp length time frame as a limited sum of phase increments.

$$g_{dig}(1\cdots N) = \sin\left(ph_0 + \sum_{n=1}^{N} 2\pi\left(f_0 + \frac{1}{2}\mu \cdot n\Delta T \right)\Delta T \right) \quad (2.22)$$

The discrete amplitude values evolving in time described in equation (2.20), are later converted into analog values through DAC devices. The quality of the DAC, defined by the resolution bits (R), constrains the quality of the analog signal. This error increases as R decreases and this is called amplitude truncation. Based on equation (2.20), the resulting signal can be written as:

$$v(t) = \sin\left(ph_0 + 2\pi\left(f_0 t + \frac{1}{2}\mu t^2 \right)\right) \quad (2.23)$$

for $f_0=0$, $ph_0=0$ and $k=\mu\pi$:

$$v(t) = \sin(kt^2) = \left| e^{jkt^2} \right| \quad (2.24)$$

Equation (2.24) acknowledges that with a discrete digital system, such as a DDS system, we are able to create a chirp waveform with defined dispersive parameters and time length.

The resulting signal is later pre-processed, meaning spectral up-conversion and frequency multiplication, in order to be integrated in the M(l)-C(s) arrangement. Frequency multiplication is needed as the available bandwidth achieved by a digital synthesizer is constrained by its maximum operation frequency. In our case, this constraint is about one third (1/3) of the needed expander bandwidth. The power consumption of digital electronics increases as the operational frequency increases and the capability of dissipating that amount of energy limits the operational frequency of the device, resulting in a bandwidth limitation. The goal of having a digitally generated expander chirp with high time-bandwidth product, can be achieved with a rational and optimal combination of different technologies, such as quadrature-modulation, high performance spectral filtering, ultra-stable frequency sources and a detailed study of the problem through numerical simulations with an ad hoc model.

2.3 Instrument configuration

The SOFIA-GREAT-CTS instrument is intended to be a complete stand-alone unit. This can be defined, as an instrument capable of distinguishing the spectral information from the input signal, then digitally acquire it, pre-process it and finally calibrate it. The final processing stage is done by a computer, which also has an important advantage of providing standard interfaces, and thus achieving a higher level of autonomy. The instrument can be summarized in four blocks, 1) the digital chirp generator, 2) the radio-frequency signal processing stage with the SAW compressors, 3) the data-acquisition with pre-processing modules and 4) the storage, calibration and final interface integrated in a computer. The described order actually defines the path of the signal processing involved in the chirp transformation, from the expanded-chirp until the spectral information of the incoming signal is obtained.

In an analog CTS, as the one developed by Hartogh [1990], 2 SAW devices for expansion and 1 for compression are used, all having the same dispersive properties (meaning bandwidth, chirp length and chirp rate). The high expander chirp bandwidth is achieved by frequency doubling of the 1st expander chirp, while the 2nd expander SAW device duplicates the chirp length, producing a quadruple time-bandwidth product expander. This arrangement carries the inconvenience of a pre-defined internal frequency scheme set by the SAW devices. The expander center frequency is $2xE_{CF}$ (E_{CF} = SAW expander center-frequency) while the compressor is C_{FC}, leading to an instrument's center frequency of $2xE_{CF} + C_{FC} = 3xSAW_{CF}$. The fact of having a fixed center frequency is of inconvenience, as backend instruments normally share the same frontend unit and therefore the same IF (intermediate frequency). With the ADCP exists the freedom of definition of the expander center-frequency (since the expander chirp is up-converted with a tuneable source), and therefore we are able to define the internal frequency scheme and the instrument's center frequency. In the case of SOFIA, the CTS spectrometer shares the IF unit with the AOS instruments at 2.1 GHz. This is possible since we use an expander chirp at $CTS_{CF} - C_{FC} = 3xE_{CF}$ (3 because we triple the up-converted digitally generated chirp signal) → $E_{CF} = (2100 – 450)/3 = 550$ MHz.

The duty cycle, the number of expander-compressor branches and the expected spectral coverage define the expander bandwidth. Having two expander-compressor branches and defining a 100% duty cycle would demand an expander bandwidth of 400 MHz (for a compressor SAW device of 200 MHz bandwidth). As we will see later in this chapter, it is an extremely complex process to produce high bandwidth signals with digital synthesis, and even more complicated if a low phase error response is expected (as in our case). Our approach is based on creating a potentially high time-bandwidth signal with smaller bandwidth, that afterwards, this signal is spectrally expanded through frequency multiplication by means of the non-linear properties of an analog amplifier in compression. Once the expander chirp is generated, it is mixed with the incoming signal, producing a spectral down-conversion of the expanded chirp, which is fed to the compressor. At the output of the compressor, the chirp transform is already done, meaning that the time domain of the output signal already contains the spectral information of the input signal. The two simultaneous expander-compressor schemes should be properly integrated in order to obtain a 100% duty cycle. These two signals are time switched at their t_0 (now representing f_0), and for a time T_c (the SAW device of the compressor has a chirp-length of $T_c = 22$ μs). This means that each branch takes a time of $2xT_c$ to perform a transformation, however, it produces only T_c worth of time of spectral information.

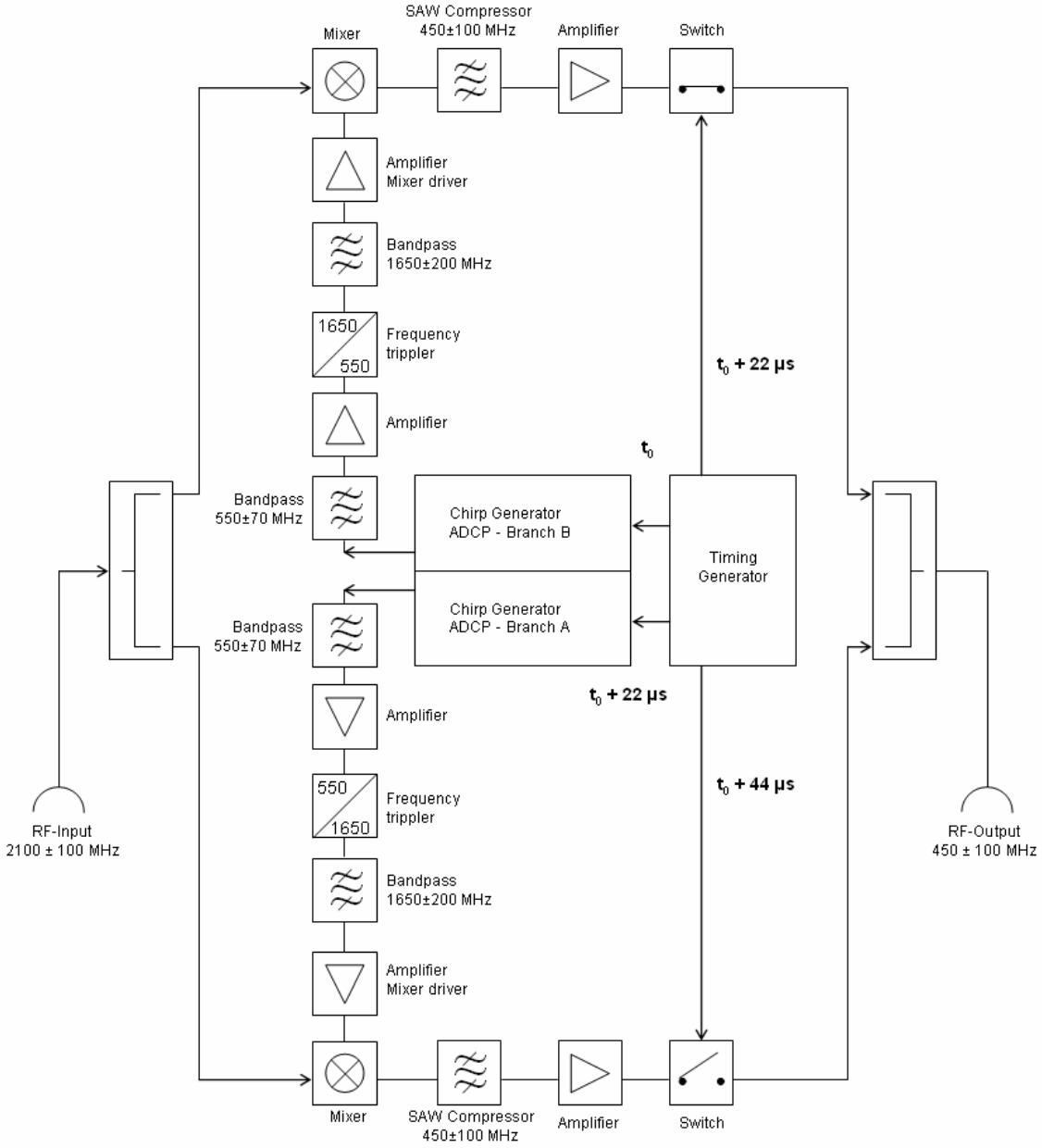

Figure 2.5: Diagram describing the configuration of the RF-block of the SOFIA-GREAT-CTS spectrometer. The incoming signal is split and sent to the down-conversion mixers. The timing generator defines the starting moment of the transform, which produces a chirp signal coming from the chirp generator. This signal is filtered, tripled and sent to be mixed with the incoming RF signal. The mixing product is convolved by the SAW compressor and according to a precise timing scheme, is then combined with the other branch.

The combined signal of the two branches is later down-converted to the baseband. It should be reminded that the compressor SAW device performs the convolution at its center frequency (450 ±100 MHz), nevertheless the available information is described only by a bandwidth of 200 MHz. For that, the signal is down-converted to 0-200 MHz, in order to be acquired by the digital electronics. An important fact about this process is that we can also retrieve the two components of the transform, that is the Real and Imaginary components by mixing the signal with a complex (a + ib) spectral source.

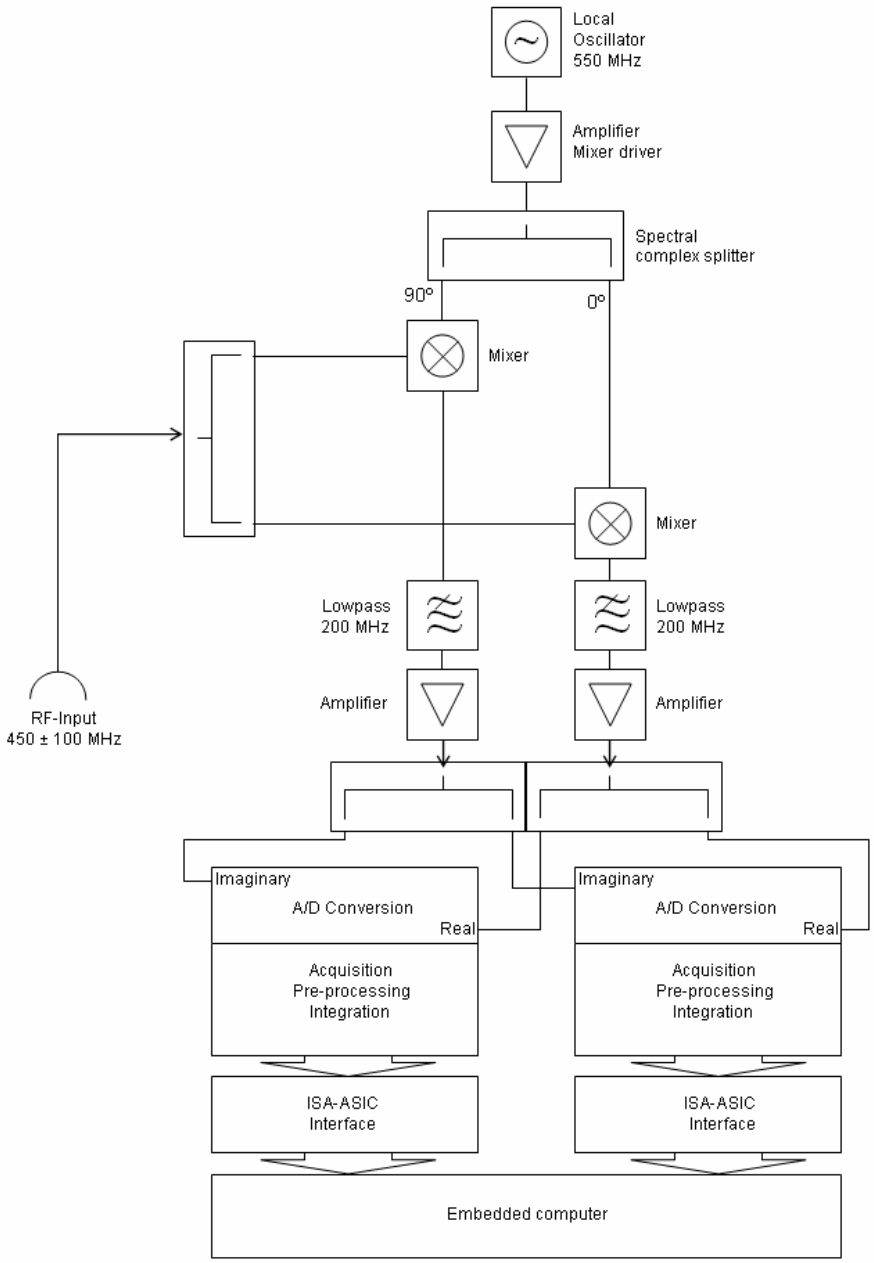

Figure 2.6: Description of the digital acquisition block. The signal produced by the RF-block is split and mixed with the complex components of the frequency source. Both components of the chirp transformation (real and imaginary) are filtered and amplified. The complex signal is fed to the analog to digital converters in order to be later pre-processed. The digitally integrated signal is transferred to the embedded computer through an ad-hoc interface.

$$\mathrm{Re}[\left(a_{signal} + i \cdot b_{signal}\right) \cdot \left(a_{source}\right)] = a_{signal}\, a_{source} \qquad (2.25)$$

$$\mathrm{Re}[\left(a_{signal} + i \cdot b_{signal}\right) \cdot \left(i \cdot b_{source}\right)] = -b_{signal}\, b_{source} \qquad (2.26)$$

Equation (2.25) describes the real part of the chirp transform, while equation (2.26) is related to the imaginary component. This complex signal, already down-converted, is sampled with a high data-rate analog to digital converters (ADC) and digitally analyzed by an ad-hoc developed pre-processor integrated in an ASIC (application-specific integrated circuit) chip with very low

power consumption and fully integrated capabilities. The pre-processor computes the power from the complex spectrum, and handles the synchronization signals needed to control the timing of the two branches. The synchronization is a critical issue as the time domain contains the spectral information; any perturbations could lead to misinterpretation and false retrieval of the spectral data. This synchronization defines the initiation of the chirp transform, the switching time and the channel spacing. Even so the different modules of the spectrometer make use of different frequencies, in all of them the phase information between the DDS clock and the data-acquisition should be identical. For that, a unique synchronization source is defined and created by the ADCP, which is then shared to the RF modules and the digital pre-processor. This will demand a more sophisticated scheme for the data-acquisition process. A solution to this matter was obtained by combining two pre-processors running in interleave and at an over sampling data rate.

The pre-processed spectral information is then transferred to an embedded computer, through a specifically developed ISA-ASIC interface that communicates with the pre-processor and reads the integrated data. As we are using two pre-processors working in interleave the measured channels are then re-organized in frequency in order to provide the real independent channels. In the actual configuration, 7500 channels are retrieved, meaning a channel spacing of 26 kHz = 200 MHz / 7500, which is then interpolated to the real channel spacing of 50 kHz → 200 MHz / 50 kHz = 4000 channels.

The use of an embedded computer as the final stage of the data acquisition modules provides several benefits. They can be summarized as: 1) the capability of performing high-level processing of the spectral data (such as calibration), 2) the computer behaves as a temporary data server, allowing sporadic data requests, and 3) the instrument provides standard interfaces for transferring data, for instance a network connection with TCP/IP protocol, a general purpose TTL interface and a serial interface.

2.4 The ADCP – Adaptive Digital Chirp Processor

2.4.1 The matching of the dispersive properties

The spectrometer performance is strongly dependent on the matching of the dispersive properties between the expander and the compressor. In the demonstration of the CTS principle in 2.1.1, a perfect dispersive matching was considered, meaning $\mu_c = -\mu_e$, but in practice there are differences in the dispersive properties, which are mainly caused by SAW imperfections and by limitations of the digital chirp generator. The dispersive properties of a SAW device are described by two parameters, the chirp rate μ, and the phase deviation; the better the quality of the SAW device the smaller is the phase deviation. Nevertheless, having a set of high-quality SAW devices will not guarantee a perfect chirp transform, and this is also because the chirp rate should be equal among the SAW devices involved in the transform. This consideration is a critical issue in the design of the CTS, especially in the case of the analog-CTS which demands a setup of 6 SAW devices (4 expanders + 2 compressors). However, in the case of the digital CTS, the fact that only two SAW devices are needed extremely simplifies the scheme; more so if the digital expanders adapt to the dispersive properties of the compressors. The dispersive mismatch produces two effects: a broadening of the sinc function and a displacement, both mainly affecting the spectral resolution. The broadening of the sinc function can be better clarified by

analyzing equation (2.8) and evolving the chirp transform assuming different chirp-rate coefficients for the expander and the compressor:

$$f(\tau) = \int_{-\infty}^{\infty} f(t) \cdot e^{-j\pi\mu_e t^2} \cdot e^{-j\pi\mu_c(\tau - t)^2} dt \qquad (2.27)$$

Considering that the transform has a finite duration T_c, and performing some algebraic transformations, the equation (2.27) can be written as:

$$f(\tau) = \frac{1}{T_c} \int_{-T_c/2}^{T_c/2} e^{-j\pi\mu_e t^2} \cdot e^{-j\pi\mu_c(\tau - t)^2} dt \qquad (2.28)$$

$$f(\tau) = \frac{e^{-j\pi\mu_c\tau^2}}{T_c} \int_{-T_c/2}^{T_c/2} e^{-j\pi(\mu_e+\mu_c)t^2} e^{j2\pi\mu_c t\tau} dt \qquad (2.29)$$

Equation (2.29) can then evolve to an ideal case when $\mu_c = -\mu_e$, meaning the first term of the integral disappears, and this would lead to the well known sinc solution

$$f(\tau) = \frac{e^{-j\pi\mu\tau^2}}{T_c} \int_{-T_c/2}^{T_c/2} e^{j2\pi\mu t\tau} dt \qquad (2.30)$$

$$|f(\tau)| = Sinc(\pi\mu T_c \tau) \qquad (2.31)$$

Nevertheless, in the case where the dispersive properties are different, we should centre our attention on equation (2.29). For that I numerically solved the equation for different levels of mismatch; defining $\xi = \Delta\mu/\mu_{comp}$ as the mismatch ratio between the chirp-rate difference with respect to the compressor device. Figure 2.7 describes the strong importance of obtaining an optimum matching of the elements involved in the transform. Even though the SAW devices used for SOFIA-GREAT-CTS are of extremely good quality (phase error less than 3º rms), they show quantitatively high differences in their chirp-rate, up to 2800 ppm.

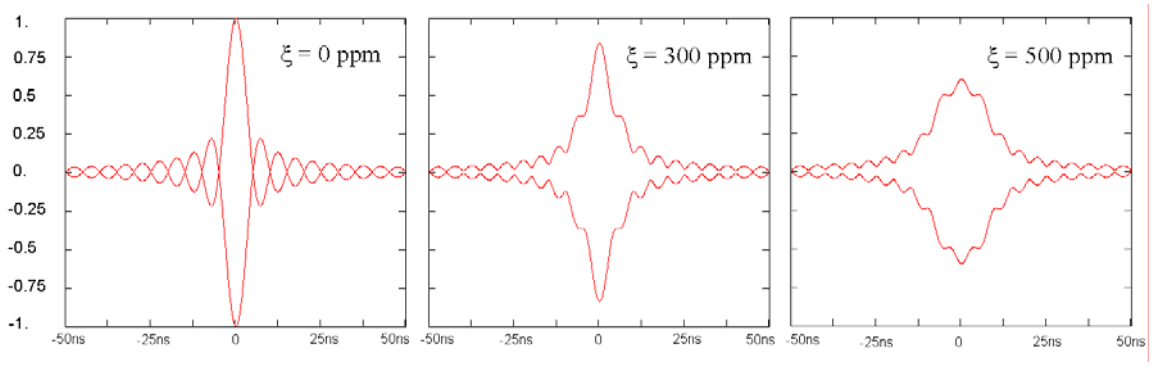

Figure 2.7: Analysis of the broadening pattern caused by an expander-compressor mismatch. The calculations are based on the SAW devices used for SOFIA with an operational bandwidth of 200MHz.

A solution to this matter can be achieved by a careful selection of the SAW devices involved in the expander-compressor setup; however, this is valid as long as there exist sufficient sets of matching filters. Another possibility is the regulation of the temperature for each device; this is actually the most convenient approach since the SAW devices are strongly sensitive to the environmental temperature. The thermal response is directly related to the mechanical properties of the substrate, lithium-niobate, and for SOFIA-GREAT-CTS's SAW devices this is dμ/dT = -0.001455 [MHz/μs·ºC] and a chirp rate μ=10.1097 [MHz/μs], giving the following thermal sensitivity formula:

$$\frac{d\xi}{dT} = \frac{1}{\mu}\frac{d\mu}{dT} = 143\left[\frac{ppm}{C^{0}_{SAW}}\right] \qquad (2.32)$$

This strong relation between temperature and dispersive properties produces a complex implementation, especially in the case of the analog CTS where 6 independent temperatures have to be regulated; an impractical approach. In the case of MIRO-CTS (analog chirp spectrometer) on board the cometary mission Rosetta, two branches of optimally selected filters (2-expanders + 1-compressor) and two independent temperatures for each branch are used. This arrangement combines the two previously described techniques in order to obtain the maximum of the setup. In the case of the digital CTS for SOFIA, only the adjustment of two independent temperatures are needed, one for each branch. However, special care should be taken to account for the isolation of the SAW devices' environmental temperature from SOFIA's fluctuating environmental temperature. Later in this chapter, the implementation of a small thermal chamber is described, obtaining a 40-fold improvement in thermal isolation.

The second effect is the displacement of the signal as a function of the chirp rate (see equation (2.3)). Considering that two branches are used and that the chirp transformed signal is latterly added, the response of both branches should be identical. Any displacements between them would lead to a reduction of the resolution. A better way of clarifying this is to analyze the numerical calculations as shown in Figure 2.8.

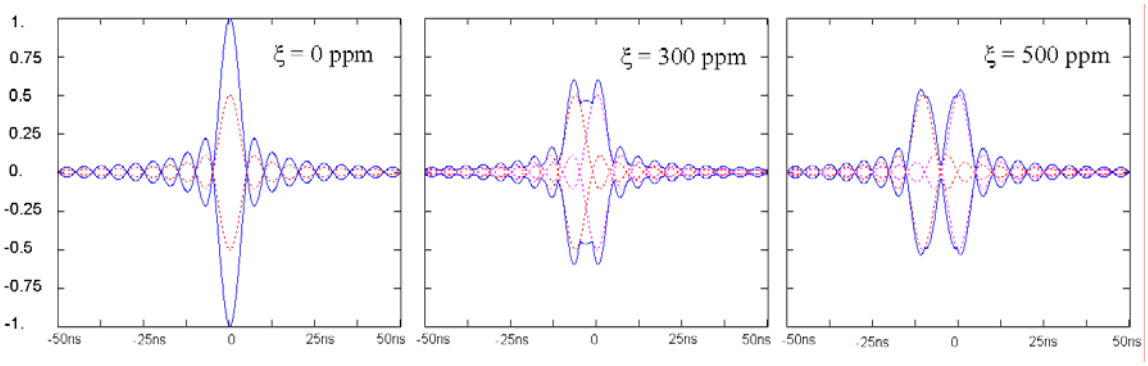

Figure 2.8: Broadening of the compressed signal introduced by response displacements between the two branches. In dotted lines (red and magenta) are described the optimally compressed pulses of each branch, while the blue line is the sum of both branches. The error ξ describes the chirp-rate deviation between the two branches.

The explanation can be obtained by analyzing the temperature derivative of equation (2.3), and the displacement can be written as:

$$\Delta t = t - t' = \frac{B}{\mu} - \frac{B}{\mu + \Delta\mu} = \frac{B}{\mu} - \frac{B}{\mu\left(1 + \frac{\Delta\mu}{\mu}\right)} = T_c - \frac{T_c}{\left(1 + \frac{\Delta\mu}{\mu}\right)} \qquad (2.33)$$

obtaining d / dT, this leads to

$$\frac{dt}{dT} = T_c - \frac{T_c}{\left(1 + \frac{1}{\mu}\frac{d\mu}{dT}\right)} \qquad (2.34)$$

$$\frac{dt}{dT} = T_c\left(1 - \left(1 + \frac{d\xi}{dT}\right)^{-1}\right) \qquad (2.35)$$

Based on the results shown in Figure 2.8 and equation (2.35), the importance of a fine matching of dispersive properties can be seen, not only between the elements in the branch but also between the different branches. One of the prime advantages of the digital CTS, with only two dispersive elements, is the simplification of the matching problem. This will lead to obtaining most of the dispersive arrangement, and therefore a higher spectral resolution. The limitation will be now defined by the dispersive quality of the SAW compressors, the digitally generated expander chirp and the temperature stability of the SAW devices environment.

2.4.2 The generation of the digital chirp

As described previously, the digital chirp can be seen as a train of analog electric signals spaced by a time ΔT with the amplitude defined by the instantaneous phase value derived from the ADCP's phase accumulator. The sampling theory provides the tools in order to study this type of signal, because the digitally generated expander chirp can be also expressed as a quantized analog signal in a discrete time domain. Each outcoming sample from the ADCP is produced by a "sampling-function" at a defined sampling frequency $f_s=1/\Delta T$, which can be represented by a train of impulses p(t). The transfer function of the sampling process is [see appendix D]:

$$H(f) = Sinc\left(\pi(f / f_s)\right) \qquad (2.36)$$

The region of the then sampled spectrum where the signal can be later reconstructed is $f_{out} < f_s/2$; initially derived by the theoretician Nyquist, $f_s/2$ is now called the Nyquist limit. At the Nyquist limit, the attenuation introduced by the sampling process is in the order of 4 dB, while the spurious f_s-f_{out} is superposed in this limit. From equation (2.36) and (G.26), it can be seen that by sampling a signal with frequency f, higher frequency components of (2f,3f,…) are also generated with an amplitude distribution following a sinc pattern. These signals, the so called "aliases" of the sampled signal, point out two considerations: 1) the given energy to produce a defined spectrum is undesirably spread, and 2) these extra components add a practical constraint when defining the operational frequency, since they should be distant from the main component in order to be easily filtered out.

In addition, there is a second effect related to the digital generation of signals: the quantization of the amplitudes. In order to quantize a sample value, this value is rounded to the nearest allowed finite value. This codification takes place since each sample is represented by a digital word.

During this instance of translating discrete values into analog values, an error is induced and depends on the amount of finite points. This effect, also known as quantization distortion, produces extra spurious components of the generated signal. A higher number of finite values would mean a finer amplitude resolution and thus a smaller quantization distortion. Since this distortion is quantifiable, the signal to quantization noise ratio is given by:

$$SN_qR_{DAC}[dB] = 1.76 + 6.02 \cdot N - 20 \cdot \log\left(\frac{V_{DAC}^{\max}}{V_{out}^{\max}}\right) \qquad (2.37)$$

Where N is the resolution of the DAC, V_{DAC} is the maximum voltage generated by the converter and V_{out} is the maximum voltage created. The first and second term of equation (2.37) are related to the DAC and describe the maximum achievable performance. In the case of the generated signal not using all of the available DAC dynamic range, the quantization distortion would become more significant, resulting in a worsening of the SN_qR ratio. This effect is represented by the third term of equation (2.37). The 8-bit DACs used by the ADCP define a SN_qR at full scale of 49.92 dB. A non-linear performance of the DAC will induce distortions leading to the production of higher-harmonic multiplexes of the generated frequency. As the DAC is combined with a time discrete process, the Nyquist criteria applies, meaning that these unwanted spurious components are mapped directly onto the 1st Nyquist region, the spectral region where the desired generated signal is located

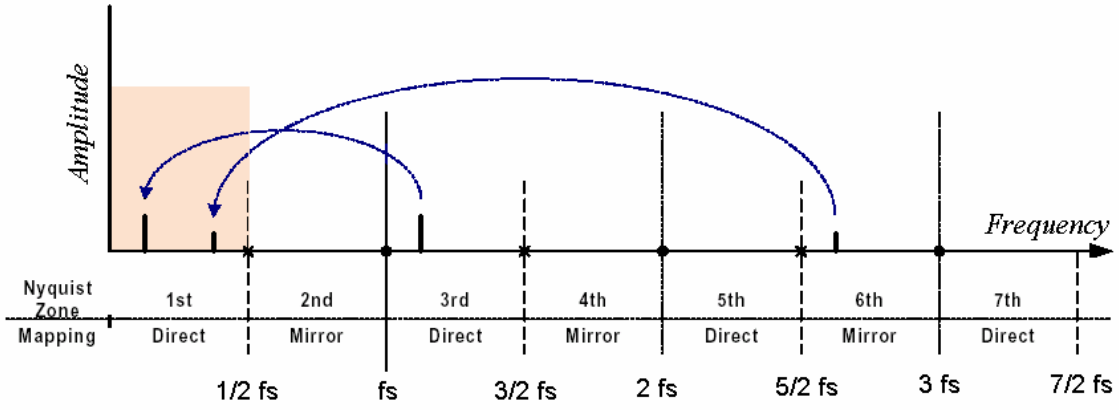

Figure 2.9: Aliasing of the upper harmonics into the 1st Nyquist region. The upper harmonics, generated due to non-linearity features in the DAC, are then transferred to the information spectral region by aliasing. Figure based on [Genttle, 1999].

The reduction of the influence of this effect to the overall performance of the ADCP processor is of extreme complexity. A possible approach to this matter is by identifying from the DAC dynamic range, the region where the non-linearity features are not so significant. However, this will also produce a reduction in the SN_qR (term 3). In the case of the ADCP processors for SOFIA-GREAT-CTS, the solution was focused on an oversampling of the signal, leading to a stronger attenuation of the aliased signals. Oversampling refers to the process of generating a signal with more samples than the one defined by the Nyquist limit. At the Nyquist limit only two points defines the signal-period, while with an oversampling of 3 it would mean that exists 6 points per period. This effect can be quantified and can be analyzed as an enhancement of the quantization S/N ratio [see Genttle, 1999]. It is described as the following:

$$SN_qR_{OS} = 10\log(f_s / f_{out}) \qquad (2.38)$$

This increment in the performance can be added to equation (2.37):

$$SN_qR[dB] = SN_qR_{DAC} + SN_qR_{OS} \quad (2.39)$$

In SOFIA-GREAT-CTS, the upper spectral range of the expander is constrained by the maximum DDS sampling frequency (350 MHz). In order to reduce the effect of aliases, the upper spectral component of the chirp signal is located at a prudent distance from the Nyquist limit, with a maximum of f_{max}=150 MHz. This frequency is selected based on the feasibility of the filtering aliased signals of the 2nd Nyquist region. With a required expander signal of B_e=400 MHz, the digital expander bandwidth before multiplication by 3 is 134 MHz, and therefore the aliased signals of the 2nd Nyquist region start at $f_s - f_{max}$ = 200 MHz and finishes at $f_s - (f_{max}-B_e/3)$=334 MHz. The relative strength of the aliases can be derived from equation (2.36), and for the 200 MHz component it is especially high (-5.3 dB). The suppression of these strong aliased signals would demand high-performance spectral filtering and, for that, low-pass filters from the Cauer family of the 4th order were designed, in order to reduce up to 40 dB these spurious components. The filtered signal is then level regulated and fed into the modulators for its up-conversion.

The digital generation process will be additionally constrained by the purity of the system clock. Any impurities in the system clock will produce distortions of the sampling rate, and herewith distortions of the out-coming signal. These variations in the system clock, normally defined as time jitter, are caused mainly by two effects: a) thermal noise, which is omnipresent and implicates that whatever circuit is used to generate the system clock, it will always exhibit some finite amount of timing jitter due to thermal noise, and b) electromagnetic interference. Interference can also be defined as coupled noise, as it is acquired by unwanted electromagnetic coupling with external sources. These sources of noise can be generated in the immediate area, as for example power-lines and near by radio-transmitters, or inboard sources such as onboard local oscillators. The level of influence of this effect depends strongly on the shielding of the digital generator, mainly from the onboard circuitry. In our case, the digital generators and its ultra-stable time source share the same circuit with the strong signals used to up-convert the generated chirp. Proper shielding is a difficult issue, more so if the different devices share the same electrical ground. The solution to this issue was to physically locate the digital generators at a prudent distance from the local-oscillators on the board.

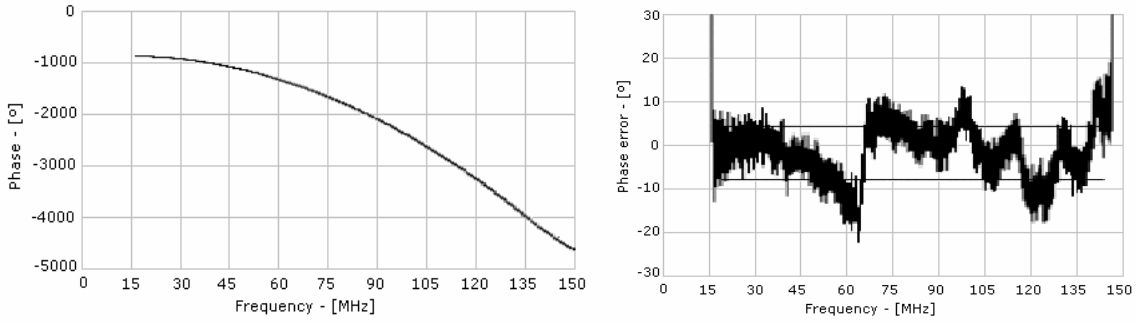

Figure 2.10: Measurement of the phase response of the ADCP for an expander chirp from 150-17 MHz. [Right] The measurement was performed with a 1 Gs oscilloscope and the phase information retrieved through a Fast-Fourier transform. [Left] Phase error from a 2nd order fitted approximation (rms phase error: 5.48º).

2.4.3 Digital signal up-conversion and quadrature modulation

In the generation of wide-band signals, as is the case of the expander signal, the bandwidth is limited by the upper and lower achievable frequencies. As was discussed previously, the upper frequency is defined by the maximum sampling rate and considerations of avoiding the high harmonic aliased signals. The main restriction defining the lower boundary for the generation of digital signals is based on the fact that this signal is later up-converted. The lower generated frequency defines the distance from the up-converter carrier and the mirror signal. As is known from the modulation theory, the amplitude modulation is expressed as a multiplication of a carrier with frequency α and the signal with frequency ω.

$$m(t) = A[1 + m\sin(\omega t)]\sin\alpha t \rightarrow m(t) = A\sin\alpha t + \frac{m}{2}A[\cos(\alpha-\omega)t - \cos(\alpha+\omega)t] \quad (2.40)$$

Equation (2.40) describes the modulated signal having three spectral components, the carrier and the two-sidebands, while m defines the modulation index. Since it is expected to produce a signal with a bandwidth of 134 MHz and the maximum achievable frequency is 150 MHz, this leads to a lower boundary of 16 MHz for the chirp signal. Considering a modulation index of 1, this would mean that at 16 MHz from the up-converted signal a twice stronger signal (the carrier) is located, while an unwanted mirror with the same energy is located at 2·16 = 32 MHz. The filtering of these unwanted harmonics is unfeasible, and would constrain the maximum achievable bandwidth. A solution to this matter can be obtained by quadrature modulation which, thanks to the use of conjugate modulation components, achieves a suppression of the carrier and the mirror.

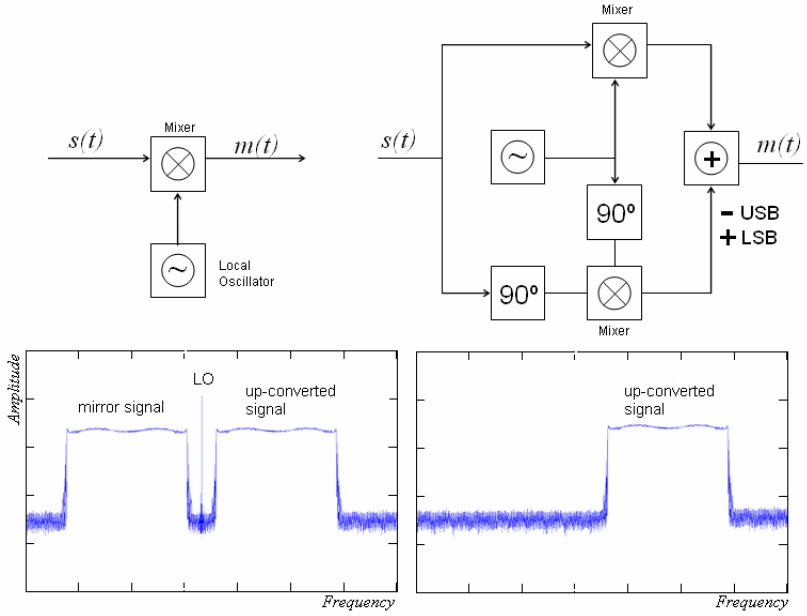

Figure 2.11:Schemes and resulting spectrum for the standard amplitude modulation and the quadrature modulation. [Left-up] Diagram describing the elements involved in an amplitude modulation. [Left-down] Resulting spectrum showing the desired up-converted signal with its mirror and the carrier. [Right-up] Quadrature modulation scheme. [Right-down] Resulting spectrum describing the suppression of the unwanted harmonics.

The convenience of this modulation method is clear, as it allows generating signals up to the DC-limit. The complexity in the implementation of this method is the fact that the signal s(t) should be delayed by 90º. This process is unachievable for wide-band signals, and the alternative is by initially generating both signals (0º and 90º). For that purpose, each DDS device provides the feature of generating two phase-orthogonal components.

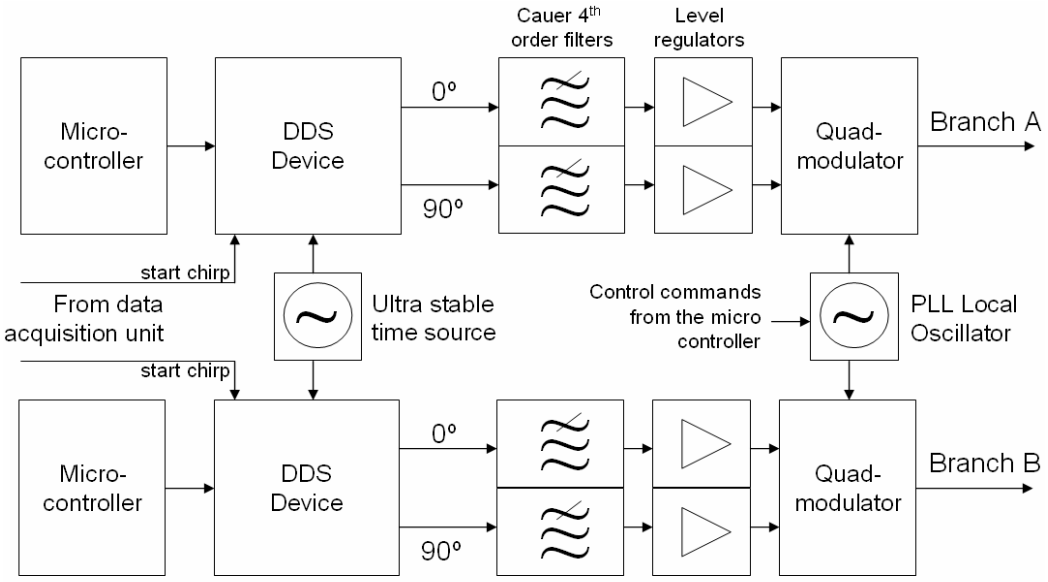

Figure 2.12: ADCP processor scheme description. The core of the ADCP processor, the DDS device, is controlled by a fast microcontroller required to produce the desired chirp-rate. The phase purity of the resulting signals are mainly defined by the sampling signal and the time jitter. Consequently, the ultra-stable system clock is located near the DDS devices. The digitally generated signal is filtered and fed into the quadrature modulator.

2.5 Controlled dispersive matching

2.5.1 Embedded control

The ADCP processor is monitored and controlled by two embedded controllers, which also provide interfaces for remote-controlling and programming. The possibility of having the computing features of a desktop computer in a single device is of great advantage, especially when strong constraints in both space and power consumption exists. The control sequence is integrated in a unique specific program (stored in non-volatile memory) that becomes the kernel of the controller. This kernel handles the following processes: a) management of the interfaces for remote control; b) controlling of the chirp-generators and their time/frequency dispersive characteristics; and c) setting and monitoring of the local oscillators involved in the up-conversion of the expander chirp and in the down-conversion of the compressed chirp.

The architecture of the microcontrollers is very similar to that of the standard computers, primarily consisting of a processing ALU (Arithmetic Logical Unit), the memory banks, the external ports and the in-between interfaces. Since microcontrollers are intended for embedded systems, their instruction set is limited and thus they provide a high computational efficiency. This is the case of the ACDP's controllers, where their ALU is based on a fast RISC (Reduced Instruction Set Computer) processor. The embedded system contains three memory banks, the program memory bank where the kernel is installed, the volatile memory bank (RAM) where the internal registers and working data are stored and a non-volatile EEPROM memory bank which is designed for storing permanent values, as is the case for the ADCP's control parameters. The

running kernel is stored in the program-memory through ad-hoc circuitry and interfaces installed in the ADCP processor. The inner configuration of the device, that is, the A/D converters, interrupts, interfaces and ports, are accessed through special registers organized in the volatile memory bank of the controller. Once the kernel has performed its initial setup, the control algorithm is driven by external events which will lead to determined actions. At start-up, the kernel of the ADCP's controller proceeds in the following manner: 1) it sets the inner configuration of the processor (ports, interfaces, interrupts, memory banks, etc), 2) it configures the local oscillator with its initial setup, 3) it sets the chirp-rate and other functionality registers of the digital generator and 4) it waits and handles the incoming events (remote commands, timing actions, etc), represented internally in the controller as computer interrupts. The main remote-control interface of the ADCP's controllers is the standard RS-232 connection, from where commands are given in a predefined format. All the desired actions and controlled devices are seen from the remote controller as registers in a memory bank of 256 positions (1-byte addressable). The interface-bank is organized in controllers and devices, the first 128 positions provide access to the first controller which controls the DDS-A and the up-conversion oscillator, while the last 128 positions are defined for the second controller with the DDS-B and the down-conversion oscillator. The communication with the PLL device required the development of several layers of interfaces, since the structure of the remote-control link is very different from the PLL's programming interface. The translation between the 8-bit address, the 8-bit data and the 2-wire remote interface with the 2-bit address, 22-bit data and 4-wire PLL interface was achieved through the implementation of cyclical buffers.

2.5.2 Control and monitor of the time/frequency dispersive properties

The capability of regulating the time/frequency dispersive properties of the expander chirp is the key element in the ADCP processor, and thus represents the main goal of the embedded controlling system. The dispersive properties of the expander-chirp, described by the chirp-rate coefficient, are internally represented in the digital generator as a binary sequence. Each SAW device has a matching sequence, a 48-bit differential phase coefficient μ_{dig} which defines the phase increment per step of the generated signal. The matching sequence is obtained from the compressor's complex conjugate chirp-rate and can be written as:

$$\mu_{dig} = 2^N \frac{\mu_{SAW} / M}{f_s} \qquad (2.41)$$

where N is the binary resolution of the digital sequence, μ_{SAW} the SAW device's chirp-rate, M the frequency multiplication factor and f_s the frequency of the sampling function. When we refer to M, we should remember that the digitally generated signal is then frequency multiplied in order to produce the desired expander bandwidth. The digital chirp-rate (μ_{dig}) together with the starting frequency (f_0), the output power (OP) and the chirp-length (T_e) are the main parameters which describe the properties of the expander chirp and are represented by binary registers in the core of the DDS device. The first 3 parameters are transferred into the phase accumulator unit of the DDS through its parallel interface, while the T_e is defined by external update signals. These signals, represented by successive impulsive signals, are generated by the time-generator (part of the data-acquisition electronics). The T_e is set by the time separation between two successive impulses, while each signal can be considered as a stop-start command.

The timing of these commands is of extreme importance, as they regulate the time-location of the expander and thus the position of the compressed pulse. For this, a common time reference source with equal phase should be used in all the electronics involved in the chirp-transformation. The implementation of this was possible with the use of a low frequency ultra-stable source at 19.44 MHz, which is then multiplied and delivered to the different electronic components.

The DDS device is fed with the 18th multiplex of this reference signal (349.92 MHz), while the data-acquisition and the time-generator circuitry with the 9th multiplex (174.96 MHz). Each phase synchronized start signal received from the time-generator to the DDS represents a clean-up of the phase accumulator register. While analyzing equation (2.19) it can be seen that the phase evolves as following:

$$ph_1 = 2\pi\left(f_0 t + \Delta f\right) \qquad \Delta f = \frac{1}{2}\mu t^2 \qquad (2.42)$$

and a phase clean-up represents $\Delta f=0$ and therefore the chirp-signal is restarted to its initial frequency f_o. Since the properties of the SAW compressor are kept constant through thermal stabilization, the 4 parameters of the digital expander chirp are also kept constant during the run. Nevertheless, the dispersive properties of the digital chirp can be modified during expansion because the controller is able to process up to 220 commands in an expansion time of 44 μs. This feature is of great advantage as it allows us to equalize the signal through modifications of the OP and to compensate dispersive deviations via constant updates of the chirp rate.

The fact that the DDS is running at its frequency limit represents also a substantial overheating effect and thus the possibility of internal damage. Hence, the DDS device is monitored through temperature sensors which provide information about the environmental temperature near the device and herewith about its juncture temperature (since the thermal conductivity of the device enclosure is known). A failure in the dissipation process will lead to an alarm procedure which will power-down the electronics and send an alert message to the remote-controller.

2.5.3 The dynamic adaptive chirp-rate

The digital chirp generator provides the possibility of producing every desired dispersive signal, especially those designed to correct dispersive deviations. This correction can be achieved through the dynamic update of the chirp-rate of the chirp signal during expansion. In order to quantify the benefits and drawbacks of this method, we should first analyze and quantify the dispersive deviations present in the expander-compressor scheme. For that, we measured the Most-Optimum-Chirp-Rate (MOCR) at different regions of the bandwidth. This measurement consists of varying the chirp rate μ_{dig} until an optimum compression of the signal is obtained. This is performed for different input frequencies. Three expander-compressor schemes were implemented for the measurements, two with different SAW compressors and one with an Arbitrary-Wave-Generator (AWG) behaving as the expander. By using different SAW devices we expect to distinguish from the measurements the effects induced by the digital expander, while the use of the AWG allows the comparison of the oversampling effects (since the AWG can be clocked up to 700 Mega-samples).

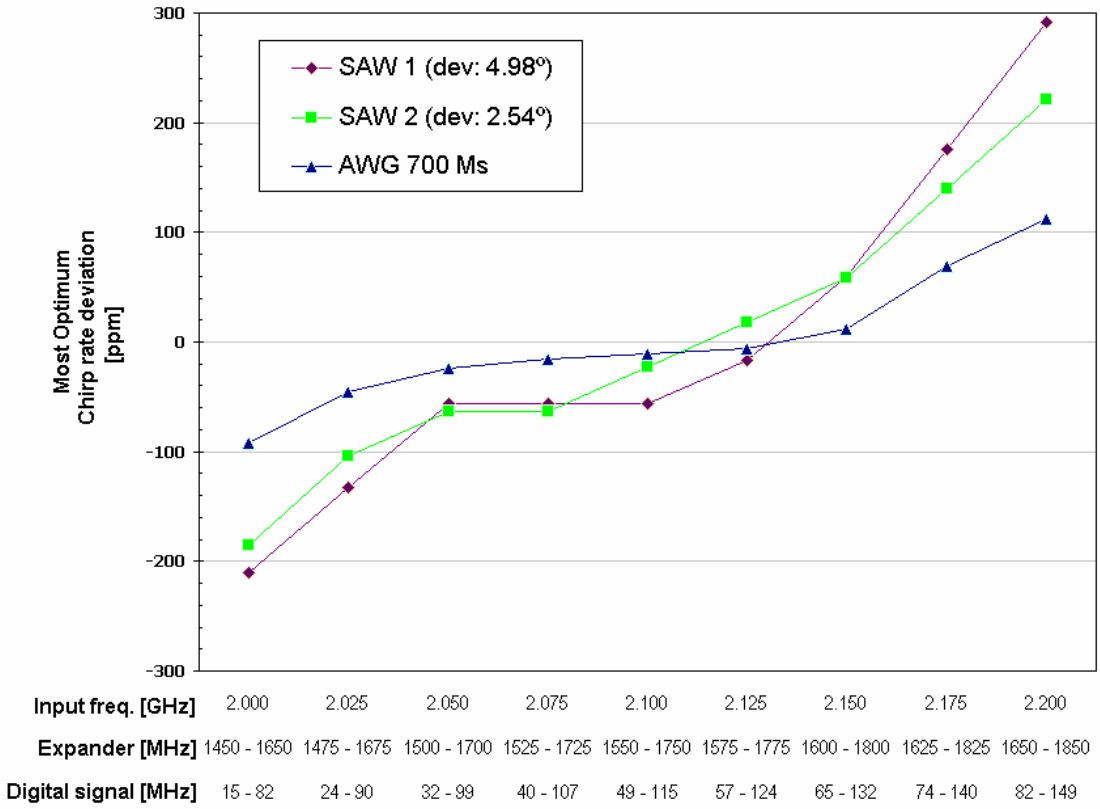

Figure 2.13: Measurements of the MOCR considering three schemes a) ADCP+SAW1, b) ADCP+SAW2 and c) AWG+SAW1. The plotted results are obtained for different input frequencies and thus different spectral regions of the expander. The digital signal represents the raw-expander signal before up-conversion and frequency multiplication. The two different SAW compressors were selected based on their phase deviation rms error, which is 4.98º for SAW1 and 2.54º for SAW2.

Figure 2.13 describes the pattern of the dispersive distortion taking place in the expander-compressor scheme. The three arrangements show the same evolution, meaning that the distortion could be related to a common issue related to the digital expander or/and to the SAW compressors. Assuming that the only difference between the ADCP processor and the AWG is the sampling frequency, we could infer that an oversampling of the digital expander will produce lower variations of the MOCR, and thus less dispersive distortion. The generated signal with the AWG is at 0.42 times the Nyquist limit while the ADCP's signal is at 0.86. This difference directly affects the phase properties of the signal since the number of sample-points per period is reduced. In the case of the differences observed between the SAW1 and SAW2, this could be related to the dispersive properties of the SAW compressors, since the SAW device with higher phase deviation shows greater MOCR variations. The monotonic evolution of the MOCR with frequency can be explained by two phenomena: a) phase deviations introduced by frequency filters, cables and active components which are especially frequency dependent, adding extra dispersive features to the processed signal; b) the digitally generated signal is more sensitive to the phase truncation effect at higher frequencies and could produce a frequency increasing dispersive distortion, similar to the pattern observed in the MOCR measurement; and c) phase deviations introduced by the compressor SAW devices. The quality of the resulting spectra for the different schemes can be studied through the properties of the out-coming sinc signal.

Properties such as the relative compression-ratio, the energy transfer fluctuations and the symmetry characteristics will let us describe and quantify the spectral performance of the scheme, and thus of the spectrometer. In order to quantify the compression of the expander signal, the ratio between the energy located in the main sidelobe and the remaining components is analyzed. For that, the theoretical limit of compression for different levels of expander-compressor mismatch was calculated, considering the results shown in Figure 2.7.

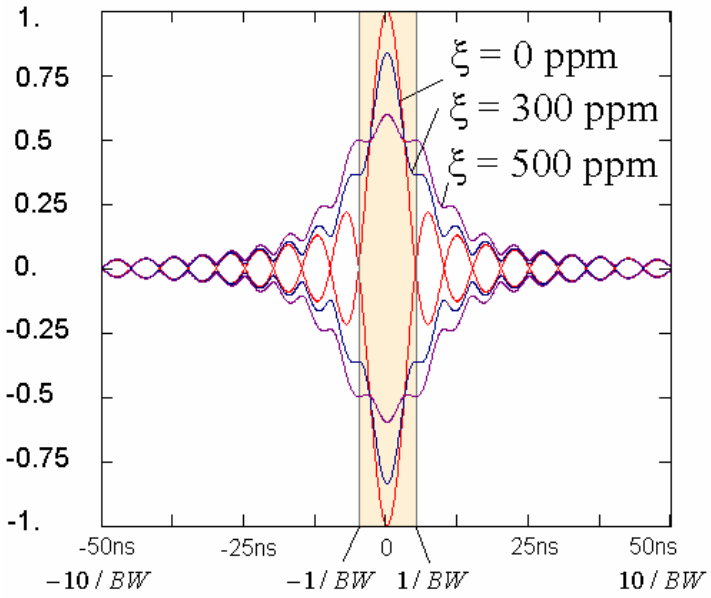

Figure 2.14: Analysis of the compression ratio for different levels of expander-compressor mismatch. The orange strip shows the main-sidelobe region.

$$S = \int_{-1/B}^{1/B} f^2(t)dt$$

$$C = \int_{-10/B}^{-1/B} f^2(t)dt + \int_{1/B}^{10/B} f^2(t)dt$$

0 ppm → S/C = 10.158 dB
300 ppm → S/C = 5.132 dB
500 ppm → S/C = 1.210 dB

By measuring the sinc-response from the expander-compressor scheme for different input frequencies, the different S/C ratios can be obtained. By comparing these values with the results shown in Figure 2.14, the level of mismatch can be retrieved. This measurement was performed for 35 frequencies in the spectrometer's bandwidth with the use of a controlled sinewave generator, a high-resolution oscilloscope and a computer arranged as the control unit.

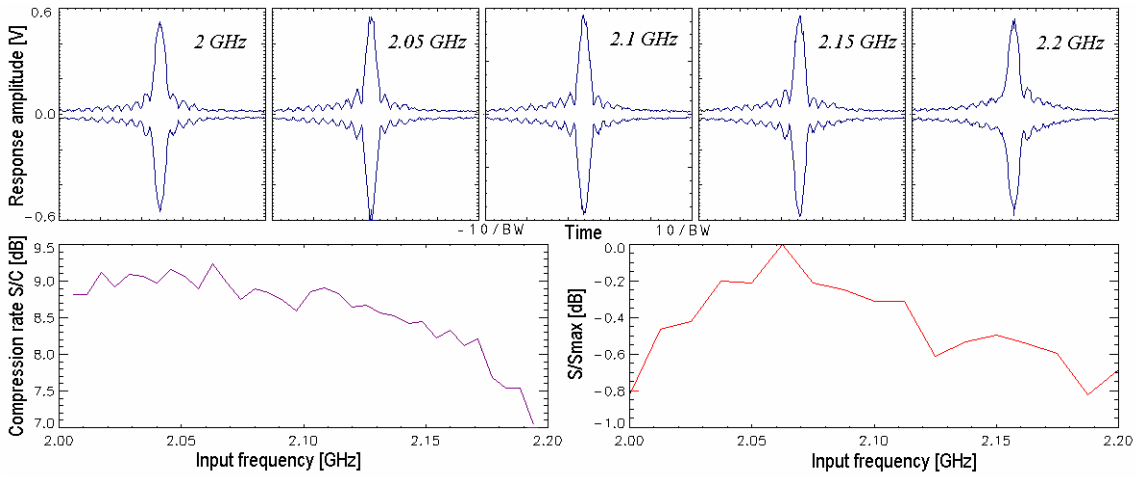

Figure 2.15: Measurements of the response of the expander-compressor scheme for different input-frequencies. The measurements consider a fixed chirp-rate which is optimized according to the dispersive properties of the SAW compressor. [Top] Envelope measurements of the sinc-response for 5 different input frequencies. [Bottom-left] Compression rate (S/C) measurements retrieved from the amplitude measurements. [Bottom-right] Variations of the energy located in the main sidelobe (S/Smax).

The properties of the response for the digital expander-compressor are encouraging, since the results show high level of compression and only 0.8 dB of signal variations for the complete bandwidth. The compression rate pattern can be correlated with the sinc response of the digital sampling process, and therefore the further the signal is generated from the Nyquist limit, the higher the phase quality and thus the compression rate. Nevertheless, a possible solution to this matter could be obtained through dynamically correcting the chirp-rate.

The dynamic modification of the chirp-rate is achieved by a precise synchronization between the dispersive correction commands and the chirp-start signal. The 20 MHz clock of the controller allows 5 million operations per second, and thus 220 computational operations per expansion time (44 μs). Considering that each chirp-rate correction demands 40 operations means 5 dispersive corrections can be done. The implementation of this feature in the controller was extremely complex due to the superposition of numerous events arriving at the device. Also the selection of the dispersive corrections demanded complex measurements and were strongly dependent on the SAW devices used for compression. Nevertheless, a test setup was arranged and the results are shown in Figure 2.16. As seen from the measurements, the compression rate and the S/Smax ratio were not significantly improved. Also in the amplitude response, the asymmetries in the sidelobes distribution can be observed. These are introduced by the compression of an expander with multiple dispersive rates. The results point out that this method of dynamically varying the chirp-rate could be of great advantage in setups with higher dispersive distortion. However in the case of SOFIA-GREAT-CTS, where the distortion is reasonable small a fixed chirp-rate setup is the most suitable solution, providing superior symmetry in the sidelobes distribution and high compression rates.

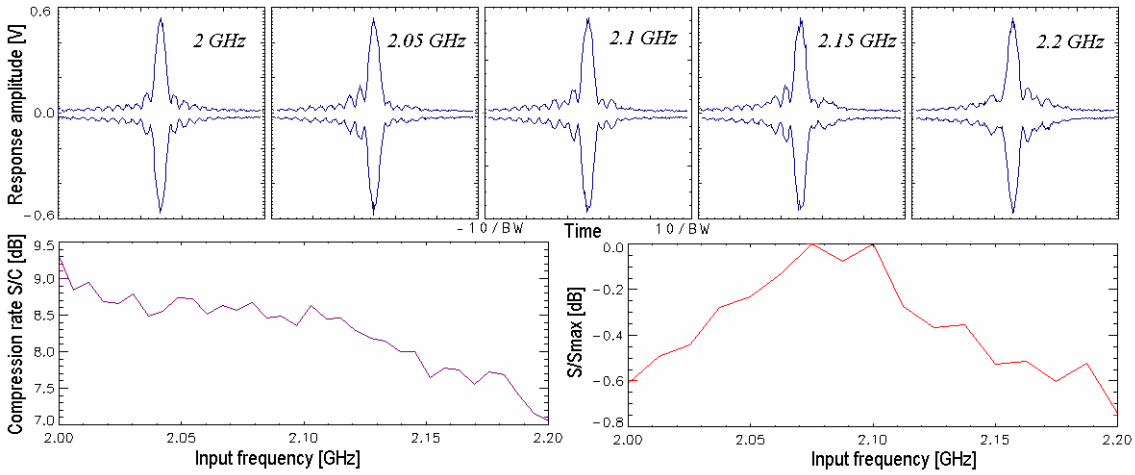

Figure 2.16: Multi-frequency measurement of the sinc-response for an expander-compressor scheme with dynamic adaptive chirp-rate. [Top] Envelope measurements of the sinc-response of 5 different input frequencies. [Bottom-left] Compression rate response over the complete CTS's bandwidth. [Bottom-right] Main sidelobe variations in logarithmic scale.

2.6 Frequency multiplication

The digitally generated chirp-signal is then frequency multiplied in order to achieve the needed expander bandwidth (2xBc = Be = 400 MHz). Since the maximum achievable bandwidth from the digital system is 134 MHz, the expander bandwidth is obtained by the generation of its third harmonic through frequency multiplication. There are several methods for frequency

multiplication. Nevertheless all of them are mostly based on phase synchronized systems. This method is valid for sinusoidal sources, where the phase evolves following a constant pattern defined by the frequency. In the case of the expander signal, where the signal follows a dispersive manner, the multiplication process should be addressed by an amplitude effect and not by a phase synchronization process. A suitable solution is the harmonic generation through the production of extra Fourier elements with a nonlinear device. Three types of nonlinear devices have been successfully used for harmonic generation: iron-core inductances, diodes, and nonlinear amplifiers. Since iron-core inductances have a nonlinear relation between magnetizing current and magnetic flux, a properly designed inductance driven by a current generator is a good source of voltage harmonics. However, the fact that they are often low-impedance devices requiring considerable driving power make them inadvisable for use in our multiplication scheme. Diodes or contact rectifiers provide a useful alternative as they can be used in various combinations to obtain waveforms with harmonic content. Since the harmonic generation in diodes is especially sensitive to the polarization state, its control adds extra complexity and less reliability, representing only a solution for very particular situations. However, an amplifier in compression provides the greatest flexibility, since the regulation of the input level gives a fine control of the spurious distribution.

In order to define the amplifiers involved in the multiplication scheme, I performed measurements on diverse amplifiers in their compression regime for different input levels, and retrieved the most optimum input level. The measurements were performed for a sinewave input and a wideband signal, considering the same spectral region that is planned for use in the CTS arrangement. The selection of the amplifiers was based on their 1dB-compression point and their intercept point IP3, both describing nonlinear effects present on the transfer function. The best solution to this matter retrieved from the measurements was a proper combination of two amplifiers in compression. Firstly a good spurious generator with low IP3 and 1dB-CP (1dB-CP=+1.8 dBm, IP3=+11.4 dBm) and secondly a high output signal (1dB-CP=+12.7 dBm, IP3=+22.9 dBm). The results of this setup for the up-converted digital chirp are shown in Figure 2.17, and portray the proper generation of the third harmonics component. Nevertheless, the non-desired odd components (2^{nd} and 4^{th}) are also generated, adding an extra source of noise and interference to the setup.

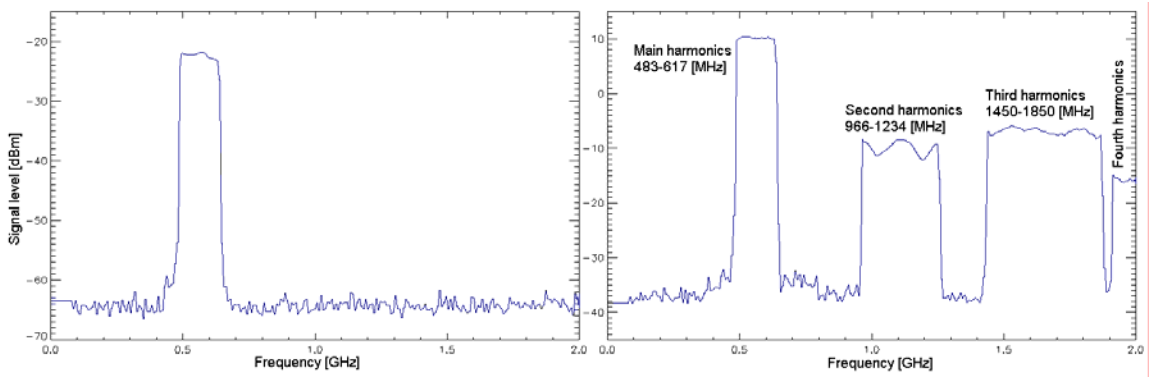

Figure 2.17: Measurements of the input and output signal for an optimized tripler arrangement. [Left] Up-converted digital expander-chirp, which is fed into the arrangement in order to be spectrally multiplied. [Right] Resulting tripled signal, showing the harmonics distribution for the even and odd components.

2.6.1 Spurious and interference

The spectral multiplication process is a very convenient method though it adds extra complexity to the arrangement as undesired spurious are added to the signal. Since the expander chirp is a double domain signal, in other words, having a defined representation in both the time domain (T_e) and the spectral domain (B), its harmonics also have a dual domain representation. This actually represents an advantage as the signal can be filtered in both domains. The spectral filtering is performed via a high performance bandpass filter (1450-1850 MHz), while the time domain filtering is intrinsically performed by the switching process taking place between both transform branches. These harmonic components are transported into the compression region by undesired effects such as intermodulation products from amplifiers, undesired feed-through in mixers and limitations in the filtering process. In order to prevent their influence, I studied the effects induced by each harmonic, which in turn assisted in the definition of the design strategies.

First harmonic components (483-671 MHz)

Due to feed-through during the mixing process of the expander and the input signals, these harmonics appear as interference in the compression region. Even though this signal is strongly attenuated by means of a bandpass filter (70 dB), its influence could play a role in the noise floor, since it is the strongest signal and is directly located in the compressor spectral region. In order to study its influence on the measured spectra I considered not only its spectral location, but also its temporal location. This signal has an expander chirp-rate which is the third of that of the compressor and since it is not yet down-converted, its dispersive properties do not follow the complex-conjugate pattern. This creates a non-compressed signal located in the following dual-domain position (f_1:483 MHz , t_1:14 μs)- (f_2:550 MHz , t_2:0 μs). From this 2D position we can conclude that even though the signal is located in the compression bandwidth, it does not affect the transformation region located between 22 μs and 44 μs. This concept is better described in Figure 2.18.

Second harmonic components (966-1234 MHz)

The second harmonic components are strongly attenuated (50 dB) by a post filtering process taking place after the multiplication. There are no important non-linear processes that could transport these harmonics into the compression region (350-550 MHz), and thus this signal does not produce any strong interference of the compressed signal.

Fourth harmonic components (1932-2468 MHz)

These components are transported into the compression region during the mixing process of the input signal with the multiplied spectra. An input-signal outside the input bandwidth could be mixed with these harmonics, and therefore gives a spectral translation of the 4^{th} harmonics into the compression region. Since both signals are spectrally located near desired signals, their filtering is extreme complex and they are substantial sources of probable interference. However, this signal is located outside the transformation region (f_1:468 MHz , t_1:12 μs)- (f_2:350 MHz , t_2:10 μs), and thus will be filtered out by the time-switching process that takes place when both transformation branches are combined.

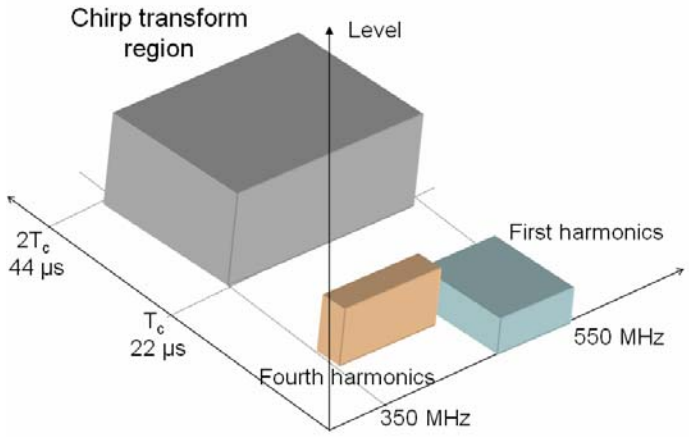

Figure 2.18: Description of the dual domain location for the translated spurious signals. The chirp-transformation region (shown in grey) describes the region where the spectral information of the signal is located, and thus the area which is actually measured. The remaining region where the 1st and 4th harmonics are located does not affect the measured signal, since this region is removed by the time switching scheme.

2.7 Spectral filtering

The spectral filters play an essential role in the signal processing taking place in the spectrometer since they remove the undesired signals generated during the chirp-transformation. Their performance is critical in defining the noise floor and dynamic range of the arrangement. The harmonics generated during frequency-multiplication and the aliased components due to digital signal generation will require high-performance filtering. This is because these harmonics are relatively strong and spectrally located near the desired band. The three processes which demand special care are: 1) the removal of the aliased signals with a steep low-pass filter at 150 MHz, 2) the input signal filtering required to constrain the input signal and the influence of the unwanted mixing of the 4th harmonics (2.0-2.2 GHz), and 3) the suppression of the unwanted harmonic components generated in the frequency-multiplication process with a bandpass filter (1.45-1.85 GHz).

2.7.1 The filter for the digitally generated aliased signals

The polynomial representation and pole distribution of the desired filter will define the filtering properties [see appendix E]. In order to transform the polynomial into a physical filtering network a transformation is needed, which is based on different "topologies". The topology defines how the different reactive elements are interconnected, and thus how each pole of the transfer equation is represented in the arrangement. The lowpass filter in charge of removing the aliased signals is taken to be a Cauer filter with a capacitor-input lowpass topology. There exist numerous publications which describe the mathematical transformations necessary in the conversion of the complex components of the poles into physical values of inductors and capacitors. Based on [Ellis, 1994], I developed an ad hoc mathematical sheet, which calculates the required order of the polynomial function and the pole distribution, based on the definition of the desired in band ripple, stopband and passband regions. These are then transformed into physical values.

Having the coefficients of the lumped elements allows the response of the network to be simulated for available inductors and capacitors. The conversion from calculated values into commercial available parts is a delicate issue, since it could produce a completely different response. Our approach consists of maintaining the properties of the resonant tanks by

simultaneously varying both lumped elements. Because the resonant frequency of the tanks is $\omega_0 = (LC)^{-½}$, an approximation to a commercial value of L → L' would lead to a corrected C'=C·L/L', and thus this leads to a better agreement between the expected and measured results. The development of filters with lumped-elements is of great convenience as they allow a straightforward tuning and their enhancement can be realised by simply correcting each separated element of the resonant node. Nevertheless, their realisation becomes impractical at higher frequencies since the capacitive and inductive parasites induced in the board are comparable to the discrete elements. For example a capacitor pad of 3 mm side in a filter operating at 200 MHz will induce up to 1.5 pF of capacitance in a multilayer board, compared to the desired capacitive value of 4 pF. Consequently, the higher frequency filters are developed via microwave structures such as resonating waveguides.

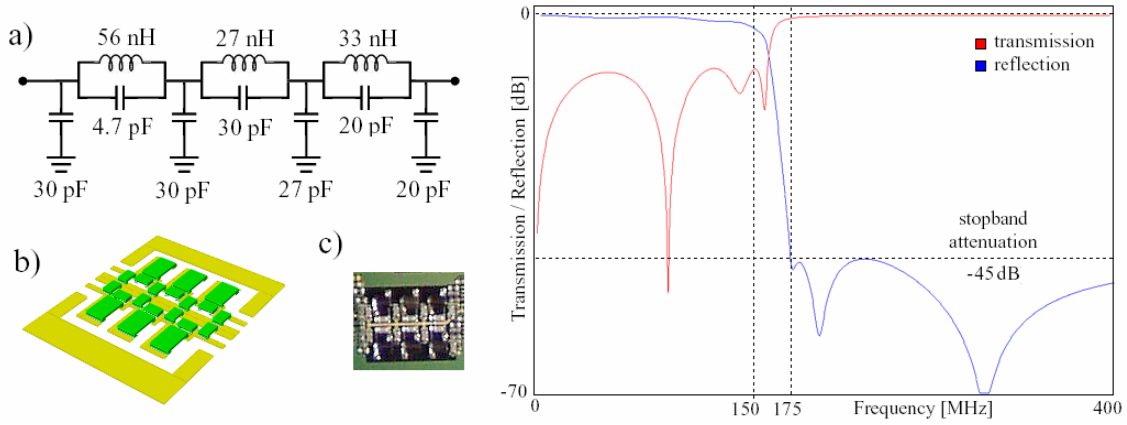

Figure 2.19: Analysis of the 150 MHz elliptical lowpass filter. a) Network description showing the capacitive coupling of the nodal resonating elements. b) 3D simulated view. c) Photograph of the filter as part of the ADCP. [Right] Response of the filter considering transmission and reflection.

2.7.2 Microwave filters with microstrip technologies

A convenient technique of tackling the difficulty of spectral filtering at GHz frequencies is through a proper control of the propagation path of the electromagnetic waves in a dielectric medium. This demanded an intensive study of the electromagnetic field distribution over quasi-dielectric thin mediums (microstrip and stripline), considering TEM and quasi-TEM field distributions [see appendix F]. The microstrip implementation is the most suitable from the manufacturing point of view but it also represents a more complex study of the electromagnetic field since it has TM and TE field distribution. However at a frequency of 6 GHz or lower, the axial E and H fields are small enough that the propagation mode can be approximated to TEM, hence the name quasi-TEM applies. This approximation is expressed by empirical reformulations of the TEM propagation equations. Based on quasi-TEM models described in [Lascari, 2001] and [Mongia et. al, 1999], a mathematical sheet was organized for the calculation of the waveguide parameters "w" (width) and "s" (separation) of the coupled lines (see Figure 2.20). The solving of the system of equations was performed by an iterative solver which requires as input parameters: the operational frequency "f", the coupling angle "ϕ", the medium height "d", the conductor thickness "t", the dielectric constant of the medium ε_r, and the odd and even impedances (Z_{0o} and Z_{0e}) [see appendix F]. The coupling angle refers to the electrical length that is actually coupled; for successive coupled lines the length is λ/4, and therefore

$\phi=\pi/2$. The length of the line is calculated based on a software provided by the material distributor, which through defining the frequency and medium structure, provides the propagation velocity and therefore the wavelength for such a medium.

Once the waveguide physical values are obtained, the filter is organized as shown in Figure 2.20 b), where the $\lambda/4$ coupled-lines of width w_n and separation s_n can be seen. The length of the setup can be easily estimated by $N\lambda/4$, which for a dielectric constant of ε_r =10 would lead to a structure size of ~70mm for a filter at 2 GHz and N=5. In order to reduce the size but keep the number of poles, and therefore the selectivity, the coupled lines are bended as shown in Figure 2.20 c) forming a "hairpin" structure, which for a structure of ~40mm provides the same number of poles, and thus the same selectivity. As can be seen, the propagation path of the resonating waveguide is kept at $\lambda/2 \rightarrow \beta l=\pi$. Nevertheless, the coupling angle ϕ is reduced and this correction is made while performing the calculations for a hairpin geometry, taking into account a coupling angle of $\phi \sim 80^{\circ}$.

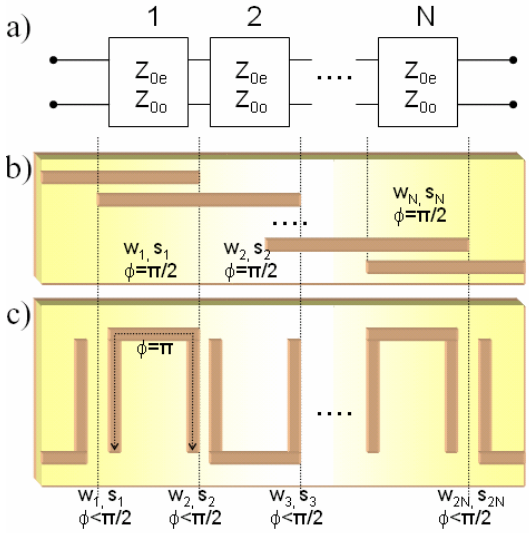

Figure 2.20: Description of spectral filtering implemented on a microstrip medium. [a] Filtering function pole's representation as electrical units of Z_{0e} and Z_{0o}. [b] Steeped impedance filter of order N. [c] Transformation into a filter of order 2N organized as a "hairpin" microstrip filter.

The analysis of the spectral filtering performance is obtained by simulating the filter geometry using the MWO software. MWO is an integrated full wave electromagnetic simulation and optimization package for the analysis and design of three dimensional microwave structures. The simulation can be approached in two manners: 1) by solving the structure as an arrangement of waveguide elements represented analytically, or 2) by solving the complete set of Maxwell equations for a TE and TM medium considering a finite element geometry. The method of dividing the structure into discrete finite elements will provide the most realistic results. This will depend on the meshing geometry which is driven mainly by the relative size of the finite elements in comparison to that of the structure. For a standard structure of 50x30 mm and considering a meshing of 1 mm, the calculation will take approximately 1 hour per analyzed spectral component. For a complete spectral analysis, this would take days. This technique is heavily inconvenient while performing the optimization of the structure, and therefore, the analysis is approached by the analytical method which still provides highly realistic results.

The optimization of the filter structure is achieved by defining the filter geometry and the optimization goals, defined by the passband-ripple, passband-insertion loss and stopband attenuation. This is a process where the filter geometry is automatically adjusted, in order to minimize the error function described by the specified performance goals. The optimization is based on the search of the global minimum by a pointer optimizer which combines the power and robustness of four widely used and accepted search methods: linear simplex, downhill

simplex, sequential quadratic programming, and genetic algorithm. The pointer selects the best combination of optimizers and step sizes in order to find the best solution in that time.
For the realisation of the 1.45-185 and the 2.0-2.2 GHz filters, a high dielectric constant material (ε_r =10) was considered in order to maximize the propagation constant β and therefore minimize λ and the filter size. Even though the manufacturing process was performed with a high resolution laser photoplotter, deviations from the simulated values were found. These were mainly due to material (ε_r,d,t) and microscopic imperfections introduced while plotting and also during the chemical treatment. The measured results were fitted to the simulated values by a proper modification of ε_r and d of the dispersive medium in order to create a bridge between measurements and simulations. The filter was regenerated once the dispersive properties of the material were corrected, obtaining remarkable agreement between measurements and simulations as shown on Figure 2.21.

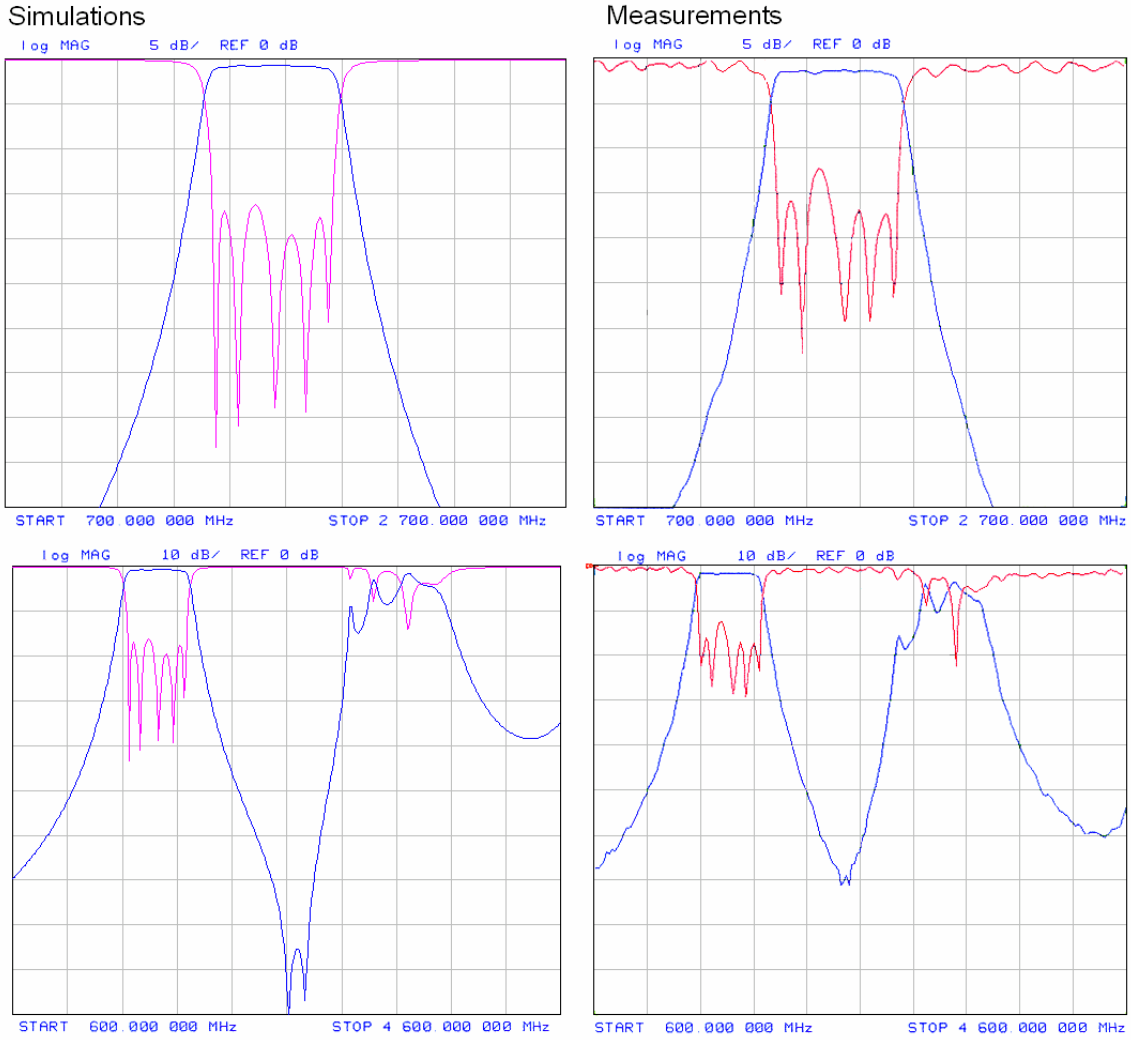

Figure 2.21: Comparison between simulations and measurements for the hairpin filter 1.45-1.85 GHz. The plots show the transmission scattering parameter s_{21} in blue, while the reflection scattering parameter s_{11} in red. [Left] Simulations results for two spectral ranges. [Right] Measurements of the 1.45-1.85 GHz hairpin bandpass filter performed with a 6 GHz network analyzer.

Since the filtering is mainly based on coupling of resonating waves, higher multiples of the resonating waves will also be accommodated in the structure and therefore they will not be strongly filtered. This effect, which can be seen in Figure 2.21 [bottom], is attenuated by using low-frequency amplifiers (f_{max}=2.5 GHz) in the signal processing chain.

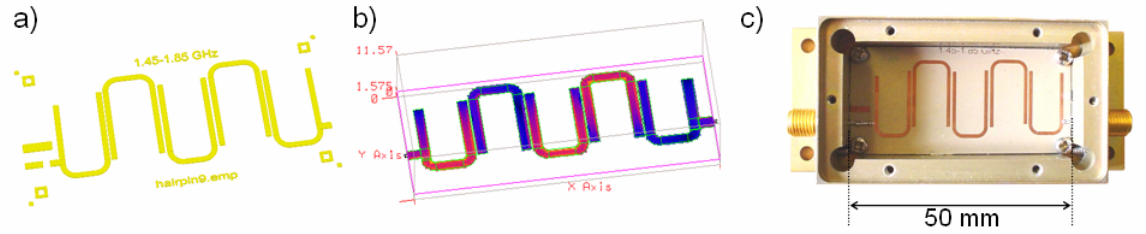

Figure 2.22: Views from the hairpin bandpass filter of 1.45-1.85 GHz. [a] 3D view of the simulated structure. [b] Superficial currents present on a resonant mode of the filter (1.65 GHz), obtained from a finite element calculation. [c] Realisation of the filter in an enclosure of 50x25 mm^2

2.8 Thermal stability

The control and study of the dissipation and thermal generation processes taking place in the spectrometer are of great importance in the analysis of the stability of the instrument. The regulation of these processes will not only provide high isolation of external thermal fluctuations, but will also prevent any possible damage of the internal components due to overheating. The three processes that are specially taken into account are: 1) the heat dissipation taking place in the ADCP kernel (DDS) in order to prevent overheating and destruction of the device; 2) the regulation of the environmental conditions of the SAW devices, since they are particularly sensitive to temperature variations, and 3) the heat exchange processes defining the internal instrument temperature and the instrument's isolation to external thermal fluctuations.

2.8.1 The ADCP Processor

The fact that the ADCP runs at high frequencies means it will risk overheating the device. Consequently, a study was taken to prevent damage to the internal silicon juncture. The performance reliability and the life expectancy of the DDS core are inversely related to the component temperature of the equipment. The relationship between the reliability and the operating temperature of a typical silicon semi-conductor device shows that a reduction in the temperature corresponds to an exponential increase in the reliability and life expectancy of the device. Each DDS device of the ADCP consumes P=5W and therefore an ambient temperature of T_A=27ºC, a thermal resistance of surface-component to the air of θ_{CA} = 2ºC/W, and a thermal resistance of the juncture to the surface-component of θ_{JC}=14ºC/W, would lead to a juncture temperature of T_J = 107º C [$T_J=T_A + P_J(\theta_{CA}+ \theta_{JC})$]. Knowing that the destruction limit of the silicon juncture is 150ºC, an improper combination of higher power consumption and a raise in the ambient temperature would lead to possible damage in the internal structure of the device. The life expectancy of the device and therefore the reliability of the instrument can also be enhanced by reducing the juncture temperature. For this, a heat sink is placed over the DDS and therefore the θ_{CA} no longer plays a role in the heat transport. All the heat is then transferred from the case to the ambient surroundings by the new dissipater with θ_D = 8 [ºC/W]. This will lead to a

juncture temperature of T_J = 77º C which is far below the device limit of 150ºC. However in order to prevent unwanted situations, the juncture temperature is constantly monitored by the ADCP microcontroller. This juncture temperature, which is retrieved by measuring the ambient temperature near the device, is compared to an operational limit (now set to 120ºC). A measurement over the limit will send an alert command by its control interface and switch off the DDS in order to prevent any possible damage in the silicon juncture of the device.

2.8.2 SAW filters

The most sensitive elements to temperature variations of the complete CTS arrangement are the SAW devices. As can be seen in equation (2.32) a variation of 1ºC will produce a deviation of approximately 140 ppm in the dispersive properties. Therefore, this would lead to an unacceptable deviation of 1400 ppm when considering an environmental variation of $\Delta T = 10ºC$ that is expected inside SOFIA's cabin. A proper thermal scheme must be implemented to prevent such unacceptable variations.

The controlled thermal environment is achieved by a combination of a controlled heater and a highly thermally isolated environment. A faster reaction of the heater and a higher isolation will lead to superior stability. Nevertheless, a faster reaction requires higher power consumption due to the fact that the energy lost E should be delivered in a time Δt through a dissipated power $P = E/\Delta t$. It is clear than when $\Delta t \rightarrow 0$ the power consumption increases enormously. It is therefore of strong importance that a good thermal isolation of the SAW devices is achieved, which is implemented in this case with a thermal chamber made of aluminium, polystyrene and polyimide film. The SAW devices are embedded in Styrofoam® (a form of polystyrene), and thermally isolated by several layers of a thin polymide film (Kapton®). The low thermal conductivity of polystyrene and the high thermal isolation provided by the polymide film and the external aluminium structure achieves high isolation, up to 40 times stronger than a non-isolated environment. The temperature inside the chamber is regulated at 40ºC above the average external temperature in order to provide superior thermal control to the heating device. The heater is a miniature proportionally controlled thermal regulator whose temperature can be programmed up to 100ºC and has a Beryllium base for good thermal conduction.

In order to test the performance of the enclosure, thermal analysis was prepared considering a varying external environment implemented with a controlled thermal chamber. Several parameters were continuously measured for different environmental temperatures: 1) instantaneous power consumption (CURR), 2) temperature of the SAW devices (SAW1,SAW2), 3) temperature of the air (AIR), and 4) temperature of the external aluminium enclosure (BOX). The measurement consisted of varying the environmental temperature from 14ºC to 47ºC in steps of 3ºC, and monitoring the evolution of the thermal enclosure parameters. The thermal chamber takes approximately 60 minutes to achieve the desired temperature, which is why there are oscillations during the stabilization of the chamber temperature, as can be seen in Figure 2.23. This in turn also leads to a fluctuating set of parameters. Since the thermal isolation can be thought of as a temperature integrative system, an increase in isolation will increase the delay between the stimulus (air temperature) and the system response (SAW's temperature). The left of Figure 2.23 shows that the external enclosure is the first that detects the external fluctuations with a ΔT_{max}=0.68ºC.

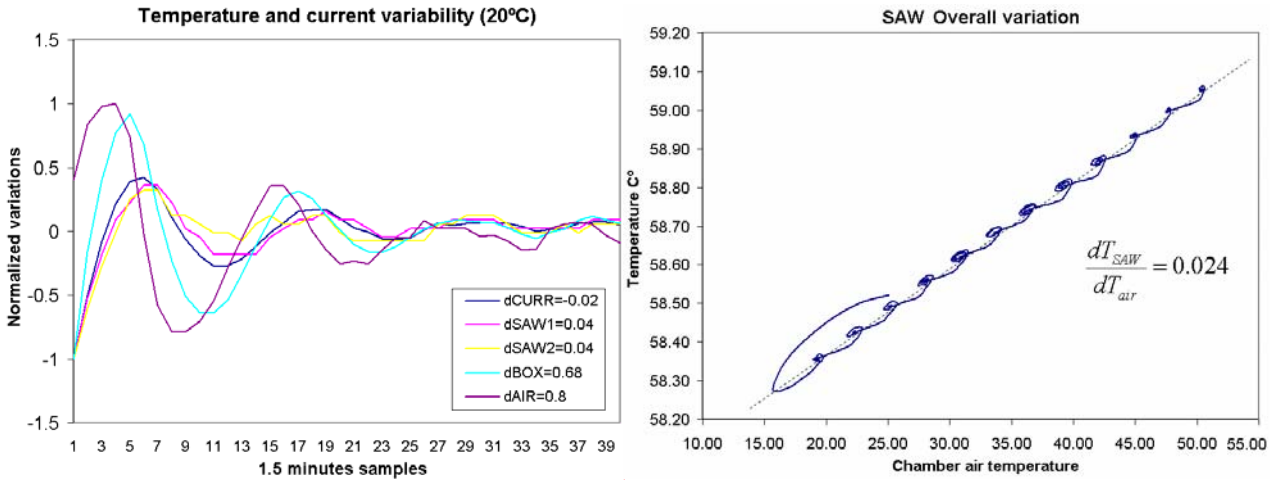

Figure 2.23: Response of the thermal enclosure to external environmental variations. [Left] Evolution of different parameters according to a fluctuating external temperature represented by the purple trace (dAir). This is an example measurement for a defined chamber temperature of 20ºC [Right] Response of the SAW's temperature with respect to the environmental temperature (chamber air temperature). A fitted line shows the dependence between both parameters, expressed by the thermal 1st derivative $dT_{SAW}/dT_{AIR}=0.024$ which produces a thermal isolation factor of 40.

The heater device later responds to these fluctuations in an inverted manner in order to maintain the thermal equilibrium of the SAW devices, which only suffer fluctuations of $\Delta T_{max}=0.04$ºC. The overall performance of the thermal enclosure, as described in Figure 2.23 right, shows that the case of SOFIA with a $\Delta T_{max}=10$ºC will lead to a SAW temperature variation of $\Delta T_{SAW}=10\cdot0.024 = 0.24$ºC. This is remarkably low and represents a maximum dispersive variation for the SAW devices no greater than 70 ppm (based on equation (2.32)) for the entire operating temperature range.

2.8.3 Spectrometer's thermal stability

The processes of cooling and heating play an important role in the design of the instrument, since they define the thermal stability of the system and therefore the overall stability of the spectrometer. Each component involved in the chirp-transformation is sensitive to temperature variations, as its electrical properties are sensitive to the environmental temperature. This is the case for amplifiers, attenuators, electrical resistors, dielectric materials of the filters, etc. The sources of heating are the surrounding environment and the dissipated power of the instruments, while the sources of cooling are diverse and is the main subject of this section. Cooling occurs via convection, conduction, radiation, evaporation and active cooling. Each cooling approach has advantages and disadvantages; for example evaporation and active cooling with refrigerant fluids are extremely convenient as they provide high efficiency and constant heat flow. Nevertheless, this method requires special pumping components which demand a higher power consumption and a complex pressure control system to prevent refrigerant leakage into the airplane's cabin. In the case of SOFIA-GREAT-CTS, the cooling approach was intended to be based mainly on passive cooling (radiation and conduction) and on forced air convection. The spectrometer's maximum DC power consumption is 38 W and for a power supply with a transformation efficiency of $\eta=0.8$, will require a total AC power of 47.5 W. In the calculations a cabin

temperature of 22ºC (T_{air}=295ºK) is assumed along with a maximum instrument temperature of 40ºC (T_{sys}=313ºK).

Dissipation by radiation

The radiation phenomena is only present in the instrument in the front and back plate, having an effective radiative area of A_{rad}=2·80e-3·440e-3=0.07 [m^2]. The Alodine 1200 chemically treated aluminium surface on the instrument's enclosure will lead to an emission efficiency of ε=0.3. Taking into account the previously defined parameters and using Stefan-Boltzmann's equation, the power dissipated by radiation on SOFIA-GREAT-CTS is determined to be only P_{rad}=2.4 W. If all the dissipation is assumed to be radiative, the inner temperature can be calculated as follows:

$$T_{in} = \sqrt[4]{\frac{P_{diss}}{\sigma \varepsilon A_{rad}} + \left(T_{air}\right)^4} \qquad (2.43)$$

where σ is the Stefan-Boltzmann's constant, T_{air} is the air temperature and P_{diss} the dissipated power. Applying the case of SOFIA-GREAT-CTS to equation (2.43) gives an inner temperature of T_{in}= 467ºK (194ºC). Since this value is highly unacceptable the dissipation is combined with a forced convection dissipation process.

Dissipation by forced air convection

The required air flow in a closed environment is proportional to the air properties, the dissipation power and the area [see Villanueva, 2002]. SOFIA's cabin environment is pressurized corresponding to an altitude of 2400 m, or 0.7 bar and the air humidity is reduced to less than 10%. This dry air condition will have a Prandtl number of Pr=0.71, a kinematic viscosity of σ=17e-6 [m^2s^{-1}] and a linear heat conductivity of λ=0.0271 [$WK^{-1}m^{-1}$]. The convection area was purposely situated near the hotter elements, i.e. the power supply and the electronics, which encompass an effective area of A_{conv}=0.35·0.15=0.05 [m^2]. Considering also that the cooling process length is the distance between the input fan and the outflow holes of the backside (l=0.35 [m]) and that the area of the fan, defined by the height of the back plate, is A_{fan}=π·0.03^2=2.83e-3 [m^2], the required air flow is then calculated to be v=12 m/s and CFM=7 [see equations in Villanueva, 2002]. As the above calculations are performed considering sea level pressure, a correction should be carried out to take into account the reduced air pressure present in the airplane cabin. A fan will move the same number of CFM regardless of altitude or temperature, however mass flow will vary due to changes in pressure/resistance. Based on [Cheng, 2002], a correction factor for the calculated air flow was obtained, which takes into account the low pressure present in SOFIA's cabin (η_{conv}=0.8). Considering the above mentioned parameters an inner temperature of T_{in}=309ºK (36ºC) is derived. This is relatively low; more so if compared to an unforced convective dissipation process, as described in the radiation section which gives an inner temperature of 194ºC. The advantages of forced air convection are clear. However, this strong cooling will also produce a lower thermal isolation of the instrument to the fluctuating external environment. As the most power consuming components are the electronic circuitry and the power supply, the air flow in the spectrometer (and thus the convective dissipation process) is organized to affect mainly these regions. The RF components on the other hand are thermally isolated through an aluminum wall, since their dissipation is relatively small and thus designed to be mainly radiative. Through the defining of different thermal regions, we are able to enhance

different dissipation processes and herewith the thermal stability of the instrument, additionally achieving higher reliability and instrument lifetime.

2.9 Structural analysis

The installation of the instrument on a flying observatory demands that is able to support the periodic high accelerations present in the airplane, and more importantly to prevent any structural damage that could become a hazardous event for the onboard scientists and flying crew. According to this, the instrument has to be certified by the FAA (Federal Aviation Administration) which is concerned with the safety of personnel associated with flight and the certification of all aspects of a flying aircraft. The certification process involves several reviews and iterations of drawings, documentation, procedures and analyses, which culminated in an extensive and detailed set of documents [see Villanueva, 2003a]. This describes all the safety issues related to the instrument, focusing mainly on structural analysis. These documents form the CADR (Critical Airworthiness Design Review) document package, which consists of six sections: 1) instrument description, 2) structural analysis, 3) electronic design note and 4) functional hazard analysis. They are developed based on FAA, NASA, and DLR regulations under the scheme of the ISO (International Standards Organization) 9001 guidelines (quality management system procedures). Since the documentation is reviewed by an external committee, all the calculations must be based on analytical formulations. Complex computational analysis such as finite element calculations are not allowed. This demanded a singular study of the numerous forces present in the instrument, especially the ones related to the supporting structure which is the spectrometer's enclosure. This section describes the structural analysis of the instrument and the mechanical stability calculations obtained from an ad hoc developed set of simple analytical formulations which take into account the geometry of the instrument and SOFIA's present accelerations.

2.9.1 Structural stability

The structural stability of the instrument is strongly defined by the distribution of the forces. The location of the instrument's centre of gravity is therefore of great importance. For that reason the complete instrument's structure was designed using CAD [see Villanueva, 2003a], obtaining not only valuable information from the expected instrument's volume and weight but about the distribution of forces on the supporting structure. The spatial and mass information of each component was used to determine the overall instrument's centre of gravity, based on the following equation:

$$\bar{x} = \frac{\sum_{i=1}^{n} m_i x_i}{\sum_{i=1}^{n} m_i} \qquad (2.44)$$

where m_i is the mass of the component, x_i is the three-dimensional location of the component, n is the number of components and x is the centre of gravity. After a careful organization of the spectrometer's components, the centre of gravity was calculated based on equation (2.44). The centre of gravity is located at x=7, y=7, z=16 [mm] away from the geometrical centre, representing a relative displacement in the axis of support (z) of (0.03,0.03), which is particularly low. The spectrometer is installed on the PI-rack (in SOFIA's cabin) as a 19" inch mountable

rack unit, where the axes of acceleration are "z" for the vertical axis (orthogonal to earth's surface), "x" for the flight coordinate (direction front-back in the plane) and "y" the wing coordinate (direction left-right in the plane). This agreement between the centre of gravity and the geometrical centre shows that the momentum forces are equally distributed on the supporting structure. Therefore, the mechanical stability of the instrument in the mounting structure is sustantially enhanced. Please refer to Figure 2.24 for a representation of the instrument's geometry.

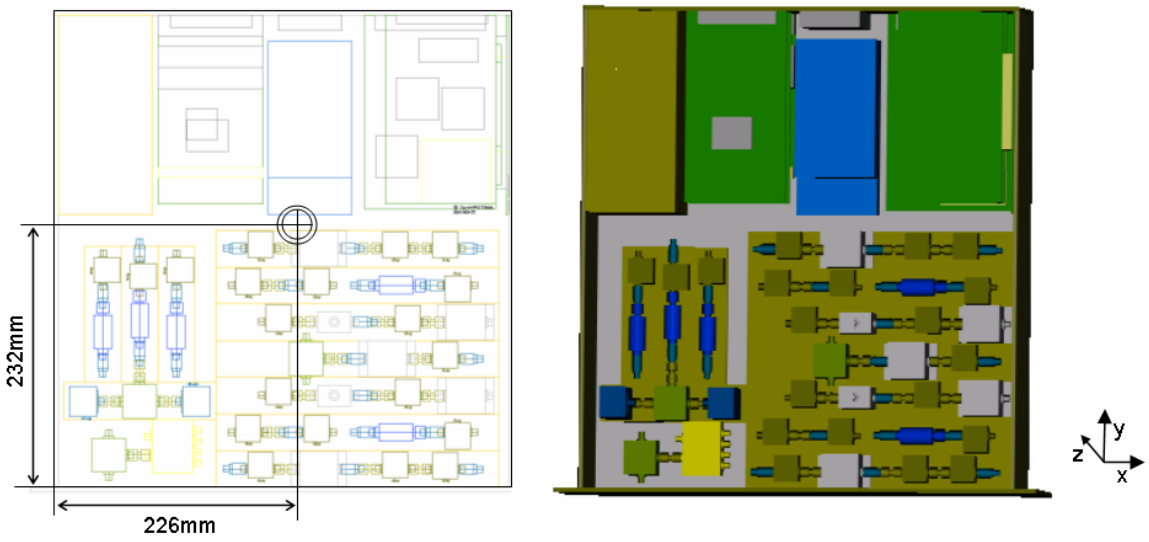

Figure 2.24: Three dimensional representation of the instrument based on a computer aided design. [Left] Centre of gravity location considering a top view (xy plane). [Right] Volume filled image of the instrument, where the various RF components in the front of the instrument can be seen, and the electronics, power supply and SAW enclosure in the back side.

2.9.2 Supporting structure analysis

The main concern of the FAA is that an unwanted structural damage does not produce a hazardous event in the airplane's cabin. Therefore, the structural analysis of the instrument is based mainly on the possibility of a component mechanically breaking free because its supporting structure breaks. The applied force is defined to be a punctual force on the containing walls of the instrument's enclosure. The present accelerations on SOFIA's aircraft are: 9g forward (+x), 1.5g reverse (-x), 6g downward (-z), 3g upward (+z) and 3g lateral (±y).
Due to safety reasons and the instrument being portable and located in a rack, the maximum load of 9g is assumed for all axes in the calculations. The calculations consider various effects: the maximum tensile strength and bearing check of the containing aluminium plates of the enclosure, and shear, tension and pull-out checks of the supporting screws. The minimal resistance moment (W) for a flat plate of uniform thickness (h) and height (b) is [see Friedrich, 1999]:

$$W = \frac{b \cdot h^2}{6} \qquad (2.45)$$

Considering a plate fixed at both ends, the resistance moment will be:

$$W = \frac{F \cdot l}{4F_{tu}} \qquad (2.46)$$

where l is the length of the plate, F the applied force and F_{tu} the tensile strength of the material. Combining the above two equations, the maximum force that can be applied to the plate can be derived.

$$F_{\max} = \frac{4}{6}\frac{bh^2 F_{tu}}{l} \qquad (2.47)$$

Once the force is obtained, the maximum weight that can be applied to this point can be determined considering a 9g acceleration:

$$M_{\max} = \frac{F_{\max}}{9g} \qquad (2.48)$$

Solving equations (2.47) and (2.48) for SOFIA-GREAT-CTS and considering a fitting coefficient of 1.15, high margins of safety were derived with more than +5 for the complete instrument's enclosure. The analysis of the maximum stress the bearing and bolts can withstand was mainly calculated based on SOFIA's handbook and high margins of safety were also obtained; more than +12 for the worst case scenario of the punctual force of the heaviest element being applied directly to only one of the supporting bolts [for all the calculations please refer to Villanueva, 2003a]. These high margins of safety combined with an equilibrated centre of gravity ensure that the instrument is secured from hazardous effects of loose parts, and therefore fulfils the structural requirements defined by the FAA certification.

2.10 Control and data-acquisition

The control and data-acquisition unit of the spectrometer is composed of the following components: 1) the instrument's electronics, 2) the interface to the main control unit and 3) the main control unit as an embedded computer. The core of the instrument's electronics is represented by the CTS-ASIC processors, which are in charge of the data acquisition and the control of the instrument's complex timing scheme. These processors, which are specially developed for the Chirp Transform Spectrometer, sample the complex signal coming from the analog part of the spectrometer, compute the spectral power and integrate the results in a periodic manner. When defining control and data-acquisition interfaces, an important requirement is to provide a standard and universal interface resulting in a superior independence between the architecture of the instrument and that of the main control unit. The instrument's electronics are connected to the control unit through an ad hoc developed interface based on ISA (Industry Standard Architecture) guidelines, a highly standardized interface for interconnecting hardware components onto a personal computer.

2.10.1 The pre-processor

The pre-processor is the first stage on the data-acquisition process and therefore is the bridge with the analog part from the spectrometer. The CTS-ASIC processors are digital components with an integrated defined logic that are parameterized via programmable registers. Each complex component of the analog signal is read by two ADC converters, which run in interleave at an operational frequency half that of the nominal sampling frequency (f_s) with a conversion resolution of 8 bits each. Therefore, each pre-processor unit reads out the complex analog signal with 4 ADC converters at a data-rate of $DR=4\cdot8\cdot f_s/2$. For the case of SOFIA-GREAT-CTS with a clock frequency of 175 MHz, a high data-rate of DR=2.8 Gbps per pre-processor is needed. The

properties of the sampled signal are constrained by the clock frequency of the sampling process and the analog-digital conversion resolution bits. The analog-digital conversion has the same physical limitations as the digital-analog conversion, as described previously for the chirp-generation process (section 2.4.2). This means that the maximum dynamic range of the sampled signal depends on the resolution bits of the conversion process, following equation (2.37), which for the case of the pre-processors with an 8-bit resolution would lead to a dynamic range of ~50 dB (operating at full scale).

The sampling frequency is constrained by two conditions: 1) it should be greater than the compressor's bandwidth in order to properly measure the signal (f_s>B), and 2) the amount of measured channels can not exceed the available channels (N) in the electronics memory (f_s<N/T_c). Since the data-acquisition unit and the ADCP share the same clock reference because of synchronization purposes, the maximum sampling frequency is defined by the maximum operational frequency of the ADCP. However, at a sampling frequency of 350 MHz the second condition is not met (f_s>N/T_c). This is because the maximum measurable channels with the CTS-ASIC processor are N=4096 and the SAW compressors chirp-length is T_c=22μs. Therefore, a sub-multiple of the ADCP clock reference is used (f_s=f_{ADCP}/2=175 MHz) which fulfils the second condition but not the first one (f_s<B). A solution to this is the arrangement of two pre-processor units running in interleave at f_s=175 MHz each, which can be seen as a unique pre-processor unit with N'=2N (8192) and f_s=350 MHZ → DR=5.6 Gbps. The implementation of this approach demanded a complex synchronization scheme that is able to handle the nanosecond time differences present between the two running pre-processors and the 8 DAC converters. The measured levels are then squared and properly interleaved in order to obtain the spectral power component, which is the sum of the squared complex components. One important capability of the CTS-ASIC pre-processor is that it also performs the integration of the signal (defined by the integration time nΔT), producing a lower transfer-rate necessary between the pre-processor and the main control unit. The high data-rate of 5.6 Gbps is therefore only handled by the set of two low-power consumption ASIC devices as shown in Figure 2.25.

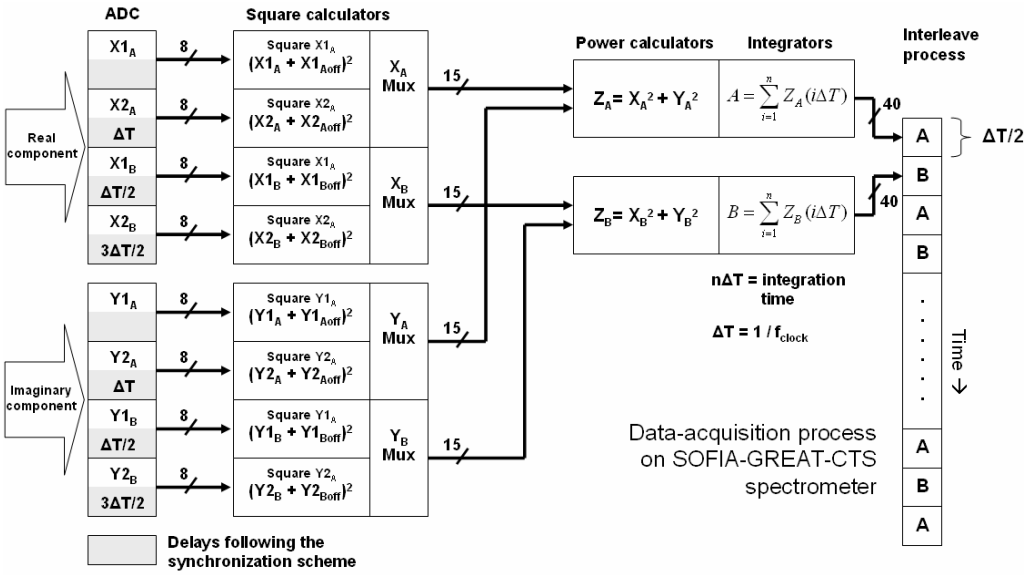

Figure 2.25: Scheme of the data-acquisition process taking place on SOFIA-GREAT-CTS. The complex analog signal is acquired through 2x4 DACs running in interleave, where the signal is later numerically squared, integrated and temporally spaced through a proper interleave mapping.

Deviations from the zero point of the ADC converters are corrected via the offset registers (X_{off}, Y_{off}). These are programmed during the start-up calibration process, leading to the removal of any unwanted continuum component of the signal. The pre-processor is also in charge of the synchronization and timing scheme of the complete spectrometer, which is one of the most critical issues since the spectral information is content in the time domain. The timing scheme handles the following issues: 1) expansion start, 2) compression end, 3) switching between branches, 4) spectral sampling synchronization and 5) integration timing. For the pre-processor, each time step ΔT represents a fraction of the time domain, and therefore a fraction of the measured spectra. The measured signal is divided in successive spectral fractions $\mu\Delta T$ (where μ is the chirp-rate of the compressor), and consequently the number of channels N should follow $N \cdot \mu\Delta T = B$. While sampling the time $N \cdot \Delta T = T_c$, the pre-processor would actually read one complete spectra and therefore the time delay between expansion start and compression end is defined to be T_c. Since the analog part of the spectrometer provides the spectral transformation of the two branches, each [expansion start - compression end] cycle is repeated for each branch and thus the time between switching signals is also T_c. The time bandwidth product of the compressor filter is defined by $TB=B \cdot T_c$, while the measured number of channels by $N=fs \cdot T_c$. Considering the case when f_s is equal to the compression bandwidth B, then TB=N and the number of channels read is perfectly equal to the time-bandwidth product of the setup. This means that each ΔT component is actually a spectral channel ($\Delta f=B/TB$). However, when there is a difference ($f_s \neq B$), as is the case of SOFIA-GREAT-CTS, each measured ΔT component is then a fraction or a sum of channels, described as the following:

$$\Delta f = \left(\frac{1}{T_c}\right)\frac{B}{fs} \qquad (2.49)$$

Since the effective sampling frequency of the setup as shown in Figure 2.25 is f_s=350 MHz > B, the measured spectral channel spacing is then Δf=26 kHz. In order to transform the measured spectral channels ($N=T_c \cdot f_s=7500$) into the real channel spacing of the expander-compressor arrangement (TB), an interpolation process is performed by the main control unit, which converts the measured 7500 channels into the real spectral channels. This oversampling of the spectra also represents an important advantage as the instrument has a higher flexibility in defining the channel spacing.

The interface between the pre-processor and the embedded control unit is performed by an ad hoc developed ISA-ASIC interface, which has as its main component a Field Programmable Gate Array (FPGA). The logic integrated in the FPGA provides read/write access to the control and data acquisition registers of the CTS-ASIC and also generates an interrupt signal on the ISA bus when an integration finished event occurs. An interface developed for the ISA architecture not only provides a standard interface to the main control unit but also a high speed one, which at a 10 MHz bus speed allows retrieving of the 7500 channels of 40 bits within 130 ms. The complete control and data-acquisition of the spectrometer is mapped through the interface as an addressable memory region on the main control unit memory space.

2.10.2 The main control unit

The main control unit, represented by an embedded computer, is the last stage in the signal processing of the spectrometer and therefore is the unit which provides external access to the instrument. The computer is a PC104-CPU module which is intended for embedded applications,

satisfying reduced space and power constraints. Within just a 91.4 x 96.5 mm PC104 standard form factor, the module has the full architecture, hardware and software compatibility similar to the PC bus. Furthermore, the unique self-stacking PC104 bus makes the modules highly integrated and flexible and eliminates the cost and bulk of backplanes and card cages. The access to the instrument is through the ISA bus, as previously described with the ISA-ASIC interface, while the external interface is the Ethernet port (with TCP/IP protocol), having as its redundant port the serial interface (RS-232). The operating system is a hard real-time Linux system that runs as its lowest priority execution thread (RTLinux). The Linux thread is made completely pre-emptible so that real-time threads and interrupt handlers are never delayed by non-real-time operations. Real-time threads in RTLinux can communicate with Linux processes via shared memory or a file-like interface, so real-time applications can utilize of all the powerful, non-real-time services of Linux. An important advantage of RTLinux is that it supports real-time interrupt handlers and real-time periodic tasks with interrupt latencies and scheduling jitter close to hardware limits, which are no more than 15μs for a standard x86 PC. This real-time capability allows instantaneous detecting events from the pre-processors, producing an enhancement of the efficiency of the observational time. This is of extreme importance in highly switching observation modes such as "position-switch + beam-switch" since a faster access to the instrument produces a higher observational switching capability.

The operating system and the main control software are integrated in a fast non-volatile memory disk (based on NAND flash technology), which provides higher reliability compared to standard hard-disk drives as flash disks do not have moving parts and thus are less sensitive to the high accelerations present in SOFIA. The main control software is a combination of a real time synchronization module, a kernel module that communicates to the spectrometer and a control application which monitors and commands the synchronization and hardware modules. For the synchronization module a real-time Linux thread is considered. This continuously runs in the background and is started by the control application when a precise timing scheme is needed, such as the case of switching observations. These timing events have direct access to the instrument through the hardware module which maps the spectrometer's data banks and command-registers via a Linux device. This software access to the instrument is implemented with a char device driver based on [Rubbini, 1998; Börner, 2003, priv. comm.] and coordinates the data read-out and the arrival of asynchronous interrupts. When an interrupt event occurs, the device driver in a real-time manner reads the data from the pre-processor via the ISA-ASIC interface, acknowledges the handling of the interrupt and stores the information (including data headers: times, integration counters …) in a cyclical buffer. This information is then handled, calibrated and eventually sent through the Ethernet port by the main control application, but in this case, in an asynchronous and non-real-time manner. This important feature provides a proper isolation between the timing of the instrument and any unwanted external time latencies, typical of network connections.

The logic of the main application is defined as "event-driven", meaning that any external command (through the Ethernet or the serial port) would lead to a defined sequence of actions. These actions are commands, such as start, stop, init…, or parameterization actions which define the actual observations conditions, for example integration time and channel spacing. For "quick-look" purposes a set of graphical commands is implemented additionally, which is of great convenience especially in the calibration and optimization phases.

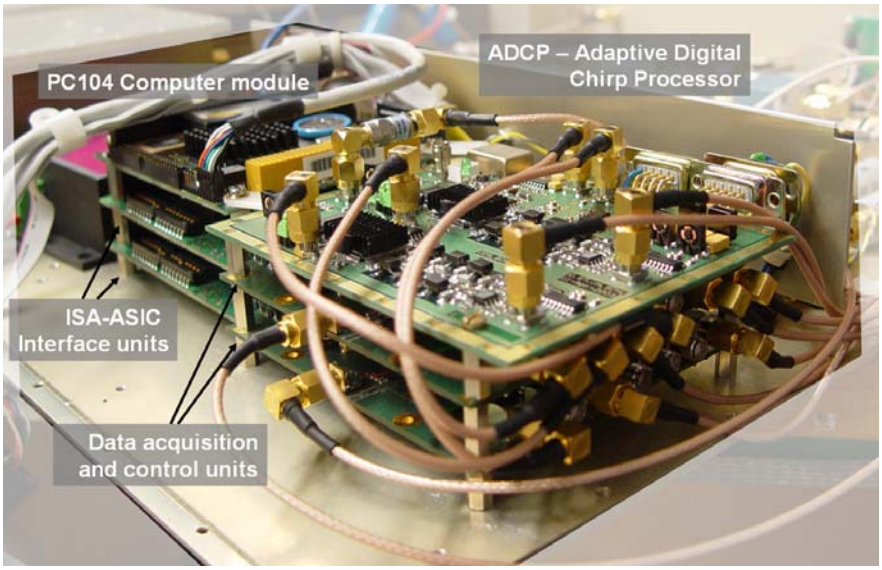

Figure 2.26: Image of the digital units of the SOFIA-GREAT-CTS instrument. Below the PC104 computer module, the two ISA-ASIC interfaces are connected through the special PC104 bus interface. The data acquisition and control units are handled by the computer by these interfaces. The ADCP processor is located above.

2.11 Characterization of the instrument

All the parameters described here are based on real measurements. Nevertheless, in each section a review of possible approaches is given in order to enhance this parameter and cater for further developments. Science is a never ending process of learning, so with SOFIA-GREAT-CTS we introduce a new method for achieving high resolution spectroscopy, but these developments are evolving constantly and we expect to introduce new and better technologies onboard SOFIA during the mission lifetime. The measurements were performed based on procedures and software tools which were previously developed and tested for the characterization of the analog MIRO-CTS [Jarchow & Hartogh, 2004, priv. comm.]

2.11.1 Dynamic range

The dynamic range describes the ratio between the maximum and the minimum input signal that produces a linear output. The maximum input signal is constrained by the compression point of the different passive and active components. This is the region where the outcoming signal is no longer linearly related to the incoming signal. The minimum input signal is mainly constrained by the noise floor, the sensitivity of the instrument and the particular properties of each device. In order to obtain a high dynamic range, a rigorous RF planning was done to prevent mismatching between the different dynamic ranges of the involved devices. The instrument can be divided into two blocks: the RF signal processor and the digital data-acquisition part. The maximum signal in the RF part is the one which drives the main mixer (+10 dBm) while the lowest is the output of the SAW device (-60 dBm). This large level difference entails a complex level handling and represents a complicated issue when evolving the signal from these extreme cases. In addition, a proper matching should exist between the dynamic range of the RF unit and the data-acquisition block in order to obtain the most of the setup. The spectral data is measured by the data-acquisition unit in an interleaving manner, and therefore the information of the dynamic range is processed taking into account the differences between the ADC devices. Each data-acquisition board reads 2 branches, with a total of 4 branches for the complete unit. The measurement requires: a) +10 dBm output power noise source with 1.5 dB flatness over the selected range [2.0-2.2 GHz], b) a high accuracy RF step attenuator with ±0.01 dB error

considering its internal calibration table at 2.1 GHz, and c) a computer behaving as the control unit. The measurement consisted of 60 iterations with 1 dB steps, an integration time of 1 second per position and a total measurement time (including overheads) of 70 seconds. The results in Figure 2.27 show a ±1dB dynamic range of 30dB and a ±0.1dB dynamic range of 18 dB with an optimum input power level of -35 dBm.

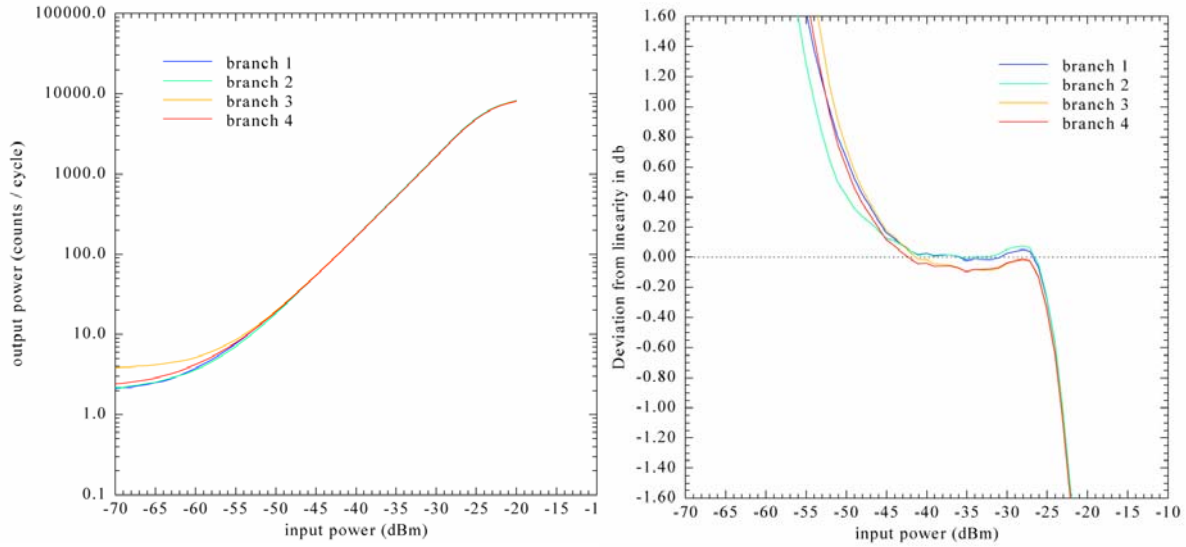

Figure 2.27: Dynamic range measurement of the SOFIA-GREAT-CTS instrument. [Left] Relation between output power counts per cycle and input power in dBm for the 2.0-2.2 GHz region. [Right] Deviation from linearity in dB derived from a linear fit of the curve on the left. The dynamic range information is provided for the 4 branches of the data-acquisition unit.

2.11.2 Bandwidth and center frequency

The instrument's bandwidth is strongly related to the time bandwidth product of the expander-compressor setup and the spectral response of the SAW devices. Higher spectral bandwidth is mainly restricted by manufacturing processes as the groove spacing decreases. The bandwidth and the center frequency are related through the fractional bandwidth (B/f_c), which for signal processing and SAW performance should not be less than 40%. A value higher than 40% causes difficulties in the RF design because the feed-through between IF, LO and RF during the mixing processes would produce unwanted interferences. It would also be extremely complicated in the design of SAW devices to maintain constant parameters (such as insertion losses, input coupling, etc) for the complete frequency range with a fractional bandwidth greater than 50%. Once the fractional bandwidth is defined, a greater bandwidth would demand a higher center frequency. At higher frequencies the structures become smaller following the rule $\lambda=v/f$ (where v~3000 m/s). At an input frequency of 1 GHz (representing a maximum spectral bandwidth of 400 MHz) the structures would be in the micron-size scale. In order to achieve high time bandwidth product devices, extraordinarily complex structures with thousand of grooves should be created. For phase deviations smaller than 1º rms, nanometers accuracy would be required. New developments in nanometer lithography are producing encouraging results that would lead in the near future, SAW devices with center frequencies of up to 5 GHz (BW = 2 GHz) and time-bandwidth products greater than 10000 for smaller bandwidths [Hartogh, 2003, priv. comm.]. The SAW devices on SOFIA-GREAT-CTS have a fractional bandwidth of 47% at a center

frequency of 450 MHz, giving an effective bandwidth of 215 MHz. The center frequency of the CTS instrument is 2100 MHz (450 MHz (compressor) + 3·550 MHz (expander)).

2.11.3 Amplitude stability

The stability of the instrument is an important parameter as the instrument is normally used to detect weak signals through long integration times. The radiometer formula is valid as long as the characteristics of the instruments are constant during the observation time. In order to quantify the stability of astronomical equipment, Allan variance measurements are normally used [e.g. Schieder & Kramer, 2001]. The Allan variance time of the instrument describes the time-frame when the characteristics of the instrument are kept constant, and the noise decreases following the radiometer formula. After the minimum of the variance (described as a time) the internal fluctuations of the instrument are stronger compared to the integrated white noise. The longer the Allan-variance time, the longer one can integrate on a single position without performing a calibration. This parameter is strongly sensitive to temperature and power supply variations. 22000 spectra with an integration time of 1.5 seconds each were obtained for the measurement of the instrument's stability, giving a minimum at k=340 and therefore an Allan variance time of 510 seconds [see Figure 2.28].

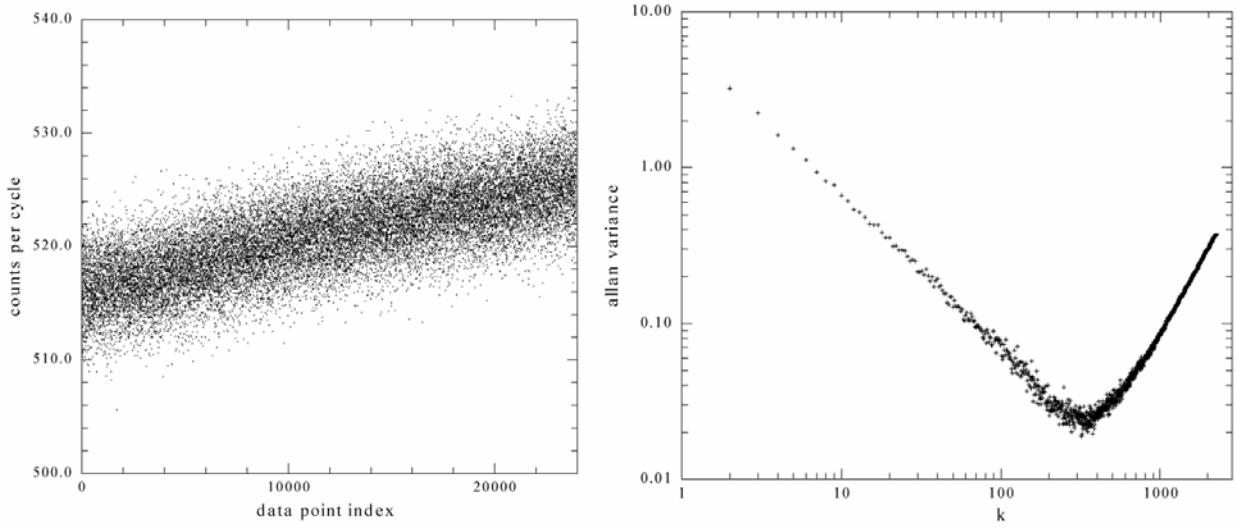

Figure 2.28: Allan variance measurement for channel 3700 of the spectrometer. [Left] Evolution of the counts per cycle for 22000 measurements with an integration time per data point of 1.5 seconds. [Right] Allan variance analysis of the measured intensity variability.

2.11.4 Spectral stability

The spectral stability of the instruments is defined by three different aspects: 1) the stability of the dispersive properties of the SAW devices, 2) the stability of the local oscillators used for up-conversion of the expander chirp and the down-conversion of the compressed signal, and 3) the stability of the propagation delay of the data acquisition electronics. For instance a fast operational gate has a propagation delay between 4 and 7 ns, and this value is strongly sensitive to the ambient temperature. With a channel spacing of $T_c/N = 5$ ns, a variation in the propagation delay of 3 ns would produce a spectral deviation higher than half a channel. This aspect can be substantially reduced by a proper matching circuitry between electronic components where the

different devices run at its optimum performance. Even though the SAW devices are thermally isolated, these devices are strongly sensitive to temperature variations. As previously described in 2.8.2, a compromise between mass, power consumption and thermal isolation was found in order to obtain the optimum approach for the SAW devices enclosure.
The measurement of the instrument's spectral stability was realized by regulating the instrument's ambient temperature with a thermal chamber and analyzing the spectral displacement. The chamber's temperature was initially set to 18ºC, and the instrument was left running in such conditions for 6 hours where a frequency of 2,099,995,600 Hz was measured for channel 3750. Later, a second measurement at 28ºC for 6 hours gave a frequency of 2,099,990,100 Hz for the same channel. Considering the above numbers, a frequency stability of 550 Hz/ºC was obtained. This gives a spectral stability of 5.5 kHz if one assumes a SOFIA cabin temperature variation of 10ºC.

2.11.5 Mass and power consumption

The fact that the spectrometer is part of an airborne mission restricts the amount of available space, weight and power consumed by the instrument. The consideration of these limitations on the development of the instrument makes the transition from the initial concept to its realization the most difficult step. These restrictions affect mainly the quality of the supply voltage (power supply size, dissipation efficiency and ripple), the isolation between blocks, the mechanical stability and the power dissipation. The constraint on the power supply led to a compromise between the efficiency of the AC/DC transformation, its size and the stability of the supply voltages. The stability of the supply voltage depends strongly on the transformation process; a small switching power supply with high efficiency would also entail extra ripples and fluctuations in the supply signal. The RF signal processing is extremely sensitive to fluctuations in the source voltage (since it consists of high gain amplifiers). A compromise between space and stability was found by the introduction of high output DC/DC ripple attenuation modules, achieving high stability and low ripple. In order to constrain the power consumption we focused on the optimization of the digital devices; mainly in the newly developed chirp generator. A power reduction of more than 50% was achieved between the first and the final prototype, consuming approximately 9 W for the Digital Chirp Generator block.

Device	Power Consumption per unit [W]	Setup DC-Power Consumption [W]
Chirp Generator with ADCP technology (Two branches)	4.7	9.4
RF Signal processing (Two branches)	2.5	5.0
SAW Filter – Temperature control chamber (Two branches)	1.2	2.4
Preprocessor and data acquisition (Two sets for interleaved measurements)	1.1	2.2
Embedded computer	5.5	5.5
Total		**24.5**

Table 2.1: SOFIA-GREAT-CTS power consumption at DC level summary.

An extra power consumption of 10 W should be considered during the heating process of the SAW filters enclosure (30 min), and therefore this will lead to a maximum power consumption of 34.5 W (24.5+10W, see Table 2.1) at DC level. Since the instrument is integrated with an 80% efficiency power supply, the nominal power consumption at AC level is 31 W and 43 W for the initial phase. Based on new developments, we expect to reduce the power consumption of the Chirp Generator by up to 4 W and be able to produce higher bandwidth Chirp waveforms (more than 1 Giga-samples).
The weight and the volume of the instrument were reduced based on an a-priori computed aided design. The dimensions of the instrument are 450mm x 470 mm x 80 mm = 0.01692 m^3 with a total weight of 9.8 kg, considering the RF block, pre-processors, data acquisition, embedded computer, SAW temperature chamber and a power supply system.

2.11.6 Spectral resolution

The spectral resolution is one of the most important characteristics of the Chirp Transform Spectrometer and represents a design parameter. The spectral resolution is defined by the common chirp length of the expander-compressor setup, described in the case of SOFIA-GREAT-CTS as the compressor's chirp length. Since the response of the spectrometer to a sinewave is a sinc function, the FWHM maximum of a squared sinc signal would be 0.88 of the distance between the maximum and the first zero crossing. Theoretically, the first zero crossing in the frequency domain at $1/T_c$ (considering that $\Delta f = \Delta t/\mu = \Delta t \cdot (BW/Tc) = (1/BW) \cdot (BW/Tc) = 1/Tc$), will give a FWHM at $0.88/T_c$. In SOFIA-GREAT-CTS, the compressor chirp-length is T_c=21.43 μs, giving a maximum expected FWHM of 41.06 kHz. In order to retrieve the instrument's spectral resolution, the response of the spectrometer to a sinewave input was measured at different frequencies. The relative HWHM (Half-Width-Half-Maximum) to the right and to the left was obtained from the measurements, and herewith the FWHM was derived. Figure 2.29 shows the measured average FWHM to be 41.68 kHz, which is particularly close to the theoretical limit. This demonstrates an optimum match between the digitally generated expander chirp and the SAW filter dispersive properties. In the context of SOFIA-GREAT, the instrument will provide a spectral resolving power higher than $\lambda/\Delta\lambda = 10^8$ at 4.7 THz.

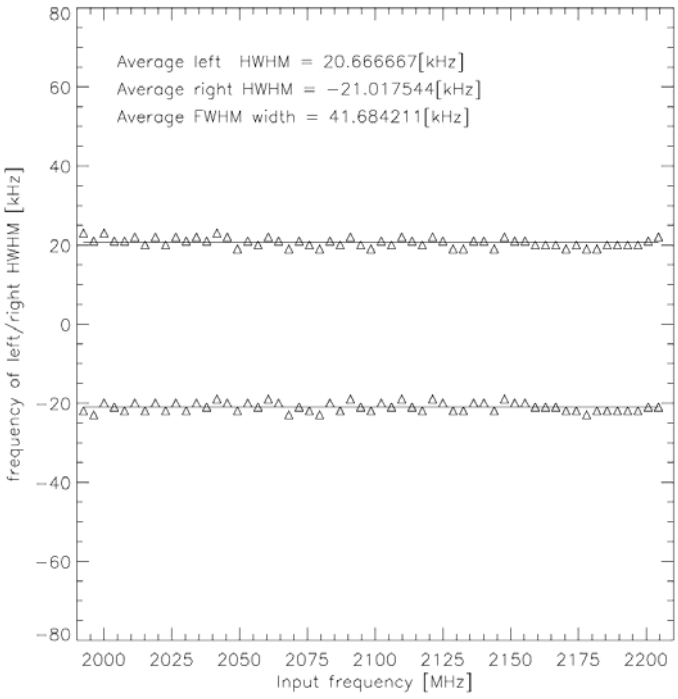

Figure 2.29: Measurement of SOFIA-GREAT-CTS's spectral resolution derived from the left and right HWHM. The measurement consists of providing a sinewave test signal to the instrument, which when measured with an oversampling mode of 8 eight points per channel spacing, leads to a measured spectrum with a resolution of 3.6 kHz. The location of the maximum and the neighbour half maximum points are retrieved from the high resolution data.

Since the source input frequency is known, the spectral scale linearity was obtained (see Figure 2.30) which is in the order of ±7 kHz for the complete bandwidth considering a channel spacing of 28.69 kHz for the 7500 channels. The flatness of the amplitude response (as shown Figure 2.31) is retrieved by the value of the peak maximum and is approximately 2 dB considering a 200 MHz bandwidth, and 3 dB considering a 215 MHz bandwidth.

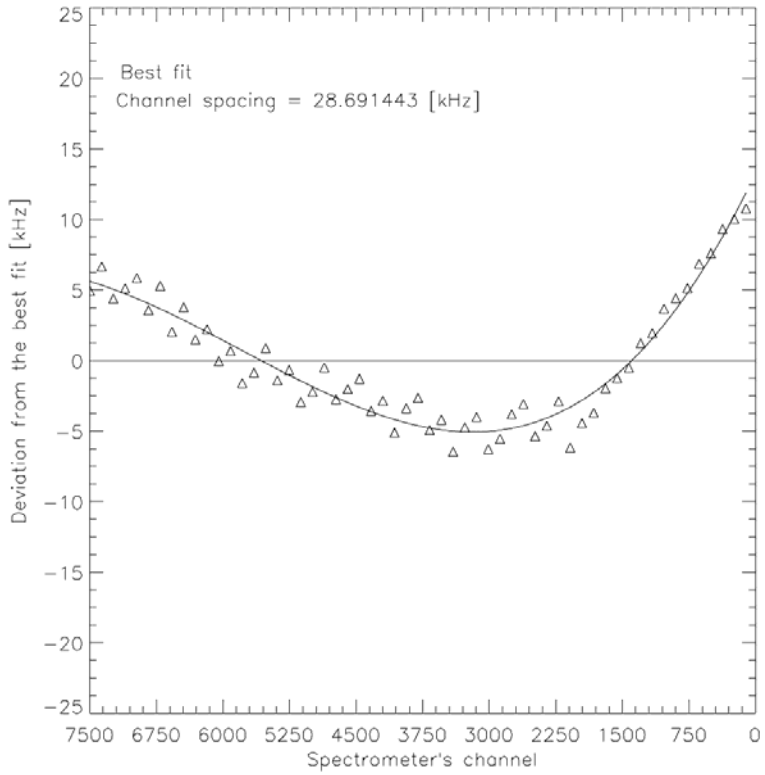

Figure 2.30: Measurement of the instrument's spectral scale linearity. The plot describes the deviations from an ideal channel spacing of 28.691443 kHz for the 7500 channels.

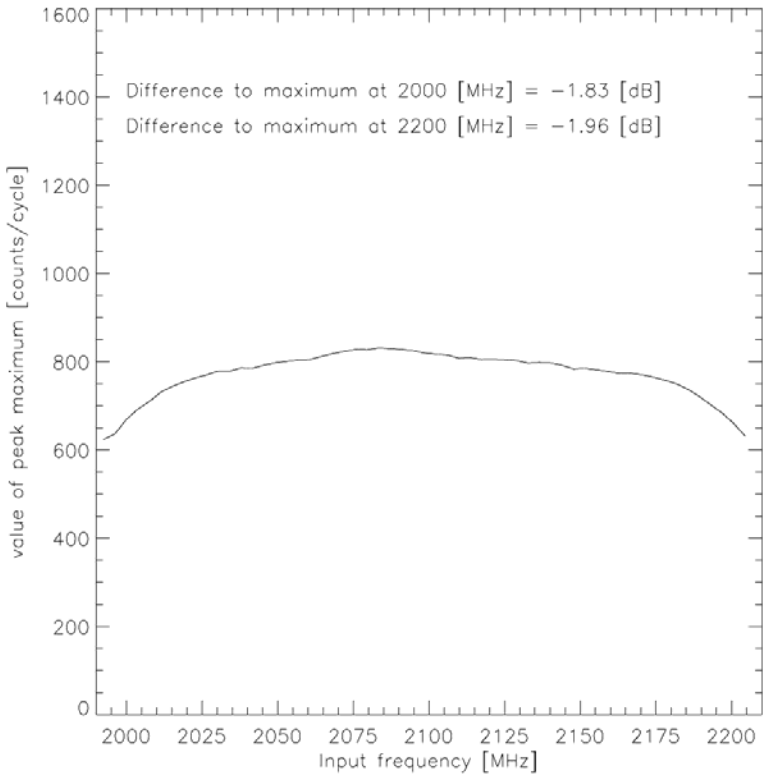

Figure 2.31: Measurement of the amplitude response considering as input a sinewave signal. The curve represents the frequency passband for the complete instrument's bandwidth (215 MHz). The relative attenuation with respect to the maximum value was calculated for the 2.0 GHz and 2.2 GHz components, obtaining values in the range of 2 dB.

Summary of SOFIA-GREAT-CTS instrument's performance	
Parameter	***Value***
Input center frequency	2.1 [GHz]
Input bandwidth [-3 dB]	215 [MHz]
Spectral resolution [FWHM]	41.7 [KHz]
Spectral resolution [noise equivalent] $1/T_c$	46.6 [kHz]
Channel spacing [7500 channels]	28.7 [kHz]
Spectral resolving power $\lambda / \lambda\Delta$ at 4.7 THz	$1.1x10^8$
Dynamic range (white noise) [±1dB]	30 [dB]
Dynamic range (white noise) [±0.1dB]	18 [dB]
Spectral stability	550 [Hz/ºK]
Input power	-35 [dBm]
Spectral density	-58 [dBm/MHz]
Absolute Allan variance	510 [s]
Frequency scale linearity	±7 [kHz]
Nominal power consumption [AC]	31 [W]
Maximum power consumption (first 30 min.) [AC]	43 [W]
Mass	9.8 [kg]
Dimensions	430 (wide) x 80 (tall) x 450 [mm^3]
Data interface	Ethernet + RS232 + TTL signals

Dispersive elements	
Parameter	***Value***
Compressor bandwidth	204.8 [MHz]
Chirp length compressor	21.43 [µs]
Time bandwidth product of the compressor	4420
Insertion loss of the compressor	-43 [dB]
Expander bandwidth through ADCP and tripling	409.6 [MHz]
Expander sampling frequency	349.92 [MHz]
Chirp length expander (through ADCP)	42.87 [µs]
Temperature isolation of the controlled chamber	0.024 [KºSAW / KºAir]
Compressor chirp rate temperature stability	3.43 [ppm / KºAir]

Table 2.2: Overall characterization of the SOFIA-GREAT-CTS instrument and the dispersive elements involved in the chirp transform.

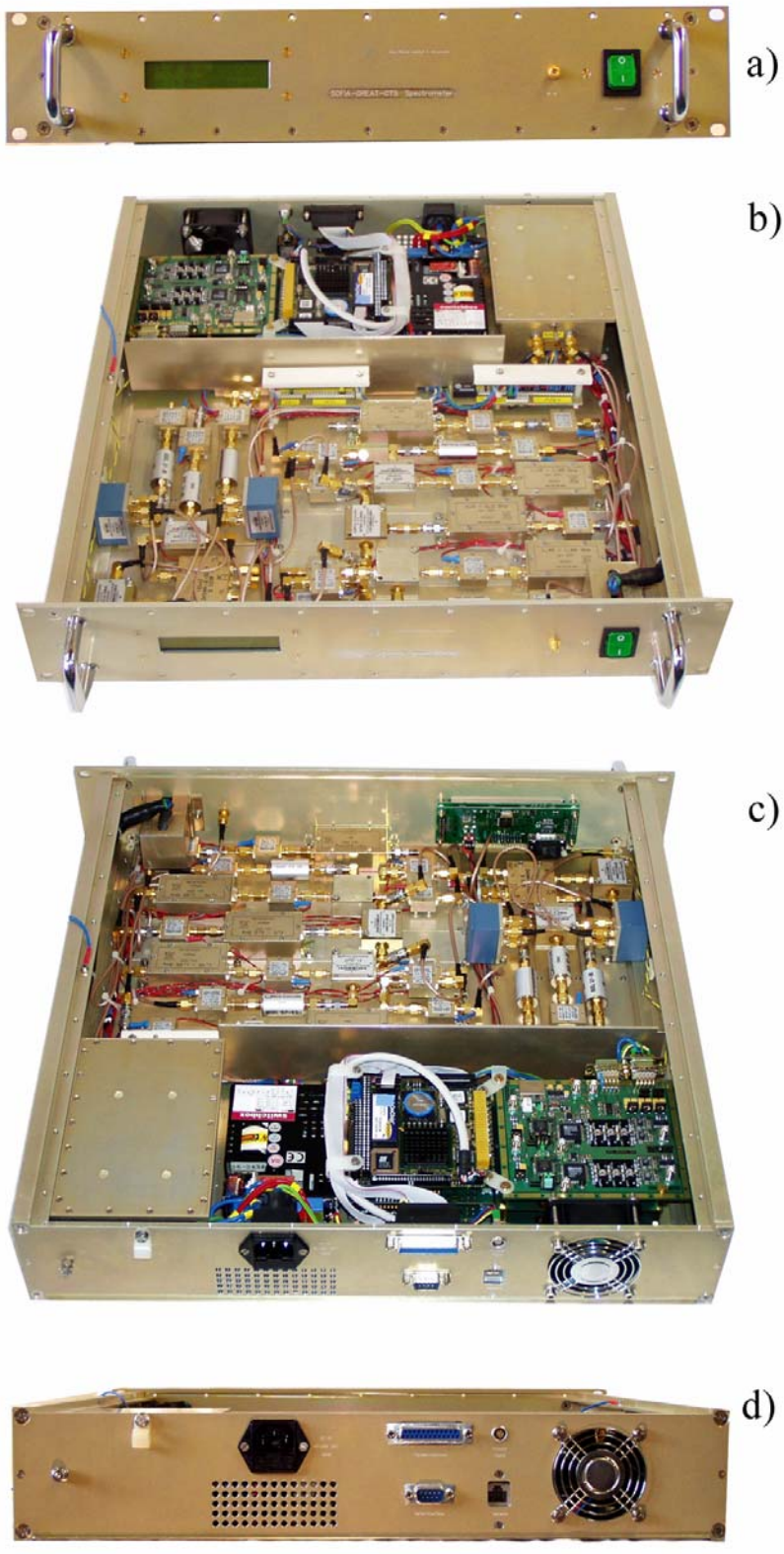

Figure 2.32: Images from the SOFIA-GREAT-CTS spectrometer. [a] Front plate with the RF-input and the display. The actual status of the CTS and the last processed command are displayed in the LCD display. [b] The RF signal processing block can be seen in the internal front part of the spectrometer. [c] The digital components (computer, data acquisition, chirp generator, pre-processors) are located in the back part of the spectrometer. The SAW filters' temperature controlled chamber is located in the bottom-left side of the spectrometer. [d] Back-side of the spectrometer with the interface-connectors.

Chapter 3

MAOAM - Atmospheric modeling

The chirp transform spectrometer is a powerful instrument for the study of planetary atmospheres, especially as part of the SOFIA mission providing highly resolved altitude profiles of temperature, wind, water vapor and minor species from ground level to 80 km. This is of great advantage in the characterization of the global chemical system of our neighboring planets, and therefore a valuable tool for the study of their atmospheres through the analysis of the forcing and tracing circulation of atmospheric chemical constituents. Consequently, I worked on a bridge between the physical phenomena of study and the spectroscopic observations via a general circulation model (GCM) for the Martian atmosphere developed under the scope of the MAOAM (Martian Atmosphere Observation And Modeling) project. The GCM is based on MART-ACC (Martian Atmosphere - Circulation and Climate Model) and is a fully non-linear, global and three-dimensional hydrodynamic Eulerian gridpoint model, which covers the troposphere, mesosphere and thermosphere up to 135 km. Improvements were made in the radiative transfer module, in the numerical stability through the reformulation of the atmospheric parameters which define the Martian atmosphere, and in the computational performance. These new developments are described in this chapter and their implications to the modeled atmosphere are compared to in-situ measurements of the lower atmosphere of Mars, and in the next chapter to ground-based observations performed with the SOFIA-GREAT-CTS.

3.1 Introduction

The first successful flyby mission to Mars, the Mariner 4 (1964), opened a new era in the study of Mars with high spatially resolved images from the Martian surface and the first in situ retrieval of the atmospheric surface pressure (4.1-6 hPa, [Kliore et al. 1965]). Further discoveries obtained from ground-based, airborne and Martian missions provided a deeper characterization of the surface and atmosphere of Mars, and represented an important motivation for new scientific developments for interpreting such unprecedented information. As a consequence, a general circulation model for the study of the general circulation and climate of the Martian

atmosphere was developed (MART-ACC, [Ebel & Berger, 1997]), which was initially intended as an interpretation tool for the attempted mission Mars96. Besides this non-successful mission which unluckily ended at the bottom of the Pacific Ocean, the general circulation model was successfully developed and forms the base of the MAOAM Martian atmospheric model. Like most of the other general circulation models such as the NASA/Ames Mars GCM [Haberle et. al, 1993] and the Paris-GCM [Hourdin et. al, 1993], MART-ACC was based on a terrestrial circulation model: the COMMA (Cologne Model of the Middle Atmosphere) model. This transformation to represent the Martian conditions was feasible since there are strong similarities between the atmospheric circulation of Mars and Earth: a) both planets have a similar planetary inclination and therefore similar annual seasonal cycle, b) both planets are fast rotating planets with almost equal sidereal rotation period and therefore the Coriolis force in both planets is comparable, c) both planetary atmospheres are nearly transparent to sunlight so that they are primarily heated by infrared radiation emanating from the surface below, and d) Martian weather exhibits similar phenomena, e.g. Hadley cells and trade winds, low pressure systems, mountain waves, slope winds, downslope windstorms and thermal tides.

The main goal of the atmospheric circulation modeling is to characterize the mechanisms by which momentum, heat, trace gases and aerosols are transported. The Martian atmosphere is mostly composed of CO_2 with only 6 [hPa] of mean surface pressure, which is especially low when compared to the terrestrial surface pressure of 1000 [hPa]. This lower pressure produces a stronger relation between the incoming radiation and the atmospheric temperature. The Martian diurnal cycle is therefore dominated by very large temperature fluctuations ($\Delta T=60ºC$) [Kieffer et al., 1977], compared to a maximum diurnal fluctuation of $\Delta T=30ºC$ on Earth. These temperature variations will give very strong thermal tides coupled to the solar heating of the atmosphere and therefore a global one-day-cycle atmospheric disturbance. Even though this phenomenon is present on Earth and therefore described by the COMMA, MART-ACC and MAOAM models, its influence on the terrestrial atmospheric circulation is relatively small. This effect on Mars however, is particularly important and a key element in the atmospheric circulation, driving tides with an amplitude of more than 0.5 [hPa] amplitude [Zurek, 1976]. Another source of strong daily tides is the surface of Mars, which is a desert. The low thermal inertia of such types of soil contributes large surface temperature variations that are driven by the daily solar cycle. Except winter, the Martian surface experiences temperature swings each day by as much as $\Delta T_{surf}>80ºC$ [Schofield et al., 1997], in response to solar heating. The effect is attenuated on Earth by the influence of the oceans which transport the heat absorbed downward during daylight. This process is inverted during the night, resulting in a superior temperature balance. Zonally symmetric circulations associated with the so-called Hadley circulation appear to be more important components of the Martian circulation than is the case on Earth. The thin Martian atmosphere combined with the lower inertia of the ground enhances this type of circulation on Mars, which on Earth is weakened in a great manner by the oceans. The overturning timescale, which is the length of the circulation structure measured in days (taking into account wind velocity and structure size), is much smaller, only 100 Martian sols or 1/6 of the Martian year, compared to a one year overturning timescale on Earth. This means a much faster wind circulation and therefore a stronger phenomenon present on Mars, which can produce wind traces of up to 150 km/h near the surface [Joshi et al., 1997].

The simulation and modeling of such a complex atmospheric system is organized by numerically solving the system of equations which drive the atmospheric circulation. The Martian atmosphere can be considered as a constrained fluid on a rotating sphere which obeys the fluid dynamics equations in a rotating frame of reference. This set of differential equations includes the horizontal equations of motion, the thermodynamic balance, the continuity equation and the hydrostatic balance equation. Each particular process that describes the Martian atmosphere is represented by a physical/mathematical term of the overall solution. However, the model can also be thought of a system composed of different specialized components, all contributing to the overall numerical calculation. The MAOAM model has three main components: 1) the dynamical module which takes into account the different atmospheric forces involved in the global circulation, 2) the radiation module which takes into account the solar heating, the atmospheric radiation and the surface energy budget, and 3) the chemical module which includes the heating and transport of the various chemical species in the atmosphere. The numerical solution of the system of equations is performed by solving the complete system of equations for each fixed point forming the three dimensional grid structure that describes the Martian atmosphere. These points are organized in a spherical coordinate system with an horizontal resolution of $\Delta x=22.5^{\circ}$ in east-west, and $\Delta y=5^{\circ}$ in north-south direction. The vertical coordinate is organized with logarithmic pressure coordinates $z=H\cdot\ln(p/p_0)$, considering a scale-height typical of Mars of $H=RT/g=10.3$ [km]. The number of vertical points in the model is 118 which results in a vertical resolution of $\Delta z=1.1$ [km]. The model provides high temporal resolution, up to 885 points per rotation period and therefore $\Delta t=100$ [s].

My present study is mainly focused on the global Martian circulation, trying to produce and constrain global circulation patterns such as Hadley cells and atmospheric waves and tides. These are subjects of high relevance in the transport of minor species which can be monitored by remote sensing from SOFIA. Therefore, the atmospheric circulation studied here considers a dustless atmosphere from where I expect to retrieve major dynamical transport structures.

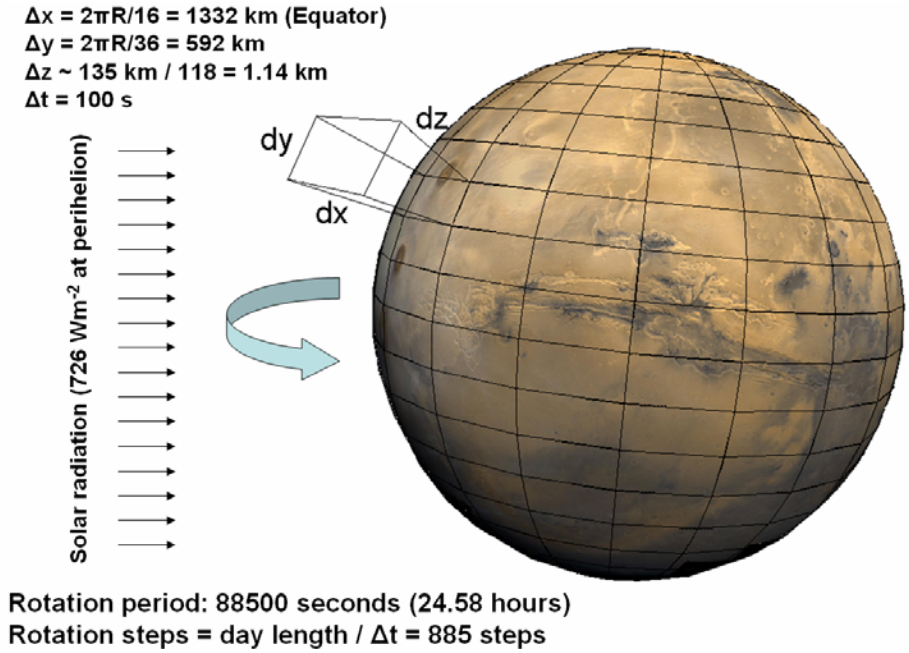

Figure 3.1: Description of the model's geometry for the four dimensions (3-space and 1-time). The model is developed to better describe the transient phenomena with a vertical resolution of approximately 1 km and a high temporal resolution of 100 seconds.

3.2 The Martian general circulation

The meridional circulation in Mars is dominated by overturning flows (Hadley cells) driven by differential atmospheric heating. At solstice, the meridional circulation can be described primarily by a big Hadley cell from the summer hemisphere to the winter hemisphere and a polar cell in the summer hemisphere, which is also thermally driven. At equinox, the circulation is represented by two symmetric Hadley cells due to the rising air flow from the warm Equatorial region.

Furthermore, the circulation on Mars is strongly affected by planetary waves induced by the strong temperature fluctuations between day and night and the marked topographic features present on Mars. This differential solar insolation induces thermal tides, producing a large scale planetary wave propagating westwards, and can be found at low-latitudes [Zurek 1976; Leovy and Zurek 1979]. Thermal tides are a larger component of the total longitudinal and temporal variability of the Martian atmosphere than at Earth. This is because of the thin atmosphere and low thermal inertia of the surface pressure, which results in a stronger relation between the incoming radiation and the atmospheric temperature. These tides were first confirmed by observations from the Viking and Mariner 9 missions, which measured transient variations of the surface pressure and of near-surface wind and temperature, consistent with a propagating atmospheric tide [Zurek et al., 1992].

Due to both the mechanical lifting of air over the Martian orography and the thermal forcing associated with topography, strong forced quasi-stationary waves are induced in the form of Rossby waves. These waves have much more significance on Mars than on Earth, especially because of the large orographic features dominant on the Martian surface. Since the restoring force for these large-scale motions is the Coriolis force, a higher Rossby number on Mars would lead to a larger atmospheric instability and enhancement of this type of atmospheric planetary waves. The Rossby number is the ratio between the inertial forces and Coriolis restoring forces. For a small Rossby number (Ro << 1), the geostrophic approximation is valid and therefore the horizontal pressure gradients are balanced by Coriolis forces associated with the horizontal winds. However, a bigger Rossby number would make the inertial forces more dominant in the flow circulation. The planetary Rossby number is expressed as the following:

$$\frac{u\,du/dx}{2\Omega v\sin\phi} \sim \frac{U^2/L}{2\Omega U\sin\phi} = \frac{U}{2\Omega\sin\phi L} \equiv Ro \qquad (3.1)$$

where u and v are the zonal and meridional winds respectively, Ω is the rotation rate of the planet, φ the latitude, L is the horizontal scale of the phenomenon and U the typical horizontal wind jet velocity.

Another important feature of the Martian circulation is high-latitude eastward-propagating travelling waves due to baroclinic instabilities in the atmosphere [Ryan et al., 1978; Barnes 1980]. Baroclinic instability produces wave motion due to vertical shear of the mean flow in the presence of Coriolis and buoyancy forces and is specially enhanced in regions where the Coriolis influence is stronger, as is the case at the polar regions.

All these atmospheric motions discussed here are also present on Earth because both planets are particularly alike even though they have differences. The diurnal and seasonal cycle in both planets is comparable and the atmospheric heating and cooling resembles similarities since both planetary atmospheres are nearly transparent to sunlight. As a result, the forces driving the

circulation and the nature of the energy exchange of the Martian atmosphere are closely alike to the ones on Earth, producing similar global atmospheric circulation patterns.

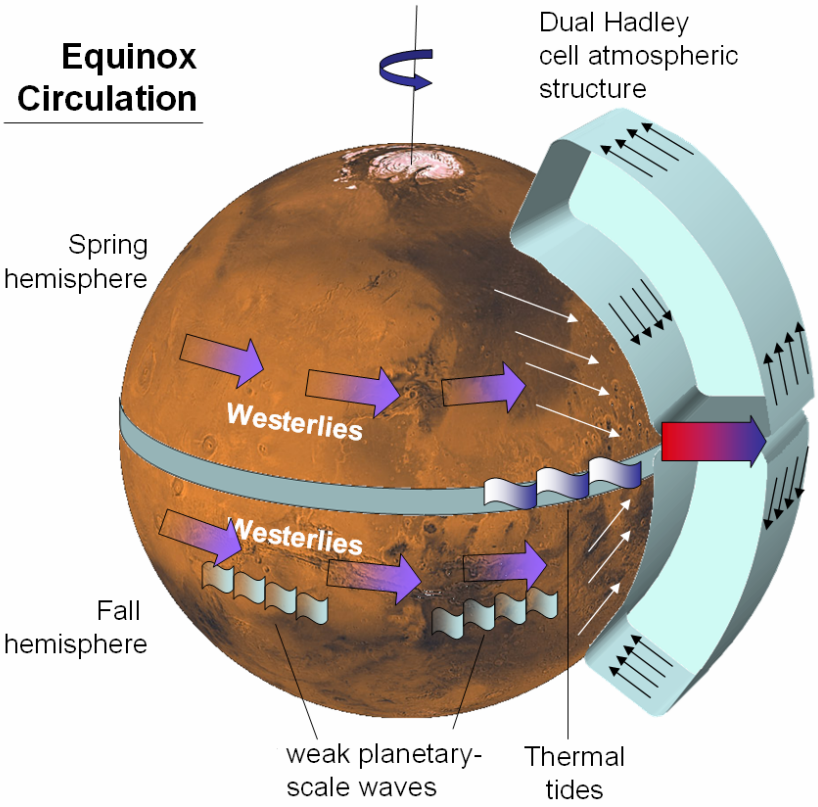

Figure 3.2: Sketch showing the main features of the atmospheric circulation at the equinox. The warm air is lifted from the equatorial regions and descends in the colder subtropics. The zonal circulation is dominated by westerlies due to the Coriolis force on the meridional circulation. Thermal tides and weak planetary scale waves describe the main transient phenomena at equinox.

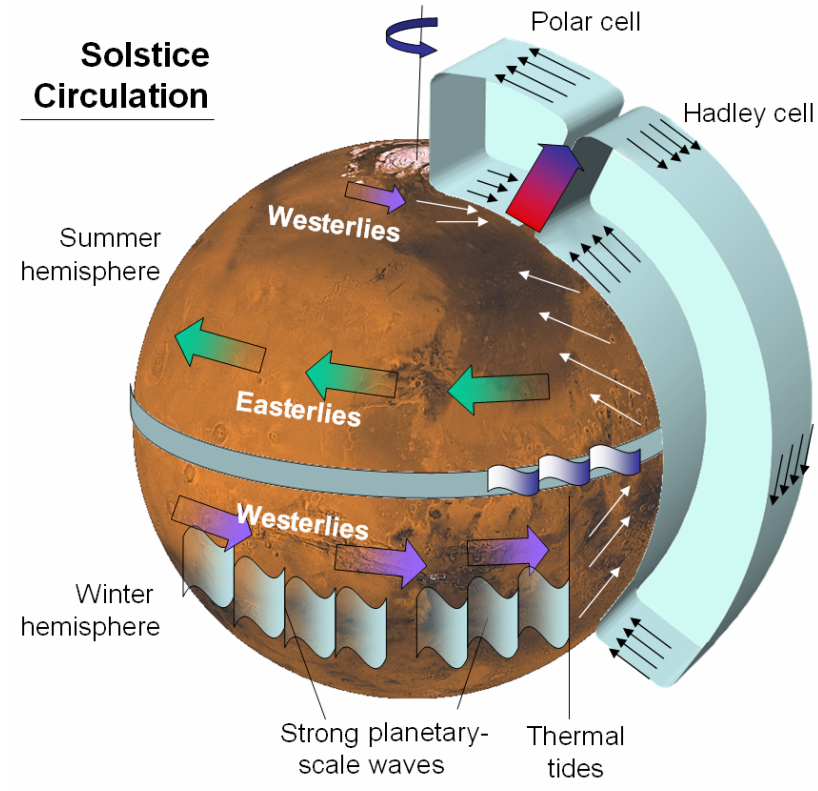

Figure 3.3: Sketch showing the main features of the atmospheric circulation at solstice. In this case the warm region is located in the summer hemisphere which leads to a main meridional circulation from the summer to the winter hemisphere, describing a Hadley cell. In the summer hemisphere, a polar cell is formed which is also thermally driven. In the winter hemisphere the zonal circulation is dominated by westerlies due to the Coriolis effect, while in the summer hemisphere this effect is inverted, leading to easterlies. Strong Rossby waves and thermal tides dominate the transient phenomena.

3.3 Mars General Circulation Model

3.3.1 The hydrodynamical system

The atmospheric modeling is achieved by numerically solving the system of equations which drive the circulation, and are based on the conservation of momentum, mass and energy. The conservation of momentum for the three dimensional field of velocities is analyzed by solving the Navier-Stokes equation of a fluid ($F_{grav} + F_{visc} + F_{press} = m \cdot a$), considering a rotating frame of reference. The vertical momentum is approximated to hydrostatic equilibrium, since on Mars the vertical accelerations are negligible if compared to the present gravity effect. This will lead to the following set of equations using the log-pressure coordinate system [Holton, 1975; Ebel & Berger, 1997]:

Zonal term | Meridional term | Vertical term

$$\frac{\partial u}{\partial t} = \underbrace{-\frac{1}{r\cos\varphi}\frac{\partial u^2}{\partial \lambda}}_{\text{Zonal term}} \underbrace{-\frac{1}{r\cos\varphi}\frac{\partial}{\partial \lambda}(uv\cos\varphi)}_{\text{Meridional term}} \underbrace{-\frac{1}{\rho_o}\frac{\partial}{\partial z}(\rho_o uw)}_{\text{Vertical term}}$$

$$\underbrace{+2\Omega v\sin\varphi}_{\text{Coriolis force}} \underbrace{-\frac{1}{r\cos\varphi}\frac{\partial \Phi}{\partial \lambda}}_{\text{Zonal planetary-potential term}} \underbrace{+\frac{uv}{r}\tan\varphi}_{\text{Curvature term}} \underbrace{+F_u}_{\text{Dissipation terms}}$$

Zonal momentum balance **(3.2)**

$$\frac{\partial v}{\partial t} = -\frac{1}{r\cos\varphi}\frac{\partial (uv)}{\partial \lambda} - \frac{1}{r\cos\varphi}\frac{\partial}{\partial \lambda}(v^2\cos\varphi) - \frac{1}{\rho_o}\frac{\partial}{\partial z}(\rho_o vw)$$

$$+2\Omega u\sin\varphi - \frac{1}{r}\frac{\partial \Phi}{\partial \varphi} + \frac{u^2}{r}\tan\varphi + F_v$$

Meridional momentum balance **(3.3)**

$$\frac{\partial \Phi}{\partial z} = \frac{RT}{m^* H}$$

Vertical hydrostatic balance **(3.4)**

The conservation of energy is derived considering the thermodynamic relation between diabatic heating and potential temperature, which assuming $T<T_0$ and small vertical winds can be formulated as the following [Holton, 1975; Ebel & Berger, 1997]:

$$\frac{\partial T}{\partial t} = -\frac{1}{r\cos\varphi}\frac{\partial (uT)}{\partial \lambda} - \frac{1}{r\cos\varphi}\frac{\partial}{\partial \varphi}(vT\cos\varphi) - \frac{1}{\rho_o}\frac{\partial}{\partial z}(\rho_o Tw)$$

$$\underbrace{+\frac{RTw}{m^* c_p H}}_{\text{Adiabatic heating}} \underbrace{+\left.\frac{\partial T}{\partial t}\right|_{sol}}_{\text{Solar radiation}} \underbrace{+\left.\frac{\partial T}{\partial t}\right|_{IR}}_{\text{Planetary radiation}} \underbrace{+F_T}_{\text{Dissipation terms}}$$

Energy conservation **(3.5)**

The mass continuity equation, considering hydrostatic balance, is:

$$0 = -\frac{1}{r\cos\varphi}\frac{\partial u}{\partial \lambda} - \frac{1}{r\cos\varphi}\frac{\partial}{\partial \varphi}(v\cos\varphi) - \frac{1}{\rho_o}\frac{\partial}{\partial z}(\rho_o w) \quad \left.\begin{array}{l}\\ \\ \\ \end{array}\right\} \begin{array}{l}\textbf{Mass}\\ \textbf{conservation}\\ \textbf{(3.6)}\end{array}$$

The symbols in equations (3.2)-(3.6) have the following definitions:

λ	longitude
φ	latitude
z	a measure of "height" = -H ln(p/p_s)
H	scale height RT_s/g
R	universal gas constant
m*	relative molecular mass
ρ_0	air density = $\rho_s$exp(-z/H)
u	eastward velocity (zonal wind)
v	northward velocity (meridional wind)
w	a measure of "vertical velocity" = dz/dt
T_s	a representative atmospheric temperature
T_0	a basic state temperature = T_0(z)
Φ	planetary potential for the vertical point $= \int_0^h g \cdot dz$
Ω	angular velocity of Mars
r	radius of Mars

F_u, F_v, F_T dissipation terms described later in this section

The lower boundary layer of the system of equations is the Martian surface, which is represented by a height matrix h(x,y) according to the surface orographic structure. The conversion from of the coordinate from height to log-pressure is obtained by defining that normal components of velocity should not exist in the lower boundary, and therefore:

$$\frac{d\Phi}{dt} = g\frac{dh}{dt} \qquad (3.7)$$

which is the mathematical formulation used in the implementation of the Martian topography on the model [Ebel & Berger, 1997].

The dissipation terms add extra physical processes to the system of equations, which are highly important in the description of the phenomena present in the Martian atmosphere. Particularly above 100 km, physical processes such as molecular heat conduction and dynamic viscosity are important sources of cooling and friction respectively. These terms are also relevant for representing micro-scale phenomena which can not be described by the macro-scale analysis performed by the model, which is constrained by the grid's structure and size. These small scale eddy phenomena characteristic of turbulent flow will produce a stronger dissipation of momentum. This can be described as a friction effect at the macro-scale. In addition, this turbulent flow contributes to the atmospheric energy balance since it produces an extra source of "turbulent" heat transport.

The contribution to the upper atmosphere (>100 km) phenomena and the small scale eddy effects in the middle atmosphere (10-120 km) are described on the dynamical set of equations by the following terms [Berger, 1994]:

$$
\begin{aligned}
F_u &= \overbrace{-\alpha_R u}^{\text{Rayleigh friction}} + \overbrace{K_{zz}\frac{\partial^2 u}{\partial z^2}}^{\text{Vertical eddy diffusion}} + \overbrace{K_d\frac{\partial^2 u}{\partial z^2}}^{\text{Dynamic viscosity}} \\
F_v &= -\alpha_R v + K_{zz}\frac{\partial^2 v}{\partial z^2} + K_d\frac{\partial^2 v}{\partial z^2} \\
F_T &= \underbrace{D_m\frac{\partial^2 T}{\partial z^2}}_{\text{Molecular heat conduction}} + \underbrace{\frac{1}{P_r\rho_o}\frac{\partial}{\partial z}\rho_0 K_{zz}\left(\frac{\partial T}{\partial z}+\Gamma_a\right)}_{\text{Turbulent heat transport}}
\end{aligned}
\qquad \left.\right\} \text{Momentum and energy dissipation terms} \qquad (3.8)
$$

where α_R is the Rayleigh friction coefficient, K_{zz} the vertical eddy diffusion coefficient, K_d the dynamic viscosity coefficient, P_r the turbulent Prandtl number of the air and D_m the molecular heat conductivity coefficient. Rayleigh friction provides a representation of the effects of gravity waves breaking in the middle atmosphere, a process that is crucial to the dynamics of this atmospheric region. On Earth, the impact of Rayleigh friction can be as large as that of the gravity-wave drag due to vertically-propagating orographic gravity waves [Kim et al., 1998]. These gravity waves induce regions of convective instability which are then limited and momentum-balanced by a wave-dissipation process driven by vertical eddy diffusion. However, above 100 km in the thermosphere, the momentum dissipation processes are dominated by dynamic viscosity, while the energy dissipation is mainly driven by molecular heat conduction.
The characterization of these parameters requires a special study, since they are particularly complex physical effects. An important contribution of [Ebel & Berger, 1997] was to include a proper parameterization of these phenomena in the MART-ACC model, which was later improved as described by [Meister, Hartogh & Villanueva, 2003] for the MAOAM model.

3.3.2 The radiative transfer analysis

The radiative transfer is a key component of the energy balance described in equation (3.5), and the driver of the atmospheric circulation. The different components of the energy balance, as described in Figure 3.4, can be divided in the short-wave energy balance driven by the solar radiation, and the long-wave energy balance which takes into account the surface and atmospheric infrared emission and absorption.

Short-wave solar absorption – Atmospheric heating

Since the chemical structure of the Martian atmosphere is mainly composed of CO_2 (95.3%), its absorption of solar radiation is the main source of atmospheric heating. The formulation of the solar heating (neglecting atmospheric scattering) is derived by considering an absorption function A, a zenith angle θ, a solar flux F_{sol}, a heat capacity of the air C_p, and an air density ρ:

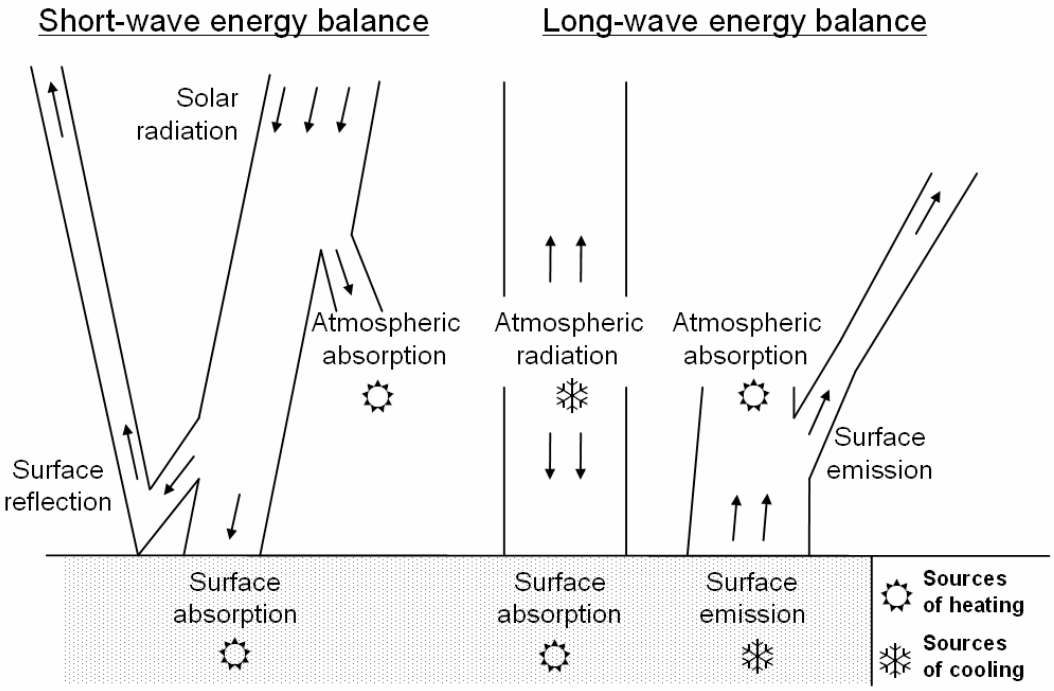

Figure 3.4: Sketch describing the energy balance analysis performed by the MAOAM model. The short-wave energy balance takes into account the solar radiation and its surface and atmospheric absorption. The long-wave energy balance considers the infrared surface-atmosphere interaction.

$$\left.\frac{\partial T}{\partial t}\right|_{sol} = -\frac{F_{sol}}{C_p \rho(z)} \frac{dA_{sol}(u)}{dz} \qquad u = f(z, \cos\theta) \tag{3.9}$$

where u is the optical path. The seasonal cycle on Mars can be represented by a correct dependence of the zenith angle and solar flux with respect to orbital obliquity and transit of Mars on its orbit. This effect is later discussed in this chapter, since an important improvement added to the MAOAM model is a characterization of the Martian seasonal cycle. The zenith angle depends on the present solar inclination δ (which is season-dependent), the latitude φ and the local time α (diurnal cycle):

$$\cos\theta = \sin\varphi \sin\delta - \cos\varphi \cos\delta \cos\alpha \tag{3.10}$$

The definition of the absorption function is complex, since it includes the information of the chemical structure on Mars and the absorption coefficients for each chemical constituent. Furthermore, the absorption coefficients compile the effects of each absorbing molecular transition from the constituent, which are temperature and pressure dependent because of line broadening. However, the solar heating radiative process on Mars can be simplified since the atmospheric absorption is mainly driven by the 2.7 μm spectral band of CO_2. This process is taken into account in the MAOAM-GCM as in the MART-ACC model [see Ebel & Berger, 1997].

Long-wave atmospheric radiation – Atmospheric cooling

The previously described absorption of the solar radiation in combination with the atmospheric emission and absorption of the long-wave radiation are the main actors in the energy balance. Their balance leads to an equilibrium temperature for the Martian atmosphere of about 210 K, with an emission maximum of 14 μm according to Planck's radiation law. The atmospheric

interaction with this long-wave radiation is mainly driven by the numerous ro-vibrational spectral lines of CO_2 present at 15 μm. The radiative transfer can be written as:

$$\left.\frac{\partial T}{\partial t}\right|_{IR} = -\frac{1}{C_p \rho(z)} \frac{d}{dz}\left(F^{\uparrow} - F^{\downarrow}\right) \qquad (3.11)$$

where $F^{\uparrow}$ and $F^{\downarrow}$ are the upward and the downward radiation fluxes respectively. This means that the isotropic radiation of the air parcel at height z is affected by an upward flux from the sources of radiation below, such as the surface and parcels of air below, and the downward radiation coming from parcels located above and the outer space. Considering local thermodynamic equilibrium (LTE), each layer of air emits radiation following the Planck function, and therefore the fluxes for a plane parallel atmosphere can be written as:

$$F^{\uparrow}(z) = \overbrace{\int_0^{\infty} \pi B_v(0) T_v^f(z) dv}^{\text{From surface}} + \overbrace{\int_0^{\infty}\int_0^{z} \pi B_v(z') \frac{dT_v^f}{dz'}(z - z') dz' dv}^{\text{From layers below}} \qquad (3.12)$$

$$F^{\downarrow}(z) = \underbrace{\int_0^{\infty} \pi B_v(z_{\infty}) T_v^f(z) dv}_{\text{From outer space}} + \underbrace{\int_0^{\infty}\int_{z_{\infty}}^{z} \pi B_v(z') \frac{dT_v^f}{dz'}(z' - z) dz' dv}_{\text{From layers above}} \qquad (3.13)$$

where T_v^f is the weighted monochromatic transmission and $B_v(0)$, $B_v(\infty)$, $B_v(z')$ are the monochromatic radiation Planck's function for the surface, the outer space and the defined layer, respectively. However, the equations described above can be simplified by considering the outer space radiation to be negligible $B_v(\infty) \approx 0$ and that most of the radiation exchange is performed by the spectral band v=15 μm where $Br = \int B_v dv$. This gives a net flux $F = F^{\uparrow} - F^{\downarrow}$ as:

$$F(z) \cong \overbrace{\pi B_r(z) T_v^f(z_{\infty} - z)}^{\text{Cooling to space}} + \overbrace{\pi\left[B_r(0) - B_r(z)\right] T_v^f(z)}^{\text{Flux exchange with surface}} + \underbrace{\int_0^{z_{\infty}} \pi\left[B_r(z') - B_r(z)\right] \frac{dT_v^f}{dz'}(|z - z'|) dz'}_{\text{Flux exchange between layers}} \qquad (3.14)$$

The "cooling to space" and the "flux exchange with surface" terms are dominant in equation (3.14). Therefore, the "flux exchange between layers" is neglected since it demands strong computational resources and does not introduce significant differences to the long-wave energy balance. The flux exchange with the surface is only significant in the lower part of the Martian atmosphere (<15 km) and is represented in the model through the Newtonian cooling approximation, where the heating rate for this effect can be written as:

$$\left.\frac{\partial T}{\partial t}\right|_{IR_{surface}} = -\alpha(z)\left[T(z) - T_{surf}\right] \qquad (3.15)$$

where α is the Newtonian coefficient of cooling, derived from approximation formulae based on comparison with line-by-line codes. However, the available information is based mainly on a terrestrial atmosphere and therefore demanded an extra parameterization which was performed for the MAOAM model. The Newtonian coefficient was parameterized as follows:

$$\alpha(z) = b\left(\frac{\rho(z)}{\rho_0}\right)^c \qquad (3.16)$$

where ρ_0 is the surface density, b and c the parameterization coefficients. These coefficients were derived based on comparison between model results from temperature fields with in-situ measurements from the MGS-TES instrument for the lower atmosphere.

The cooling to space effect is analyzed in a similar manner as that for solar heating, defined by an absorption function. Since the energy balance analysis is based on a LTE atmosphere, a stronger absorption of the CO_2 at 15 μm produces a stronger emission on the same band because Kirchhoff's law is applicable in LTE mediums. Therefore, a stronger absorption/emission produces a stronger cooling; meaning that the term T_v^f can be described by the absorption function of CO_2 at 15 μm. Following the same procedure performed by [Liou and Sasamori, 1975] for the solar absorption at 2.7 μm, Ramanathan et al. [1976] obtained an empirical formula which describes the absorption of CO_2 at 15 μm (for $u/\delta>1.5$):

$$T_v^f(u) \approx 0.753 A_0 \delta \left\{\ln\left(u/\delta\right)^{3/2} + 1.21\right\} \qquad u = S\beta\omega(z)/A_0 \qquad \delta = 2\pi^{1/2}\alpha_D / D \qquad (3.17)$$

where ω is the column above the layer and u is the dimensionless optical depth. The parameters for the CO_2 band at 15 μm are tabulated on [Ramanathan et al., 1983] where S is the band intensity, β is the line-width parameter, D is the mean line spacing and $\alpha_D = f(T^{1/2})$ is the Doppler line half-width. A_0 is the bandwidth parameter, which is described considering terrestrial conditions. However, there is no information about this parameter for the Martian atmosphere at present. Since A_0 represents a spectral broadening effect, this value was parameterized on the MAOAM model following a formulation based on [Liou, 1992; Elsasser & Culbertson, 1960]:

$$A_0 = k\left(\frac{p_0}{p}\right)^n \left(\frac{T}{T_0}\right)^m \qquad (3.18)$$

where k, n and m are the proportional, pressure and temperature scaling factors respectively. The parameterization was performed based on the same procedure for the retrieval of the Newtonian cooling coefficients. The total long-wave heating/cooling rate equation can be now written as:

$$\left.\frac{\partial T}{\partial t}\right|_{IR} = \left.\frac{\partial T}{\partial t}\right|_{IR_{surf}} + \left.\frac{\partial T}{\partial t}\right|_{IR_{space}} \qquad (3.19)$$

$$\left.\frac{\partial T}{\partial t}\right|_{IR} = -\alpha(z)\left[T(z) - T_{surf}\right] - \frac{\pi}{C_p \rho(z)} B_r(z) \frac{dT_v^f(z_\infty - z)}{dz} \qquad (3.20)$$

The surface energy balance

Figure 3.4 shows the surface interaction to be an important component of the general energy balance. Like the atmosphere, the balance between the absorbed and emitted radiation by the

surface defines its equilibrium temperature. The surface absorbs and emits radiation which then warms the atmosphere. It also interacts with the atmosphere due to the turbulent heat-exchange process between the surface and the lower atmosphere. The energy balance of the surface can be described as the following:

$$\frac{\partial T_s}{\partial t} = \frac{1}{C_s}\left(Q_s + F^{\downarrow} - F^{\uparrow}\right) \qquad (3.21)$$

where C_s is the heat capacity of the ground, Q_s is the absorbed short-wave radiation, and $F^{\uparrow}$ and $F^{\downarrow}$ are the emitted and absorbed long-wave radiation respectively. The interaction with the short-wave radiation by the surface is specially defined by the surface's bolometric albedo at surface wavelengths A, which for the case of Mars is A=0.25 [see appendix G], producing the following incoming short-wave diabatic heating by solar radiation:

$$Q_s = u(1-A)F_{sol} \qquad (3.22)$$

where F_{sol} is the solar flux (as for the atmospheric heating on (3.9)), and u is the optical path. The optical path is derived in the same manner as for the atmospheric heating where $u=1/\cos\theta$ for zenith angles lower than 75º. This is corrected for greater angles following the correction proposed by [Smith & Smith, 1972] which takes into account the sphericity of the planet. The atmospheric absorption is neglected in this case since the Martian atmosphere is mostly optically thin for the solar radiation.

The remaining component of equation (3.21), the net radiation flux, can be expressed in terms of the emissivity ε and the atmospheric T_{air} and surface temperatures T_{surf}: [Pollack et al., 1993]

$$F = F^{\downarrow} - F^{\uparrow} = \sigma\left(\varepsilon_{air} T_{air}^4 - \varepsilon_{surf} T_{surf}^4\right) \qquad (3.23)$$

where σ is the Stefan-Boltzmann constant. The air temperature is taken from the lower layers of the atmosphere in the range of 0 to 3 km. The emissivity of the air is defined to be ε_{air} =0.0609 [Pollack et al. 1993] (assuming a dustless atmosphere) and the surface emissivity is taken to be ε_{surf} =0.9 [Cuzzi & Muhleman 1972].

Furthermore, a dry convective adjustment is applied to the potential temperature next to the surface if the temperature profile is calculated to be statically unstable. This adjustment is only necessary within the boundary layer where the mixing by the turbulent fluxes can cause the flow to become statically unstable.

3.3.3 The seasonal cycle

The seasonal cycle is driven by the motion of the planet on its orbit, where the distance to the sun and the declination are affected. On Earth with an eccentricity close to zero, the variations of the solar flux with respect to the distance to the sun are small, and therefore the seasons in both hemispheres are comparable in length and incoming solar flux. Nevertheless, on Mars with an e=0.093 the insolation factors are far from simple hemispherical alternation. Considering an elliptical orbit, such as is the case of Mars, the distance to the sun (r) can be written as: [Roy, 1988]

$$r = R_0 \frac{(1-e^2)}{(1+e\cdot\cos f)} \qquad (3.24)$$

where R_0 is the mean solar distance to the Sun and f is the true anomaly, where f=0º and f=360º represents perihelion and aphelion respectively. This orbital variation of the distance produces variations on the received solar flux, because the apparent size of Mars with respect to a constant solar flux density is altered. Thus, the solar flux can be now written as:

$$F_{sol} = F_0\left(\frac{R_0}{r}\right)^2 \qquad (3.25)$$

where F_0 is the mean solar flux. This seasonal dependence of the solar flux is integrated in the energy balance through equations (3.9) and (3.22) for the solar atmospheric heating and the surface heating respectively. The second parameter for the analysis of the seasonal cycle is the planet declination. This depends on the planet obliquity φ (25.19º) and the true anomaly at equinox for the northern spring ψ (109.41º):

$$\delta = \varphi \sin(f - \psi) \qquad (3.26)$$

where (f − ψ) is known as the solar longitude (L_S), and L_S=0º northern spring, L_S=90º summer, L_S=180º autumn and L_S=270º winter. The value ψ is retrieved from the mean orbital parameters, where ψ=90º+l-w, where l is the mean longitude (355.45º) and w is the longitude of perihelion (336.04º) [Williams, 2004]. Variations in the planetary declination are introduced in the model through the calculation of the zenith angle as described in equation (3.10). However, the derivation of the true anomaly (f) which respect to time (t) is complex, since the planet velocity on its elliptical orbit is not constant and no matter which form is used, there is no known mathematical formula which gives f in terms of t. However, solutions can often be approximated to any degree of accuracy by a computationally iterative process. The process involves an auxiliary angle, the eccentric anomaly E which like f and M grows by 360º each orbit. At perigee, all three anomalies equal zero. The eccentric anomaly can be approximated to:

$$E = M + e \cdot \sin M \cdot (1 + e \cdot \cos M) \qquad (3.27)$$

where M is the mean anomaly, which describes the continuous pass of time on Mars ($M=360º \cdot t/t_{year}$). For the derivation of f on MAOAM, the formulations considered are based on the ones described by Schmitt [2004; Duffett-Smith, 1989], which are based on the eccentric anomaly E:

$$f = 2 \cdot \tan^{-1}\left(\tan\left(\frac{E}{2}\right)\sqrt{\frac{(1+e)}{(1-e)}}\right) \qquad (3.28)$$

For high eccentricity orbits such as is the case of Mars, an error of less than 0.1% in the true anomaly is obtained by a direct calculation of the equations mentioned above. This value is more than acceptable.

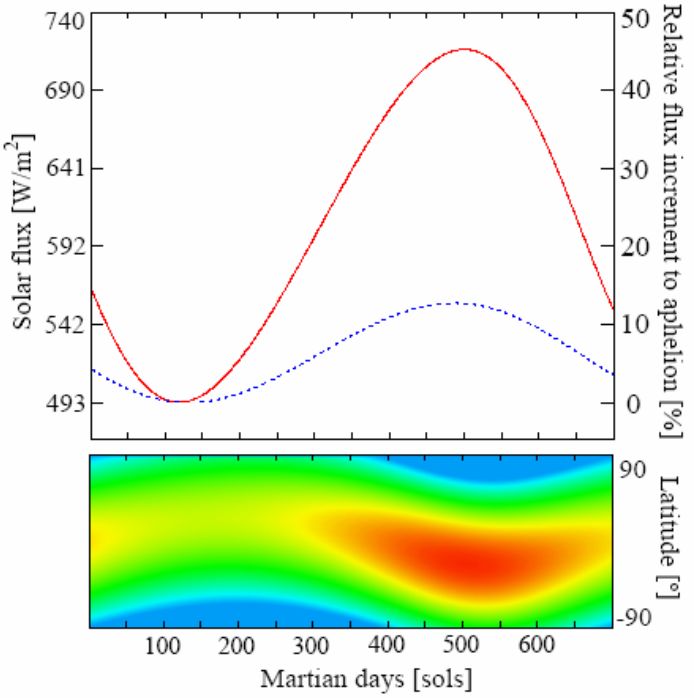

Figure 3.5: Variations of the solar flux according to the annual cycle. The solid line describes the solar flux variation and relative flux increment with respect to aphelion, considering the Martian orbital eccentricity of e=0.0932; while the dotted line considers Earth's eccentricity of e=0.0167. The contour plot shows the simultaneous effect of the solar flux and the planet declination in the heating of the Martian atmosphere.

3.3.4 The non-LTE effect and a possible implementation

When considering the radiative processes related to the energy balance, most of the above mentioned formulations are based on Local Thermodynamical Equilibrium or LTE. On a LTE medium the population of the energetic levels follows a Boltzmann distribution with the excitation temperature (T_{ex}) equal to the kinetic (T_{kin}) or background temperature. The population distribution among the energy levels is therefore determined predominantly by collisions. Under such conditions the source radiative function for such a medium is defined by Planck's function. However, the energy interaction and therefore the radiation source function, are also affected by other contributions not only collisions. The equation of state for a two level analysis can be written as:

$$(A_{21} + B_{21}\bar{y} + C_{21})n_2 = (B_{12}\bar{y} + C_{12})n_1$$
$$n_1 + n_2 = n \qquad (3.29)$$

where n_1 and n_2 are the number of molecules at the energy levels 1 and 2 respectively, y is the radiation field at frequency $\nu=(E_2-E_1)/h$, A_{21} is the spontaneous emission coefficient, B_{12} is the stimulated emission coefficient, B_{21} is the stimulated absorption coefficient, C_{21} is the collisional de-excitation coefficient, and C_{12} is the collisional excitation coefficient. In a LTE medium, the collisional coefficients are dominant on the equation of state (C>>A,B) and therefore the excitation temperature is defined by the background temperature. At low density atmospheric regions where the excitation by radiation is more significant, such is the case of the upper atmosphere of Mars, this consideration is no longer valid. This means that the radiation coefficients start to play a role on the radiation balance and therefore the source function is no longer the Planck's function. In such a case, the population of the energetic levels can not be represented by the equilibrium Boltzmann distribution with $T_{ex}=T_{kin}$ and is therefore called non-LTE.

The deviation of the excited energy levels from a Boltzmann distribution has implications on the cooling and heating rates for the Martian atmosphere, especially above 90 km. As described by López-Puertas & Taylor [2001] the main contribution to the atmospheric cooling is at the emission of carbon dioxide at 15 μm, which is dominant on the cooling process, and therefore the cooling rates are strongly sensitive to this transition. In the case of the heating, the absorption of solar radiation by carbon dioxide at 4.3 and 2.7 μm is much higher above 100 km when considering non-LTE. The calculation of the radiative transfer including non-LTE is extremely complex, and demands a precise formulation of the atmospheric chemistry, its isotopic ratios and most importantly all the information of the possible molecular transitions. This is normally described as a "line-by-line" calculation, where the contribution of the overall heating rate is calculated for each particular isotope and its molecular transitions.

An initial study of such problems was addressed by Kutepov et al. [1991], who developed an iterative method based on the use of approximate transfer operators for the solution of the non-LTE molecular band radiative transfer in planetary atmospheres. Important contributions were later obtained by Lopez-Puertas & Lopez-Valverde [1995], who obtained the radiative energy balance of CO_2 non-LTE infrared emissions for the Martian atmosphere. The resolution of the radiative transfer considers a population following a Boltzmann distribution as its initial parameter. The system then evolves taking into account molecular collisions along with spontaneous and emissive radiation effects, until it converges to the equilibrium state where all the equations of state as described in (3.29) are fulfilled. The solution in the case of Kutepov et al. [2003] is accelerated by using the lambda iteration method, which converges much more rapidly than a straightforward iteration between the transfer equation and the equations of statistical equilibrium.

Even though the algorithm is accelerated, the calculation of the heating and cooling rates for an atmospheric column demand more than 4 hours on a fast computer (Pentium IV – 2.66GHz), considering the most important isotopes and molecular transitions. This means that a GCM with 16x36 columns would demand 16x36x4hours = 2304 hours, or roughly one third of a year for a global computation. In addition, this calculation should be performed 10000 times for the convergence of the dynamical core of the MAOAM-GCM, meaning that under present conditions, a run of the GCM assuming non-LTE conditions would demand more than 26 centuries of computation. The question that immediately arises is how this important phenomenon can be integrated without demanding such high computational power. The solution can be addressed by approximations or look-up tables. However, this demands a careful characterization of the non-LTE effects and their implications to the atmospheric circulation.

Consequently, work is continuing on the parameterization of the non-LTE features on the radiative transfer module. Our approach mainly consists of deriving an analytical formula that would explain with a certain degree of accuracy the deviation of the non-LTE with the LTE calculation. This could be later applied to correct the LTE heating and cooling rates in a similar vein as described by Forget et al. [1999], who derived a simple parameterization to include the effect of non-LTE on the heating by CO_2 solar radiation absorption at 1.2-2.0, 2.7, 4.3 μm wavelengths. This simplified parameterization is based on the one-dimensional model calculations performed by Lopez-Puertas and Lopez-Valverde [1995]. The parameterization takes into account the optical path u, the relative distance to the sun r and the atmospheric pressure p:

$$\left.\frac{\partial T}{\partial t}\right|_{sol}(p,r,u)=\frac{\partial T}{\partial t}(p_0,r_0,0)\times\frac{r_0^2}{r^2}\sqrt{\frac{p_0}{p}u'}\left(1+\frac{p_{NLTE}}{p}\right)^{-1} \quad (3.30)$$

where $\partial T/\partial t(p_0,r_0,0)$ is the heating rate corresponding to zero solar angle, a mean distance to the sun r_0=1.52 AU, and an average pressure p_0=700 Pa. The non-LTE effect is described by the term with the relative pressure (p_{NLTE}/p), where p_{NLTE} is the pressure limit that describes the limit between LTE and non-LTE. According to the value derived by Forget et al. [1999], p_{NLTE}=0.0075 Pa. In addition, the optical path is corrected to take into account atmospheric refraction and then $u'=((1224u^2+1)/1225)^{½}$. Even though the approximation is significantly simplified, it provides an important constraint for the derivation of a more sophisticated formulation.

The updated study performed by López-Valverde et al. [1998] of non-LTE for general circulation models of the Martian atmosphere, shows that the LTE/non-LTE solar heating rate is mainly described by the atmospheric pressure. In this study, the solar zenith angle and the thermal distribution were found to be not important for the non-LTE departure from LTE, allowing approximations based on a pressure scaling function. From the published departure values provided by Lopez-Valverde et al. [1998], I obtained an empirical formula to correct the LTE solar heating rates as calculated in (3.9) but now taking into account non-LTE effects. The results are shown on Figure 3.6 and the approximation is described as the following:

$$\left.\frac{\partial T}{\partial t}\right|_{sol}^{non-LTE}=\left(1+\left(\frac{p_{NLTE}}{p}\right)^k\right)^{-1}\left.\frac{\partial T}{\partial t}\right|_{sol}^{LTE} \quad (3.31)$$

where k is the pressure scaling factor, and p_{NLTE} (similar to that assumed by Forget et al. [1999]) is the pressure limit for non-LTE effects. Both coefficients were fitted to the line-by-line calculation, obtaining k=1.6 and p_{NLTE}= 0.00016 Pa.

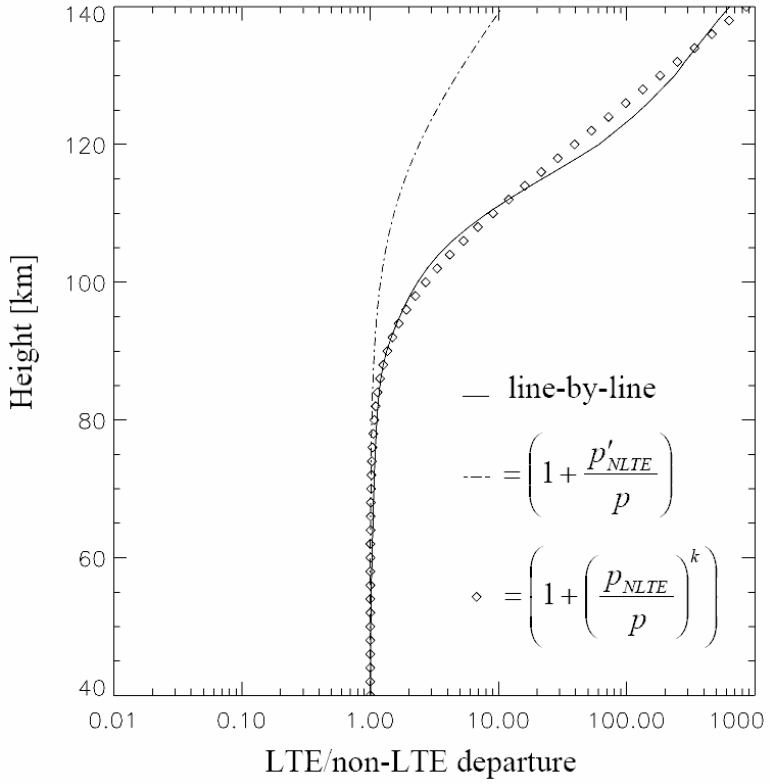

Figure 3.6: Comparison between the departure of non-LTE/LTE solar absorption of CO_2 on the Martian atmosphere, as derived by López-Valverde et al. [1998] (solid line), with a simple approximation as described by Forget et al. [1999] in dotted lines, and the new approximation as described in equation (3.31) dotted with diamonds. The assumed coefficients are p'_{NLTE}=0.0075 Pa, p_{NLTE}=0.00016 Pa and k=1.6.

This approximation is particularly convenient for the radiative transfer calculations performed on the GCM, where strong computational constraints exist. Figure 3.6 shows that below 80 km the solar heating can be calculated with high accuracy using an LTE approach, while above this limit, it is mandatory to take into account non-LTE effects in order to obtain realistic results.

The treatment of the thermal infrared absorption and emission of the atmosphere is particularly more complex. A parameterization based only on a pressure scaling as considered for the solar absorption does not truly represent the thermal cooling rates since they are also strongly affected by the thermal structure. There have been several attempts to include this effect as a departure formulation from the LTE assumption. There are the four cases of a) Fels & Schwarzkopf [1981], who developed a technique with three a-priori thermal profiles, from where the cooling rates were later obtained based on interpolation of the present thermal distribution in-between the a-priori given profiles; b) Kutepov & Fomichev [1993], who developed a approximated recurrence formula for the calculation of the cooling rates based on a 2-level model for the CO_2-626 isotope fundamental bands; and c) Lopez-Valverde & Lopez-Puertas [2001], who obtained a parameterization of the CO_2 cooling at 15 μm considering a 3-level system (two excited states coupled with one collisional exchange).

The study derived by Lopez-Valverde & Lopez-Puertas [2001] revealed the complexity of such a parameterization. Nevertheless, they arrived to a possible parameterization which is the one integrated on the LMD-AOPP GCM model of the Martian Atmosphere. Similarly, we are currently working on a characterization based on an a-priori tabulation of the main elements in the non-LTE calculation for the cooling rates. These parameters are the population of the excited levels and the rate coefficients of absorption, extinction, emission and collision for the selected bands, which are dependent on the atmospheric temperature, pressure and chemical diversity. Most of the results obtained by Lopez-Valverde & Lopez-Puertas [2001] are based on the main CO_2 isotope, the 626 isotope. However, Figure 3.7 shows that the contribution of the 636, 628, 627 and 638 isotopes also plays an important role in the calculation of the heating and cooling rates. The non-LTE calculation was performed with the ALI-RET model developed by Kutepov et al. [2003], including the main ro-vibrational bands of the CO_2 isotopes, assuming a zenith angle of 39º and a thermal profile as shown in figure G.1 [see appendix G] for clear conditions.
We expect to obtain a sensitivity matrix which could provide correlation factors derived from non-LTE radiative transfer calculations. These correlation factors, dependent on the thermal gradient and the zenith angle, could be then tabulated and implemented in the radiative transfer module of the MAOAM-GCM. The study should also take into account the feasibility of linear interpolation between the tabulated coefficients and the present atmospheric values. Further research on this topic is needed to provide a more realistic characterization of the upper atmosphere of Mars, which at present is not deeply understood due to the lack of observations. With the upcoming introduction of SOFIA, highly resolved vertical maps of the upper atmosphere are expected. This would reveal the thermal structure of the Martian atmosphere and more importantly its chemical diversity.

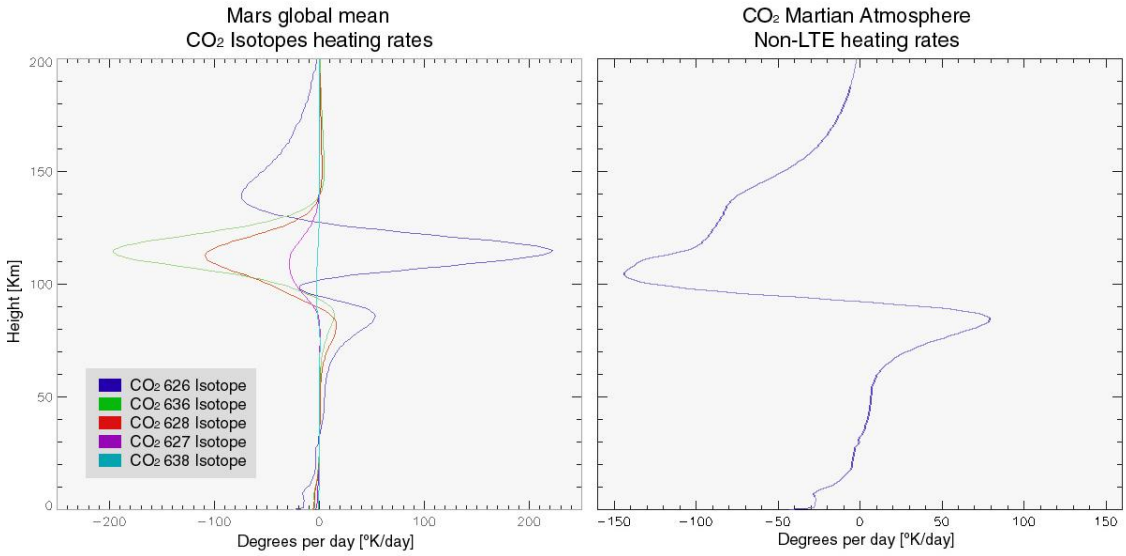

Figure 3.7: Heating and cooling rates as calculated with the ALI-RET non-LTE radiative transfer model, assuming a standard Martian atmosphere and a zenith angle of 39º. [Right] Contribution of the different isotopes to the overall heating and cooling rates. [Left] Integrated non-LTE heating rates.

3.3.5 The numerical implementation

Taking as base the MART-ACC model, the MAOAM-GCM was completely re-formulated and re-programmed. The new model is prepared for multiple grid structure and allows the implementation of a parallel computing scheme since the computation is properly segmented for the different processors. Furthermore, through a proper combination of optimization techniques the model was importantly enhanced in performance: MAOAM-GCM is three times faster than MART-ACC, taking 0.14s per time step and 23 minutes for a complete run (when considering a Pentium IV, 2.66 GHz).

To solve the set of prognostic differential equations, the leapfrog scheme is adopted. The leapfrog scheme refers to the centred time difference which is used in conjunction with centred space differences, meaning that the future is calculated based on the past value plus the central difference of the spatially surrounded points at present, or:

$$\frac{\partial a}{\partial t} = \ldots + c\frac{\partial a}{\partial x} + \ldots \qquad (3.32)$$

$$a_{x,t+1} = a_{x,t-1} - c\frac{\Delta t}{\Delta x}\left(a_{x+1,t} - a_{x-1,t}\right) \qquad (3.33)$$

This scheme is an explicit three-time-level scheme which is second order in space and time. The leapfrog scheme has the important property that there is no damping on the solution, nevertheless it is a dispersive numerical method [Kurihara, 1965]. The stability of the solution is defined by the criterion of Courant-Friedrichs-Levy (CFL), which dictates that the time step must be smaller than the time for some significant action to occur, and preferably, a considerably less time of $\Delta t \leq \| \Delta x/c \|$, where c is the fastest speed of waves allowed in the atmosphere.

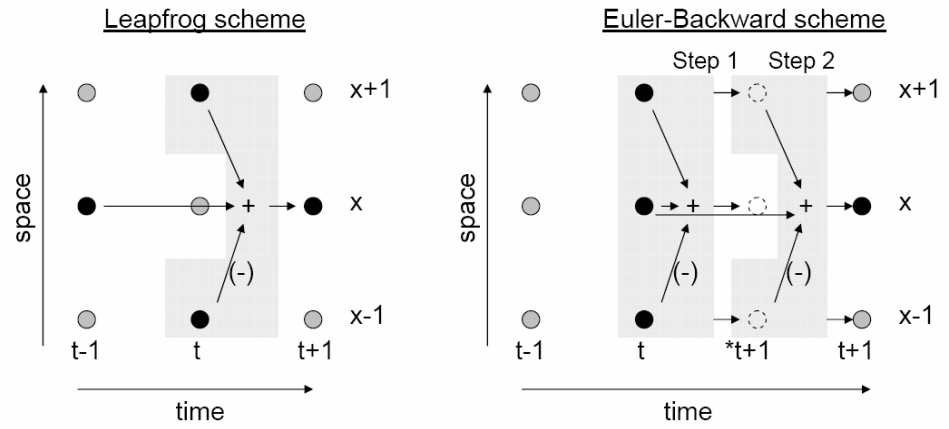

Figure 3.8: Sketch describing the numerical calculation for the Leapfrog scheme (right) and the Euler-Backward scheme (left).

However, it can be seen that the solution for odd and even time steps for the same space point are completely independent, and therefore would lead to a non-stable state while solving non-linear differential equations. Consequently, the solution of the system of equations is completed with a correction step which considers the present state (t) and is based on the Euler-backward scheme. This correction procedure includes two numerical operations, one to obtain the intermediate future *t+1 and the second to calculate the future value. Referring to equation (3.33), the Euler-Backward scheme can be written as:

$$a_{x,*t+1} = a_{x,t} - c\frac{\Delta t}{2\Delta x}\left(a_{x+1,t} - a_{x-1,t}\right) \qquad (3.34)$$

$$a_{x,t+1} = a_{x,t} - c\frac{\Delta t}{\Delta x}\left(a_{x+1,*t+1} - a_{x-1,*t+1}\right) \qquad (3.35)$$

The use of this diffusive numerical scheme leads to the attenuation of the amplitudes and phase velocity changes of the solved wave phenomena [Friese, 1998; Kurihara, 1965]. A proper combination of the numerical properties of each scheme is organized in the MAOAM-GCM, in order to obtain a stable and correct solution. The implemented scheme calculates 48 steps using the Leapfrog scheme, followed by 3 steps according to the Euler-Backward scheme.

The main concern while solving the system of equations is the stability and convergence of the solution. As previously described by the CFL criterion, the stability of the system is driven by the phenomenon properties (represented by c) and the grid structure (Δt and Δx). A higher spatial resolution ($\Delta x' < \Delta x$) will produce a reduction on the stability of the system, or a higher computational cost for the calculation of the same time period due to the reduction of $\Delta t' < \Delta t \cdot (\Delta x'/\Delta x)$. However, the computational stability can be significantly improved without affecting the grid structure by the removal of small wavenumber ($L/\Delta x$) phenomena that are not completely resolved by the grid structure. This is done through numerical filtering; such as the case of Shapiro filters that remove small scale grid noise without affecting the physical structures of the fields. The computational scheme of the Shapiro filter can be expressed as: [Shapiro, 1971]

$$a_x^{filter} = a_x + \frac{(-1)^{n-1}}{4^n}\delta_{2n}(a_x) \qquad (3.36)$$

$$\begin{aligned}
&\delta_2(a_x) = a_{x-1} + 2a_x - a_{x+1} \\
&\delta_4(a_x) = a_{x-2} + 4a_{x-1} + 6a_x - 4a_{x+1} + a_{x+2} \\
&\delta_4(a_x) = \delta_2(\delta_2(a_x)) \\
&\delta_8(a_x) = \delta_4(\delta_4(a_x)) \\
&\dots
\end{aligned} \qquad (3.37)$$

where 'a' is the prognostic variable and n is the order of the filter ($n=2^j$, j=0,1,2…) that defines the damping strength. In the MAOAM-GCM, the prognostic variables are filtered on the horizontal plane with Shapiro filters of order 2.

3.3.6 The stability study

As previously discussed, the balance between c (representing the model parameters), Δx and Δt is highly delicate, and demands a proper study of the computational schemes and the physicochemical parameters that drive the atmospheric circulation.

Stability is a critical issue in the modeling of the meteorological evolution of the atmosphere, especially on Mars with its enormous orographic features that are complex to resolve even with high spatial resolution. The MAOAM-GCM model was integrated with the new topographic measurements from the MGS-MOLA instrument that revealed the structure of the Martian surface with an unprecedented precision of 1 m and a horizontal resolution of 160x160 m^2 [Smith et al., 2001]. This data was transformed afterwards into the model grid scheme through a cubic interpolation method, from where a grid of 16 points longitude and 36 longitude were retrieved, as shown on Figure 3.9. The orographic structures observed in the figure are quantitatively high and they produce strong numerical instability features while the complex set of hydrodynamical equations is solved.

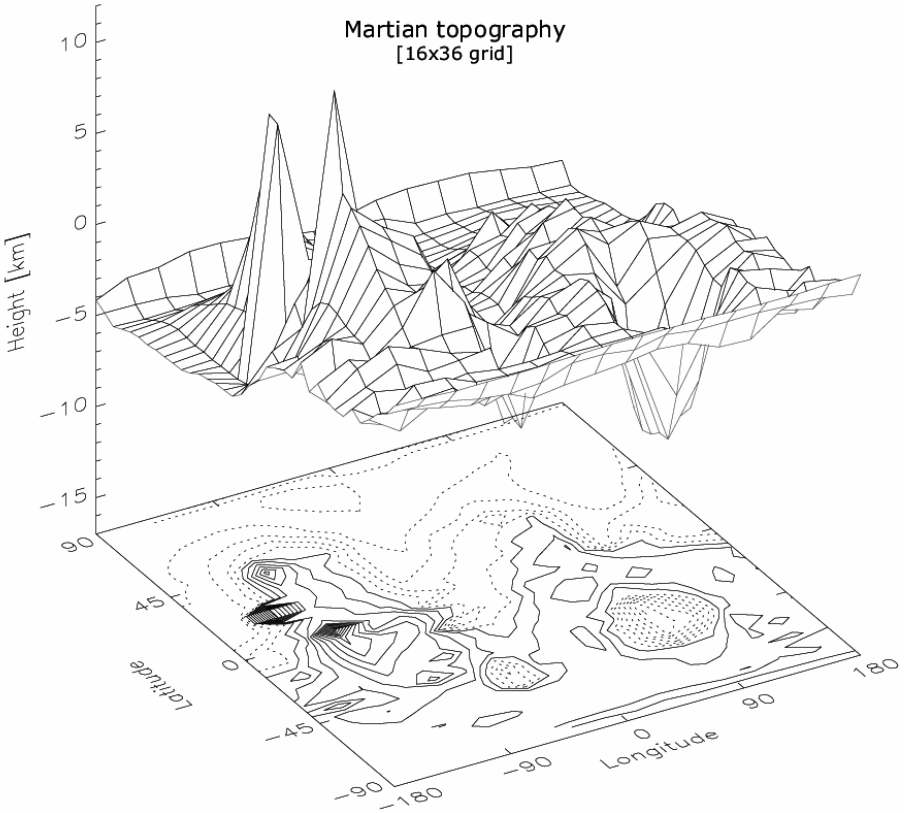

Figure 3.9: Martian topography assuming the grid structure as on the MAOAM-GCM. The plotted topography is based on the MGS-MOLA high resolution data, which is then transformed into the grid scheme through a cubic convolution interpolation method. The contour plot is organized in 1km layers.

For instance, analyzing the maximum gradients for the 16x36 topography map gives $\partial h/\partial x|_{max} \approx$ 14 [km/gridpoint] and $\partial h/\partial y|_{max} \approx$ -12 [km/gridpoint]. Thus if a parcel of air travels longitudinally, it would encounter an orographic wall of 14 km in its next simulated spatial state. Furthermore, this would result in a strong baroclinic excitation for the circulating flow of air, which can be derived by the following equation:

$$\frac{\partial p}{\partial x} = p_0 \left(\frac{\partial}{\partial x} e^{-h/H} \right) \qquad (3.38)$$

where p_0 is the pressure at the reference point (~6 hPa) and H is the scale height (10.3 km), obtaining $\partial p/\partial x|_{max} \approx 6.5$ [hPa/gridpoint] and $\partial p/\partial y|_{max} \approx 5.1$ [hPa/gridpoint].

These previously calculated topographic gradients point out the complexity of integrating the Martian topography on the simulations that is why on the MAOAM-GCM the so called "smoothed" topography is considered. This provides a description of such an important effect without strongly affecting the stability of the numerical computation. An alternative approach would be through a progressive increment of the model topography from null at the initial stage until its final value. This is similar approach as for the solar heating. The initial state of the model is an isothermal atmosphere with zero vertical and horizontal wind fields. Therefore, to prevent the strong transition between this static initial stage and the stimulated evolving state, the atmosphere receives solar radiation in a progressive manner providing a smooth transition. This evolution, $Fsol' = Fsol \cdot e^{-t/\tau sol}$, is regulated by the "progressive-solar" constant τ_{sol}, which is defined according to the numerical stability of the computational scheme and the evolution of the modelled circulation from its initial stage. This method is substantially consistent when performing numerical simulations, more so if the starting conditions do not have a correlation with the resulting atmospheric circulation. By setting an isothermal atmosphere as an initial state, more realistic simulations are intended, since the atmosphere evolves as driven by the hydrodynamical set of equations until its convergence state, and not by the a-priori defined atmospheric state.

The study of the stability also revealed that the characterization of the diffusive and friction parameters were critical for the convergence and robustness of the computation. This is particularly the case for the Rayleigh friction coefficients (α_R), the vertical eddy diffusion coefficients (K_{zz}), the dynamic viscosity coefficients (K_d) and the molecular heat conductivity coefficients (D_m), as part of the diffusive terms in equation (3.8). These atmospheric parameters describe phenomena which are dependent on the pressure scale, and therefore are characterized by vertical profiles. On the MART-ACC however, they were poorly described in their vertical distribution with only 5 points for the parameterization of the 118 vertical gridpoint layers. T They were therefore refined through analytical formulations dependent on the atmospheric pressure and independent of the grid structure. This refinement of the parameters reduces the vertical gradients of their distribution and produces smooth transitions between subsequent vertical points. Keeping the previous section in mind, the stability of the solution is defined by the CFL criterion and therefore small gradients between subsequent vertical points would significantly enhance the convergence of the solution. Figure 3.10 shows the comparison between the stability enhanced MAOAM-GCM with its predecessor MART-ACC, as calculated for the same circulation state.

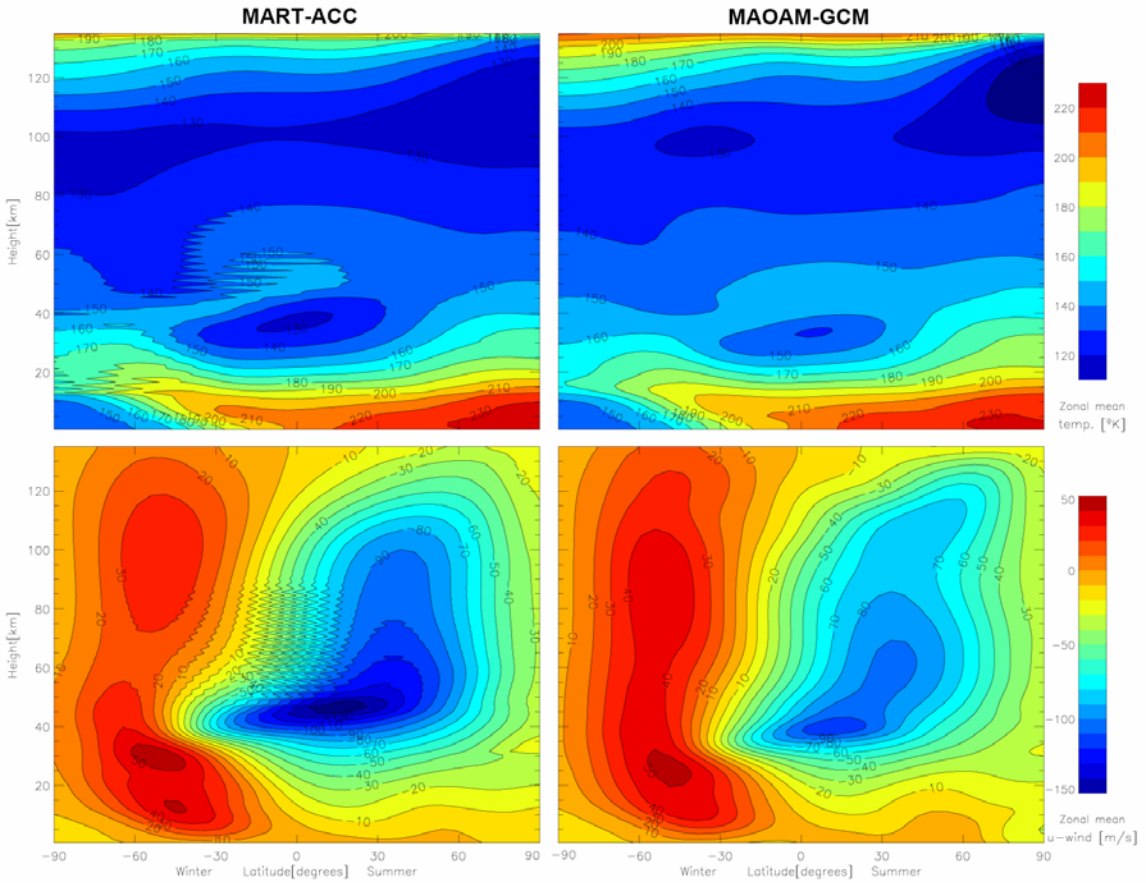

Figure 3.10: Comparison between the MART-ACC model and the MAOAM-GCM with stability improvements. The numerical instabilities, represented by ripples in the solution, are not present on the MAOAM-GCM and most importantly, the improved computation does not affect the modeling of the general circulation since both results are comparable.

3.4 Simulation results

3.4.1 Validation and comparison to in-situ measurements

The new global maps of temperatures on Mars retrieved by the orbiting MGS spacecraft with its onboard TES instrument open a new era in the modeling of the Martian atmospheric circulation. Although the results are limited to a maximum altitude of 60 km, they are important for the validation of the numerical model in the lower atmosphere of Mars. The TES instrument is an infrared interferometer/spectrometer with six detectors which simultaneously take spectra covering the spectral range of 6-50 μm [Smith et al., 2001]. The observations are performed in the nadir mode with a unique spatial resolution and altitude coverage below 40 km, or in the limb mode, provide a refined vertical coverage of the temperature profiles up to 65 km. The thermal information is retrieved from the spectral information of the CO_2 band at 15 μm, which is the most representative band since main radiative interaction with the atmosphere takes place at this spectral range.

The zonal wind field is estimated from the temperature field by assuming a gradient wind balance in the horizontal and hydrostatic balance in the vertical, and therefore a balance between the horizontal pressure gradient force, the Coriolis force, and the centrifugal forces [Smith et al., 2001]. This means that the wind field is not directly measured by the instrument but retrieved from a hydrodynamical model based on the thermal structure of the atmosphere. This approximation is very accurate outside the deep tropics for zonal mean eastward wind 'u' on Mars, and is described by the following equation [Leovy, 2001]:

$$\frac{\partial(u/T)}{\partial z} = -g\left(f^{*}T^{2}\right)^{-1}\frac{\partial T}{\partial y} \qquad f^{*} = (2\Omega\sin\phi + u\tan(\phi/a)) \qquad (3.39)$$

To solve equation (3.39), information of the horizontal surface wind is needed. Since wind near the surface is generally much weaker than the wind aloft, the surface wind is set to zero for the retrieval [Smith et al., 2001, Levoy, 2001]. Even though the resulting wind field is approximated, is important for the validation of the hydrodynamical core of the general circulation model. Consequently, the zonal wind field as shown on Figure 3.11 is retrieved assuming equation (3.39) stated above and the thermal distribution as measured by the MGS-TES instrument.

The results from the MAOAM general circulation model, as shown on Figure 3.11, are obtained taking into account the new improvements on the numerical computational scheme, the implemented orographic features as measured by the MOLA instrument, and an enhanced radiative transfer module for the short/long-wave energy balance. The improvements on the computational scheme as previously described provide a stable solution free from numerical instabilities. In addition, through the study of the radiative energy balance and its parameterization for the atmospheric interaction with the short/long-wave radiation, there are important agreements between the measured atmospheric parameters as obtained by the MGS-TES instrument and the numerical results as obtained with the MAOAM-GCM. The flux exchange with the surface is represented by the Newtonian cooling approximation, where the cooling coefficients were reformulated and newly parameterized as described in section 3.3.2. In the case of atmospheric radiation represented by the cooling to space approximation, the study was based on the broadening of the long-wave band at 15 μm and its influence on the atmospheric emission. Furthermore, the atmospheric absorption of the solar radiation, the main heating source, has a proper characterization for the seasonal cycle considering Martian orbital parameters, such as its high eccentricity.

During the observations of TES at southern summer solstice (LS=270º), the MGS-TES instrument measured an important dust opacity [Smith, 2001] related to a seasonal dust storm. At present however, the MAOAM-GCM is a dustless atmospheric model, and hence the transport of dust and its interaction by the atmospheric radiation is not fully described. The phenomenon was preliminarily addressed by considering an extra term in the radiation module related to a stronger absorption and emission of short/long-wave radiation on the lower atmosphere where the dust content is mainly located. Work is continuing to add a complete characterization for the dust distribution and transport and most importantly its influence on the radiative transfer, which would later enrich the numerical modeling of the Martian atmosphere.

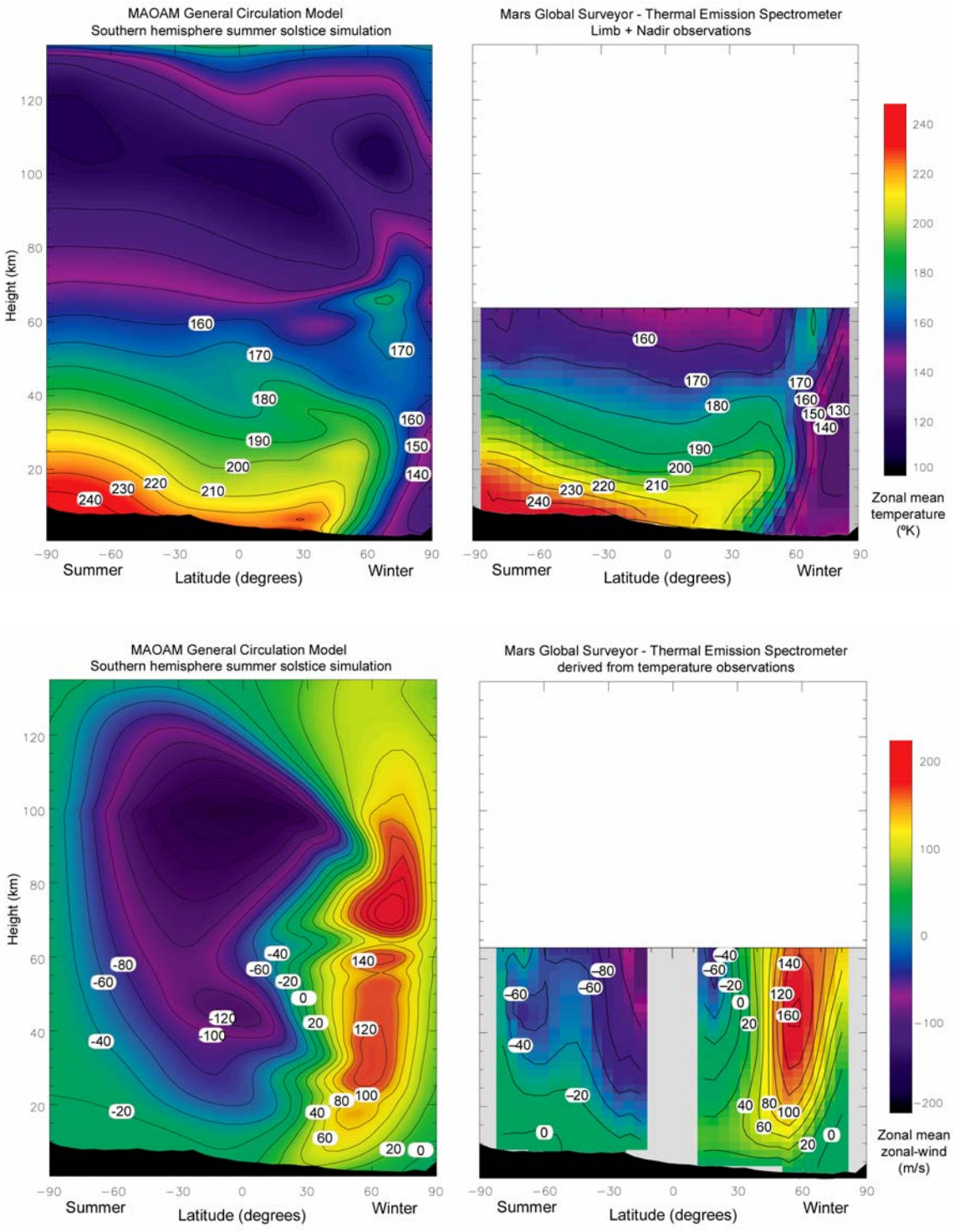

Figure 3.11: Comparison between simulation results from a modelled Martian atmosphere with the MAOAM-GCM and in-situ measurements as observed by the MGS-TES instrument. The thermal distribution is retrieved from the observed variation of radiance across the 15 μm CO_2 band [Smith et al., 2001]. The zonal wind field is derived from the thermal structure following equation (3.39), and it is therefore not directly measured. The MAOAM-GCM simulation is calculated assuming the reformulated parameterization of the radiative transfer module as described in section 3.3.2 and a stronger radiative interaction in the lower atmosphere driven by a static dust content.

3.4.2 The diurnal cycle

The rotation of the planet and its subsequent insolation variation drives the diurnal cycle. Because of the low density of the Martian atmosphere, the temperature of the surface is controlled primarily by solar heating and infrared cooling to the atmosphere and space, rather than heat exchange with the atmosphere. The lower few kilometers of the atmosphere during the day, and the lower tens to hundreds of meters during night, are in turn controlled by heat exchange with the surface. This region is also controlled by absorption of infrared radiation from the sun and the surface, and re-radiation to the surface, space and the rest of the atmosphere.

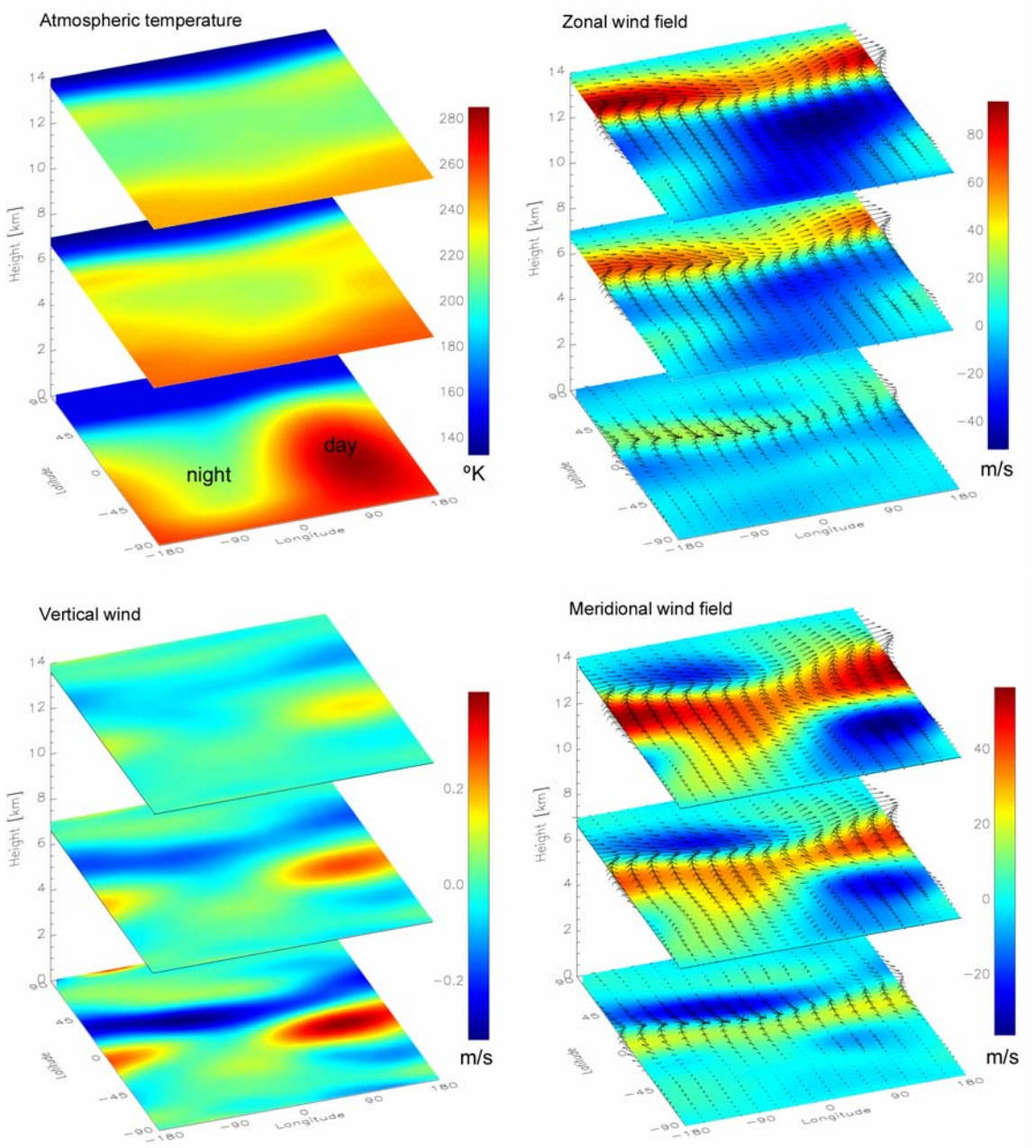

Figure 3.12: Simulation results considering the diurnal cycle for the different atmospheric parameters. The combined horizontal vector field is super-imposed on the meridional and zonal wind plots. The vertical wind field plot shows the strong dependence of the upward / downward flow with respect to the diurnal cycle on the lower atmosphere.

For the same conditions, (wind speed, clear sky, etc.), the denser atmosphere of Earth near the surface exchanges more heat with the surface than does the thinner Martian atmosphere. Hence, the terrestrial atmosphere has more influence on the surface temperature than does the Martian atmosphere.
Furthermore, the diurnal cycle on Mars is particularly strong due to the dry and arid structure of the Martian surface which has a very small thermal inertia, and therefore has reduced capability of accumulating heat. This produces strong fluctuations of the surface temperature during day and night as plotted on Figure 3.12, and are comparable to the daily variations measured by the Mars Pathfinder mission of $\Delta T=\sim 60^{o}K$ [Schofield et al., 1997]. The lower layers of the atmosphere with a stronger radiative interaction with the surface are more affected by the cycle. Above 15 km however, the circulation and atmospheric parameters are mostly driven by the zonal mean structure as observed on the atmospheric temperature and zonal wind results in Figure 3.12.
The warming of the parcels of air with direct solar insolation induces an adiabatic expansion, and therefore a rising flow of air in the day-side, as shown on the vertical wind plot. This is then balanced by an adiabatic compression and a subsequent downward flow in the night side.

3.4.3 The seasonal cycle

The seasonal cycle on Mars is comparable with the one on Earth due to the similar obliquity of both planets, with a planet obliquity $\varphi=25.19^{o}$ on Mars and $\varphi=23.5^{o}$ on Earth. As was previously described on Mars however, the eccentricity of its orbit induces strong variations in the solar insolation during the full orbital transit. These variations are as large as 45% between aphelion and perihelion, which is much higher than on Earth with an insolation difference of less than 15% as shown in Figure 3.5. Furthermore, the thinness of Martian atmosphere leads to a more rapid atmospheric equilibrium time when compared to Earth. This effect together with the high obliquity orbit makes the constantly illuminated summer polar regions warmer than the rest of the hemisphere, and causes the meridional temperature gradient to reverse.
The seasonal dependence of the planet declination affects the zenith angle and therefore the optical path of the incoming absorption path following equation (3.10). At the equinoxes, where $Ls=0^{o}$ (northern hemisphere spring) and $Ls=180^{o}$ (northern hemisphere autumn), the solar insolation warms the equatorial region leading to a zonal mean dual Hadley circulation pattern with a significant hemispheric symmetry. The topographic anisotropy between both hemispheres with the lowlands on the north and the highlands on the south contributes to a differentiation between both hemispheres.
A dustless atmosphere was considered for the simulations at the equinoxes ($LS=0, 180^{o}$) and at the northern hemisphere summer ($LS=90^{o}$), while at the southern hemisphere summer ($LS=270^{o}$) dusty conditions were considered. The assumed seasonal dependence of the dust content is based on measurements, for instance, those obtained by the MGS-TES instrument [Smith et al., 2000]. These measurements show a systematic variation of the dust content with season and a higher dust opacity (and more frequent regional dust storms) in the southern hemisphere summer. Even though there is an interannual variability of the dust opacity and each dust storm is unique and evolves in its own way [see Martin et al., 1986; Richardson et al., 1993; Clancy et al. 1996], the simulations are intended to reproduce the evolution of the atmosphere considering a typical annual Martian cycle.

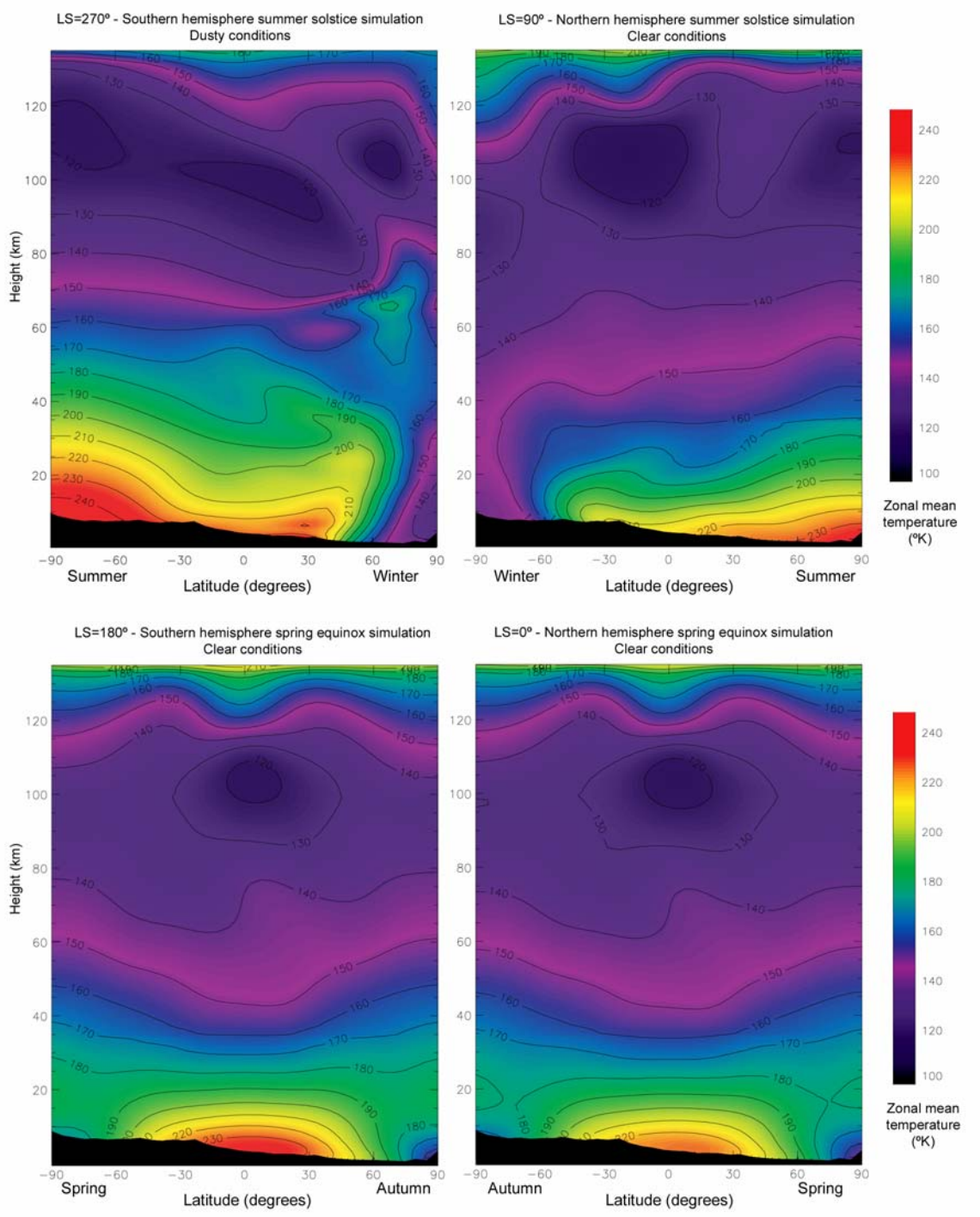

Figure 3.13: Seasonal variation of the Martian thermal structure. At southern hemisphere summer, the dust content introduces an extra source of heating in the lower regions of the atmosphere which extends to higher regions, following the general circulation. There is a strong dependence of the incoming solar flux with season because of the high eccentricity of the Martian orbit. This leads to a warmer equator at LS=180º when comparing equinoxes and a colder northern summer when considering solstices.

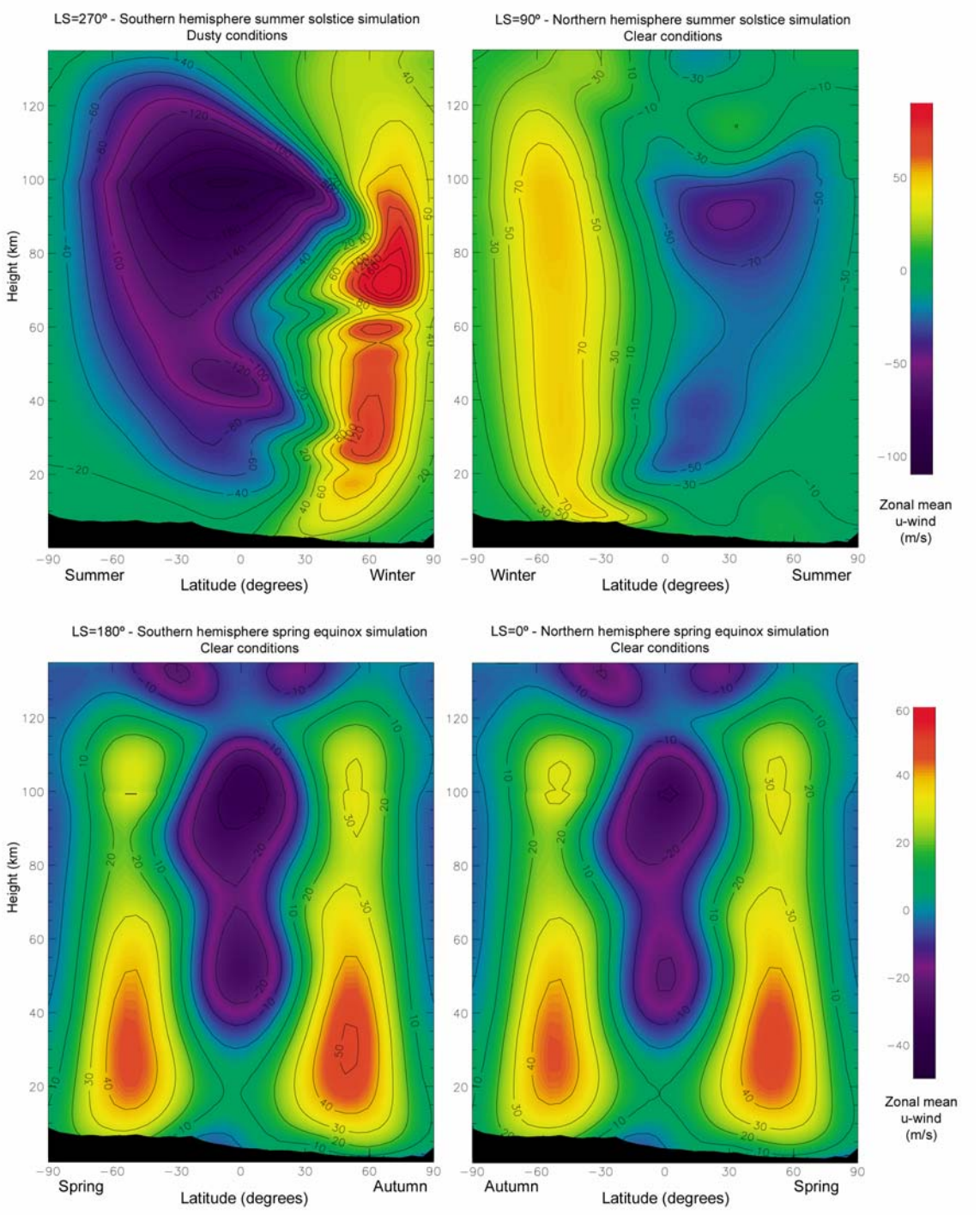

Figure 3.14: Zonally averaged zonal wind as a function of the seasonal cycle. At both equinoxes the zonal circulation is comparable. At solstices however, the wind field is particularly affected by the insolation difference and the dust content with a zonal velocity field at southern hemisphere summer. This can be as much as twice than on the opposite solstice. The intensive (blue, negative value) zonal winds in the southern hemisphere for LS=270º are mainly caused by the daily-mean northward directed temperature gradients. Furthermore, the strong dependence of the zonal winds on the altitude and latitude allows one to conclude that baroclinic and barotropic waves are generated in the atmosphere.

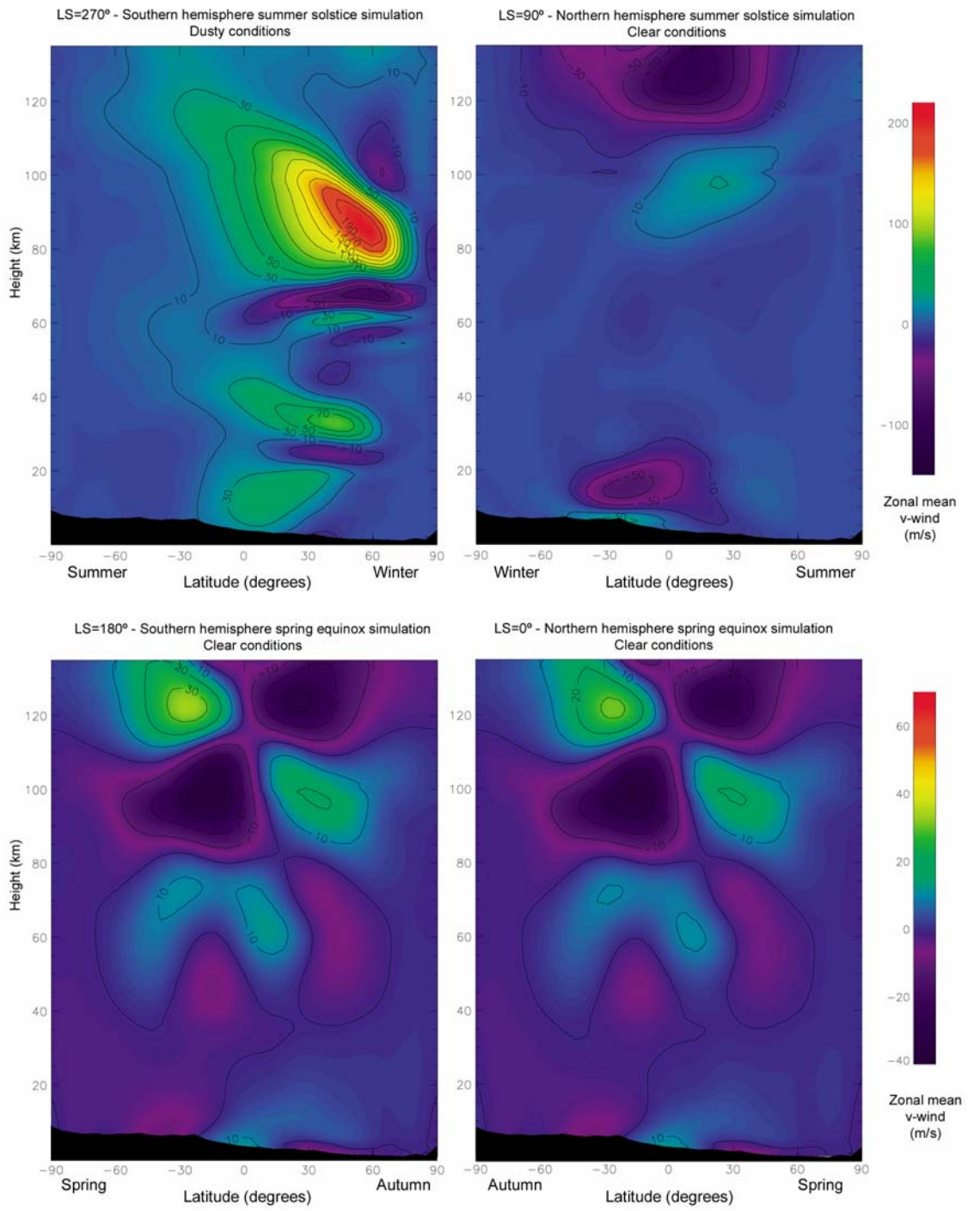

Figure 3.15: Zonally averaged meridional wind fields calculated for the different Martian seasons. The wind pattern below 40 km suggests the existence of one and two Hadley cells at solstice and equinox respectively. At higher altitudes, the velocity field is dominated by the non-linear terms in the momentum balance equation (3.13), which are much larger than the Coriolis term. Therefore, the atmosphere becomes non-geostrophic at these altitudes.

Chapter 4

Planetary observations with CTS

This chapter describes the remote sensing capabilities of the CTS spectrometer in the retrieval of atmospheric distribution of species in our neighbouring planets, especially on Mars, and the spectral analysis of weak cometary emission lines. The measured spectral information is influenced by the atmospheric transmission, the receiver's characteristics, the spectrometer's performance, and the telescope's structure and optical system. The observation could be then divided into a "pre-analysis", where based on estimated parameters the feasibility of the detection is quantified and a "post-analysis" where the observed data is properly calibrated and processed to represent the physical phenomena of study. In this chapter, observations of HCN on comet Encke are shown from where outgassing properties are derived. Finally, a 6-point map of the CO broadened line at 345 GHz was obtained in coincidence with Mars-Earth close approach on September 2003. From the observations, temperature profiles were retrieved and then compared to simulation results of the MAOAM general circulation model.

4.1 Spectral remote sensing

4.1.1 Sensitivity of the observation and system temperature

The sensitivity of the measurement, and therefore the minimum detectable signal, can be expressed by the radiometer formula. The expected noise is calculated by:

$$\langle \Delta T \rangle = \frac{T_{sys} + \eta_{FF} T_{source}}{\sqrt{\eta_{obs} B \cdot t / 2}} \qquad (4.1)$$

where B is the spectral channel width, t the observational time on the source, T_{sys} the system temperature, η_{FF} the filling factor of the telescope's beam, T_{source} the brightness temperature of the observed source and η_{obs} is the observational efficiency coefficient which depends on the observation mode. For dual-beam switched observations (the usual case for submillimeter and far-IR observations), the telescope constantly switches between the source and cold sky,

therefore $\eta_{obs}=t_{source}/t=0.5$. The filling factor coefficient is related to the source size on the sky and the beamwidth of the telescope which is quantified by the FWHM of the antenna gain pattern. The FWHM represents 50% of the collected energy when considering a Gaussian beam pattern, and is related to the size of the main dish (D) and the observational wavelength (λ) as described in equation C.1 [see appendix C]. Consequently, the filling factor for a source's size θ_{source} [rad] can be written as:

$$\eta_{FF} = 1 - 2^{-\left(\frac{\theta_{source}}{1.22\ \lambda / D}\right)^2} \qquad (4.2)$$

The system temperature summarizes all the sources of noise in the measurement, being the receiver temperature (T_{rec}) as analyzed in equation (1.7), the contribution by the atmosphere on the line of sight (T_{sky}) and the contribution from the telescope itself (T_{tel}). This last term can be expressed as:

$$T_{tel} = \left(1-\eta_{tel}\right)T_{tel}^{physical} \qquad (4.3)$$

where η_{tel} is the telescope efficiency after accounting for ohmic losses and spillover, and $T_{tel}^{physical}$ is the mean radiation temperature of the telescope enclosure (270ºK HHSMT, 220ºK SOFIA). The atmosphere is also an important noise source. This contribution represented by T_{sky} is dependent on the atmospheric temperature and the transmission, written as:

$$T_{sky} = \left(1-\eta_{sky}\right)T_{sky}^{physical} \qquad (4.4)$$

where η_{sky} is the atmospheric transmission dependent on the atmospheric opacity τ and the zenith angle (θ) by $\eta_{sky}= e^{-\tau}/\cos\theta$, and $T_{sky}^{physical}$ is the average temperature of the atmosphere (~255ºK). The noise temperatures quantified above would then affect the overall measurement in the following manner for an observation on source (P_{on}) and off source (P_{off}):

$$P_{on} = G_T\left(\eta_{SB}\cdot\eta_{ant}\cdot\eta_{sky}\cdot T_{source} + T_{rec} + T_{sky} + T_{tel}\right) \qquad (4.5)$$

$$P_{off} = G_T\left(T_{rec} + T_{sky} + T_{tel}\right) \qquad (4.6)$$

where η_{SB} quantifies the effect of a single or a double side band detection, and G_T is the temperature to intensity gain factor. For a DSB (Double-Side-Band) detection, the observed line is located in only one of the sidebands of the two detected, and hence $\eta_{SB}=1/2$. For a SSB (Single-Side-Band) detection $\eta_{SB}=1$. The source temperature (T_{source}) can be then derived from equation (4.5) and (4.6):

$$T_{source} = \frac{P_{on}-P_{off}}{P_{off}}\left[\frac{T_{rec}+T_{sky}+T_{tel}}{\eta_{SB}\cdot\eta_{sky}\cdot\eta_{tel}}\right] \equiv \frac{P_{on}-P_{off}}{P_{off}}T_{sys} \qquad (4.7)$$

$$T_{sys} = \left[\frac{T_{rec}+T_{sky}+T_{tel}}{\eta_{SB}\cdot\eta_{sky}\cdot\eta_{tel}}\right] \qquad (4.8)$$

4.1.2 Spectral line formation and line-strength

The ability to detect the molecular signature will depend on the sensitivity of the observation and the spectral line strength, which is related to the amount of molecules radiative active on the source and the efficiency of collecting this radiation by the telescope. The submillimeter and far-IR spectral ranges are mostly populated by rotational and vibrational molecular lines. Vibrational-rotational bands are observed in the gaseous state, where the molecules are able to rotate freely. In condensed (liquid or solid) phases, one only observes the vibrational transitions of the molecules. In addition, not every molecule has rotational lines, as is the case of molecular nitrogen due to its lack of an electric or magnetic dipole moment. However, most of the molecules are detectable through observations at its rotational transitions, for instance water vapor, hydrogen cyanide (HCN), carbon sulfide (CS) (which was observed on comet C/2000 WM1 with a CTS instrument [Hartogh et al., 2001]) and carbon monoxide (CO) (used for remote sensing of the Martian atmosphere).

Since the molecules are embedded in an environment with a certain temperature and pressure, the emitted/absorbed radiation as a monochromatic spectral component is spectrally broadened in relation to the environment of formation. This radiation is also affected by the natural broadening arising from the uncertainty in energy of the states involved in the transition. In low temperature and high pressure environments, the spectral lines are mainly broadened by collisions, normally called "Lorentz broadening" or pressure broadening. The pressure broadening is dependent on the pressure (p) and the temperature (T) as follows:

$$f_L(\nu-\nu_0)=\frac{\alpha/\pi}{(\nu-\nu_0)^2+\alpha^2} \qquad (4.9)$$

$$\alpha(p,T)=\alpha_0\frac{p}{p_0}\left(\frac{T_0}{T}\right)^{1/2} \qquad (4.10)$$

where f_L is the shape factor, ν_0 describes the center of the line, α is the half-width of the Lorentz line shape, and α_0 is the reference half-width for STP: T_0 = 273K; p_0=1013 hPa. The second broadening effect is the Doppler broadening typical of comets, which follows a Gaussian distribution. This effect is uniquely defined by the temperature (T):

$$f_D(\nu-\nu_0)=\frac{1}{\alpha_D\sqrt{\pi}}\exp\left[-\left(\frac{\nu-\nu_0}{\alpha_D}\right)^2\right] \qquad (4.11)$$

$$\alpha_D=\frac{\nu_0}{c}\left(\frac{T}{m/2k}\right)^{1/2} \qquad (4.12)$$

where f_D is the shape factor, α_D is the half-width of the Doppler line shape, c is the speed of light, m is the mass of the molecule and k is Boltzmann's constant. The combination of Doppler and Lorentz broadening is represented by the Voigt profile, which describes the interrelation of both non independent phenomena by the following integral:

$$f_\nu(\nu-\nu_0)=\int_{*\infty}^{\infty} f_L(\nu'-\nu_0)\,f_D(\nu-\nu').d\nu' \qquad (4.13)$$

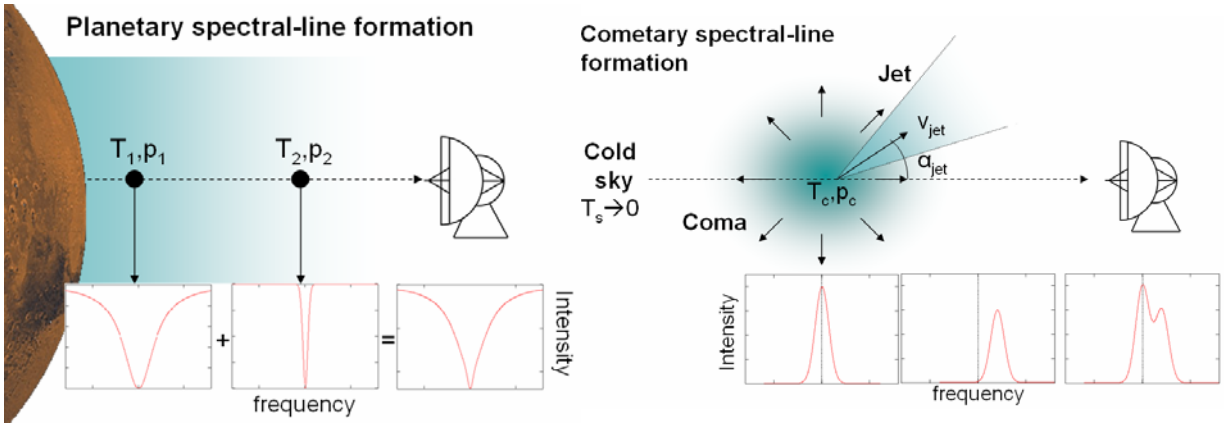

Figure 4.1: Sketch describing the line formation on planets and comets. The shown analysis is overly simplified but still provides a description of the processes involved in the absorption and emission features observed by the chirp transform spectrometer. In comets, the line profile shown contains the emission from the isotropic outflow and the emission from an active region with an outgassing velocity v_{jet} and a relative angle α_{jet}.

Depending on the environmental pressure and temperature, the broadening of the spectral line provides an important link between spectral displacement and different formation regions of the line. For instance, the line intensity of a planetary broadened line at the wings provides information about the lower layers of the atmosphere, while points near the center give that about the upper regions. Through this relation and with a highly resolved spectral distribution, the distribution of the chemical constituent can be retrieved at different points in the line of sight. The estimation of the line strength requires an a-priori knowledge of the astrophysical object of study and its chemical diversity. For planetary spectral lines shown in Figure 4.1 [left], the expected line intensity and line shape are calculated through a forward radiative transfer model [e.g Jarchow, 1999], which takes into account an estimated a-priori distribution of the constituent.

4.1.3 Optimum observation modes with the CTS spectrometer

There are different methods and sequences of observation. They differ on the calibration process, the way the optical system is arranged for source and reference, and the way time is distributed. Dual beam switch mode is the suitable choice for the study of small angular sources (less than 4 arcmin), such as planets and comets typical of research done with a CTS spectrometer. In the beam-switch part, the secondary mirror of the Cassegrain optical system ‘chops’ between ‘off’ and ‘on’ positions at frequencies between 5 and 0.1 Hz. This periodic measurement will cancel the atmospheric variability that affects the atmospheric transmission and therefore the calibration process. As the optical path in the instrument is changed by the motion of the chop mirror, the standing waves in the system will differ between two chop directions resulting in an imperfect cancellation of the baseline ripples. Consequently, the observation is combined with a position switch measurement every 20-60 seconds, where the telescope is moved in order to bring the source into the other beam.

Chopping frequency

The chopping frequency is constrained by the mechanical setup, the stability of the instrument, the atmospheric variability, the dead times during switching and the available data-transfer rate. The stability and resolution of the backend will primarily define the maximum time allowed on the source. The minimum of the Allan variance, which for the SOFIA-GREAT-CTS is t_A=510 s, indicates the turn-over from the dominance of the radiometric noise to the dominance of the drift effects. A time on source longer than 510s (f_{chop}≈2 mHz) is therefore useless and not recommended. The dead time between the stabilization of the switching process and the data transfer rate constrains the maximum chopping frequency. The data read-out of the electronics in SOFIA-GREAT-CTS and the latency of the TCP/IP interface with the main control computer of the telescope give a maximum chopping frequency of 3 Hz. In HHSMT, the dead time between two chopping positions is approximately 35 ms and therefore chopping at 3 Hz reduces the observational efficiency by 10%. Consequently in most of the cases, a chopping frequency of 1 Hz is recommended, which still provides reasonable atmospheric cancellation.

Chopping throw

The chopping throw, or the spatial distance between reference and source, is restricted by the secondary chopping mechanism and the extension of the source. For the study of extent sources, a distance of 4 arcmin will be unable to provide a cold sky reference and therefore position-switch would be an alternative. For the study of cometary bodies, the reference should be located at a prudent distance from the extended coma and tail. This can be approximately determined from the molecular lifetime (τ_m) or the time before the molecule is photodissociated, the outflow velocity ($v_{outflow}$) and the geocentric distance (d_{geo}) as follows:

$$\theta_{\min}^{chop} = \arctan\left(\frac{\tau_m \cdot v_{outflow}}{d_{geo}}\right) \qquad (4.14)$$

4.2 Cometary observations with SOFIA-GREAT-CTS

The high resolution provided by the chirp transform spectrometer represents a significant advantage in the study of narrow spectral line features observed on comets. The low collision rate between molecules and the low temperatures in comets lead to small broadening effects on the line, and these are only resolved with resolutions higher than 0.5 km/s. Observations with a SOFIA-GREAT-CTS prototype were organized at the submillimeter telescope (HHSMT) in Arizona, USA. The study of hydrogen cyanide (HCN) at the submillimeter spectral range was combined with a simultaneous observation of its photodissociated daughter molecule cyanide (CN) at the optical range. The selected object was the periodic comet Encke/2P which has a small period of 3.3 years and was particularly bright during the observational period, with an optical brightness of ~7. Dynamically old comets like Encke frequently show dust jets or fans which are believed to be driven by the remaining active regions on an otherwise already rather inactive nucleus. An outgoing fan of hydrogen cyanide (with a particular angle φ to the observer) leads to a Doppler shifted component ($\Delta v = -v_{jet} \cdot \cos\varphi$), which produces the observed line profile as an asymmetrical spectral profile.

4.2.1 Definition of the observational campaign and feasibility of the measurement

In order to define the detection limits and S/N ratio, we initially calculated the expected T_{sys} and noise for different integration times and atmospheric opacity. The following considerations were taken into account:

- Position-beam-switch observations (t/2=ON, t/2=OFF).
- The observations will be performed in double sideband
- The spectral resolution is 46.6 kHz (SOFIA-GREAT-CTS resolution).
- For short integration periods (1h) and non-optimum atmospheric conditions, a broader spectral resolution is chosen (BR: $B=2^3 \cdot CTS=372.8$ kHz or 0.315 km s^{-1} at 354 GHz, see radiometer formula 4.1)
- The receiver temperatures are: SIS-230 (100ºK) and SIS-345 (125ºK).
- The terrestrial atmosphere is considered to be at 250ºK.
- A Gaussian antenna pattern with a main dish of 10 meters in diameter.

Molecule	Transition	Frequency [Ghz]	Tsys [K] τ=0.1	Noise [mK] t=1h BR	Noise [mK] t=30h BR	Tsys [K] τ=0.05	Noise [mK] t=30h
Hydrogen cyanide (HCN)	J 3 → 2	265.886	295.43	16.13	2.94	253.01	7.13
Hydrogen cyanide (HCN)	J 4 → 3	354.505	541.67	29.62	5.39	388.90	10.96
Formaldehyde (H_2CO)	J 5_{15} → 4_{14} ortho line	351.768	527.39	28.79	5.26	382.35	10.78

Table 4.1: Estimated observation sensitivity for hydrogen cyanide and formaldehyde at submillimeter wavelengths.

- The line intensity is estimated considering a similar observation of HCN(4-3) of comet Hale-Bopp by Hofstadter et al. [1997], who obtained a 17ºK line for HCN(4-3) and a 4ºK line for $H_2CO(5_{15}\text{-}4_{14})$.
- The production rate of CN on comet Hale-Bopp as derived by Woodney et al. [2002] is $2\text{x}10^{28}$ s^{-1}, while it is much weaker on comet Encke ($1.5\text{x}10^{25}$ s^{-1}) as derived by A'Hearn et al. [1977] at comparable heliocentric distances.
- When considering that the HCN/CN and H_2CO/HCN ratios are the same as Hale-Bopp, the line intensity can be then estimated as:

$$T_b^{Encke} = T_b^{Hale-Bopp}\left(\frac{Q_{HCN}^{Encke}}{Q_{HCN}^{Hale-Bopp}}\right) \qquad (4.15)$$

- The length-scale of HCN at 1 AU heliocentric distance is 10^4 km.
- The HCN(4-3)/HCN(3-2) ratio is defined to be = 3.
- The observations of Hale-Bopp were performed at 1.33 AU (θ_{HB}=10") while comet Encke at 0.27 AU (θ_{Encke}=51").

Molecule	Transition	FWHM [arcsec]	η_{FF} Hale-Bopp	η_{FF} Encke	Signal [mK]	S/N τ=0.1 t=1h BR	S/N τ=0.1 t=30h BR	S/N τ=0.05 t=30h
Hydrogen cyanide (HCN)	J 3 → 2	28.37	-	1.00	27.78	1.72	9.45	3.89
Hydrogen cyanide (HCN)	J 4 → 3	21.28	0.153	1.00	83.33	2.81	15.46	7.60
Formaldehyde (H_2CO)	J 5_{15} → 4_{14} ortho line	21.44	0.150	1.00	20.15	0.70	3.83	1.87

Table 4.2: Estimated signal to noise ratios derived from a previous observation of hydrogen cyanide at Hale-Bopp and an optical observation of cyanide on comet Encke.

Signal-to-noise ratios higher than 3 are considered to be feasible detections and are marked on the table with an underline. The weakness in the production rates on comet Encke would demand a sacrifice on the spectral resolution for short integration times. Therefore for atmospheric opacities (τ) higher than 0.1, the broader resolution mode (BR) is considered.

4.2.2 Observations and data reduction

These submillimeter observations were obtained over a period of two weeks near perihelion (1 month before) on November 2003. During this period the comet showed a strong variability in its activity as observed in both spectral ranges: optical for CN and submillimeter for HCN.
Table 4.3 lists all the observation used in this analysis. Due to the high orbital speed of the comet, the heliocentric distance of the comet varied by more than 20% during the observational time-frame of 10 days. This is important in monitoring the HCN production rate evolution at pre-perihelion.
The HCN J 3-2 spectral information was detected with a double side-band Steward Observatory Radio Astronomy Laboratory (SOLAR)-SIS receiver while the upper transition of HCN J 4-3 and H_2CO with the MPIfR-SIS receiver at 1 mm wavelength. The detected signal was then spectrally analyzed with the SOFIA-GREAT-CTS prototype (Δf=46.6 kHz, BW=215 MHz), and a set of two AOSs (AOSA: Δf=934 kHz, BW=1000 MHz; AOSB: Δf=913 kHz, BW=1000 MHz). The data was reduced using CLASS (Continuum and Line Analysis Single-dish Software) for calibration, statistical analysis, integration and alignment of multiple time-frame observations and retrieval of the line intensities. The velocity resolution for the data set was organized in order to provide sufficient signal-to-noise ratios. In the case of the CTS, the short term data sets were analyzed with a resolution of 458 kHz (approximately 7 points per HCN J 4-3 line), and 229 kHz for the daily and total averaged data. While the AOS data was analyzed with an oversampling mode of 480 kHz.
From the calibration process performed by the telescope, effects such as the beam efficiency are not included, and hence the detected antenna temperatures are calibrated to radiation intensity or brightness temperature by comparing them to known sources. The selected sources were the "Egg nebula" (CRL2688) an importantly large cloud of dust and gas ejected by its central red giant star, the "v_cyg" star and the Mira type star "chi_cyg".

Date	r [AU]	Δ [AU]	Sun - Encke - Earth pos. angle [º]	HCN J 3 → 2 [UT time]	HCN J 4 → 3 [UT time]	H_2CO J 5_{15} → 4_{14} [UT time]
November 19/20	0.959	0.263	85.05		19:30-04:00	
November 20/21	0.943	0.265	88.54	19:30-04:30		
November 21/22	0.926	0.268	92.02	17:30-04:00		
November 23/24	0.892	0.275	98.92	20:30-03:30		
November 24/25	0.877	0.279	102.33	18:30-00:30		
November 26/27	0.839	0.291	109.05		19:30-02:00	
November 27/28	0.823	0.296	112.34		18:00-02:00	
November 28/29	0.804	0.303	115.60		18:00-01:30	
November 29/30	0.789	0.309	118.80		17:00-20:30	20:30-02:00
Nov. 30/ Dec. 01	0.770	0.318	121.96		17:30-01:30	

Table 4.3: Time summary of the cometary observations performed on Encke/2P from the Heinrich Hertz Submillimeter Telescope (HHT) on Mount Graham.

The study of morphological variability from the line profile is extremely complex on comet Encke because of its weakness in production rates. This poses a true challenge for high resolution spectroscopy, where noise increases as a function of the spectral resolving power. The approach consisted of setting observational data-sets which were overlapping in time and were helpful in the improvement of the sensitivity. Depending on the observational conditions, technical feasibility and weather conditions, between 4 and 8 data sets were obtained per observation day. The time spacing between the data-sets is 1 hour and each set represents a 2-hour integration time observation. It is difficult to distinguish outgassing asymmetries from these data-sets since the maximum obtained S/N ratios are in the range of 4. A better description of the spectral line shape and therefore the processes involved in the line formation, is retrieved from the daily analysis and considering a broader resolution with S/N ratios in the order of 10. However, information about the jet in a fixed position of the comet would be lost with such long integration times, because the time period is comparable with the presumable rotation period of the comet. In the case of the formaldehyde observations, we could not obtain S/N ratios higher than 3 for the integration period of 5:30 hours and therefore this spectral data is not analyzed further.

4.2.3 Analysis

The production rate can be derived from the integrated line intensity by modeling the physical effects involved in the outgassing phenomena. For the retrieval, Szutowicz [2004, priv. comm.] developed a cometary outgassing model which relates the measured line intensity to an estimated production rate. The cometary outflow is described by the Haser model [Haser, 1957] which assumes isotropic outgassing with a constant radial velocity, and that the parent species coming off the nucleus are exponentially decaying to the daughter-species. Using these assumptions, the local density in the coma is given by:

$$n(r_n) = \frac{Q}{4\pi \cdot r_n^2 \cdot v_{\exp}} e^{-\frac{r_n}{v_{\exp}\tau}} \qquad (4.16)$$

where Q is the production rate, r_n is the distance to the nucleus, v_{exp} is the outflow velocity and τ is the molecular lifetime. An outflow velocity of v_{exp}=0.8 km s^{-1} was assumed. The molecular lifetime for HCN was derived from empirical values. These were obtained from the radial dependence of isotropic CN comae observed by Fink & Combi, 2004, who obtained a mean lifetime (τ) of 26250 s for a heliocentric distance of 1 AU and a scale length of 21000 km. This leads to a photodissocation rate of $\beta_0 = (1/r_h^2\tau) = 3.81\text{x}10^{-5}$ s^{-1}, where r_h is the heliocentric distance. By integrating $n(r_n)$ over the field of view, the molecular column density $< N >$ can be derived in relation to the optical opacity. The relation between the column density and observed integrated line intensity is given by [Bockelee-Morvan et al., 2004]:

$$\int T_B dv = \frac{hc^3 A_{J,J-1}}{8\pi k \nu_{J,J-1}^2} u_J \langle N \rangle \qquad (4.17)$$

where $A_{J,J-1}$ is the Einstein coefficient of spontaneous emission, $\nu_{J,J-1}$ is the rest frequency and u_J is the relative population of the upper level, which are estimated based on HCN values given by [Bockelee-Morvan, 1984].

The main daughter product by photodisassociation of HCN is CN, which was measured simultaneously from the optical range. Jockers [2004, priv. comm.] derived similar production rates for the daughter molecule in the range of Q_{CN}=0.9-1.1 x 10^{25}, which is comparable with the submillimeter observations of the parent molecule of Q_{HCN}=1.0-2.9 x 10^{25}.

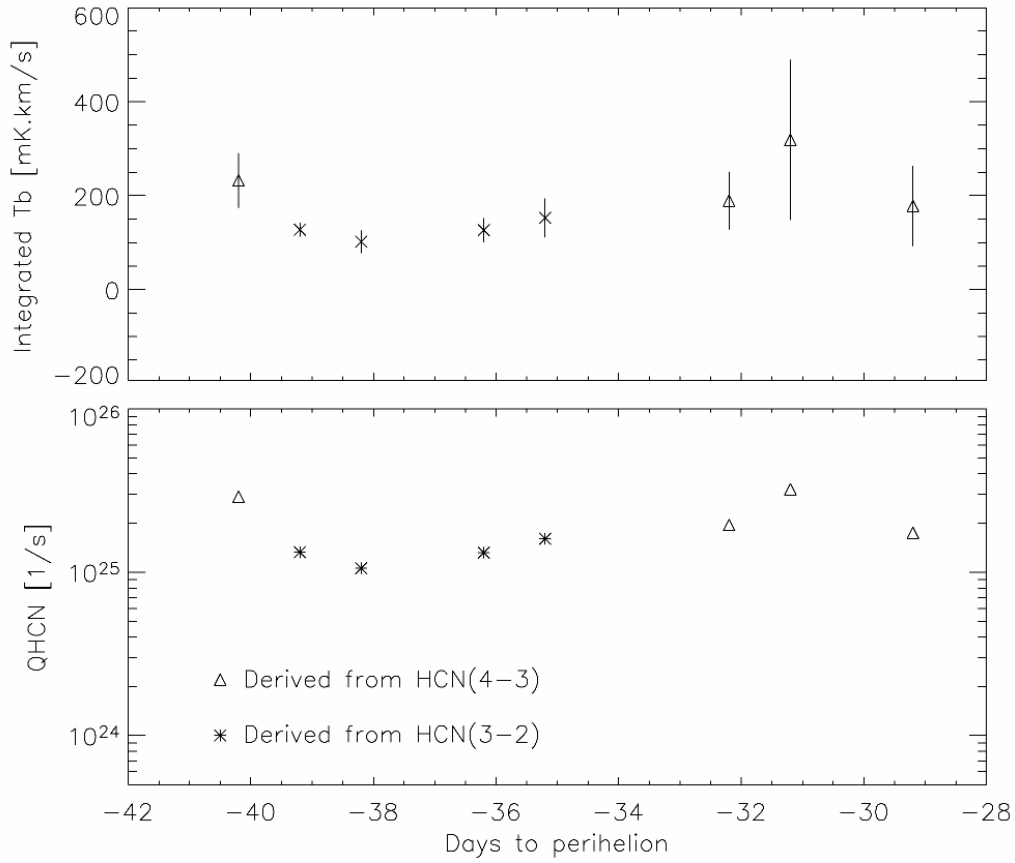

Figure 4.2: HCN production rates on comet Encke derived from the integrated line intensity. [Top] Integrated line intensity considering a line width of ±2 km/s. [Bottom] Retrieved production rates based on a cometary outgassing model developed by Szutowicz [2004, priv. comm.].

By analyzing the evolution of the spectrum as a function of the position angle (PA), an evolving Doppler shift of the emission feature can be seen [Figure 4.3]. This feature can be presumably explained by considering that most of the molecules come from a single active region, that is activated by the direct solar insolation. In other words, we should expect different relative velocity displacements at different position angles, following $-v_{jet} \cdot \cos(PA)$ where v_{jet} is the outflow velocity of the active region [see sketch on Figure 4.4]. This is consistent with our observations because firstly at November 19th with PA=85º, the measured spectrum of HCN J 4-3 showed a slight blue-shifted displacement. In the following days, the line then became increasingly red-shifted (PA>90º) in a monotonic manner. We could then conclude that on comet Encke, the outgassing phenomena is far from being isotropic and is mainly driven by a single solar related active region.

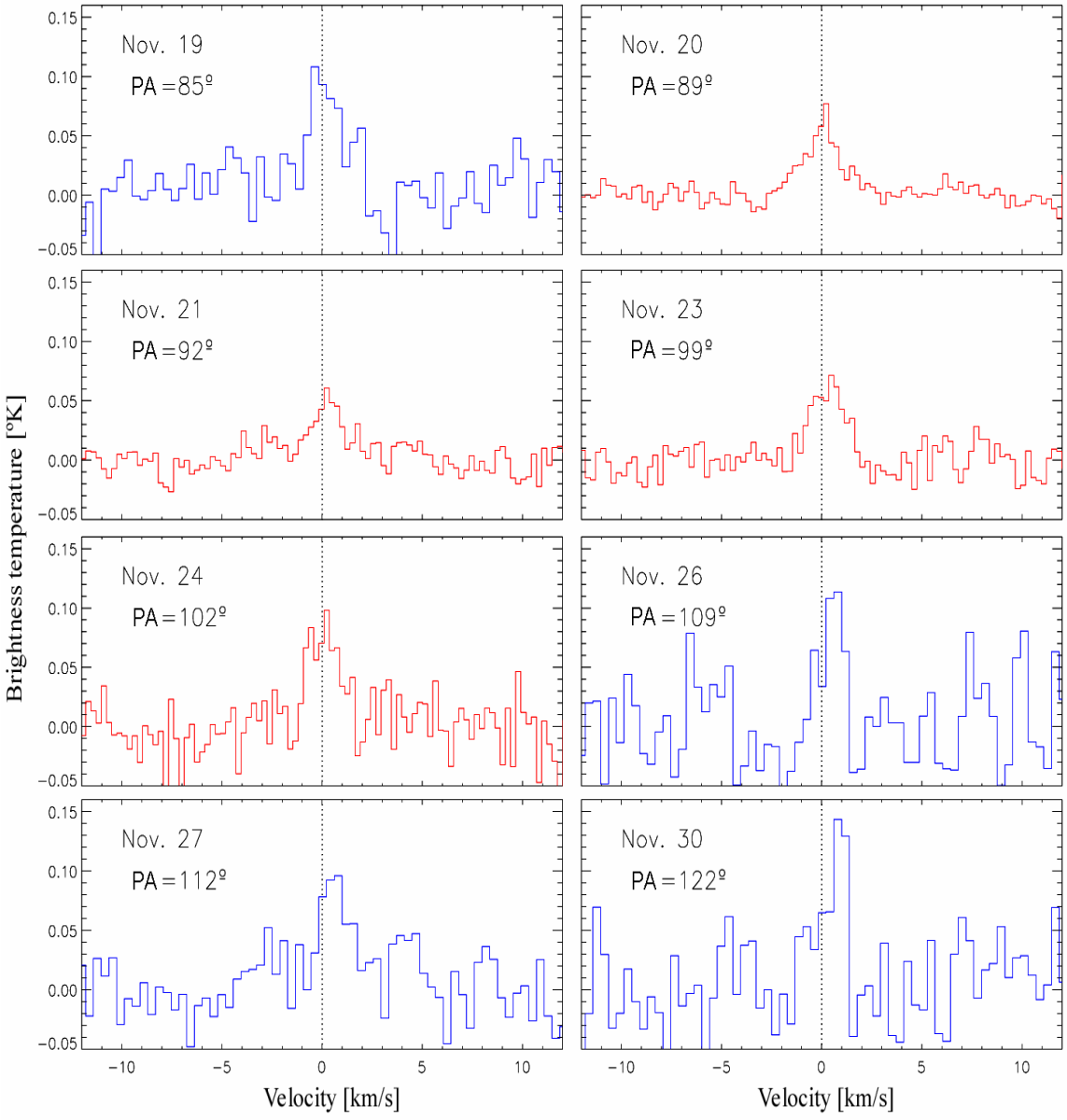

Figure 4.3: Observational summary for detections with S/N ratios higher than 3. The HCN(3-2) observations are plotted in red with a velocity resolution of 0.2587 km s^{-1} and the HCN(4-3) observations in blue with a resolution of 0.3880 km s^{-1}. Each set represents a complete observational day with integration times in the range between 3 and 6 hours.

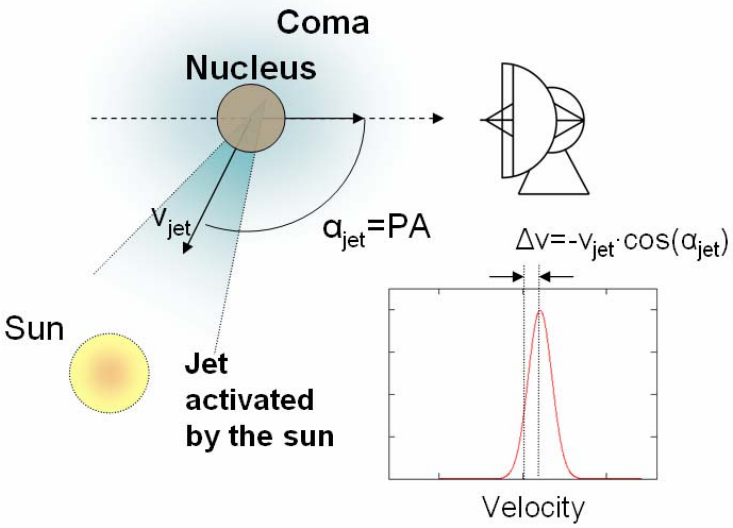

Figure 4.4: Diagram describing the emission features for a comet with an outgassing process mainly driven by a single solar related active region. In this particular case, the relative angle "active region"-"observer" is related to the Sun-Target-Observer angle (or position angle). This angle varied in our cometary observations on comet Encke from 85º in November 19th to 122º in November 30th.

4.3 Mapping the thermal distribution of Mars' atmosphere with SOFIA-GREAT-CTS

The distance between Earth-Mars was particularly small (0.40-0.44 AU) on September 2003. This short distance represented an angular size of ~22" for the complete Martian disk as seen from Earth. We therefore organized an observational campaign for remote sensing of the atmosphere of Mars. Considering the measured atmospheric opacity at HHSMT for September for the previous five years (1987-2002), we expected good observation conditions with τ=0.2 and τ=0.1 at 225 GHz. To take advantage of the complete observation run, we observed the CO J 2-1 transition in days with τ greater than 0.2, while with better conditions (τ between 0.2 and 0.1), we focused on the CO J 3-2 transition. Temperature profiles from the complete disk can be retrieved from the CO J 2-1 observations, while the CO J 3-2 line allowed the retrieval of a 6 point map of the global atmospheric thermal distribution up to 80 km.

The feasibility analysis and observation procedure was similar to that previously described for comet Encke. However in our study of planet Mars, the stronger signal provided a higher flexibility in the integration time. We considered a Martian continuum of 220 K for the analysis. The expected line intensity was derived based on an atmospheric transmission model and Martian synthesized spectral lines [Jarchow et al., 1997], considering a Gaussian antenna pattern with a main dish of 10 meters and an a-priori temperature profile based on Viking's entry temperature profile [see details in Villanueva, 2003b].

4.3.1 Observations and data reduction

The first week of the observation campaign was primarily used for the setup and integration of the SOFIA-GREAT-CTS instrument into the submillimeter telescope. First light on the instrument took place on September 19th and Mars was observed until September 30th. At this time stage, Mars had already passed its closest approach (August 27th 2003). However, Mars had a high angular diameter in the range of 20 to 23 arcsec during the observational period.

Date	Δ [AU]	Mars apparent diameter [arcsec]	Sun - Mars - Earth PA angle [º]	CO J 2 → 1 [UT time]	CO J 3 → 2 [UT time]
September 19	0.410	22.92	18.80	02:00-08:30	
September 20	0.412	22.75	19.30	02:00-09:00	
September 21	0.415	22.57	19.82		02:30-08:30
September 22	0.418	22.39	20.53	03:30-07:30	
September 26	0.433	21.64	23.13	03:30-08:30	
September 27	0.437	21.47	23.76		03:30-05:30
September 28	0.440	21.52	24.37	02:30-09:00	
September 30	0.448	20.86	25.53	04:00-08:00	

Table 4.4: Time summary of the observations on planet Mars with the SOFIA-GREAT-CTS instrument at the HHSMT telescope.

Most of the observations were based on the CO J 2-1 transition as shown in Table 4.4, mainly due to technical problems in the telescope with the 345 GHz receiver. The combination of optimum weather conditions ($\tau \approx 0.1$) and the fact that the 345 GHz receiver was up and running on September 21st allowed us to perform the planned 6 point map of Mars. With a similar frontend-backend setup previously described for comet Encke, the CO J 2-1 spectral information was detected with the double side-band SOLAR-SIS receiver at 230 GHz while the upper transition of CO J 2-3 with the MPIfR-SIS receiver at 345 GHz. The backend instruments were the two AOS instruments (AOSA, AOSB), the SOFIA-GREAT-CTS prototype and an analog CTS with a bandwidth of B=180 MHz and a spectral resolution of Δf=47 kHz. Initial data processing including continuum calibration, integration and alignment were performed with CLASS. The calibrated data was integrated and organized in daily data sets, with the exception of the map observation on September 21st which was separated in 6 data sets, one per data point for the Martian map. The integration time in each point was 1 hour, plus 10 minutes overhead for re-focusing and re-pointing of the telescope. This led to a total observational time of 7 hours for 6 points, which was in concordance with the time when Mars was 15 degrees above the horizon. With Mars having an apparent diameter of 22.57 arcsec and a FWHM of the beamsize of 21.74 arcsec at 345 GHz, the spatial spacing between the points in the map was defined to be 10 arcsec in right ascension and declination.

Strong baseline features were induced in the observed spectra of the order of 10% of the line intensity due to an unwanted response of the telescope's optical system. While studying the optical path of the incoming radiation from the main dish to the receivers, a correlation was found between the observed baseline period and the distance between the sub-reflector and the receivers flange. This led us to believe that this effect was introduced by a standing wave between these two points. A 3 step process was implemented for the removal of the baseline features: 1) we retrieved temperature profiles from the original spectrum obtaining also a synthesized spectrum and the residuals from the inversion process; 2) we performed an spectral analysis on the residuals, where we selected the three strongest components obtaining a set of three frequencies, phases and amplitudes; 3) and finally we subtracted the previously identified baseline components from the original spectrum. This process was repeated in all our reduced data sets, obtaining significant improvements in most of the cases.

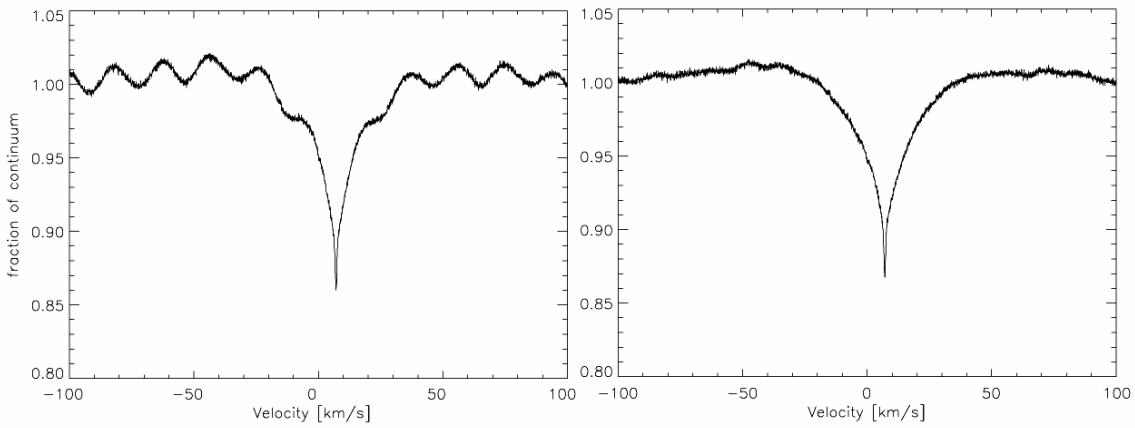

Figure 4.5: Original measured spectrum with baseline features (introduced by a disarrangement in the optical system of the telescope) and the corrected spectrum. The removal algorithm obtains the three strongest spectral components from the baseline pattern and subtracts them from the original spectrum.

4.3.2 Analysis

Remote sensing on Mars through CO rotational transitions is very convenient since they provide substantial line opacities at wavelengths where LTE emission is dominant. With a bandwidth of 215 MHz, the obtained high resolution spectra allow the retrieval of atmospheric temperature profiles from ground level to an altitude of 80 km [see averaging kernels in Figure 4.6]. The pressure profile used in the inversion is derived considering hydrostatic equilibrium with a Martian surface pressure of 6.75 hPa [see appendix G] and an altitude independent CO-volume mixing ratio of 8×10^{-4} (see Clancy et al., 1990 and Lellouch et al., 1991) is assumed. In order to properly weight the contribution of the observed spectrum as a function of the Martian disk, the whole disk and the limb area are subdivided into 16 circular rings of an appropriately chosen size [Jarchow et al., 1997].

For calibration purposes, the intensity of the observed line is represented as a fraction of the continuum. Continuum is defined as the observed brightness temperature at a frequency far from the line center. It can be concluded by observing the spectral line shape and the residuals of the inversion process shown in Figure 4.6, that the baseline pattern was successfully removed from the observed spectra since there is no indication of a sinusoidal component.

Mars was in its southern summer solstice (LS=270º) on September 2003, which as previously mentioned, is also normally associated with dust storms. The retrieved temperature profiles describe features which can also be seen in the simulations of the MAOAM-GCM for this season (see Figure 3.13 top-left); for instance, the lower temperatures in the middle atmosphere for the equatorial region and the colder lower atmosphere for the winter hemisphere. As shown in Figure 4.6 c) and d) for the equatorial region, the fraction of continuum for the line center is comparably smaller (~0.6) than for the southern hemisphere and northern hemisphere (~0.7). This high absorption shows the equilibrium temperature to be much lower than the continuum temperature at the upper region of the atmosphere, where the broadening of the line is minimum. This feature is then retrieved by the inversion algorithm, obtaining lower temperatures in the 70-80 km atmospheric region for the equatorial region as shown in the middle column from Figure 4.7.

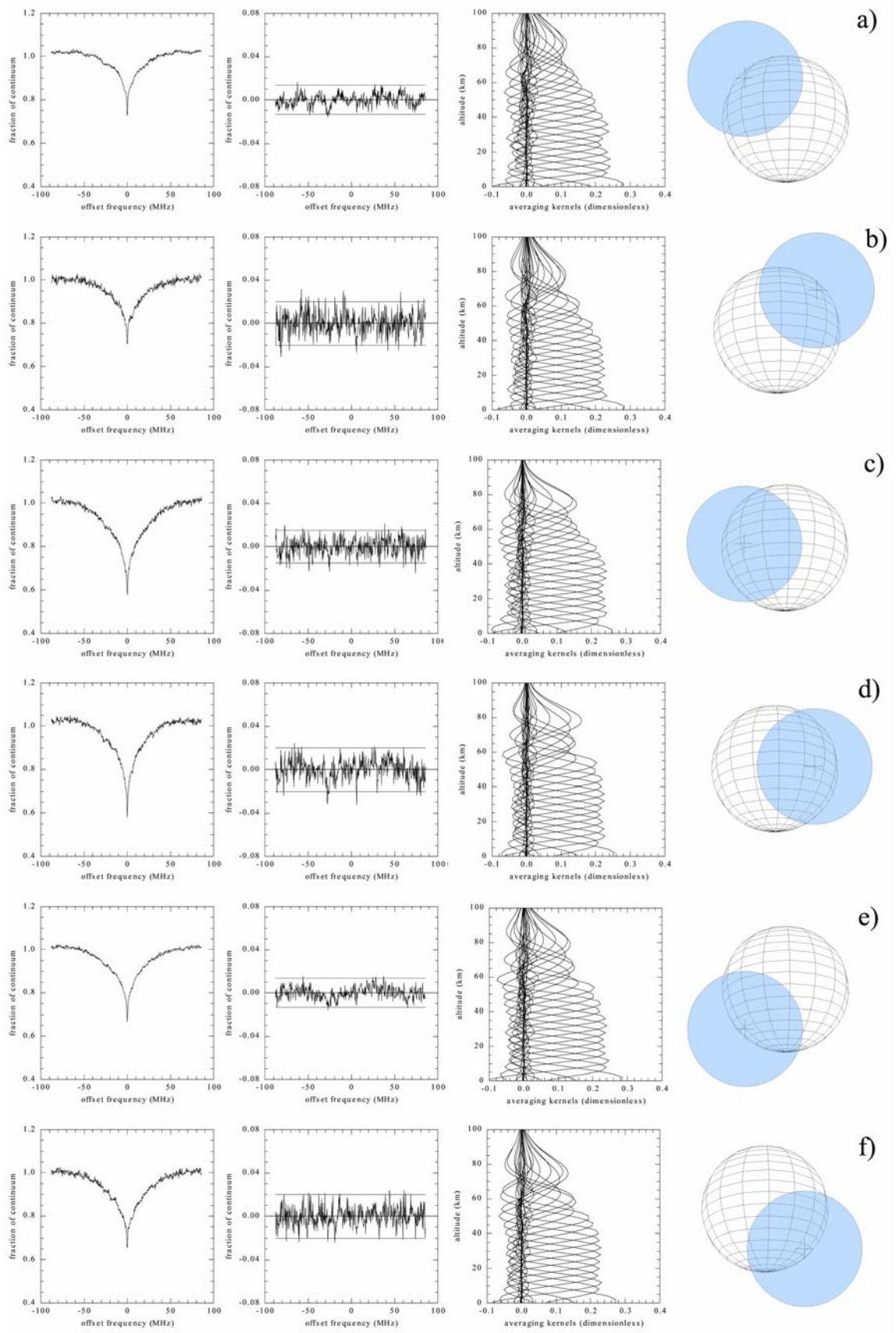

Figure 4.6: [Left] Observed CO J 3-2 broadened spectral line. [Middle-Left] Residuals from the inversion calculation. [Middle-Right] Averaging kernels of the inversion. [Right] Location of the observational beam with respect to the Martian disk.

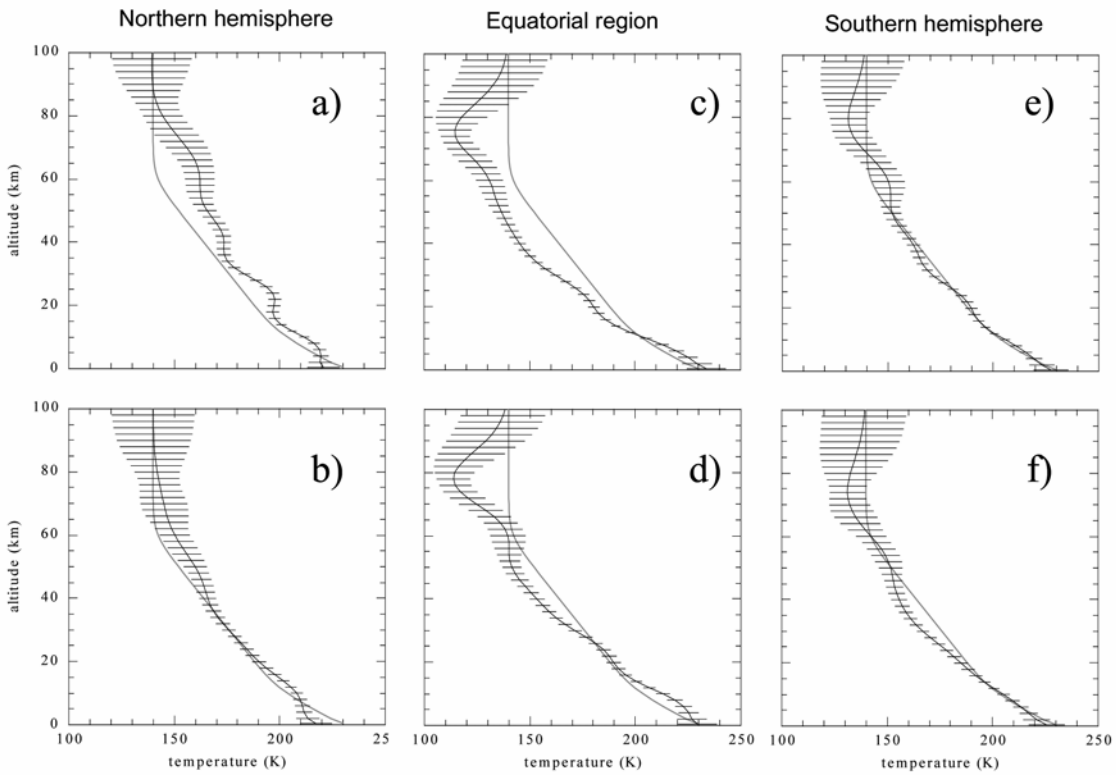

Figure 4.7: Retrieved temperature profiles (in black) from the 6-point submillimeter observation, as measured with the SOFIA-GREAT-CTS instrument in the HHSMT telescope. The grey profile indicates the a-priori profile used in the inversion while the horizontal bars the error bars for the retrieved profile.

Summary and conclusions

The primary purpose of this work is the development of a high resolution spectrometer which introduces new technologies that enhance its spectral resolution and dynamic range. The new instrument is particularly convenient for remote sensing of our neighbouring planets, for example Mars, since the highly resolved spectral data can then be used to retrieve information about the planetary atmospheric structure and most importantly about the chemical diversity. Hence, this study was combined with the development of a general circulation model (GCM) for the Martian atmosphere, which will assimilate further atmospheric information retrieved with the upcoming SOFIA (Stratospheric Observatory For Infrared Astronomy) mission.

The instrument's principle is based on the chirp transform, a form of analog Fourier transform using pulse-compression techniques from radar systems, where two dispersive elements (an expander and a compressor) are arranged to perform the transform. The dispersive elements are SAW (Surface Acoustic Wave) filters, which have a chirp signal or a linear frequency modulated signal as its impulse response. These dispersive properties are implemented on the SAW filter as microscopic structures on the surface of a crystal base and consequently these properties are strongly sensitive to micro-scale mechanical deformations and to the expansion of the substrate underneath. The dispersive matching between the expander and the compressor constrains the properties of the transform. Hence, I have studied and developed a new design called the "Adaptive Digital Chirp Processor (ADCP)". The principle behind the ADCP is to incorporate the dispersive properties of the compressor into a digital system that produces the optimum matching expander waveform. The benefits of the arrangement are various, the main being: high spectral resolution as the expander-compressor dispersive matching is optimized and high dynamic range as the out-coming chirp waveform has a large signal-to-noise ratio. The newly developed instrument has an operational bandwidth of 215 MHz, a channel spacing of 28.6 kHz and a spectral resolution (FWHM) of 41.68 kHz, which would mean a spectral resolving power ($\lambda/\Delta\lambda$) higher than 10^8 when analyzing signals with the high frequency band receiver on the GREAT instrument (4.7 THz). The measured dynamic range, which describes the region where the input is linearly related to the output, is also remarkably high: 18 dB considering a ±0.1dB compression point and 30 dB for a ±1dB compression point.

The stability of the instrument is a critical issue and therefore a design parameter since the spectrometer is normally used to detect weak signals through long integration times (the radiometer formula is valid as long as the characteristics of the instruments are constant during the observation time). The spectrometer's stability is quantified in amplitude (or gain) by means of the Allan variance time, and in frequency by analyzing the spectral displacement as a function of the environmental temperature. A compromise between mass, power consumption and thermal isolation was found, obtaining a high spectral stability of 550 Hz/ºK and an Allan variance time of 510 s.

The Martian GCM is developed under the scope of the MAOAM (Martian Atmosphere Observation And Modeling) project, which is based on MART-ACC (Martian Atmosphere - Circulation and Climate Model) and is a fully non-linear, global and three-dimensional hydrodynamic Eulerian gridpoint model, covering the troposphere, mesosphere and

thermosphere up to 135 km. The new model introduces improvements in the numerical computational scheme, a refined topographic map as measured by the MOLA (Mars Orbiter Laser Altimeter) instrument onboard the MGS (Mars Global Surveyor) spacecraft and an enhanced radiative transfer module for the short/long-wave energy balance. These improvements resulted in a higher correspondence between modeling and measurements, since simulation results are in strong agreement with the atmospheric parameters obtained by the MGS-TES (Thermal Emission Spectrometer) instrument. The flux exchange with the surface is described by the Newtonian cooling approximation, where the cooling coefficients were reformulated and newly parameterized. In the case of atmospheric radiation, represented by the cooling to space approximation, the study was based on the broadening of the long-wave CO_2 band at 15 μm and its influence on the atmospheric emission. Furthermore, the atmospheric absorption of the solar radiation, the main heating source, has a proper characterization for the seasonal cycle considering its particular orbital parameters such as the high eccentricity.

Planetary and cometary observations were performed with a running prototype of SOFIA-GREAT-CTS at the Heinrich Hertz Submillimeter Telescope (HHSMT) in Arizona. To coincide with the Mars-Earth closest approach on September 2003, a 6-point map of the CO broadened line at 345 GHz was obtained. From the observations, temperature profiles were retrieved which at present are being used as data assimilation inputs for our GCM. The instrument's high spectral resolution was also used to study cometary bodies with narrow spectral emission features. In November 2003, important outgassing asymmetries were detected on comet Encke/2P, derived from the measured HCN J 3-2 and J 4-3 rotational lines.

As a consequence of the successful results obtained with the new instrument, many new developments are planned. For instance: a CTS for the project WASPAM (Wasserdampf- und Spurengasmessungen in der Atmosphäre mit Mikrowellen) at the Arctic Lidar Observatory For Middle Atmosphere Research (ALOMAR) and a large bandwidth spectrometer (400 MHz) for the VEnus Sounder for Planetary ExploRation (VESPER). VESPER will provide the first comprehensive and synoptic study of Venus with the required sensitivity and duration to test major models of the dynamics, chemistry and circulation throughout the atmosphere of Venus.

Appendix A – High frequency local oscillators

Introduction

In order to pump the detectors into the detection regime, a powerful source is needed with frequency, phase and amplitude properties in order to behave as a reference for the mixing process. There are many ways to produce a coherent reference source; the differences are related to the frequency stability, power output and structural constraints such as weight and volume. In the Terahertz regime, a large limitation is the output power, as it is very difficult from a technological point of view to amplify signals at these high frequencies. Thanks to the use of HEB detectors, the power needed by the coherent reference is significantly smaller than that needed by SIS and Schottky detectors. GREAT actually uses three types of LOs: the BWO used at the low-frequency receiver (1.9 THz), the solid-state LO at the mid-frequency receiver (2.7 THz) and the FIR-laser LO for the high-frequency receiver (4.6 THz).

The Backward Wave Oscillator (BWO)

The backward wave oscillator generates a defined oscillation through the passing of travelling electrons in a wiggled waveguide tube. The electrons are generated by a glowing filament at the heater element. The voltage applied to the heater electrodes should be AC in order to avoid electro-migration (deterioration of the filament) [Rabanus, 2003]. Next to the heater, there is a negative kV source, the cathode. In the other side of the travelling tube, a plate with a potential close to 0 V behaves as anode. Since the potential of the electron source is negative with respect to the anode, the electrons are accelerated between the electron source and the anode. While the electrons travel to the anode, they pass through a wiggled waveguide, producing an alternate electromagnetic field. As the structure of the travelling tube is fixed, the speed of the electrons, will define the frequency of oscillation of the field. At the anode the electrons are then collected.
A PLL (Phase-Locked-Loop) unit stabilizes the potential between the cathode and the anode in order to fine-tune the acceleration of the electrons, the speed and also the output frequency. The kinetic energy of the electrons is dissipated in the anode, resulting in a high temperature of the BWO. Therefore in GREAT a water cooling system is required to dissipate this heat power. Another issue is to keep the electron beam focused and in proximity to the slow wave structure, a strong magnet field (approx. 1 Tesla) is required, achieved in GREAT by the use of a permanent magnet, since the field is intrinsically steady and no extra cooling is required. The BWO oscillator used by GREAT, developed by KOSMA, is located in the SI flange and provides the coherent reference source for the 1.9 THz receiver module.

The Solid-State Oscillator (SSO)

Also known as Gunn oscillators, the solid-state oscillators started to be used in the GHz region, and achieved the THz regime thanks to the use of frequency multipliers. A generated wave in a resonant cavity travels across the cavity and it is reflected at its walls. By the properties of constructive and destructive interference of waves, the length between the resonant walls defines the resonant frequency. The oscillator element in charge of producing the resonant wave is a

Gunn diode biased into the “negative-resistance” regime. This property behaves as a special form of feedback for the oscillator system. In practice, the diode will take time t_d to react to any change in the voltage across it. This time is determined by the physical processes which cause the Gunn Effect. This means that the oscillator can only operate at a maximum frequency of $f_{max} = 0.5/t_d$. On the other hand, the size of the cavity will define the lower frequency; hence it is ensured a single-value oscillation frequency. In fact GaAs Gunn diodes have a maximum achievable frequency of 140 GHz, and therefore in order to obtain the 2.9 THz needed by the MPIfR receiver, several submillimeter frequency multipliers are used.

The FIR ring lasers oscillator

The FIR ring lasers oscillators are based on two lasers in tandem, with the FIR laser excited by the (9-10 μm) CO_2 pump laser [e.g. Güsten et al., 2003]. Their operation is reliable and with the use of appropriate stabilization loops, reasonable output power and high frequency stability is obtained. The limited output power obtained with FIR ring lasers constrains the applications, making them unavailable to be used with Schottky diode mixers, but suitable with hot-electron-bolometer detectors. Also this type of LO is not continuously tuneable and commonly it is difficult to find a sufficiently strong laser line near the observing frequency, limiting its flexibility in RF- and IF-coverage. Nevertheless a convenient and strong laser line is located at 4.6 THz, needed by the high-frequency receiver module developed by DLR. One of the main advantages of this technology is that it has already been tested in airborne missions with an earlier version, developed by MPIfR, on the KAO observatory.

Appendix B - Detectors in the THz regime

Introduction

Most of the submillimeter coherent detectors are based on superconducting mixers, while another possibility could be the use of Schottky diode mixers, which for THz applications employ whisker contacted Schottky diodes in a quasi-optical corner cube mount. The non-linear properties of such solid-state devices make them the first choice in order to achieve the expected coherent detection. But at THz frequencies, they have several disadvantages. To minimize coupling losses to a Schottky diode, its anode diameter has to be much smaller than 1 μm; yielding mechanically unstable contacts as the newest diodes have anode diameters of ¼ μm [Schwaab et al., 1998]. Also the strong attenuation in the transmission line and the conversion losses in the Schottky diode, limits even further the use of this technology.

The two successful coherent detector technologies for mm/submm/IR astronomy are based on superconducting structures. These are the SIS (Superconductor-Isolator-Superconductor) and the HEB (Hot-Electron-Bolometer) detectors. The SIS mixing is based on the principle of photon-assisted tunnelling. The structure of a SIS junction is basically a sandwich of two superconductors with a very thin (10-20Å) insulating barrier in between. Electrons can tunnel across the barrier when the energy provided by the bias voltage V exceeds the superconducting gap energy: $eV > 2\Delta$. Applying a bias voltage next to the superconducting gap energy, a photon can supply the missing energy, if $h\nu + eV > 2\Delta$. In this mode, the junction behaves like a photodiode, providing one electron of tunnelling current per photon absorbed. In theory the SIS mixers can approach the quantum limit, if the photons can be coupled efficiently to the tunnel junction. This requires efficient waveguides probes or planar antennas and tuning circuits in order to compensate for the parallel-plate capacitance of the junction. At higher frequencies, the tuning circuit becomes very difficult to develop, because the RF impedance of the junction capacitance scales inversely with frequency and the losses in the superconductors or metals used for the tuning inductor increase.

In a bolometric detector (HEB), the mixing process relies on the thermal dissipation through a superconducting nanobridge of the energy provided by the signal (RF) and a reference source (LO). The output reflects the beat between the incoming RF-and LO-signals. The basic geometry of a HEB device consists of a thin-film nanobridge of superconductor with thick normal-metal contacts at each end, at a bath temperature of Tb. These contacts provide efficient coupling between the detector and the planar antenna. The mixing is achieved by the slow thermal dissipation of the photons heating the nanobridge. This thermal response will define the output bandwidth, smaller thermal response time τ_{tg} would mean higher output bandwidth. Applying the input signal with the LO signal to the detector it can be expressed as:

$$V(t) = V_{LO}(t).\cos(\omega_{LO}t) + V_{ANT}(t).\cos(\omega_{ANT}t) \quad \text{(B.1)}$$

and as the power dissipated in the film depends on the incoming power and the capability to dissipate this energy, one can obtain as output the V_{IF}, as the following

$$P(t) = V(t)^2 / R_N \rightarrow V_{IF} \propto \sqrt{P_{LO} P_S} \cos\left[\left|\omega_{LO} - \omega_{ANT}\right| t\right] \quad \text{(B.1)}$$

where P is the dissipated power, R_N is the dissipation load, P_{LO} the power of the LO and P_{ANT} the incoming radiation. As the thermal response time is large compared to the incoming switching time of the antenna and LO power, the bolometric detector will dissipate the time averaged value of the incoming power. This means that the detection is performed thanks to the quadratic response of the process and the low-pass filtering of the f_{LO}-f_{ANT}, thanks to the slowness of the dissipation process.

There are actually two ways to dissipate this heat, either by the production of phonons which carry the heat out to the substrate (phonon-cooled HEB mixers) or by the diffusion of hot electrons out of the end of the bridge (diffusion-cooled HEB mixers). In the Phonon-cooled type, the energy in the nanobridge is removed by means of electrophonon collisions and the subsequent escape of nonequilibrium phonons into the substrate [e.g. Hübers et al., 2002]. In this dissipation process, most of the energy is transferred from electrons to the substrate, defining that the thermal dissipation time depends mainly on the electron-photon time (τ_{ep}). The 3dB IF bandwidth for this type of mixer is $f_{3dB} = 1/(2\pi\ \tau_{ep})$. This time depends strongly on the thermal properties of the material and its thickness. With NbN films of about 3nm, a thermal dissipation time of ~30 ps can be achieved, meaning a bandwidth of 5 GHz [Mc Grath et al., 2002]. Another possibility in order to dissipate this energy is by a rapid diffusion of hot electrons out of the ends of the nanobridge into the normal metal contacts. This means that the metal contacts also behave as a heat sink for the dissipation process. In order for this effect to dominate over the phonon-cooling mechanism, the length of the nanobridge should be small: $L < 2(D\ \tau_{ee})^{1/2}$, where τ_{ee} is the electron-to-electron energy exchange and D the diffusion constant.

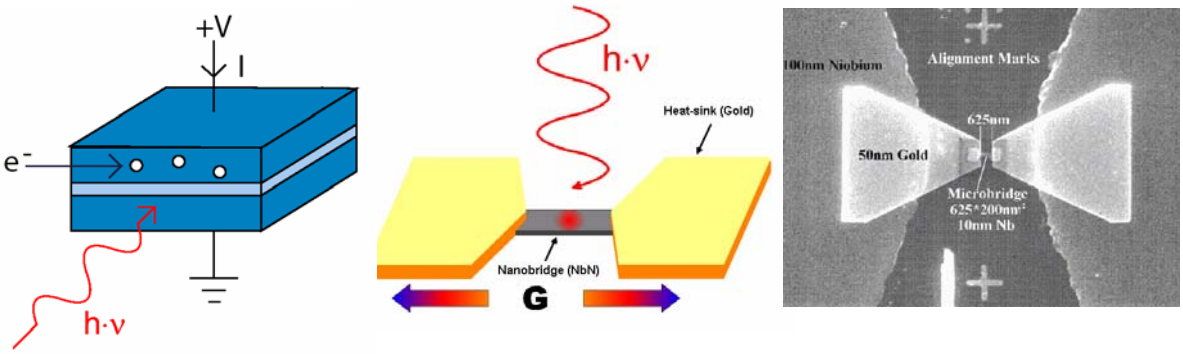

Figure B.1: [Left] A superconducting (SIS) tunnel junction. Junctions are made from a Superconductor/Insulator/Superconductor trilayer. A typical junction area is ~ 1 μm^2 [based on Zmuidzinas, 2003]. [Middle] Diffusion cooled hot electron bolometer structure. In the niobium nanobridge the electron bath is heated above the lattice energy level, cooling it by the diffusion of the heat into the gold contact-pads [image based on KOSMA, 2004]. [Right] SEM (Scanning Electron Microscopy) micrograph image of HEB built in KOSMA's microstructure lab [from Güsten et al. 2003].

Appendix C – The SOFIA observatory

Operation

The length of a flight is limited to 7-9 hours, including landing and take-off. As the telescope cavity is opened when the cruising altitude has been reached, this means an average of 7 hours of usable astronomical time per flight. Most observation flights will start and end at the same airport: NASA Ames Research Center, Moffet Field, south of San Francisco for the northern hemisphere and Christchurch, New Zealand for the Southern Hemisphere. The unvignetted elevation range of the telescope is 20º to 60º (which excludes zenith observations), constrains the observation time and forces a detailed planning for maximizing the scientific information return per target. SOFIA flight tracks will cover mainly the western region of USA, as far as Denver. They may however, cross the Canadian border frequently. The Tropopause comes further down at higher latitudes, allowing longer flights at lower altitudes.

The telescope structure

The telescope structure has to be constructed to support extreme conditions. The narrowness in the aircraft cavity, the disturbances during operation by vibrations and wind, the low air pressure, the large temperature differences between the bulkhead and the cabin and the airworthiness regulations required 10 years of studies in order to define the final telescope concept. By the arrangement of the optics in the open cavity and the science instrument in the pressurized and acclimatized cabin, the telescope has to protrude the aircraft bulkhead and has to handle the pressure differences of about 0.5 bar and temperature differences of about 70 ºK [Krabbe & Röser, 1999]. A pressure window for the science instruments and several pressure-seals at the spherical bearing and at the vibration isolation system prevent pressure leakages. These pressure seals have to support axial forces up to 30 kN, which is about twice the weight of the telescope. A thermal insulation layer covers the cavity side of the bulkhead, which in combination with the Nasmyth tube made of carbon fiber composites, allows handling these large temperature differences in this load carrying structure.

The stability of the pointing is one the most critical conditions to be achieved. This parameter is defined mainly by the connection of the telescope to the aircraft. There are two different types of excitations, an excitation in the base of the telescope by the aircraft vibration and an excitation on top of the telescope by the wind, which is produced by airflow across the open door of the cavity and related aero-acoustic effects. The mechanical system for SOFIA is based on proven technologies already developed for the KAO observatory and it consists of four subsystems:

- *The vibration isolation*: consists of springs and dampers with an additional active component for counteracting the wind torque. The overall subsystem can be imagined as a PID (Proportional, Derivative, Integral) system with the springs behaving as the integrative block in-charge of absorbing the vibration waves coming from the aircraft, while the active counteracting block behaves as the proportional-derivative in charge of correcting the system for unbalanced external vibrations.

- *The rotation isolation*: consists of a spherical bearing. Initially it was intended to be developed as in the KAO telescope with air bearings but the need to handle high loads demanded hydrostatic bearings. The disadvantage of this implementation is the incorporation of the extra seals in order to prevent oil or oil-vapor leakage.

- *The torque motor*: must position the telescope to the required 0.2 arcsec pointing and transport the wind torque from the telescope to the aircraft structure. It's based on a three-phase brushless DC motor.

- *The telescope structure*: includes the mirror support, the telescope frame, the secondary support, and the Nasmyth tube and is made of carbon fiber in order to reduce the weight of the block and to provide longer flight autonomy. The structure is balanced by counterweights on the opposite side of the telescope bearing and the telescope itself is balanced within the aircraft by counterweights in the front of the aircraft.

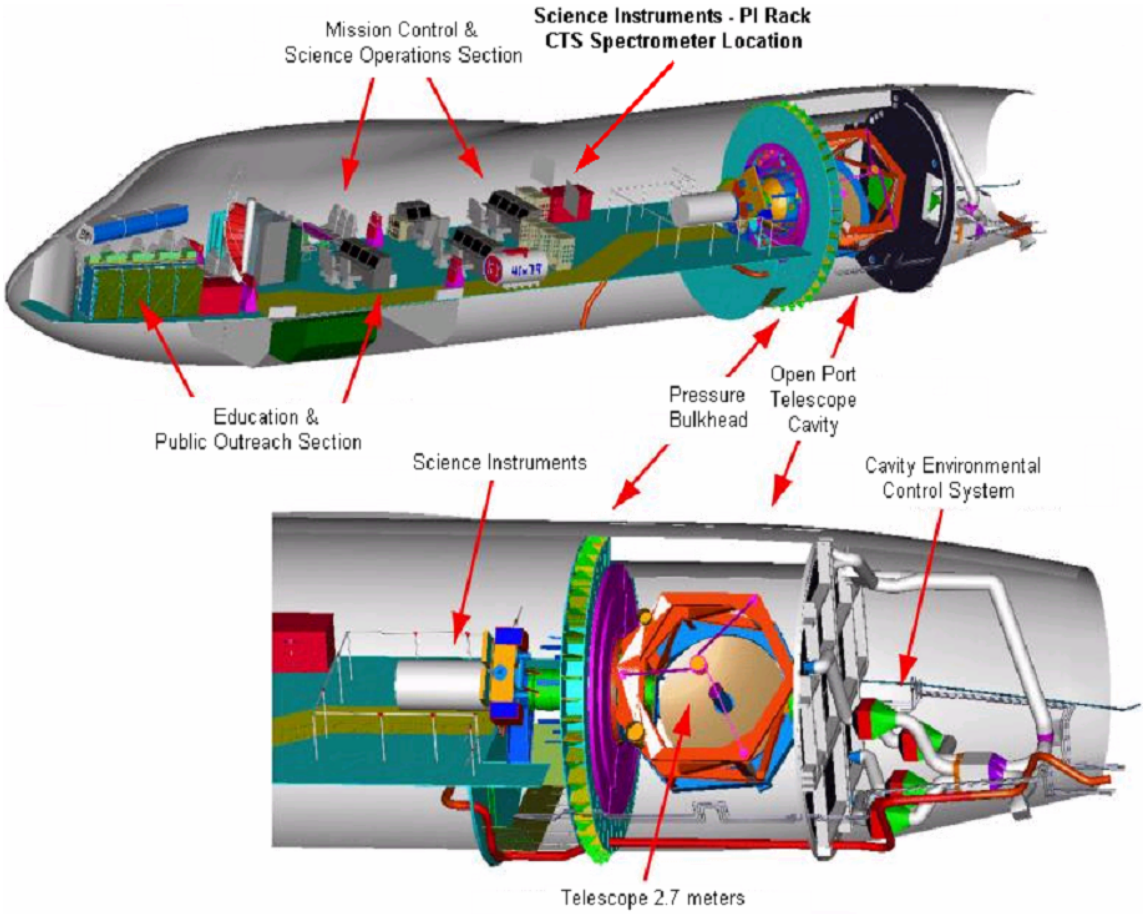

Figure C.1: Inner view of the SOFIA airborne observatory. It can be seen in orange the 2.7 meters main mirror of the telescope. In the head of the Boeing 747, just below the Public Outreach Section, the counterweights are located. They compensate the weight of the telescope structure. [Van der Wal, 2003].

The optical system

The optical system consists of a Cassegrain telescope with an effective aperture of 2.5 m and a so-called Nasmyth focus providing access to the focus from the cabin. The main block of the optical system is the primary mirror, a monolithic 2.705m diameter Zerodur mirror, which is light weighted down to 850 kg by a special double arch shape and by hexagonal holes bored from the back. This yields a light weighting factor of 80%. Zerodur, which belongs to the glass-ceramic composite class of materials, has both an amorphous (vitreous) component and a crystalline component. This substrate material has a near-zero coefficient of thermal expansion and an excellent thermal shock resistance. The mirror is then supported by an axial 18-point whiffle tree support and a 6-point lateral support.

The secondary mirror is also a light weighted structure made from Silicon Carbide with a diameter of 0.35 m. The supporting structure and a moving mechanism provide focus, center, and tilt adjustment as well as chopping capability. In order to compensate the fast movement while chopping, the mirror moments are totally dynamically compensated by reaction masses. A high-performance chopper actuator allows efficient background suppression, with a maximum switching frequency of 20 Hz and chop amplitudes up to ± 5 arcmin. The radiation is reflected on the primary mirror, then in the secondary mirror and finally is pointed into the Nasmyth tube to the science instruments with the tertiary mirror. There are in fact two tertiary mirrors, a dichroic IR-mirror with a 10-15 nm thickness gold coating that reflects the infrared radiation to the SI (Science Instruments)-flange, while passing more than 50% of the visible radiation into the next tertiary mirror (a visible radiation mirror). The material Fused Silica is used for the dichroic IR-mirror because of its good thermal behavior, low specific mass and excellent transmission.

In order to achieve the high pointing stability of 0.2 arcsec rms, not only a fast reacting and vibrations-isolated telescope structure is needed but also a good tracker subsystem. The SOFIA telescope uses three cameras (WFI, FFI, FPI) and a tracker control processor as a tracker subsystem. The reflected signal obtained from the optical tertiary mirror is sent to the visible Nasmyth focus, which is then fed into the focal plane imager (FPI). The other two cameras, the WFI (wide field imager) and the FFI (far field imager), are mounted directly on the telescope structure in the cavity and attached to the main telescope. All these cameras are fed through different special sets of optics, described in table C.1. The primary mirror assembly pointing error contribution is approx. 0.1 arcsec rms, 0.05 arcsec rms for the secondary mirror assembly, the tertiary mirror assembly, as well as for the focal plane imager, while the Nasmyth tube, the metering structures and the vibration and rotation isolation systems contribute in approx. 0.15 arcsec rms. The square root sum of all errors can be estimated to be $(0.1^2 + 0.05^2 + 0.15^2)^{½} \approx 0.2$ arcsec rms which meets the SOFIA requirements.

The principle behind the pointing system design is based on the bearing of the telescope in its center of mass, for insulation from aircraft rotational excursions. The overall stabilization of the three degrees of freedom of the telescope is controlled with a feed-back system. The angular position of the telescope is sensed by a gyro and the demanded torques to the telescope are applied by the torque drive. The system is managed by the position controller. As was discussed before, the *vibration isolation* assembly suppresses the external impartation of aircraft vibrations and translational excursions.

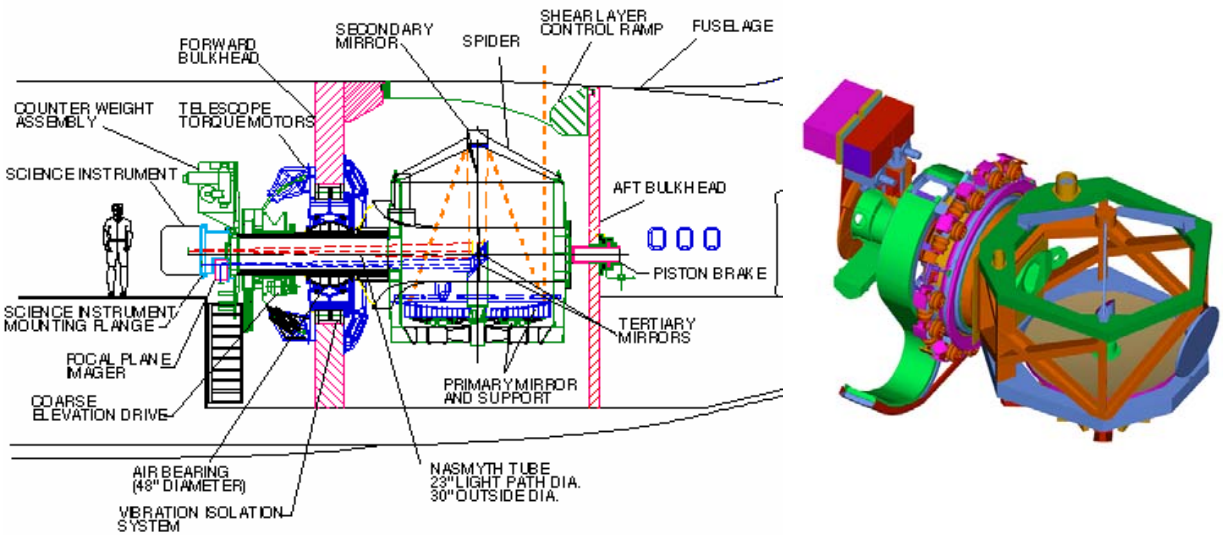

Figure C.2: [Left] The SOFIA telescope is designed based on a Cassegrain system with two Nasmyth foci, the nominal IR focus (IR light path shown in orange) and the additional visible light focus (visible light path shown in blue). The upper tertiary mirror, the IR-dichroic mirror, reflects the incoming IR radiation into the IR Nasmyth tube, while allowing 50% the visible path to pass through into the visible tertiary mirror. [Right] View of the SOFIA optical structure. The primary and secondary mirrors are located in the cavity and mounted into a carbon fiber support structure. The telescope as shown here weights about 16 metric tons [pictures SOFIA-DLR, 2004].

Optical system summary describing the parameters of the SOFIA telescope			**Optical data of the visual imaging and guiding cameras**	
Parameter	Value		Parameter	Value
Entrance pupil diameter	2500 mm		CCD Chip size	1024x1024 px
Nominal focal length	49141 mm		Pixel size	14 µm x 14 µm
Entrance pupil position	22514 mm			
Exit pupil distance	6849 mm		FFI optical system	Schimdt-Cassegrain
Image height	± 57.2 mm		primary diameter	254 mm
Back focal length	6849 mm		night sensitivity	13 mag (after 2.1sec)
Unvignetted field of view	± 4 arcsec		point stability	0.35 arcsec
Aperture stop location	Secondary			
Aperture stop diameter	352 mm		WFI optical system	Petzval
			font lens diameter	70 mm
Primary mirror focal length	3200 mm		night sensitivity	8 mag (after 0.5 sec)
free optical diameter	2690 mm		pointing stability	1.8 arcsec
conic constant	-1			
center hole diameter	420 mm		FPI optical system	Eyepiece + CONTAX
material	Zerodur			1.4/85 mm lens
mass	~850 Kg		night sensitivity	16 mag (after 1.7 sec)
			pointing stability	0.035 arcsec (nominal)
Secondary mirror focal length	477 mm			
free optical diameter	350 mm			
conic constant	-1.298			
material	C / SiC			

Table C.1: [Left] Description of the overall optical system with information from the sub-blocks involved in the Cassegrain telescope disposition [Krabbe & Röser, 1999]. [Right] The visible radiation coming from the visible-tertiary mirror is then used by the FPI focal plane imager as part of a guiding system. Also it's described the optical performance of the cameras that are directly attached to the ring of the telescope, the WFI and the FFI camera, also involved in the guiding system.

Pointing accuracy

In order to quantify the pointing accuracy a numerical study is required, which would combine the previously described telescope's structure and a precise knowledge of the applied forces. For that, Schönhoff [et al., 1998] performed simulations considering excursions and vibrations applied to the telescope. The integrated mathematical model included the optics, the structural dynamics and the control loops. The value that describes the pointing is the ray-trace of the telescope. Deformations, vibrations, excursions etc, in each of the three involved mirrors (primary, secondary and tertiary) will deform or deviate the ray-trace leading to pointing errors. All the blocks involved in the pointing system are modeled based on frequency responses of the linear simulation model and power spectral densities of the disturbances. The main block of the model, the telescope structure, is represented by a modal reduced FEM (finite element analysis) with 100 degrees of freedom. The simulations also describe the improvements in the image stability achieved thanks to the use of an IMC (Image Motion Compensation) system. This system is planned to be integrated to SOFIA in the first years of the mission. The results are promising, the combination of the rigid body motion and the flexible deformation represent no more than 1 cumulative arcsec rms, at aerodynamic disturbances with frequencies lower than 100 Hz. In the simulations it can also be seen the importance of the IMC system, reducing the cumulative rms by more than 30% for most of the aerodynamic disturbances.

Spatial resolution

The spatial resolution depends mainly on the properties of the primary mirror, its effective diameter and its diffraction limit. The primary optimal mirror is 2.69 m, while the effective diameter or entrance pupil diameter is 2.5 meters. The FWHM (Full Width Half Maximum), of the telescope beam, can be written as

$$\Theta = 1.22\frac{\lambda}{D} \tag{C.1}$$

Formula (C.1) is considering a Gaussian beam pattern. Also this is considering that all the radiation coming to the primary mirror is reflected and focalized to the secondary mirror. This consideration is valid as long as the incoming radiation is below the diffraction limit. This limit defines the point when the primary mirror is non-longer behaving as a perfect mirror and for SOFIA this value is about 10 μm. Below 5 μm, the viewing is dominated by wind turbulences within and in the outer vicinity of the cavity. The spatial resolution achieved between 30 μm and 300 μm will be unprecedented and the highest achieved so far. It will be more than 4 and 3 times higher than of ISO (Infrared Space Observatory) and SIRTF (Space Infrared Telescope Facility), respectively.

Atmospheric transmission

The atmospheric transmission versus wavelength is especially good at 14 km altitude, allowing the detection of extraterrestrial radiation from a vast region of the spectrum. The SOFIA telescope will collect radiation from the visible (0.2 – 0.7 μm) with transmission around 0.9, the photographic infrared (0.7 – 1.1 μm) with transmission around 0.9, the near infrared (1.1 – 4 μm) with transmission around 0.7, the mid infrared (4- 40 μm) with transmission around 0.9, the far infrared (40 – 300 μm) with transmission around 0.7, and the submillimeter range (300 – 1000 μm) with transmission around 0.9.

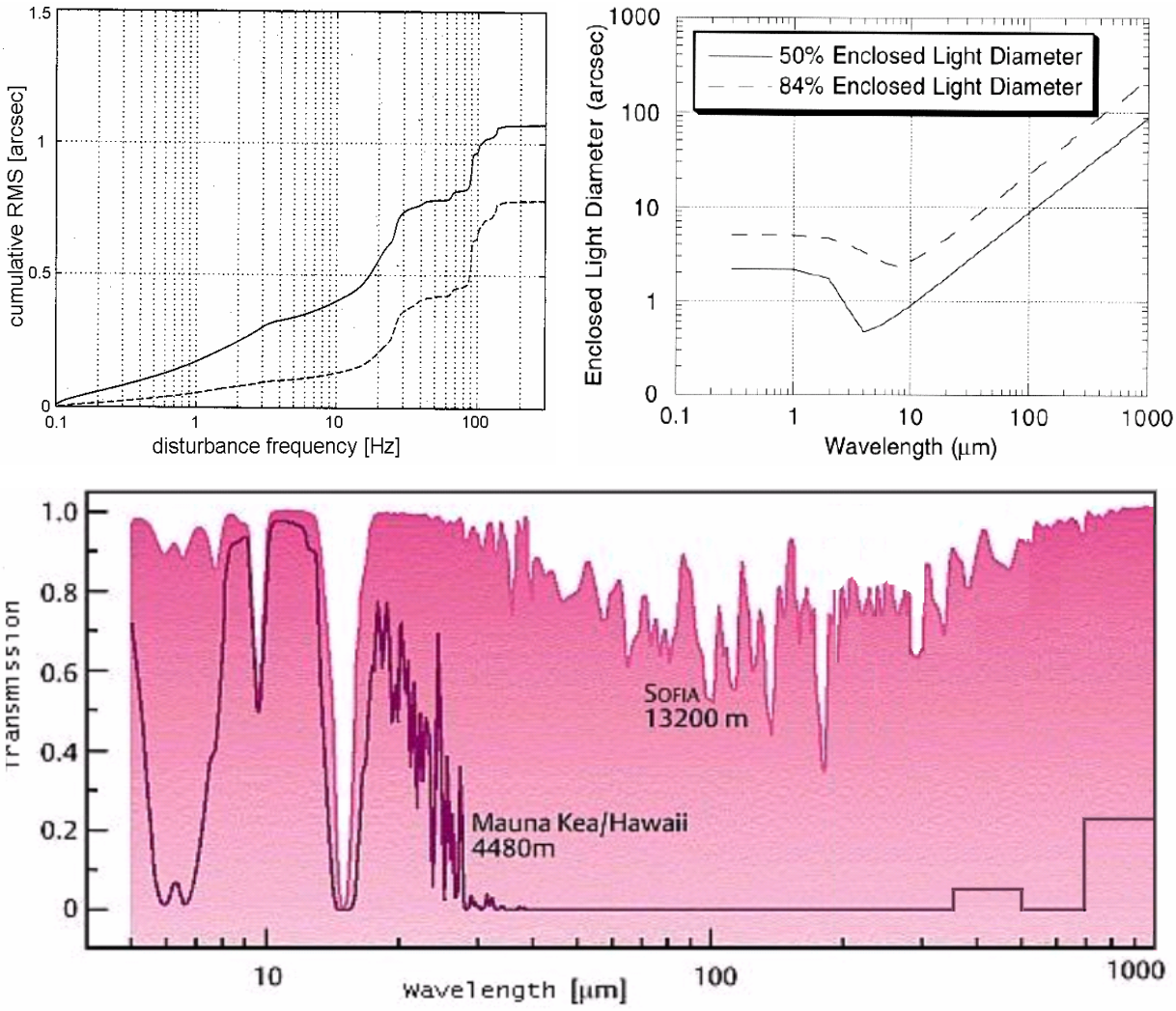

Figure C.3: [Top-Left] Pointing stability results for specified aerodynamic disturbances [Schönhoff et al., 1998]. [Top-Right] Spatial resolution versus wavelength, for FWHM (50% enclosed light diameter) and 84% enclosed light diameter [Krabbe & Röser, 1998]. [Bottom] Atmospheric Transmission versus Wavelength for SOFIA flying at an altitude of 13.2 km in comparison with one of the most important ground based telescope, Mauna Kea, at an altitude of 4.5 km [based on Churchwell et al., 2000 and Krabbe & Röser, 1999].

Appendix D – Digital sampling theory

The sampling function can be represented by a Fourier series since is periodic in time T, having a sampling frequency $f_s=1/T$. This can be written as:

$$p(t) = \sum_{n=-\infty}^{\infty} C_n \cdot e^{j2\pi f_s t \cdot n} \tag{D.1}$$

Where C_n is the n^{th} Fourier coefficient of p(t) and it is described as:

$$C_n = \frac{1}{T} \int_{-T/2}^{T/2} p(t) \cdot e^{-j2\pi f_s t \cdot n} dt \tag{D.2}$$

The sampled signal can be defined as the multiplication of the sampling signal p(t) and the continuous signal x(t) → $x_s(t) = x(t) \cdot p(t)$

$$x_s(t) = \sum_{n=-\infty}^{\infty} C_n \cdot x(t) \cdot e^{j2\pi f_s t \cdot n} \tag{D.3}$$

Obtaining the Fourier transform of the sampled signal will lead to:

$$X_s(f) = \int_{-\infty}^{\infty} \sum_{n=-\infty}^{\infty} C_n \cdot x(t) \cdot e^{j2\pi f_s t \cdot n} \cdot e^{-j2\pi f t} dt \tag{D.4}$$

Interchanging the integral and the sum operators and applying the frequency displacement theorem will lead to:

$$X_s(f) = \sum_{n=-\infty}^{\infty} C_n \int_{-\infty}^{\infty} x(t) \cdot e^{-j2\pi t (f - n f_s)} dt \tag{D.5}$$

$$X_s(f) = \sum_{n=-\infty}^{\infty} C_n \cdot X(f - n f_s) \tag{D.6}$$

Equation (D.6) states that the sampled signal's spectrum is periodic in the frequency domain and that this periodicity is defined by the sampling signal, represented by f_s and C_n. The output spectrum of a sampled spectrum can be analyzed by deriving the impulse response of the sampling process. For that, x(t) is considered to be an impulse δ(t) leading to a Fourier transform (F) of the form:

$$X_s(f) = F\{p(t) \cdot x(t)\} \tag{D.7}$$

$$H(f) = \int_{-\infty}^{\infty} p(t) \cdot \delta(t) \cdot e^{-j2\pi f t} dt \tag{D.8}$$

where H(f) is the transfer function of the digital generation process. If an impulsive signal is multiplied with the sampling signal, this would mean that only the first element of the sampling process provides information, deriving to the following

$$H(f) = \frac{1}{T}\int_{-T/2}^{T/2} 1 \cdot e^{-j2\pi f_s t}\,dt \tag{D.9}$$

$$\left|H(f)\right| = \left|Sinc\left(\pi(f / f_s)\right)\right| \tag{D.10}$$

Equation (D.10) describes the transfer function in the spectral domain for the sampling process.

Appendix E – Filter synthesis

The filtering is induced by a proper combination of reactive properties of physical devices, which can be lumped elements, microwave structures, etc. These reactive elements have a representation in the complex Laplace domain, while its disposition and number of elements would define the characteristics and performance of the arrangement. The response of this arrangement is studied by its impulsive time response h(t), since the convolution of the incoming signal a(t) with h(t) would derive to obtain the resulting output b(t)=a(t)*b(t). In the Laplace domain this is simplified by the multiplication of the Laplace transformation of a(t) and h(t) that would derive to obtain B(s)=A(s)·H(s). The structure of H(s) can be seen as a division of s-polynomials, which can be written as:

$$H(s) = \frac{\sum_{i=0}^{nz} s^i a_i}{\sum_{j=0}^{np} s^j b_j} \qquad \text{(E.2)} \qquad\qquad H(s) = \frac{\prod_{i=1}^{nz} (s - z_i)}{\prod_{j=1}^{np} (s - p_j)} \qquad \text{(E.1)}$$

where np is the s-order of the numerator or the number of "zeros", nz is the s-order of the denominator or the number "poles", a and b the coefficients of the polynomials, z the "zeros" and p the "poles" of the filter transfer equation. In a reactive passive element, the performance is mainly driven by its poles distribution, while the different pole distribution or polynomial structure will define the spectral filtering performance for the scheme. The main two parameters which define the performance of the filter and thus the polynomial structure are the group-delay and the spectral selectivity. For example, a Bessel filter whose transfer function is derived from a Bessel Polynomial provides maximal flat time delay but has poor selectivity. On the other hand an elliptical filter whose transfer function is derived with the use of Jacobean elliptic functions provides low phase linear performance and induces ripple in both passband and stopband; nevertheless it offers the most spectral selectivity per pole. Elliptical filters would mean a sacrifice in the phase properties of the out-coming signal but will also represent stronger removal of the aliases signals. The nearer the digital signal is generated from the Nyquist limit, the nearer and stronger the aliases components are. A higher spectral selectivity of the low-pass filter would also mean the capability to produce higher bandwidth signals. A 6th order elliptical filter allows up to 50 dB attenuation of the 2nd Nyquist region to be obtained, which is much higher than the 5 dB attenuation provided by a Bessel family filter. The frequency transfer equation for an elliptical filter can be described based on equation (E.2) as:

$$\left|H(j\omega)\right| = \frac{1}{1 + \varepsilon^2 U_n\left(\frac{\omega}{\omega_c}\right)} \qquad \text{(E.3)}$$

where ε is the ripple factor and $U_n(x)$ is the Jacobian elliptic function. This function does not exist in "closed" form, and so is tabulated. In the case of the RF filtering processes, it is considered Chebyshev filters since they provide improved selectivity (compared to Bessel, Butterworth ...) but they show better delay and phase linear performance compared to elliptical filters. Its transfer equation is written as following:

$$|H(j\omega)| = \frac{1}{1+\varepsilon^2 T_n\left(\frac{\omega}{\omega_c}\right)} \quad \text{(E.4)} \qquad T_n\left(\frac{\omega}{\omega_c}\right) = \cosh\left(N\cosh^{-1}\left(\frac{\omega}{\omega_c}\right)\right) \quad \text{(E.5)}$$

where $T_n(x)$ is the Chebyshev polynomial, ω_c the cut-off frequency, N the order of the polynomial, and n and integer between 1 and N. The information of the physical reactive elements is retrieved from the poles (s_n) and can be written as:

$$\alpha_n = \pm\frac{1}{N}\sinh^{-1}\frac{1}{\varepsilon} \quad \text{(E.6)}$$

$$s_n = \omega_c\left(\sinh\alpha_n \sin\gamma_n + j\cosh\alpha_n\cos\gamma_n\right) \quad \text{(E.7)}$$

$$\gamma_n = \frac{(2n-1)\pi}{2N} \quad \text{(E.8)}$$

The design parameters which characterize the pole distribution of the spectral filters, meaning cut-off frequency (ω_c), bandpass ripple (ε) and order of the polynomial (N), are defined according to the spectral filtering requirements. These requirements are represented by the spectral location and expected attenuation of the stopband and passband regions, as shown in figure E.1.

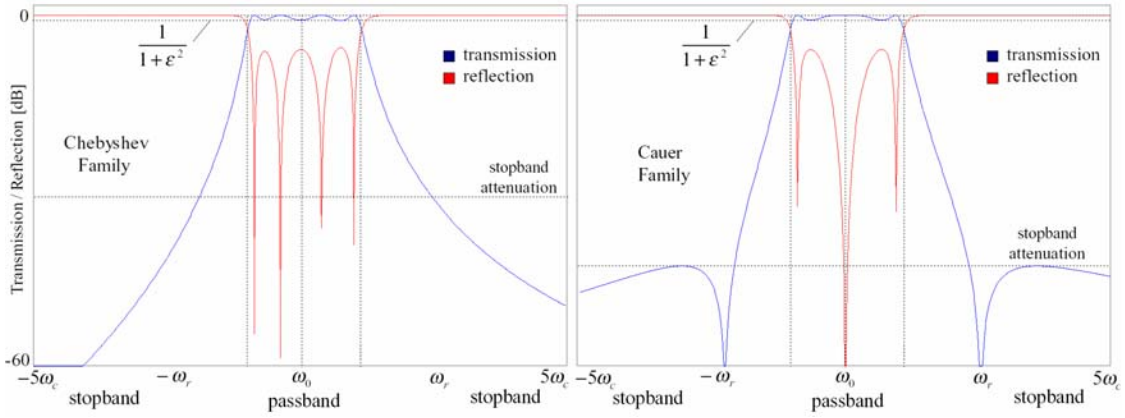

Figure E.1: Ideal response for a Chebyshev (left) and a Cauer (right) 4th order family filters. The transmission is represented by |H(jw)| while the reflection by 1-|H(jw)|, both in logarithmic scales.

Appendix F – Resonating waveguides for spectral filtering

Filters at GHz can be realised by a proper control of the propagation path of the electromagnetic waves in a dielectric medium. Structures in such a medium will affect the properties of a travelling electromagnetic wave and thus they will produce constructive and destructive modes which will derive to a form of spectral filtering. In order to constrain the wave trajectory, the propagating path should always have a ground plane separated by a dielectric medium (having ε_r as dielectric constant). This dielectric medium could be embedding the propagating path, as is the case of stripline, or only below the path being the case of microstrip. The propagating conditions, even though both mediums share the same dielectric constant, are extremely different. For stripline the propagation mode is TEM since the conducting trace is surrounded by similar dielectric material. However, for microstrip line the propagation mode is a combination of TM and TE modes. This is due to the fact that the upper dielectric of a microstrip line is usually air while only the bottom is the dielectric medium. The electrical properties of the path will vary according to the location of the ground plane, the dielectric medium and the path structure. This is the case of the characteristic impedance of the line, which varies with the dielectric thickness (approximately in an inverse square relationship) and inversely with the line width. The realisation of filters in microstrip or stripline medium can be addressed in two manners: by a stepped impedance scheme or by a proper arrangement of coupled lines. The stepped impedance scheme is based on transforming each lumped element into a transmission line of a defined length and width. The width as it is previously described will modify the characteristic impedance of the line, and thus its electrical representation, while the length describes the properties of a shorted or open shunt line. However, in the development of bandpass filters the setup of coupled lines is of greater convenience, as each waveguide structure represents a group of lumped elements and therefore it will lead to simpler arrangements. The realisation is achieved by transforming each pole of the filtering function into a resonant coupled line, and since it can be considered as an independent electrical entity, it is described by its impedance or ABCD matrix, meaning that each pole has a coupled line representation. The input impedance of a coupled line as described in figure F.1 can be written as following:

$$Z_{in} = \frac{\sqrt{(Z_{0e} - Z_{0o})^2 + (Z_{0e} + Z_{0o})^2 \cos^2(\beta l)}}{2\sin(\beta l)} \qquad \text{(F.1)}$$

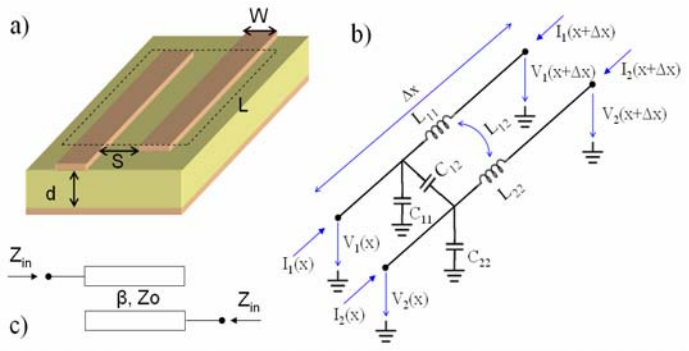

Figure F.1: Representation of a coupled-line in a microstrip medium. [a] Sketch of a coupled line of separation "s", line-width "w", material thickness "d" and coupled length "l". [b] Electrical representation of the coupled line, where C_{12} and L_{12} describe the coupling coefficients. [c] Waveguide representation of the coupled-line, with electrical length βl and characteristic impedance Zo.

where Z_{0o} is the odd mode impedance (the impedance seen when testing the impedance of one side of a pair of lines when the other is driven in equal and opposite polarity), Z_{0e} the even mode impedance (the impedance measured testing one of a pair of lines which are driven by identical signals), and βl the electrical length of the line. The properties of the odd impedance depend on the coupling elements shown in figure F.1 c) (C_{12} and L_{12}), while the even impedance properties depend on the non-coupled elements of each line, meaning C_{11}, L_{11}, C_{22}, L_{22}.

The calculation procedure consists of retrieving the prototype coefficients from the poles, which represent the normalized conductances of the filter, then the calculation of the network parameters for each pole, meaning Z_{0o}, Z_{0e} and the coupling coefficients and finally the transformation of these parameters into real physical units of width, separation and lengths of the coupled resonating transmission lines. The prototype coefficients g_n depend on each filter family and for a Chebyshev polynomial function they can be written as:

$$a_n = \sin(\gamma_n) \qquad \text{(F.1)}$$

$$\xi = \sinh\left(\ln\left(\frac{\sqrt{1+\varepsilon^2}+1}{\sqrt{1+\varepsilon^2}-1}\right)\right) \qquad \text{(F.3)}$$

$$b_n = \xi^2 + \sin^2\left(\frac{k\pi}{N}\right) \qquad \text{(F.5)}$$

$$g_1 = \frac{2a_1}{\xi} \qquad \text{(F.2)}$$

$$g_n = \frac{4a_{n-1}a_n}{b_{n-1}g_{n-1}} \qquad \text{(F.4)}$$

The transformation from normalized conductances into network values will mostly depend on the topology of the filter. Since a coupled line of λ/2 can be modelled by two transmission lines of λ/4 and an admittance inverter (J) in between, the odd and even impedances can be expressed as: (see demonstration on [Makimoto & Yamashita, 2000])

$$Z_{0e}^n = Zo\left(1 + J_n Zo + \left(J_n Zo\right)^2\right) \qquad \text{(F.6)}$$

$$Z_{0o}^n = Zo\left(1 - J_n Zo + \left(J_n Zo\right)^2\right) \qquad \text{(F.7)}$$

where J_n are the admittance inverter coefficients and Z_o is the characteristic impedance of each transmission line conforming the coupled line. It can be derived from the circuit theory for a ladder circuit (represented in the normalized conductances g_n), that J_nZ_o can be expressed as following:

$$Z_0 J_n = \sqrt{\frac{\pi b}{2g_n}} \quad \text{for n=1} \qquad Z_0 J_n = \frac{\pi b}{2\sqrt{g_{n-1}g_n}} \quad \text{n=2..N-1} \qquad Z_0 J_n = \sqrt{\frac{\pi b}{2g_{n-1}g_n}} \quad \text{n=N} \qquad \text{(F.8)}$$

where b is the fractional bandwidth (b = (f2-f1) / $(f2 \cdot f1)^{½}$) of the desired filter. From the above equations, the odd and even impedances can be derived, as shown on equations (F.6) and (F.7). Once the electrical parameters are defined, they are converted into coupled line parameters. This can be written as, assuming a stripline medium (see [Mongia et. al, 1999]):

$$s = d\left[1 - \left(\frac{Z_{0e}}{Z_{0o}}\right)^2 + \sqrt{\left(\left(\frac{Z_{0e}}{Z_{0o}}\right)^2 - 1\right) + 1}\right] \qquad \text{(F.9)}$$

where d is the material thickness and s the separation between lines. The width of the coupled lines is also dependent on the obtained odd and even impedances, and it can be expressed as following:

$$w = d\left[\frac{\left(1-(s/d)^2\right)}{4\left(Z_{0e}/Z_{0o}\right)\sqrt{\varepsilon_r}}\right] \tag{F.10}$$

Appendix G – The Martian physical and chemical structure

The main chemical constituent of the Martian atmosphere is CO_2, as measured by the Mariner probes. Furthermore, a precise characterization of the Martian atmosphere chemical composition was first established by the Viking landers, which measured a chemical mixing ratio of 95.3% for carbon dioxide, 2.7% for nitrogen, and 1.6% for argon, with small quantities of oxygen (0.15%) and water vapor (0.03%) [Nier et al., 1976]. Despite the paucity of water vapor, the lower atmosphere of Mars is saturated with water vapor due to the low pressure. Hence fogs and water ice clouds are relatively common on the Martian surface. Nevertheless, latent heat release due the condensation of water vapour is negligible when compared to radiative effects due to the low air column. Water vapor is also an important regulator of the ozone composition, since the photolytic hydrogen component from the broken down water molecule catalytically destroy ozone, limiting ozone to the winter high latitudes where the cold limits the abundance of water vapor in the atmosphere [Barth et al., 1992]. On Earth, because of solar heating due to the ozone layer the temperature increases above 18 km forming the stratosphere, as seen on figure G.1. This region is of high static stability, and therefore a chemical container for the terrestrial biosphere. This stability is produced since a rising air parcel would cool (due to adiabatically expansion), making it heavier than its lighter environment and so it will sink back to its previous stable state following a Brunt-Väisälä oscillation. On Mars, besides its low composition of ozone, airborne dust absorbs incoming solar radiation and thermal emission from the surface, developing high static stability regions. However, the Martian atmosphere results in only a weak greenhouse effect that raises the surface temperature only by about 7°C, compared to 35ºC as on Earth. Consequently, most of Mars is well below the freezing point of water for most of the year. Even considering exceptional atmospheric warming in summer latitudes or hypothetical volcanic activity, the water ice would turn directly into water vapor due to the low atmospheric pressure present on Mars.

The almost non-present condensation/sublimation of water, which is normally associated to non-adiabatic processes, since heat is released from the parcel with the water-drop, derives to classify the Martian atmospheric lapse rate as a dry adiabatic lapse rate (DALR) described by $\Gamma_a = g/C_p \approx$ 4.3 K/km. While in the middle atmosphere of Mars, the temperature declines more gradually with height and can approach the limiting condensation temperature of carbon dioxide condensation (130-150 ºK) [Leovy, 2001]. This is consistent with measurements of the Martian atmospheric temperature profile obtained on the entry of the Viking landers and observations performed from the Mariner 9 orbiter, as it is shown on Figure G.1.

The upper region of the Martian atmosphere, normally described as the thermosphere because of its high temperatures, is a highly ionized and low pressure environment. The lower boundary is usually cold, only ~130 ºK, but it increases dramatically with altitude and at 150 km it is already 280 ºK (as derived from measurements of Mars Pathfinder during entry [Schofield, 1997]) and still rising. However, there is no feeling of heat because these high temperatures are embedded on a non-LTE (Local Thermodynamic Equilibrium) environment, meaning that because of low pressures, collisions between molecules are too infrequent to thermalize the absorbed photons before they are pre-radiated, decreasing local heating. In such conditions, molecular heat

conduction is an important effect of atmospheric cooling and is mandatory on the modeling of the Martian atmosphere above 100 km. Consequently, this effect together with dynamic viscosity and ion-drag, characteristic of a low pressure, highly thermalized and ionized environment, are included on the modeling of the upper atmospheric region with the MAOAM model.

An important contribution to the overall atmospheric circulation is driven by the Martian orography and the soil's properties, such as the thermal inertia, superficial dust particles structure and the surface albedo. Diurnal cycles on the surface temperature are strongly dependent on the thermal and physical properties of the top several centimetres of the "soil" and thermal inertia is the key property in controlling these temperature oscillations. It is defined as a combination of thermal conductivity k, density ρ, and specific heat capacity C_p:

$$I = \sqrt{C_p k \rho} \tag{G.1}$$

In the case of Mars, it can be considered as a mixture of coarse and fine sand as described by [Mellon, 2001], leading to a heat capacity of 800 [$JKg^{-1}K^{-1}$] (assuming a basaltic mineral composition), to an average density of 1700 [Kgm^{-3}] and a thermal conductivity of 0.05 [$Wm^{-1}K^{-1}$] (assuming Martian atmospheric pressures in the interstice of the porous materials), and therefore a representative thermal inertia of I≈260 [$Jm^{-2}K^{-1}s^{-½}$]. This is consistent with the newly highly spatial resolved global characterization of the Thermal inertia obtained by the TES mapping mission night-time temperature data [Mellon et al., 2000]. The surface albedo is also an important parameter which resembles several characteristics of the surface, such as the superficial dust particles structure and soil's composition.

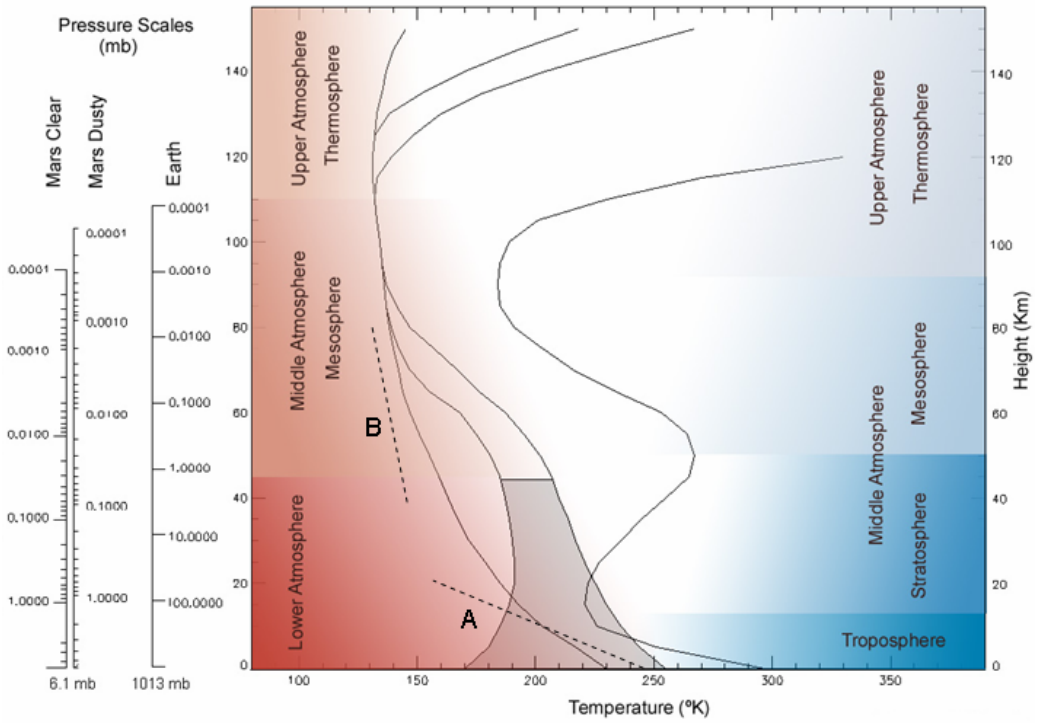

Figure G.1: The standard atmosphere for Earth and representative temperature profiles for Mars during clear conditions (solid line) and dusty conditions for day and night (filled lines). Based on Viking entry profiles, Mariner 9 and Viking radio-occulation measurements and Mariner 6 and 7 UVS data for the upper Martian atmosphere. The A and B dotted lines represent the dry adiabatic lapse rate and CO_2 condensation respectively. Temperature profiles based on Zurek, 1992.

The albedo is the ratio between the received radiation and the emitted (1: total reflection, 0: total absorption). The first global retrieval of the Martian albedo was performed by the Viking IRTM (Infrared Thermal Mapper) instrument, which measured an average albedo of ~0.2 [Kieffer et al., 1977; Martin, 1981]. These measurements were contrasted with the recent observations from the MGS (Mars Global Surveyor) orbiter, through the onboard instruments TES (Thermal Emission Spectrometer) and MOLA (Mars Orbiter Laser Altimeter), obtaining comparative results. However, the main achievement from the MOLA instrument was the first precise characterization of the Martian topography with an absolute vertical accuracy less than 10 m and an unprecedented spatial resolution [Smith et al., 1999]. The full range of topography on Mars is about 30 km, which is comparable to the terrestrial of ~20 km. However, the presence of oceans on Earth derives a topography range of 9 km, meaning that on Mars the orographic features are more than three times stronger. Due to this larger variability in surface altitudes, the effect of topography on the general circulation is expected to be more important on Mars than on Earth. On the other hand, the thermal effects of ocean-continent differences, which produce large zonally asymmetric circulation systems on Earth, are absent on Mars. The most impressive topographic features are the Olympus Mons, which it is the greatest volcano of the solar system with 27 km altitude (above the gravity surface of Mars) and a 600 km diameter base and the massive Hellas impact basin in the southern hemisphere which is 4 km below the surface level and has an impressive diameter of 2100 km.

Parameters governing the general circulation on Earth and Mars	**Earth**	**Mars**
Mean distance to the Sun [AU]	1.0	1.52
Orbital eccentricity	0.0167	0.093
Inclination of rotation axis to ecliptic	23.5º	25º
Rotation rate Ω [$10^{-4}s^{-1}$]	0.729	0.709
Planetary radius [km]	6380	3390
Surface gravity [ms^{-2}]	9.81	3.72
Surface density [km^{-3}]	1.2	0.02
Representative scale height RT/g [km]	7.6	10.3
Dry adiabatic lapse rate Γa = g/Cp [Kkm^{-1}]	9.8	4.3
Specific heat constant at constant pressure Cp [$m^2s^{-2}K^{-1}$]	1000	860
Representative thermal inertia [$Jm^{-2}K^{-1}s^{-½}$].	1500	260
Representative surface's bolometric albedo	0.29	0.25
Typical zonal wind at jet level U [ms^{-1}]	30	80
Planetary Rossby number, Ro	0.05	0.2
Full range topographic variation from oceans level [km]	9	30

Table G.1: Comparison between Mars and Earth for the parameters governing the general circulation. Special attention should be placed on the high orbital eccentricity, the high planetary Rossby number, the low representative thermal inertia and the full range topographic variation present on Mars. The values are based on [Leovy, 2001], [Smith et al., 1999], [Williams, 2004] and [Mellon, 2001].

Bibliography

Aellig C. P., N. Kämpfer, C. Rudin, R. M. Bevilacqua, W. Degenhardt, P. Hartogh, C. Jarchow, K. Künzi, J. J. Olivero, C. Croskey, J. W. Waters, and H. A. Michelsen, Latitudinal distribution of upper stratospheric ClO as derived from space borne microwave spectroscopy, Geophys. Res. Lett., 23, 2321–2324, 1996. A'Hearn, M.F., Millis, R. L., Sopka, R., Periodic Comet Encke, IAU Circ., No. 3091, 1, 1977.

Barth, C.A., Stewart A.I.F., Bougher S.W., Hunten D.M., Bauer S.J., NAGY A.F., Aeronomy of the current Martian atmosphere, Mars, Chapter Nº 30, Univ. Arizona press, ISBN 0-8165-1257-4, 1992.

Barnes, J. R., Time spectral analysis of mid latitude disturbances in the Martian atmosphere, Journal of the Atmospheric Sciences, vol. 37, p. 2002-2015, 1980.

Berger, U., Numerische Simulation klimatologischer Prozesse und thermische Gezeiten in der mittleren Atmosphäre, Dissertation, Universität zu Köln, 1994.

Bjoraker, G.L., Mumma, M.J., Larson, H.P., Isotopic Abundance Ratios for Hydrogen and Oxygen in the Martian Atmosphere, *B.A.A.S.*, 21, 991, 1989.

Bockelee-Morvan, D.; Crovisier, J.; Baudry, A.; Despois, D.; Perault, M.; Irvine, W. M.; Schloerb, F. P.; Swade, D., Hydrogen cyanide in comets - Excitation conditions and radio observations of comet IRAS-Araki-Alcock 1983d, Astronomy and Astrophysics, vol. 141, no. 2, p. 411-418, 1984.

Bockelée-Morvan D., Crovisier J., Les molecules des cométes, La Recherche, 271, 1272-1278, 1994.

Bratt P., Engeler W., Levinstein H., Mac Rae A., Pehek J., A status report on infra-red detectors, Infrared Physics, vol 1, pp. 27-38, 1961.

Cheng, Jingquan, Forced Air Cooling at High altitude, ALMA-NRAO Memo Nº203, 2002.

Churchwell, E.B., Erickson E.F, Glassgold, A.E., Haas, M.R., Hollenbach, D.J., Joseph, R.D., Meixner, M., Shull, M.J., Tielens, A., Townes C.H., Werner M.W., Young E.T., High-Resolution Far-Infrared Spectroscopy on SOFIA, A White Paper for NASA's Space Science Roadmap Revision, 2002

Clancy, R. Todd; Muhleman, Duane O.; Berge, Glenn L., Global changes in the 0-70 KM thermal structure of the Mars atmosphere derived from 1975 to 1989 microwave CO spectra, Journal of Geophysical Research, vol. 95, Aug. 30, p. 14543-14554, 1990.

Crewell, S., K. Künzi, H. Nett, T. Wehr, and P. Hartogh, Aircraft measurements of CLO and HCL during EASOE 1991/92, Geophys. Res. Lett., 21, 1267–1270, 1994.

Crovisier J., Encrenaz T., Comet Science, *Cambridge University Press*, ISBN 0 521 64591 3, 2000.

Cuzzi J.N., Muhleman D.O., The microwave spectrum and the nature of the sub-surface of Mars, Icarus, 17, 548, 1972.

Darlington, S., "Demodulation of wide band low-power FM signals", Bell Syst., Tech. J., 43, 339-374, 1964.

de LaNoe J., J. Ovarlez, H. Ovarlez, C. Schiller, N. Lautié, P. Ricaud, J. Urban, D. G. Feist, L. Zalesak, N. Kämpfer, M. Ridal, D. Murtagh, K. Lindner, U. Klein, K. Künzi, P. Forkman, A. Winnberg, P. Hartogh, and A. Engel, Water vapour distribution inside and outside the polar vortex during THESEO, in: Stratospheric ozone 1999 (edited by N. R. P. Harris, M. Guirlet, and G. T. Amanatidis), no. 73 in Air pollution research report, Saint Jean de Luz, France, 2000.

Duffett-Smith P., *Practical Astronomy with your calculator* 3rd edition, Cambridge University Press, ISBN 0521356997, 1989.

Ebel A., Berger, U., Martian Atmosphere – Circulation and Climate – MART-ACC, Schlussbericht, Universität zu Köln, 1997.

Elliot, J.L., Stellar Occultation Studies of the Solar System, NASA Technical report CR-97-206750; NAS 1.26:206750, 1998.

Ellis, M.G., Electronic Filter Analysis and Synthesis, Electronic Filter Analysis and Synthesis, ISBN: 0890066167, 1994

Elsasser W.M., Culbertson M.F., Atmospheric radiation tables. Meteor. Monogr., 4, 1-43, 1960.

Emrich A., Digital Autocorrelation Spectrometer Study and Implementation, Final Report, ESTEC, 1996.

Feist D. G., C. P. Aellig, N. Kämpfer, P. M. Solomon, J. W. Barrett, S. Zoonematkermani, P. Hartogh, C. Jarchow, and J. W. Waters, Validation of stratospheric ClO measurements from the Millimeter-wave Atmospheric Sounder (MAS), J. Geophys. Res., 105(D7), 9053–9062, 2000.

Fels S.B. & Schwarzkopf M.D., An efficient accurate algorithm for calculating CO_2 15 μm band cooling rates, JGR, 86,1205-1232, 1981.

Fink, U.; Combi, M. R., The effect of using different scale lengths on the production rates of Comet 46P/Wirtanen, Planetary and Space Science, Volume 52, Issue 7, p. 573-580., 2004.

Forget, François; Hourdin, Frédéric; Fournier, Richard; Hourdin, Christophe; Talagrand, Olivier; Collins, Matthew; Lewis, Stephen R.; Read, Peter. L.; Huot, Jean-Paul, Journal of Geophysical Research, Volume 104, Issue E10, p. 24155-24176, 1999.

Friedrich, W., Metall- und Maschinentechnik, Verlag Dümmlers, ISBN 3427510336, 1999.

Friese E., Sigma: Ein dynamische Kern für Modelle planetarer Atmosphären, Diplomarbeit, Institut für Geophysik und Metereologie der Universität zu Köln, 1998.

Genttle, Ken, A Technical Tutorial on Digital Signal Synthesis, Analog Devices application notes, 1999.

Guesten, R.; Camara, I.; Hartogh, P.; Huebers, H. W.; Graf, U. U.; Jacobs, K.; Kasemann, C.; Roeser, H. P.; Schieder, R. T.; Schnieder, G.; Sievertz, O.; Stutzki, J.; Villanueva, G.; Wagner, A.; van der Wal, P.; Wunsch, A., GREAT: The German Receiver for Astronomy at Terahertz Frequencies, Proceedings of the SPIE, Volume 4857, pp. 56-61, 2003.

Haberle R.M., Pollack J.B., Barnes J.R., Zurek R.W., Leovy C.B., Murphy J.R., Lee H., Schaeffer J., Mars atmosphere dynamics as simulated by the NASA/Ames general circulation model, J. Geophys.Res. 98, 3093-3123, 1993.

Haser, L., Distribution d'intensite dans la tete d'une comete, Bull. Acad. R. Sci. Liege, vol. 43, p. 740-750, 1957.

Hartogh, P., Chirptransformations-Spektrometer für die passive Millimeterwellenradiometrie: Messung der 142 GHz Emissionslinie des atmosphärischen Ozons, PhD thesis, MPAE-W-66-90-13, 1990.

Hartogh, P., Microwave Investigation on Mars Express (MIME), Part I: Scientific and Technical Plan, Proposal in response to AO ESA/SI(97), 1998.

Hartogh P., The MPAE ground-based microwave spectrometer (WASPAM), in: *Stratospheric Processes and their Role in Climate (A project of the WMO/ICSU/IOC World Climate Research Programme): SPARC Assessment of Upper Tropospheric and Stratospheric Water Vapour* (edited by D. Kley, J. M. R. III, and C. Phillips), vol. WCRP-113, pp. 45–46, SPARC Office, BP 3, 91371 Verrieres le Buisson Cedex, France, 2000.

Hartogh P., Jarchow C., Millimeter wave detection of mesospheric ozone using a high resolution chirp transform spectrometer backend, in: Proc. IGARSS'94 on Surface and Atmospheric Remote Sensing: Technologies, Data Analysis and Interpretation, Pasadena, Vol. III (edited by T. I. Stein), pp. 3–5, Publ. Services IEEE, Piscataway, N. J., 1994.

Hartogh P., Jarchow C., Groundbased detection of middle atmospheric water vapor, in: Global Process Monitoring and Remote Sensing of Ocean and Sea Ice, EUROPTO-Series 2586, pp. 188–195, SPIE, Bellingham, 1995.

Hartogh, P., Jarchow C., Martin R., Ground-based Microwave Measurement of the Martian Atmosphere, AGU Fall meeting, Poster P32B-10, 1997.

Hartogh, P., Jarchow C., Biver N., Crovisier, J., Bockelee-Morvan D., Villanueva G., Investigation of the composition of comet C/2000 WM_1, Submillimeter Telescope Observatory proposal, 2001.

Hartogh P., C. Jarchow, G. R. Sonnemann, and M. Grygalashvyly, On the spatiotemporal behaviour of ozone within the upper mesosphere/mesopause region under nearly polar night conditions, J. Geophys. Res., 109, D18303, doi:10.1029/2004JD004576, 2004.

Hartogh, P., Hartmann, G.K., A high-resolution chirp transform spectrometer for microwave spectrometers, Meas. Sci. Technol., 1, 592-595, 1990.

Hofstadter, M.D., Hartogh, P., McMullin, J.P., Martin, R.N., Jarchow, C., Peters, W., A Search For Variability in the HCN to H2CO Ratio in Comet Hale-Bopp , Earth, Moon and Planets 78, 53, 1999.

Holton, J.R., The dynamic meteorology of the stratosphere and mesosphere, Meteorological monographs 15(37), American Meteorological Society, Boston, 1975.

Hourdin F., Phu Le Van, Forget F., Talagrand O., Meteorological variability and the annual surface pressure cycle on Mars, J. Atmos. Sci., 50, 3625-3640, 1993.

Hübers, H.-W., Semenov A., Richter H., Birk M., Krocka M., Mair U., Smirnov K., Gol'tsman G., Voronov B., Terahertz Heterodyne Receiver with a Hot-Electron Bolometer mixer, Proceedings of the Far-IR, Sub-MM, and MM detector Technology Workshop, J.Wold and J.Davidson, eds., 2002.

Ivanov A.B, Muhleman D.O., Reflected signal analysis and surface albedo in the Mars orbiter laser altimeter (MOLA) investigation, Lunar and Planetary Science 32, Paper Nº 1917, 2001.

Jarchow C., Bestimmung atmosphärischer Wasserdampf- und Ozonprofile mittels bodengebundener Millimeterwellen-Fernerkundung, PhD thesis, MPAE-W-016-99-06, 1999.

Jarchow, C., Hartogh P., Martin R., Ground based observations of the Martian atmosphere at different solar longitudes with a Chirp Transform Spectrometer, AGU Fall Meeting, 1997.

Jack M.A., Grant P.M., Collins J.H, The Theory, Design and Applications of Surface Acoustic Wave Fourier-Transfrom Processors.

Jackson B. D., Klapwijk T. M., The current status of low-noise THz mixers based on SIS junctions, Physica C, Volume 372, p. 368-373, 2002.

Jockers K., Hartogh P., Jarchow C., Villanueva G., Slawomira Szutowicz, HCN outgassing properties of the nucleus of comet 2P/Encke, Submillimeter Telescope Observatory proposal, 2003.

Joshi, M. M.; Haberle, R. M.; Barnes, J. R.; Murphy, J. R.; Schaeffer, J., Low-level jets in the NASA Ames Mars general circulation model, Journal of Geophysical Research, Volume 102, Issue E3, p. 6511-6524, 1997.

Kakar, R. K.; Walters, J. W.; Wilson, W. J., Mars - Microwave detection of carbon monoxide, Science, vol. 196, p. 1090-1091, 1997.

Kieffer, H. H.; Martin, T. Z.; Peterfreund, A. R.; Jakosky, B. M.; Miner, E. D.; Palluconi, F. D., Thermal and albedo mapping of Mars during the Viking primary mission, Journal of Geophysical Research, vol. 82, p. 4249-4291, 1977.

Kim Y.J., Farrara J.D. and Mechoso C.R., Sensitivity of AGCM Simulations to Modifications in the Ozone Distribution and Refinements in Selected Physical Parameterizations, Journal of the Meteorological Society of Japan, 1998.

Klapwijk T.M., Dieleman P., Graauw M.W.M., Pushing the operating range of SIS mixers into the THz regime, Supercond. Sci. Technol. 10, 876-879, 1997.

Klauder, J.R., Price A.C., Darlington S., Albersheim W.J., The theory and design of chirp radars, Bell Syst., Tech. J., 39, 745-808, 1960.

Kliore A.J., Cain D.L., Levy G.S., Eshleman V.R., Fjeldbo G., Drake F.D., Occultation experiment: Results of the first direct measurement of Mars' atmosphere and ionosphere. Science 149:1243-1248, 1965.

KOSMA, Diffusionsgekühlte Hot-Electron-Bolometer, URL: http://www.ph1.uni-koeln.de/workgroups/astro_instrumentation/sis/deutsch/jheb_d.html, 2004.

Krabbe, A., Astronomical research with SOFIA at the DLR Institute for Space Sensor Technology, SOFIA Astronomy and Technology in the 21st century, *Wissenschaft und Technik Verlag*, ISBN 3-89685-558-1, 1998.

Krabbe, A. Röser, H.P., SOFIA - Astronomy and Technology in the 21st Century, Reviews of Modern Astronomy, Vol.12, Astronomische Gesellschaft, ISBN 3-9805176-2-4, 1999.

Kurihara, Y., On the use of implicit and iterative methods for the time integration of the wave equation. Monthly Weather Review, 93(1), 33-46, 1965.

Kutepov, A. A.; Formichev V.I., Application of the second order escape probability approximation to the solution of the NLTE vibration-rotational band radiative transfer problem, J. Atmos. Terr. Phys. 55, 1-7, 1993.

Kutepov, A. A.; Kunze, D.; Hummer, D. G.; Rybicki, G. B., The solution of radiative transfer problems in molecular bands without the LTE assumption by accelerated lambda iteration methods, Journal of Quantitative Spectroscopy and Radiative Transfer, 46, 347-365, 1991.

Kutepov, A. A.; Feofilov, A. G.; Gusev, O. A.; Kostiuk, Th.; Maguire, W.; Livengood, T. A., Probing Earth's Middle Atmosphere: Non-LTE Processes And Infrared Heterodyne Spectroscopy, A Preliminary Study, EGS XXVII General Assembly, Nice, 21-26 April 2002, abstract #5003, 2002.

Larson, H. P.; Weaver, H. A.; Mumma, M. J.; Drapatz, S., Airborne infrared spectroscopy of Comet Wilson (1986l) and comparisons with Comet Halley, *Astrophysical Journal*, 338, 1106-1114, 1989.

Lascari, L., Microwave Filters, Impedance-Matching Networks, and Coupling Structures, Public release tools.rfdude.com, 2001.

Laskar, J.; Gastineau, M.; Joutel, F.; Levrard, B.; Robutel, P.; Correia, A., A New Astronomical Solution for the Long Term Evolution of the Insolation Quantities of Mars, *35th Lunar and Planetary Science Conference*, abstract no.1600, 2004.

Lellouch, E.; Rosenqvist, J.; Goldstein, J. J.; Bougher, S. W.; Paubert, G., First absolute wind measurements in the middle atmosphere of Mars, Astrophysical Journal, Part 1, vol. 383, p. 401-406., 1991.

Leovy, C. B.; Zurek, R. W., Thermal tides and Martian dust storms - Direct evidence for coupling, Journal of Geophysical Research, vol. 84, p. 2956-2968, 1979.

Leovy C., Weather and climate on Mars, Insight Review Articles, Nature, 412, 2001.

Liou, K.N., Radiation and Cloud Processes in the Atmosphere, Oxford Monographs on Geology and Geophysics, 20, 1992.

Liou, K.N., Sasamori T., On the transfer of solar radiation in aerosol atmospheres, J. Atmos. Sci., 32, 2166-2177, 1975.

Lopez-Puertas M. and Lopez-Valverde M., Radiative energy balance of CO_2 non-LTE infrared emissions in the Martian atmosphere, Icarus, 114, 113-129, 1995.

Lopez-Puertas M. and Lopez-Valverde M., CO2 non-LTE cooling rate at 15 μm and its parameterization for the Mars atmosphere, ESA Contract – Martian Environment Models, ESA 11369/95/NL/JG, 2001.

Lopez-Puertas M., Taylor F.W., Non-LTE Radiative Transfer in the Atmosphere, Series on Atmospheric, Oceanic and Planetary Physics Vol.3, World Scientific Publishing, ISBN 981-02-4566-1, 2001.

Lopez-Valverde M.A., Edwards D.P., Lopez-Puertas M., Roldan C., Non-local thermodynamic equilibrium in general circulation models of the Martian atmosphere. 1. Effects of the local thermodynamic equilibrium approximation on thermal cooling and solar heating, JGR, 103-E7, p16-799-16811, 1998.

Maines J.D., Paige E.G.S., Surface-Acoustic Wave Devices for Signal Processing Applications, Proc. IEEE, 64, 639-652, 1976.

Makimoto, M., Yamashita, S., Microwave Resonators and Filters for Wireless Communication. Theory, Design and Application, Springer-Verlag Berlin Heidelberg, ISBN: 3540675353, 2000.

Martin T. Z., Mean thermal and albedo behavior of the Mars surface and atmosphere over a martian year. Icarus 45, 427–446, 1981.

McGrath W., Skalare A., Karasik B., Superconductive Hot-Electron bolometer mixers, Proceedings of the Far-IR, Sub-MM, and MM detector Technology Workshop, J.Wold and J.Davidson, eds., 2002.

Meister C.V., Hartogh P., Villanueva G., Hydronamic Model of the Martian Atmosphere between Near-Surface Layers and an Altitude of about 130 km, MPAE Internal publication, MPAE-W-100-02-06, 2003.

Mellon M.T., Thermal Inertia and Rock Abundance, Exploring Mars with TES Conference, Tempe-Arizona, 2001.

Mellon, M.T., Jakosky B., Kieffer H., Christensen P., High resolution thermal inertia mapping from the Mars Global Surveyor Thermal Emission Spectrometer, Icarus, 148, 437-455, 2000

Mumma M. J., Novak R. E., DiSanti M. A., Bonev B., Dello Russo N., Magee-Sauer K., Seasonal mapping of HDO and H2O in the Martian Atmosphere, *Sixth International Conference on Mars*, paper: 3186, 2003.

Mongia, R., Bhartia, P., Bahl, I.J., RF and Microwave Coupled-Line Circuits, Artech House Publishers, ISBN: 0890068305, 1999.

Nair, H., Allen, M., Anbar, A.D., Yung, Y.L., Clancy, L.T., A photochemical model of the Martian atmosphere, *Icarus*, 111, 124-150, 1994.

Nedoluha G.E, Hartogh P., Upper stratosphere comparisons (WVMS and WASPAM), in: *Stratospheric Processes and their Role in Climate (A project of the WMO/ICSU/IOC World*

Climate Research Programme): SPARC Assessment of Upper Tropospheric and Stratospheric Water Vapour (edited by D. Kley, J. M. R. III, and C. Phillips), vol. WCRP-113, pp. 139–141, SPARC Office, BP 3, 91371 Verrieres le Buisson Cedex, France, 2000.

Nett, H., S. Crewell, P. Hartogh, and K. Künzi, Detection of stratospheric minor constituents using an airborne submillimeter wave sensor, in: *Proceedings of the International Symposium on Environmental Sensing*, vol. 1715 of *SPIE, Optical Methods in Atmospheric Chemistry*, pp. 468–475, Berlin, 22-26 June 1992, 1993.

Newman, C.E., Lewis, S.R., Read P.L., Forget, F., Modeling the Martian dust cycle 2. Multiannual radiatively active dust transport simulations, *J. Geophys. Res.,* 107, 5124, 2002.

Nier, A. O.; Hanson, W. B.; Seiff, A.; McElroy, M. B.; Spencer, N. W.; Duckett, R. J.; Knight, T. C. D.; Cook, W. S., Composition and structure of the Martian atmosphere - Preliminary results from Viking 1, Science, vol. 193, Aug. 27, 1976, p. 786-788, 1976.

Putley, E.H., The Detection of Sub-mm Radiation, Proceedings of the IEEE, p 1412, 1963.

Pollack J.B., Haberle R.M., Murphy J.R., Schaeffer J.R., Lee H., Simulations of the general circulation of the Martian atmosphere. II Seasonal pressure variations, Journal of Geophysical Research, 98-E2, 3149-3181, 1993.

Rabanus D., Functional Description of GREAT's BWO (Local Oscillator of the 1.9 THz channel), GA5300001, GREAT-FAA certification documents, 2003.

Ramanathan V., Radiative Transfer within the Earth's Troposphere and Stratosphere: a simplified Radiative-Convective Model, Journal of Atmospheric Sciences, Vol.33,p1330,1976.

Ramanathan, V., Pitcher E.J., Malone R.C., Blackmon M.L., The response of a spectral general circulation model refinements in radiative processes, J. Atmos. Sci., 40, 605-621, 1983.

Rauer, H., Harris, A., Knollenberg J., Kührt, E., Observations of comets and asteroids with SOFIA, SOFIA Astronomy and Technology in the 21st century, *Wissenschaft und Technik Verlag*, ISBN 3-89685-558-1, 1998.

Reindl L., SAW-Based Radio Tag and Sensor Systems, Proc. International Conference on New Piezoelectric Materials and High Performance Acoustic Wave Devices, 71-78, 2002.

Reindl, L., Shrena I., Richter H., Peter R., High Precision Wireless Measurement of Temperature by Using Surface Acoustic Waves Sensors, Proc. 2003 IEEE Freq. Contrl. Symp., 2003.

Richardson, M.I., Water ice clouds in the Martian atmosphere: General Circulation model experiments with a simple cloud scheme*, J. Geophys. Res.,* 107, 5064, 2002.

Rodin, A.V., Wilson, R.J., Seasonal modulation of water and dust cycles in the Martian atmosphere by planetary waves synchronization, *EGS-AGU-EUG Joint Assembly*, abstract 13202, 2003.

Römer K., Villanueva G., Clement M., The IF-LO processor for the SOFIA-GREAT Instrument, CADR FAA-Certification package, GA4600002, 2003.

Roy A.E., Orbital motion, Institute of Physics Publishing – Bristol, ISBN 0-85274-229-0, 1988.

Rubini, A., Linux Device Drivers, O'Reilly, ISBN: 1-56892-292-1, 1998.

Rylyakov A.V., Polonsky S.V., All-Digital 1-Bit RSFQ Autocorrelator for Radioastronomy Applications: Design and Experimental Results, IEEE trans. ap. supercond., 8, 14-19, 1998.

Ryan, J. A.; Henry, R. M.; Hess, S. L.; Leovy, C. B.; Tillman, J. E.; Walcek, C., Mars meteorology - Three seasons at the surface, Geophysical Research Letters, vol. 5, p. 715-718., 1978.

Schieder R., Kramer C., Optimization of heterodyne observations using Allan variance measurements, A&A, 373, 746-756, 2001.

Schmitt S.R., Computing planet positions using mean orbital elements, http://home.att.net/~srschmitt/planetorbits.html, 2004.

Schubert J., Semenov A., Gol'tsman G., Hübers H-W., Schwaab G., Voronov B., Gerschenzon E., Noise temperature of an NbN hot-electron bolometric mixer at frequencies from 0.7 to 5.2 THz, Supercond. Sci. Technol., 12, 748-750, 1999.

Schofield J. T., Barnes J. R., Crisp D., Haberle R. M., Larsen S., Magalhães J. A., Murphy J. R., Seiff A., Wilson G., The Mars Pathfinder Atmospheric Structure Investigation/Meteorology (ASI/MET) Experiment, Science, 278, Issue 5344, 1752-1758 , 1997.

Schönhoff, U., Nordmann, R., Kärcher, H.J., Wandner, K., Eisenträger, P., End-to-End Simulation for the Verification of the Pointing Accuracy of the SOFIA telescope, SOFIA Astronomy and Technology in the 21st century, *Wissenschaft und Technik Verlag*, ISBN 3-89685-558-1, 1998.

Schwaab G.W., Bründerman E., Hübers H.W., Krabbe A., Röser H.P., Sirmain G., A 2.6 THz Heterodyne Receiver Array for SOFIA, SOFIA Astronomy and Technology in the 21st century, *Wissenschaft und Technik Verlag*, ISBN 3-89685-558-1, 1998.

Seele C., Hartogh P., A case study of the middle atmospheric water vapour transport during the February 1998 stratospheric warning, Geophysical Research letters, 27, 3309-3312, 2000.

Shapiro R., The use of linear filtering as a parameterization of atmospheric diffusion, J. Atmos. Sci., 28, 523-531, 1971.

Sievertz, O., Functional description of AOS, GREAT-FAA certification documents, GA7100001, 2003.

Slobodonik A.J. Jr., Surface Acoustic Waves and SAW materials, Proc. IEEE, 64, 581-595, 1976.

Smith, M.D., The annual cycle of water vapour on Mars as observed by the Thermal Emission Spectrometer, *J. Geophys. Res.*, 107, 5115, 2002.

Smith, D.E., M.T. Zuber, S.C. Solomon, R.J. Phillips, J.W. Head, J.B. Garvin, W.B. Banerdt, D.O. Muhleman, G.H. Pettengill, G.A. Neumann, F.G. Lemoine, J.B. Abshire, O. Aharonson, C.D. Brown, S.A. Hauck, A.B. Ivanov, P.J. McGovern, H.J. Zwally, T.C. Duxbury, The global topography of Mars and implications for surface evolution, Science, 284 1495-1503, 1999.

Smith, D.E.; Zuber, M.T.; Frey, H.V.; Garvin, J.B.; Head, J.W.; Muhleman, D.O.; Pettengill, G.H.; Phillips, R.J.; Solomon, S.C.; Zwally, H.J.; Banerdt, W. B.; Duxbury, T.C.; Golombek, M.P.; Lemoine, F.G.; Neumann, G.A.; Rowlands, D.D.; Aharonson, O.; Ford, P.G.; Ivanov, A.B.; Johnson, C.L.; McGovern, P.J.; Abshire, J.B.; Afzal, R.S.; Sun, X., Mars Orbiter Laser Altimeter: Experiment summary after the first year of global mapping of Mars, Journal of Geophysical Research, Volume 106, Issue E10, p. 23689-23722, 2001.

Smith III, F.L., Smith C., Numerical Evaluation of Chapman's grazing incidence integral ch(X,X), J.Geophys. Res., 77, 3592-3597, 1972.

SOFIA, www.sofia.arc.nasa.gov, 2004.

Song L., P. Hartogh, C. Jarchow, and G. Hansen, Comparison of ozone profiles derived from ground-based microwave and lidar measurements, in: Microwave Remote Sensing of the Atmosphere and Environment (edited by D. L. Wu, Y.-Q. Jin, and J.-S. Jiang), vol. 3503 of Proceedings of SPIE, pp. 202–209, Washington, 1998.

Stephens G.L., The parameterization of radiation for numerical weather prediction models (Review), Mon. Wea. Rev., 112, 1069-1082, 1984.

Stevens M.H., J. Gumbel, C. R. Englert, K. U. Grossmann, M. Rapp, and P. Hartogh, Polar mesospheric clouds formed from space shuttle exhaust, Geophys. Res. Lett., 30(10), 1546, doi:10.1029/2003GL017249, 2003.

Tokano T., Spatial in homogeneity of the Martian subsurface water distribution: implication from a global water cycle model, *Icarus*, 164, 50-78, 2003.

van der Wal, P.B., GREAT Functional description, GREAT-FAA certification documents, GA1000001, 2003.

Villanueva, G., SOFIA-GREAT-CTS CADR FAA-Certification package, GA4600002 - GA7200001 - GA7200002 - GA7200003 - GA7200004 - GA7200005 - GA7200006 - GH7200001, 2003a.

Villanueva, G., Hartogh, P., Jarchow, C., Mapping of the Martian temperature and wind distribution, Submillimeter Telescope Observatory proposal, 2003b.

Villanueva, G., Dissipation processes on enclosed equipment units, MPAE-Internal publication, 2002.

von Zahn U., G. Baumgarten, U. Berger, J. Fiedler, and P. Hartogh, Noctilucent clouds and the mesospheric water vapour: the past decade, Atmos. Chem. Phys.,, 4, 2449–2464, 2004.

Weaver, H.A., Mumma, M.J., Larson, H.P., Airborne infrared investigations of water in the coma of Halley's Comet, ESA 20th ESLAB Symposium of the Exploration of Halley's Comet, 1, 329-334, 1986.

Wehr T., S. Crewell, K. Künzi, J. Langen, H. Nett, J. Urban, and P. Hartogh, Remote sensing of ClO and HCl over northern Scandinavia in winter 1992 with an airborne submillimeter radiometer, J. Geophys. Res., 100, 20957–20968, 1995.

Williams, D.R., Mars Fact Sheet, NASA Goddard Space Flight Center Mars website, http://nssdc.gsfc.nasa.gov/planetary/factsheet/marsfact.html, 2004.

Woodney, L. M.; A'Hearn, M. F.; Schleicher, David G.; Farnham, Tony L.; McMullin, J. P.; Wright, M. C. H.; Veal, J. M.; Snyder, Lewis E.; de Pater, Imke; Forster, J. R.; Palmer, Patrick; Kuan, Y.-J.; Williams, Wendy R.; Cheung, Chi C.; Smith, Bridget R., Morphology of HCN and CN in Comet Hale-Bopp (1995 O1), Icarus, Volume 157, Issue 1, pp. 193-204, 2002.

Zmuidzinas J., Coherent detection and SIS Mixers, Proceedings of the Far-IR, Sub-MM, and MM detector Technology Workshop, J.Wold and J.Davidson, eds., 2002.

Zmuidzinas J., Edgar M., Caltech Submillimeter Interstellar Medium Investigations Receiver, SOFIA-PADR - Preliminary Airworthiness Design Review document, 1999.

Zurek R.W., Comparative aspects of the climate on Mars: an introduction to the current atmosphere, Mars, Chapter Nº 24, Univ. Arizona press, ISBN 0-8165-1257-4, 1992.

Zurek, R.W., Diurnal Tide in the Martian Atmosphere; J. Atmos. Sci., 33, 321-337, 1976.

Acknowledgements

It is hard to express in this short letter my gratitude to all the people who contributed in the development of this thesis, their help is invaluable. I especially thank my supervisor at MPS, Dr. Paul Hartogh, for constantly posing new challenges and for his guidance and sharing of knowledge in the subject. Also, I feel honored and grateful to the Max-Planck-Society for giving me the chance as an international student to research in such a motivating topic.

I thank my supervisors at the Albert-Ludwigs-Universität Freiburg, Prof. Dr. Leonhardt Reindl and Prof. Dr. Oskar von der Lühe for their interest in my work and their patience.

I would like to thank the scientific and technical members of the group at the institute, who significantly contributed in the realization of the instrument with their accurate and reliable work; thanks Christopher, Eckhard, Peter, Helmut and Markus. Many thanks to Michael Clement, with whom I shared long days and long nights in the laboratory solving critical issues, which at that moment seemed to be unsolvable.

When I recall my moments at the HHSMT telescope in Arizona I can only feel enthusiasm, having the possibility to observe the immensity of the sky with one of the best submillimeter telescopes. Thanks Bill, Patrick and the two Bobs, as well as the other staff members.

Even though I spent my years of studies in a city with 1000 inhabitants and 100 meters away from the Institute, my friends at the Institute made me feel like home. They were not only good companions but people from whom you can learn and enjoy life. Thanks a lot my friends: Santo, Juan, Luciano, Francisco and Hebe.

Finally and most important, I thank my parents, Luis y Celina, my brother Emiliano and his wife Clarisa for their tremendous and unconditional support in the completion of my goals and dreams. To my beloved, Katharina Rudolf, who was always there to accompany me with her love, patient and wisdom. I am deeply grateful to you Katharina!!