QUASI–OPTICAL SIS RECEIVERS AND ASTROPHYSICAL OBSERVATIONS AT SUBMILLIMETER WAVELENGTHS

Thesis by

Thomas H. Büttgenbach

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

1993

(Submitted April, 1993)

Acknowledgements

I would like to express my appreciation to all the people at Caltech for making it a great place and for creating an enjoyable and productive atmosphere in which to be a graduate student. The small size of the Caltech community has allowed me to easily communicate with people from a broad range of scientific backgrounds, and I have benefited greatly from the very high quality of my fellow students, faculty and administration. I cannot possibly thank everybody personally who has been of help to me during my stay here, but would like to extend my gratitude to the following people.

Foremost I am indebted to my advisor Tom Phillips, who has strongly supported my research and has also provided me with much guidance in personal and professional matters. He has always been tolerant of my erratic time schedule and extracurricular activities.

Throughout the years of labor in the lead and bismuth mines of the submillimeter receiver group, Ron Miller from AT&T Bell Labs has tirelessly supported my efforts, for which I will always be thankful. I have also enjoyed his warm hospitality during my stays at his home.

Mike Wengler got me started on this business of SIS receivers and Dan Watson introduced me to the mysteries of the submillimeter group lab, for which I would like to thank both of them.

Jocelyn Keene has suffered with me through many sometimes frustrating but always enjoyable observing runs at the CSO. I also enjoyed our many hours of common work at trying to figure out what we did in Hawaii. Nick Scoville has reassured me through his delightful character that life in science after graduate school can be fun.

I am indebted to the following people for their collaborative efforts, and support in the lab: Mei Bin, Gabriel Rebeiz, Jonas Zmuidzinas, Peter Zimmermann, Pat Schaffer, Dan Watson, Brian Ellison, Chris Walker, Hermann Rothermel, Peter Siegel, Geoff Blake, Evelyne Roueff, and Bruce Veidt. I also appreciate the considerable assistance of Richard Borup, Dave Vail, Walt Schaal, and Marty Gould in the construction of the receivers. I am indebted to the people at JPL, foremost Rick LeDuc and Jeff Stern, also Paul Maker, Rich Muller and Bruce Bumble, for their expertise and efforts in Niobium SIS junction fabrication. The staff at the Caltech Submillimeter

Observatory, Hawaii, has always been extremely helpful and supportive of my efforts in testing new receivers.

Without the many fruitful and enlightening discussions with Jonas Zmuidzinas, Robert Schoelkopf, Taco, Peter Schilke, Gene Serabyn, Thomas Rose, Frank Lewen, Todd Groesbeck, Darek Lis, John Carlstrom, to name a few, I would have never been able to complete my work.

I would like to thank Gisbert Winnewisser, Peter Zimmermann, Dieter Wohlleben, and Janos Haidu, all at the University of Cologne, for supporting me in studying abroad.

My life at Caltech has been enriched by my office mates Taco, Jeff Stern, Eric Aslakson, Alice Quillen, and my roommates Nils Halverson and Stephanie Puddy. Since I moved into Taco's office my computer problems seem to have mysteriously disappeared and I have greatly enjoyed the many late night discussions. Stephanie Puddy I owe special thanks for introducing me to life outside Caltech and Nils Halverson, my current roommate, for helping me to restore my sanity.

During my stay at Caltech I picked up a new passion that has truly expanded my horizons and my acknowledgments would not be complete without thanking the friends that fly along the same airways as I do: Sheri Trujillo, Kevin Condroski, Nils Halverson, Ernie Camacho, Steve Taylor, Yoshi Yamada, and especially Eric Wierman, who not only instructed me to fly, but also inspired my love for flying. Eric's mind boggling ideas have broadened my outlook on life. Sheri's companionship and always positive spirit have turned many cloudy days into fun ones. Among my friends I have met during my California days but who prefer the solid ground are Claudia Muth, Debra George, Christoph Schmidhuber, Kalle and Julie Voss. All of them have made my life more balanced.

None of this, of course, would have been possible without the love and support of my parents, Ilse Katharina and Klaus, who have always given me the freedom and encouragement to pursue my interests, and my brothers Frank and Jochen.

I would like to acknowledge personal financial support from the Studienstiftung des Deutschen Volkes, Fulbright Foundation, Stiftung zur Förderung Selbständiger Junger Menschen of the German Foundation for Science, Schlumberger Foundation, and the Stemple Memorial Fund.

Abstract

This thesis describes the development of a new generation of submillimeter wave receivers aimed at future integrated array receivers and presents astronomical observations taken with a single element of such a new receiver in the submillimeter wavelength band.

The technological development presented in this thesis was driven by the need to develop heterodyne receivers based on superconductor–insulator–superconductor (SIS) detectors, which have proven to be the most sensitive detectors in the millimeter and low submillimeter band, and that are suitable for the construction of array receivers. A quasi–optical approach was chosen in order to take advantage of the planar photolithography used in SIS diode manufacturing. This allows straightforward integration with planar antennas. The introductory chapter describes an initial receiver design, based on a planar logarithmic spiral antenna, that was an improvement over existing quasi–optical designs using planar antennas.

This receiver was the first submillimeter SIS receiver ever to be used for astronomical observations. However, it also showed that two major problems still had to be overcome to develop this technology to the mature state where array receivers can be designed. The beam of an individual element, launched by the planar antenna, has to be of high quality – comparable to that of a waveguide horn antenna – to provide good coupling to a telescope. Second, the impedance of the SIS detector, dominated by its capacitance, has to be matched to the antenna's impedance over the operating range of the receiver.

The optics problem was solved by introducing a novel antenna design called the hybrid antenna. The hybrid antenna is defined as a dielectric lens–antenna as a special case of an extended hemispherical dielectric lens that is operated in the diffraction limited regime. It is a modified version of the planar antenna on a lens scheme developed by Rutledge. The dielectric lens–antenna is fed by a planar–structure feed antenna and the combination of the two is termed a hybrid antenna. Beam pattern and aperture efficiency measurements were made at millimeter and submillimeter wavelengths as a function of extension of the hemispherical lens and different lens sizes. An optimum extension distance is found experimentally and theoretically for which excellent beam patterns and simultaneously high aperture efficiencies can be achieved. At 115 GHz the aperture efficiency was measured to be $(76 \pm 5)\%$ for a diffraction limited beam with

sidelobes below -17 dB. The hybrid antenna is diffraction limited, space efficient in an array due to its high aperture efficiency, and is easily mass produced, thus being well suited for focal plane heterodyne receiver arrays. A single element hybrid antenna receiver yielded coupling efficiencies between the receiver and the Caltech Submillimeter Telescope of values approaching those achieved by the best waveguide horn based receiver systems.

The problem of tuning the SIS junction's capacitance was solved by using a novel superconducting transmission line circuit, called the end–loaded–stub, together with a real impedance transformer. This Nb based circuit is integrated with the planar feed antenna of the hybrid antenna and a $Nb/AlO_x/Nb$ SIS detector. A goal for the matching circuit design was to achieve an octave of bandwidth and computer modeling of the device correctly predicted the measured bandwidth and characteristic frequencies to within 8%. The bandwidth measurements were carried out by using the SIS diode in direct detection mode with a Fourier Transform Spectrometer. A good match was obtained from 200 GHz to 475 GHz between the antenna and a relatively large area (1 μm^2) tunnel junction with $\omega R_n C \approx 2-4$. Noise measurements were made in heterodyne mode at 318 GHz, 395 GHz, 426 GHz and 492 GHz, yielding uncorrected double sideband receiver noise temperatures of 200 K, 230 K, 220 K and 500 K, respectively. These results are comparable to state of the art waveguide receivers.

The combination of a hybrid antenna with the integrated tuning circuit thus lays the foundation for the development of integrated SIS focal plane receiver arrays.

The final chapter describes observations of the neutral carbon C I($^3P_1 \rightarrow {}^3P_0$) line at 492 GHz with a single element quasi–optical SIS receiver. The goal was to investigate the large– and small–scale distribution of C I in the interstellar medium (ISM). Observations of neutral atomic carbon in the galaxy IC 342, made at the Caltech Submillimeter Observatory, are presented, which represent the first extragalactic submillimeter detection of C I. The C I emission from IC 342 was found to have a main beam temperature of about 1 K, which is a factor of two greater than that for ^{13}CO(2 $\rightarrow$ 1) in the same size beam. The integrated line intensity for the central 15″ of IC 342 was found to be $(6.2 \pm 1.2) \times 10^{-6}$ erg s^{-1} cm^{-2} sr^{-1}. A 45″ cut from the center to the south showed that the intensity ratio of C I to CO(2 $\rightarrow$ 1) is constant at about 1.4. The C I to CO abundance ratio is about 15% and the total C I abundance relative to hydrogen is 7×10^{-6},

yielding about $\geq 10\%$ of all carbon in the gas phase to be in form of neutral atomic carbon.

Comparing the IC 342 data with COBE observations of the Milky Way it is found that the intensity ratio of C I to CO($2 \rightarrow 1$) is approximately the same for IC 342 as it is for the Milky Way, despite the fact that the IC 342 measurements are from the center a moderate star burst galaxy, while the COBE data are an average over the entire Milky Way. From the comparison between the Milky Way and IC 342 and the cut through IC 342 it can thus be concluded that there is a strong correlation of CO with C I emission on large scales.

The small–scale structure of neutral carbon in the ISM is studied with a $3'$ by $4'$ ($\Delta\alpha \times \Delta\delta$) map, sampled at $20''$ intervals with a $15''$ beam, in the Orion Molecular Cloud (OMC). The map covers the embedded infrared source IRc2, the southern source (OMC–1 south, FIR 3,4) and the ionizing stars of the Trapezium. C I is found to be widespread with a typical intensity of about 6×10^{-6} erg s^{-1} cm^{-2} sr^{-1}and some enhancement at the interface between the ionized and neutral molecular medium with a peak intensity of 9.6×10^{-6} erg s^{-1} cm^{-2} sr^{-1}. A typical gas phase abundance of neutral carbon relative to hydrogen of $\approx 3 \times 10^{-6}$ is found, decreasing two orders of magnitude in the cores of the condensations.

Theoretical model calculations of ion chemistry of molecular clouds with external UV illumination were performed for this thesis, based on a code by Le Bourlot, which allow for an explanation of the observed relation of neutral carbon to CO and hydrogen on both, large and small, scales. It is suggested, based on the observational data presented and the good agreement of the observations with the ion chemistry model, that the C I ($1 \rightarrow 0$) emission emanates from the bulk of gas in molecular clouds, *i.e.*, from regions similar to those emitting in the low J ^{13}CO lines. This requires the bulk of the observed medium to be in the chemically low–density regime where charge transfer reactions of atomic species dominate over protonation reactions (started by H_3^+) of oxygen bearing molecules. The fine structure emission lines from neutral carbon are thus a more important coolant for the ISM than the rotational transitions of CO for low to moderate hydrogen densities. This is in contrast to previous PDR models that only had significant neutral carbon abundances in a thin PDR transition layer on the surface of molecular clouds irradiated by FUV.

Table of Contents

List of Figures

List of Tables

CHAPTER I.

INTRODUCTION TO SUBMM WAVELENGTH RECEIVERS AND ASTROPHYSICS

1.1 Astrophysical Motivation for Submm Receiver Development and Thesis Overview

The submillimeter wavelength band† (1 mm to 0.1 mm or equivalently 300 GHz to 3 THz) is currently at the frontier of radio astronomy. It is one of the last, mostly unexplored, spectral ranges in astronomy with important insights into many diverse astrophysical topics such as star formation and proto–planetary disks, galaxy formation, infrared luminous galaxies at cosmological distances, planetary atmospheres, and cosmo–chemistry. There are several reasons for the importance of submillimeter radio astronomy to these fields.

First, the photon energy of submillimeter waves expressed in units of temperature ($T = h\nu/k_B$) is about 10 to 100 K, which is well matched to the conditions in most interstellar molecular clouds, as shown in figure 1.1. This is important, for example, in star–formation, where molecular clouds collapse under their own gravitational field. The collapse would be stopped by the rising thermodynamic pressure due to adiabatic heating, if it were not for radiative processes carrying away thermal energy, thus allowing the collapse to continue to form a star.

Second, dust is a constituent of the interstellar medium (roughly 1% by mass) and its emission peaks in the mid–infrared (typically 10 to 100 μm). The dust emission opacity is usually very high in the mid–infrared, thus limiting the depth to which these objects can be studied and making mass determinations difficult or impossible. However, in the submillimeter the optical depths are low enough for radiation to penetrate through the cloud containing the dust but still high enough to allow the dust emission to be detectable.

---

† This band is between the infrared (IR) band and the millimeter wave band and when detector techniques typical for the infrared band are used, *e.g.*, bolometers, it is also often called the far infrared (FIR) band. However, when radio techniques steming from millimeter wave technological developments are used, the very same frequency range is referred to as the submillimeter band as will be done in this thesis.

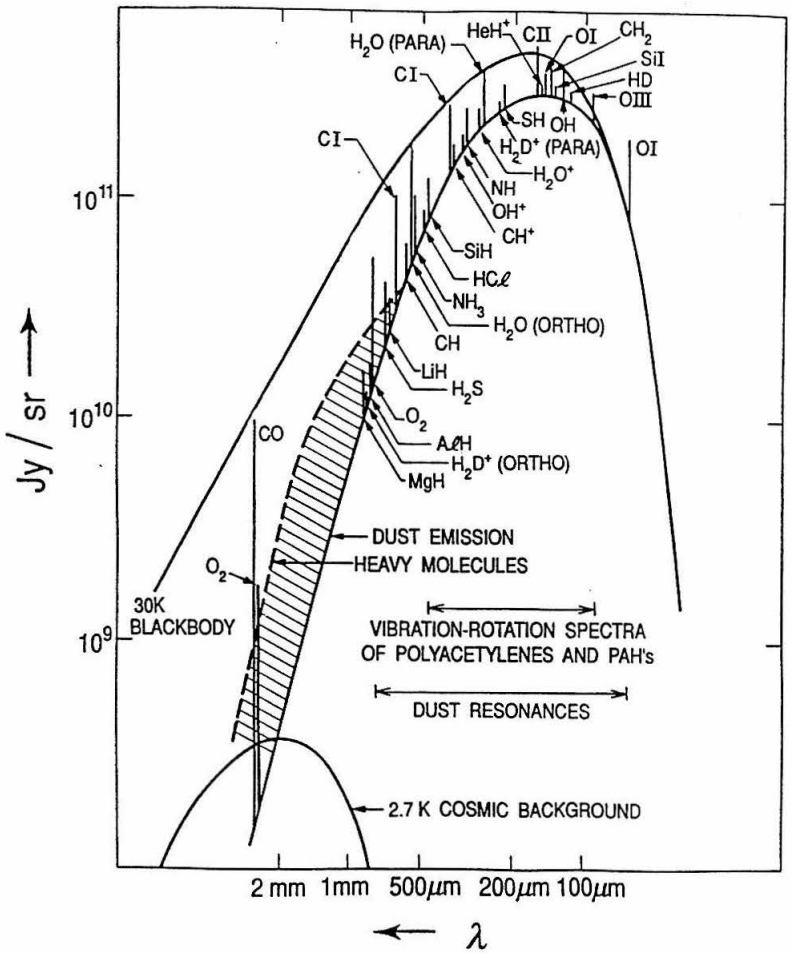


Figure 1.1 Synthesized emission spectrum of a typical interstellar cloud with a physical temperature of 30 K (from Phillips and Keene [1.1]). Rotational emission lines from heavy molecules (most notably CO) dominate the long wavelength range while light molecules (mostly hydrides) combined with atoms (Carbon and Oxygen) and dust dominate at shorter wavelengths.

For the field of cosmo–chemistry the submillimeter band is the most important frequency range since many interesting molecular species, such as hydrides, have a wealth of transitions in that band as can be seen in figure 1.2, which is a composite spectrum from 325 to 360 GHz towards the Orion molecular cloud. In addition to molecules, some atomic species, such as neutral and singly ionized carbon which are important coolants for the interstellar medium (ISM), have transitions in the submillimeter. Thus, spectroscopic studies of the interstellar medium in the submillimeter wavelength range should significantly contribute to the understanding of chemical processes in the ISM, which provides insight into very fundamental problems of chemistry since

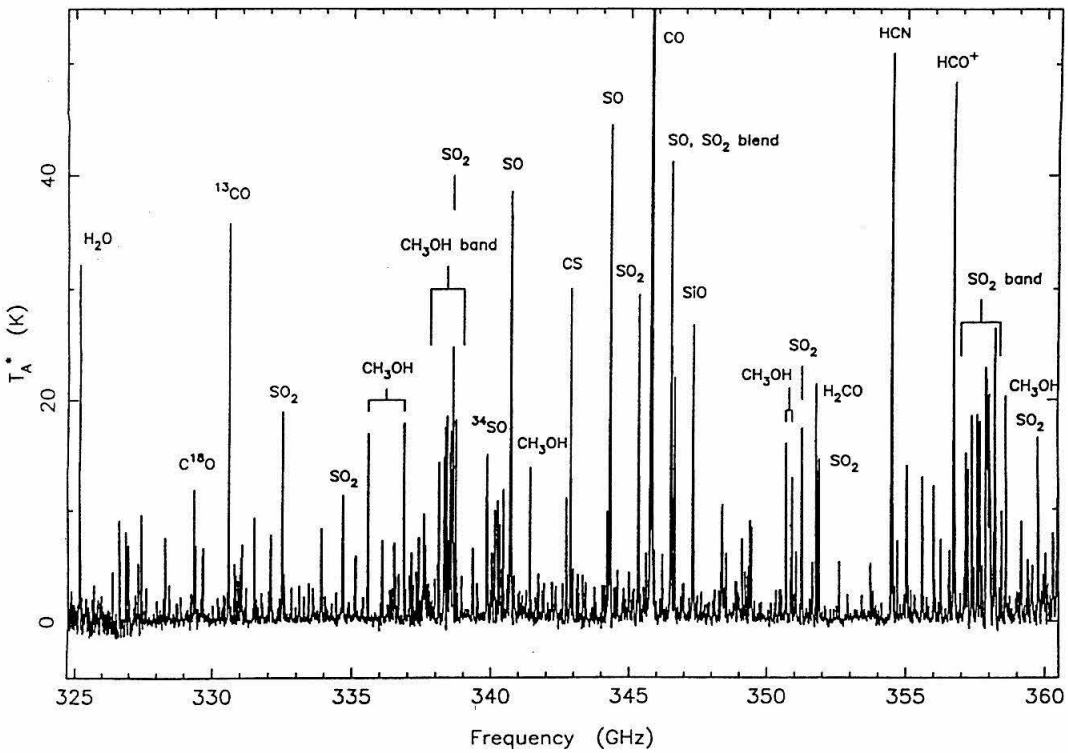


Figure 1.2 Submillimeter spectrum from 325 to 360 Ghz towards the Orion molecular cloud taken at the Caltech Submillimeter Observatory by Groesbeck *et al.* [1.2]. The spectrum shown is a composite of over 100 individual spectra showing the diversity of chemical species in molecular clouds.

the ISM provides for a very unique chemical laboratory not realizable on earth.

In addition to physical parameters like temperature and density of the gas and dust, spectroscopic techniques allow detailed information about radial velocities, chemical abundances, turbulence, magnetic fields etc. The spectroscopic techniques are best realized by using heterodyne receivers with virtually no limit on the spectroscopic resolution.

If the submillimeter band is important, why then has it not been until recently that its exploration has started? The main obstacles are of technological nature. To have good spatial resolution and high sensitivity to point–like sources, large aperture radio telescopes had to be constructed with sufficiently high surface accuracy to be useful in the submillimeter band and then erected at sites where the earth atmosphere is sufficiently transparent. Since water vapor in the earth atmosphere is mostly responsible for the absorption of submillimeter waves, high altitude sites are required. Only in 1987 did the first submillimeter wave radiotelescope, with an aperture of 10.4 m (Caltech Submillimeter Observatory, CSO), go into operation on Mauna

Kea, Hawaii at an altitude of 4200 m, followed shortly thereafter by the 15 m aperture James Clark Maxwell Telescope (JCMT) on the same site. Today there are several submillimeter radio telescopes operating worldwide and several new ones, including an interferometric array, are under construction. Since 1975 the Kuiper Airborne Observatory (KAO) has allowed observations at altitudes of up to 13.7 km with little atmospheric degradation in large windows in the submillimeter band (see figure 1.3), although many of the most important molecules, *e.g.*, H_2O and O_2, cannot be studied from even those altitudes. Additionally, the available observing time is short and the 0.9 m aperture of the telescope small. A future upgrade (Stratospheric Observatory for Infrared Astronomy, SOFIA) with a larger aperture (2.5m) is under investigation. The ideal platforms for submillimeter observations are, of course, space–borne since they allow for observations with no atmospheric degradation. One space–borne mission is currently being built by NASA with an 0.55m aperture (SWAS). SWAS will have no cryogenic capability thus less sensitive receivers and is meant as a limited exploratory mission to survey five lines, including one of H_2O and one of O_2, which have never been observed. It is scheduled for launch in 1996. An upgraded, cooled focal plane, 2.5m aperture mission (SMIM) is in the planning stages. ESA is planning a 4m aperture (FIRST) satellite with cooled receivers to be launched by the end of the decade.

Submillimeter telescopes need to be equipped with the very best receivers and spectrometers possible. Bolometric receivers are available by adapting techniques from the infrared band. However, the development of very sensitive receivers allowing for spectroscopic observations was – and at frequencies above 500 GHz (0.6 mm) still is – a mostly unsolved technological challenge. Heterodyne receivers, as used in the centimeter and millimeter wave bands, require significant changes to be applicable in the submillimeter wavelength band and thus development of new technologies and materials is needed. Furthermore, a technological push towards array receivers, to increase the effectiveness of telescopes when imaging astronomical sources, is imperative.

This thesis describes the development of a new generation of submillimeter wave receivers aimed at future integrated array receivers followed by a presentation of astronomical observations and their astrophysical interpretation for data taken with a single element of such a new receiver in the submillimeter wavelength band.

The technological development presented in this thesis was driven by the need to develop

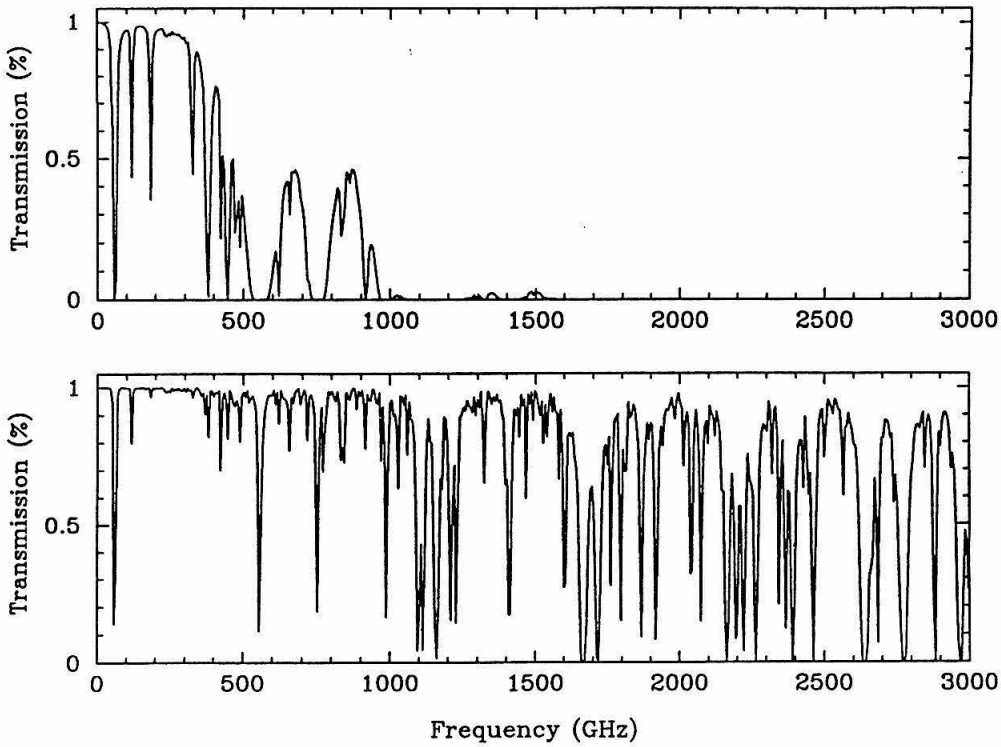


Figure 1.3 Atmospheric transmission comparison for (top) a very high ground based site such as Mauna Kea (4200m) with a precipitable water vapor of 1mm and (bottom) from an airborne platform like the Kuiper Airborne Observatory at 12 km altitude (from Grossman [1.3]).

receivers based on superconductor–insulator–superconductor (SIS) detectors that are suitable to be integrated into the focal plane of the telescope to form array receivers. SIS diodes are favorable as detectors since they have proved to be the most sensitive detectors in the millimeter and low submillimeter band. The route chosen in this thesis is based on work done by Rutledge *et al.* [1.4] and Wengler *et al.* [1.5], who used planar antennas with a detector integrated in the apex of the planar antenna, which is quasi–optically coupled to the telescope. The term "Quasi–optical" receiver thus refers to receivers using optical–style lenses and planar antennas to couple the radiation into the detector rather than waveguide structures as used in conventional microwave receivers. Two major problems had to be overcome to develop this technology to sufficient maturity that array receivers could be designed. First, the beam of an individual element, launched by the planar antenna, had to be of high quality – comparable to that of a waveguide horn antenna – to provide good coupling to a telescope. Second, the impedance of the SIS detector had to be matched to the antenna's impedance over the operating range of the receiver, since quasi–

optical receivers, unlike waveguide receivers, have no mechanically adjustable tuners. Serving as an example of a typical single element quasi–optical receiver, the design and performance of a quasi–optical receiver based on a planar–logarithmic–spiral antenna is shown below in this chapter. This receiver was the initial design as part of this thesis project on which the new developments discussed in the following chapters are based. It was the first submillimeter SIS receiver ever to be successfully used for astronomical observations. Preceding this example of a quasi–optical receiver is a more detailed discussion of heterodyne receiver principles, the physics of SIS detectors based on the semiconductor picture, and the quantum theory of mixing with SIS detectors. Chapter II will describe the development of a new antenna, called the hybrid antenna, which changes the optical system of the receiver significantly and allows this antenna to be used as a single element of an integrated focal plane array receiver yielding high aperture efficiencies and excellent beam patterns. A broadband impedance matching structure, based on a novel superconducting transmission line circuit, is introduced and discussed in chapter III. Theoretical predictions and comparison with direct and heterodyne measurements are presented for this circuit. Chapter IV concludes this thesis with an application of a quasi–optical receiver in radio astronomy. Observations and astrophysical interpretation of the small scale structure of neutral carbon emission (C I) in Orion and the first extragalactic detection of neutral carbon at a frequency of 492 GHz (0.61 mm) in the galaxy IC 342 are shown. The data is interpreted in the framework of a photo–dissociation region (PDR) model combined with a cosmo–chemical model based on ion chemistry.

1.2 Heterodyne Receiver Principles

In radio astronomy, and many other fields, the radiation received by a radio antenna is too high in frequency and too weak in signal strength to be analysed directly by devices such as spectrometers. A heterodyne receiver is an electronic device that down–converts the signal from the radio frequency (RF) to a much lower intermediate frequency (IF) where appropriate instru-

mentation for direct analysis of the signals is technologically feasible. The analysis includes the strength of the radiation as a function of frequency, *i.e.*, spectroscopy, or the phase of the radiation for use in interferometric measurements, or both simultaneously. The heart of a heterodyne receiver is the mixer, *i.e.*, a device that performs the frequency down–conversion without losing amplitude and phase information, thus also preserving frequency information. Figure 1.4 shows the schematic for a typical heterodyne receiver. For signal frequencies above about 100 GHz the amplifier in front of the mixer is not feasible since the technology to build such amplifiers does not exist currently. This makes the mixer the first element that processes the very weak signal. The next section will show why this makes it the most critical component for signal to noise considerations in a millimeter or submillimeter heterodyne receiver.

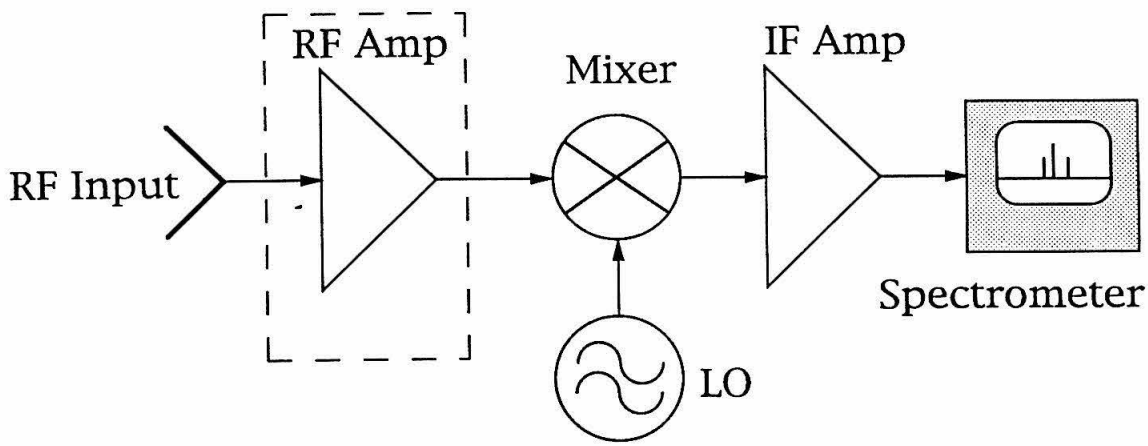


Figure 1.4 Schematic view of a heterodyne receiver. The RF amplifier in front of the mixer is currently technologically not feasible for frequencies above 100 GHz. Thus the mixer element is the first to process the weak received signal.

a) Operating principles of a mixer

As stated above, the mixer is the central device of a heterodyne receiver since it down–converts the signal frequency to a much lower intermediate frequency without losing amplitude and phase information. A mixer requires a reference signal relatively close in frequency to that of the signal and an electronic device performing the mixing of the signal with the reference signal to yield the down–converted signal at the intermediate frequency. The reference frequency signal is monochromatic and much stronger than the signal. It is generated within the receiver and thus

called the local oscillator (LO) signal. The electronic device most commonly used in a mixer is a diode with a non–linear current to voltage characteristic. Two types of diodes are frequently used in radio astronomical receivers: Schottky diodes, which are semiconductor devices, and Superconductor–Insulator–Superconductor (SIS) diodes. The work presented in this thesis is solely based on SIS devices and they will be discussed in more detail in the following sections. However, for the classical considerations following now the only relevant characteristic is the non–linearity of the current–voltage characteristic of any device used in a mixer. If such a device is supplied with a dc bias voltage V_b and a small time varying voltage $V(t)$ the current through the device can be expanded in a Taylor series :

$$I(V_b + V(t)) = I(V_b) + \frac{dI(V_b)}{dV}V(t) + \frac{d^2I(V_b)}{2dV^2}V^2(t) + \ldots$$

$$\equiv I(V_b) + K_1V(t) + K_2V^2(t) + \ldots \qquad (1.1)$$

Let the time varying voltage be the sum of the local oscillator signal with amplitude A and frequency ω_{LO} and the signal frequency with amplitude B and frequency ω_S,

$$V(t) = A\sin(\omega_{LO}t) + B\sin(\omega_S t) \quad , \qquad (1.2)$$

then

$$\begin{aligned} I(t) = {} & I(V_b) + K_1(A\sin(\omega_{LO}t) + B\sin(\omega_S t)) && \text{[original signal]} \\ & + K_2\left(\frac{A^2}{2} + \frac{B^2}{2}\right) && \text{[rectified signal]} \\ & - K_2\left(\frac{A^2}{2}\cos(2\omega_{LO}t) + \frac{B^2}{2}\cos(2\omega_S t)\right) && \text{[first harmonic]} \\ & + K_2\,A\,B\cos[(\omega_{LO} - \omega_S)t] && \text{[difference frequency]} \\ & - K_2\,A\,B\cos[(\omega_{LO} + \omega_S)t] && \text{[sum frequency]} \\ & + \ldots \quad . && (1.3) \end{aligned}$$

Thus the current contains the original signal, rectified signal, harmonics of the local oscillator and signal, sum and difference of the local oscillator frequency and signal frequency. In general the current spectrum contains frequencies f

$$f = |mf_{LO} \pm nf_S| \quad m, n = 0, 1, 2, \ldots \qquad (1.4)$$

The rectified current can be used to monitor the RF power seen by the diode, in order to avoid saturation or destruction. The difference frequency signal, or intermediate frequency (IF), is easily separated out from all the other frequencies, since the LO frequency is very close to the signal frequency and therefore the IF is very low. The receiver will, of course, down–convert not just one single frequency but rather a frequency band, which is defined by the bandwidth of the IF circuit. The resulting IF bandwidth is sometimes called the receiver's instantaneous bandwidth when compared to bolometric receivers. The amplitude and frequency of the local oscillator is kept constant, thus the IF amplitude at a certain frequency is proportional to the signal amplitude with the frequencies related through (1.4). However, there is an ambiguity apparent from (1.4), which is that the IF contains signals with frequencies $f_{IF} = f_{LO} - f_S$ and $f_{IF} = f_S - f_{LO}$. These two possible signal frequencies are the two sidebands of a receiver called the lower sideband and the upper sideband,† respectively. Figure 1.5 shows the relation of the two sidebands, the local oscillator frequency and the IF band. A receiver that allows both sidebands to be present in the IF band is called a double sideband (DSB) receiver, whereas a receiver that employs some means of blocking one sideband before the mixer is a single sideband (SSB) receiver.

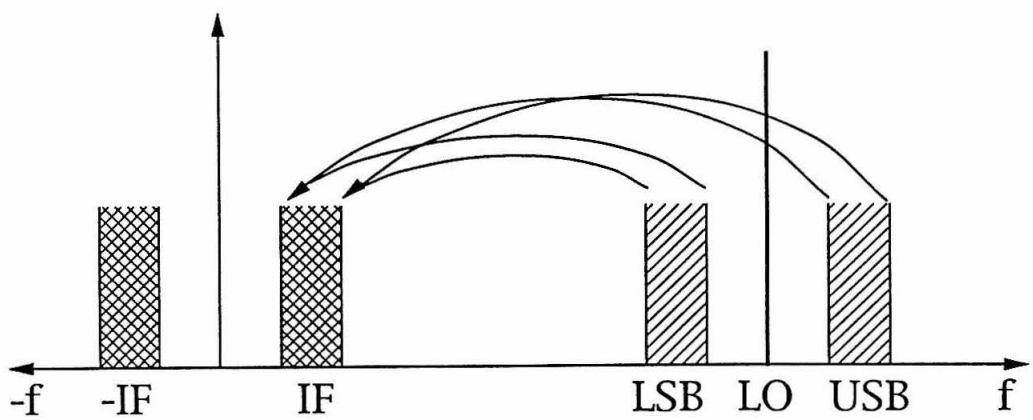


Figure 1.5 Down–conversion process of a heterodyne receiver from the upper and lower sideband into the intermediate frequency band. The negative frequency IF band cannot be distinguished from the positive frequency band in the simple circuit of figure 1.4.

† If the signal under investigation, *e.g.*, an emission line from a molecule, is in one particular sideband this sideband is also referred to as the signal sideband. The other sideband is then called the image sideband.

b) Sensitivity determination of a heterodyne receiver

An important figure of merit for any receiver system is its sensitivity to the signal band, *i.e.*, how much noise is added to the signal by the receiver. In millimeter and submillimeter radio astronomy the strength of a signal is typically quoted in units of brightness or radiation temperature and it is thus useful to quote the sensitivity of a receiver in the same units. The standard technique used to measure the sensitivity – or equivalently noise temperature T_r – of a receiver involves black body radiation of two sources of known temperature, subsequently referred to as a hot and cold load. In the Rayleigh–Jeans approximation ($h\nu \ll k_B T$), the power P from a black body at temperature T emitted in a frequency band of width B is given by

$$P = B\ k_B\ T_{\text{blackbody}}\ , \tag{1.5}$$

where k_B is the Boltzmann constant. The power in the receiver's IF band is the sum of power from the black body source and the receiver's own noise contribution at equivalent temperature T_r, referenced to the input of the receiver, multiplied by the gain of the receiver, G,

$$P^{\text{IF}} = G\ B\ k_B\ (T_r + T_{\text{blackbody}})\ . \tag{1.6}$$

Taking two measurements, one with a hot and one with a cold load at the input of the receiver, yields the ratio

$$Y = \frac{P^{\text{IF}}_{\text{hot}}}{P^{\text{IF}}_{\text{cold}}} = \frac{T_r + T_{\text{hot}}}{T_r + T_{\text{cold}}} \tag{1.7}$$

of the power in the IF band. The receiver's noise temperature can then be calculated from (1.7),

$$T_r = \frac{T_{\text{hot}} - Y\ T_{\text{cold}}}{Y - 1}\ . \tag{1.8}$$

All noise temperature measurements quoted in this thesis were performed using this technique with hot and cold loads at room temperature (295 K) and liquid nitrogen temperature (78 K), respectively.

c) Noise contribution of individual receiver stages to the total noise

The noise temperature of a receiver describes the sensitivity of the complete receiver system. However, sometimes it is interesting to investigate the noise contribution of individual stages

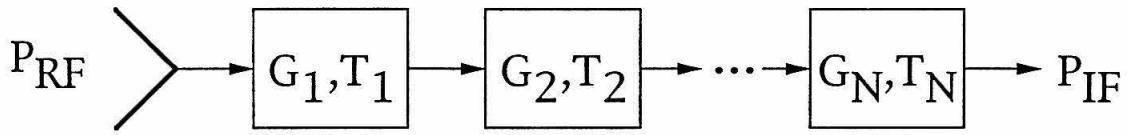


Figure 1.6 Chain of components of a receiver system each component having its own noise temperature T_i and gain G_i.

within the receiver to the total noise temperature. Figure 1.6 shows a chain of components each with its own noise temperature T_i and gain G_i (or loss, *i.e.*, a gain of less than 1).

From (1.5) the total IF power for N stages then is

$$P_{IF} = k_B B(G_1G_2G_3...G_NT_1 + G_2G_3...G_NT_2 + ... + G_NT_N) \quad , \tag{1.9}$$

which can be rewritten for the total receiver noise temperature

$$T_r = T_1 + \frac{T_2}{G_1} + \frac{T_3}{G_1G_2} + ... + \frac{T_N}{G_1...G_{N-1}} \quad . \tag{1.10}$$

This shows that the first stage is the most important stage if all stages have a gain larger than unity. For receivers that have no amplifier before the mixer, the mixer's noise contribution is thus most important. Additionally, since mixers typically have no gain (and often considerable loss), the amplifier following the mixer is also critical, since its noise contribution increases as the loss of the mixer increases. Recall from equations (1.1) and (1.3) that the mixer's conversion of the RF to the IF band is proportional to $K_2 = \frac{d^2I(V_b)}{2dV^2}$, which implies that the non–linear device utilized in the mixer should have as strong a non–linearity in its current–voltage characteristic as possible to maximize the gain (or minimize the loss) of the mixer. SIS diodes have very strong non–linearities, which is one of the reasons why they have become the most successful devices for millimeter and low submillimeter frequencies. The physics of SIS devices will be discussed in the following section.

d) Practical and Theoretical Limitations of Mixers

Typically, the main design goal of a mixer is to make it as sensitive to the signal as possible. Due to the quantization of the electro–magnetic radiation field the Heisenberg uncertainty principle dictates that there is a fundamental limitation to a simultaneous measurement of amplitude and phase of a signal. A mixer linearly converts a high frequency signal to a lower frequency signal, preserving amplitude and phase information. General linear amplifier theories, such as Caves' [1.6], predict a quantum limited noise temperature for a single sideband (SSB) mixer of $T_M \geq h\nu/k_B$.

However, in practice it is hard to achieve quantum limited performance for several reasons. The mixing element itself is typically a tunneling device, such as an SIS or Schottky barrier diode, which introduces shot noise due to the leakage current of the diode. Any lossy element with a gain of less than unity, *i.e.*, loss $L = G^{-1}$, in the path of the signal to the mixer will not only reduce the amount of signal but also introduce noise. The lossy element acts as a black body at physical temperature T and emissivity $\epsilon = 1 - G$.

In addition to the mixer's noise temperature T_M, its conversion gain or loss G_M, will determine the overall performance of the receiver as shown above by equation (1.10). It was seen that the conversion gain depends on the non–linearity of the mixer's diode. However, the diode's impedance may not be matched to the incoming signal, resulting in signal power being reflected. This reduces the conversion gain further. Tunnel barrier diodes have a capacitance C that, if not tuned by an inductive element, can reduce the conversion gain significantly. The higher the frequency of operation the more critical that capacitance will be. A figure of merit for diodes used in mixers is thus the ratio of the signal frequency $f = \omega/2\pi$ to roll–off frequency $1/2\pi RC$, where R is the real part of the diode's impedance. When this ratio, $\omega\, RC$,† is less than unity, the device is sufficiently matched, while for $\omega\, RC > 1$ it will need some means of inductive tuning. SIS diodes can have moderately high $\omega\, RC$ products in the millimeter wavelengths range, increasing to high $\omega\, RC$ products in the submillimeter wavelengths range, where tuning circuits are required. A design of such a tuning circuit will be presented in chapter III.

† traditionally called the "$\omega\, RC$ product", even though it is a ratio of two frequencies

1.3 SIS Detectors in the Semiconductor–Picture

A Superconducting–Insulator–Superconducting (SIS) diode† is a sandwich of two superconductors separated by a very thin insulator acting as a tunneling barrier. The origin of the current–voltage characteristic and the mechanism by which radiation is detected in an SIS diode is best understood in the semiconductor–picture.

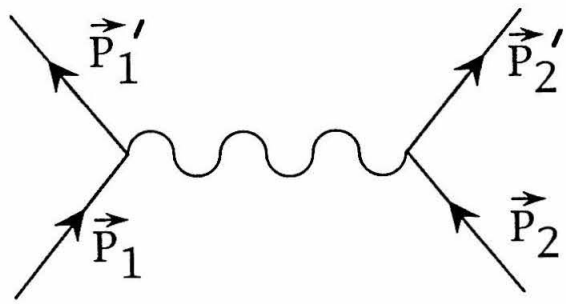


Figure 1.7 Feynman diagram of the exchange of a phonon between two electrons in a superconductor forming a Cooper pair. The total momentum is conserved.

In 1957 Bardeen, Cooper, and Schriefer [1.7] published a microscopic theory of superconductivity referred to as the BCS theory. Free electrons in a metal feel the repulsive Coulomb force between them. The BCS theory adds an attractive force between two electrons, based on the exchange of a phonon, thus leading to pair formation. Such a bound pair of electrons is called a Cooper pair. The exchanged phonon is a quantum of thermal energy in the lattice vibration and its exchange conserves the total momentum of the system as illustrated in figure 1.7. The average distance between the two electrons where the repulsive forces equal the attractive forces is the coherence length ξ_{Cooper} of the Cooper pair, which is typically ξ_{Cooper} = 5 to 100 nm for conventional superconductors like Pb, Nb or NbN, and less than 1 nm for high–T_c ones, like YBaCuO. The volume between the two electrons in a Cooper pair contains about 10^6 to 10^7 conducting

† SIS actually describes the structure of the device rather than the physical effect in use for the device. Tunneling of both single quasi–particles and Cooper pairs takes place in the device, but when the single quasi–particle is in use the nomenclature SIS is used. When Cooper pairs are active it is called a Josephson junction.

electrons of which a fraction are also correlated in Cooper pairs. The partners in a Cooper pair have opposite spin and opposite momenta so that a Cooper pair can be described by $(\vec{p}\uparrow, -\vec{p}\downarrow)$, or with $\vec{p} = \hbar\vec{k}$ as $(\vec{k}\uparrow, -\vec{k}\downarrow)$. However, this description only holds for a no–current condition so that the total momentum is zero. The most important effect of the pair formation is that the pair as a unit acts as a single Boson. All Cooper pairs can therefore be in the same state, *i.e.*, they do not have to obey the Pauli exclusion principle for Fermions. The quantum–mechanical wave function of all Cooper pairs in one superconductor is therefore simply

$$\Psi(\vec{r}) = |\Psi(\vec{r})|e^{i\theta(\vec{r})} \quad , \tag{1.11}$$

i.e., the phase coherence extends through the whole superconductor.

In addition to Cooper pairs there are still normally conducting electrons left so that a band picture of a superconductor needs to include both species. In a normal metal the energy as a function of wave number above the conduction band edge is roughly quadratic since the electrons can be treated as free particles with kinetic energy $\frac{\hbar^2}{2m}k^2$. The energy it takes to move a conducting electron with energy E_1 above the Fermi energy to an energy state E_2 is

$$\delta E = E_2 - E_1 = E_2 - E_{\text{Fermi}} + (E_{\text{Fermi}} - E_1) \quad , \tag{1.12}$$

i.e., it takes $\hat{E}_1 \equiv E_{\text{Fermi}} - E_1$ to create a hole excitation and $\hat{E}_2 \equiv E_2 - E_{\text{Fermi}}$ to create an electron excitation. $\hat{E}_1$ and $\hat{E}_2$ are called reduced energies. Figure 1.8a shows an $\hat{E}$–k diagram of reduced energies versus momenta. The important feature here is that the depicted graph touches the zero line for reduced energies, *i.e.*, an arbitrary small amount of energy is sufficient to create a hole and electron like excitation. In a superconductor there is a binding energy Δ for each electron in a Cooper pair, which modifies the $\hat{E}$–k diagram to that depicted in figure 1.8b. The binding energy per electron Δ prevents arbitrary small energies from creating excitations, *i.e.*, the metal now has a bandgap–like structure with the gap energy 2Δ, typically between 1 and 10 meV. To create an electron like excitation thus requires a minimum energy of 2Δ, in analogy to the band gap of a semiconductor.

a) The DC current–voltage characteristic of an SIS diode

To calculate the tunneling probability across the barrier in an SIS diode one needs to know the density of states as a function of energy in the two superconductors. Assuming that the density of

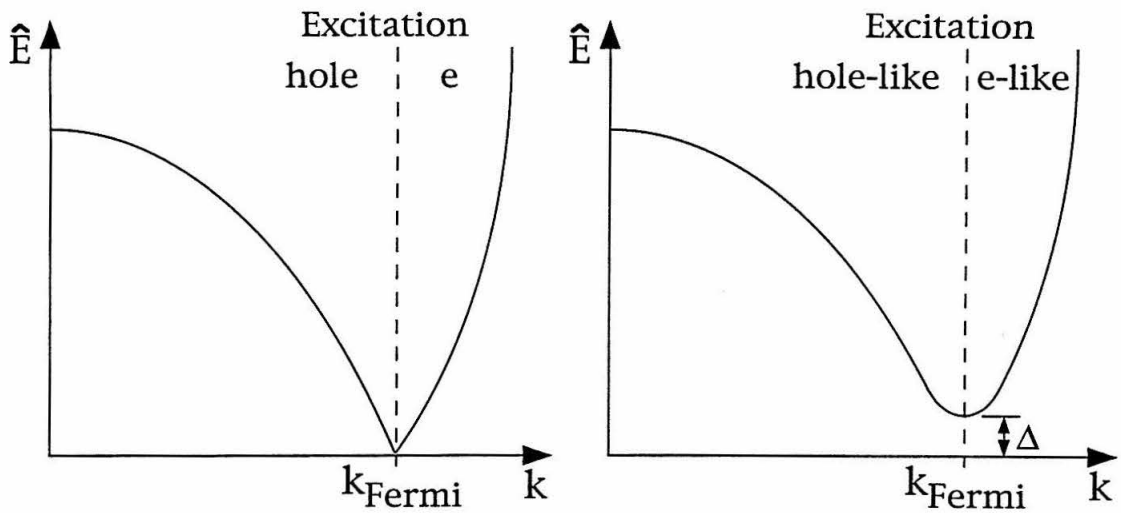


Figure 1.8 Reduced energies as a function of wavenumbers. a) Reduced electron and hole excitation energy of a normal metal. b) Reduced electron–like and hole–like excitations for a superconductor with the binding energy Δ of an electron in a Cooper pair.

states as a function of $\vec{k}$ does not change significantly when transitioning from the normal metal state above the superconductor's critical temperature T_c to below T_c, *i.e.*,

$$\left(\frac{dn}{dk}\right) dk = \text{const.} = \frac{dn}{d\hat{E}_{\text{normal}}} \frac{d\hat{E}_{\text{normal}}}{dk} dk = \frac{dn}{d\hat{E}_{\text{superc.}}} \frac{d\hat{E}_{\text{superc.}}}{dk} dk \ , \tag{1.13}$$

so that

$$\frac{dn}{d\hat{E}_{\text{superc.}}} = \frac{dn}{d\hat{E}_{\text{normal}}} \frac{d\hat{E}_{\text{normal}}}{d\hat{E}_{\text{superc.}}} \ . \tag{1.14}$$

From BCS theory (also compare figure 1.8a and 1.8b) the reduced energies available for quasi–particle excitations are

$$\hat{E}_{\text{superc.}} = \sqrt{\hat{E}^2_{\text{normal}} + \Delta^2} \ , \tag{1.15}$$

which yields for (1.14)

$$\frac{dn}{d\hat{E}_{\text{superc.}}} \simeq \left(\frac{dn}{d\hat{E}_{\text{normal}}}\right)_{\hat{E}_{\text{normal}}=0} \frac{\hat{E}_{\text{superc.}}}{\sqrt{\hat{E}^2_{\text{superc.}} - \Delta^2}} \ . \tag{1.16}$$

There are no states for quasi–particles allowed for $|\hat{E}_{\text{superc.}}| < \Delta$. Figure 1.9 shows the density of states according to (1.16).

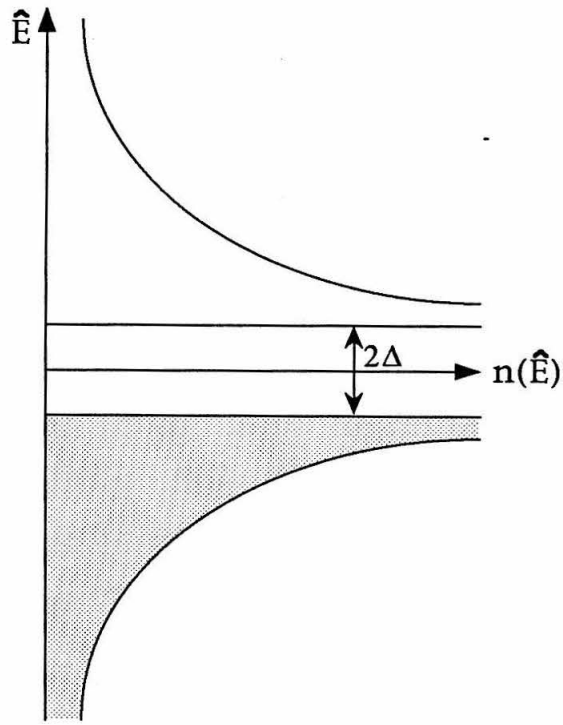


Figure 1.9 Density of states $n(\hat{E})$ of a superconductor as a function of reduced energy $\hat{E}$. The size of the energy gap is 2Δ, which is the minimum energy it takes to create an excitation from the ground state ($T = 0$). The shaded region indicates that the states are filled with electron like quasi–particles.

In an SIS diode two superconductors are brought together close enough so that the electronic quantum–mechanical wave functions overlap, *i.e.*, quasi–particles can tunnel from one superconductor to the other. The tunneling barrier in a typical SIS diode is only a few nanometers thick yielding a potential barrier of less than 1 eV. Figure 1.10a shows the density of states as a function of reduced energy for the two superconductors with an applied bias voltage across the junction of $V_b < 2\Delta/e$.

The current–voltage characteristic of an SIS diode follows from the energy band model. Since energy has to be conserved, no quasi–particle can tunnel from one superconductor to the other until the bias voltage applied across the junction reaches the value $V_b = 2\Delta/e$ at which point there will be a very sudden onset of current. The exact shape of the I–V curve, shown in figure 1.10b, can be calculated from an analysis of the tunneling probabilities. At a temperature $T > 0$

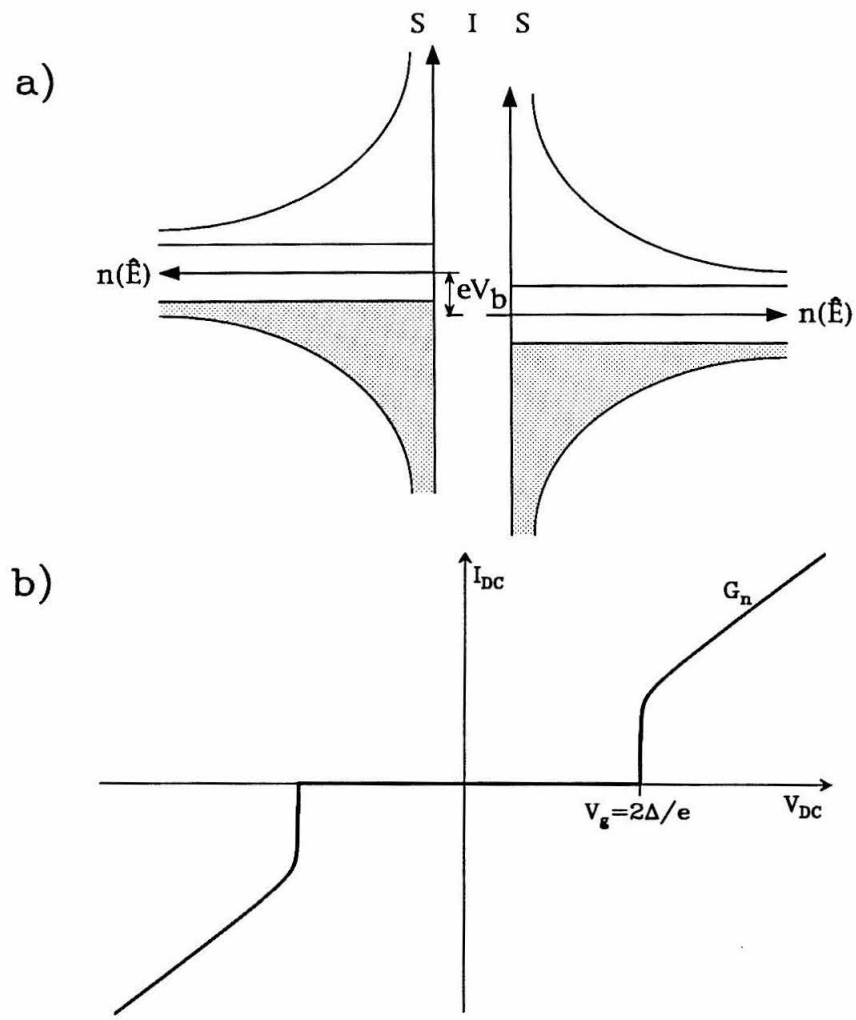


Figure 1.10 a) Density of states $n(\hat{E})$ as a function of reduced energy $\hat{E}$ of an SIS diode, *i.e.*, two superconductors in close proximity. A voltage V_b has been applied moving the energy band structure of one superconductor with respect to the other. b) The current voltage characteristic of an SIS diode based on numerical integration of equation (1.22) for the ground state ($T = 0$). The slope above $V_{\text{gap}} = 2\Delta/e$ is the normal state conductance G_n of the SIS diode.

the probability of exited states is described by the Fermi function $f(\hat{E})$

$$f(\hat{E}) = \frac{1}{1 + e^{\frac{\hat{E}}{k_B T}}} , \tag{1.17}$$

where $\hat{E}$ is the reduced energy of the state. The number of electrons per energy interval in the right–hand superconductor is

$$\frac{dN_{eR}}{d\hat{E}} = \frac{dn_{\text{R}}(\hat{E})}{d\hat{E}} f(\hat{E}) , \tag{1.18}$$

i.e., the density of states times the probability that the states are filled. The right–hand superconductor is assumed grounded while the left–hand one has an additional potential energy of eV_b. The number of holes per energy interval in the left–hand superconductor is

$$\frac{dN_{hL}}{d\hat{E}} = \frac{dn_L(\hat{E} - eV_b)}{d\hat{E}} \left[1 - f(\hat{E} - eV_b)\right] , \tag{1.19}$$

i.e., the density of states at reduced energy $\hat{E} - eV_b$ times the probability that the states are not filled. Integrating over all possible energies then yields the current flow from the right to left

$$I_{RL}(V_b) = -e\ C \int_{-\infty}^{\infty} \frac{dn_R(\hat{E})}{d\hat{E}}\ f(\hat{E}) \frac{dn_L(\hat{E} - eV_b)}{d\hat{E}} \left[1 - f(\hat{E} - eV_b)\right]\ d\hat{E} \tag{1.20}$$

with the constant C depending on the characteristics of the tunnel barrier. Similarly the current flow from the left to right is

$$I_{LR}(V_b) = -e\ C \int_{-\infty}^{\infty} \frac{dn_L(\hat{E} - eV_b)}{d\hat{E}}\ f(\hat{E} - V_b) \frac{dn_R(\hat{E})}{d\hat{E}} \left[1 - f(\hat{E})\right]\ d\hat{E}\ . \tag{1.21}$$

The resulting current I_{DC} then is the difference of the two currents with $\frac{dn_R(\hat{E})}{d\hat{E}} = \frac{dn_L(\hat{E})}{d\hat{E}} \equiv \frac{dn(\hat{E})}{d\hat{E}}$

$$I_{DC}(V_b) = e\ C \int_{-\infty}^{\infty} \frac{dn(\hat{E} - eV_b)}{d\hat{E}}\ \frac{dn(\hat{E})}{d\hat{E}} \left[\ f(\hat{E} - V_b) - f(\hat{E})\right]\ d\hat{E}\ . \tag{1.22}$$

Using (1.16) for the density of states the DC current–voltage characteristic can then be calculated by numeric integration of (1.22). The results for the system in the ground state ($T = 0$) are shown in figure 1.10b. The calculated current–voltage characteristic looks very similar to measured ones (see figures 1.17 and 3.4), except for a strong current at zero bias voltage seen in the measurements. This current is due to tunneling of Cooper pairs, also called Josephson currents [1.8], and is not part of the theoretical treatment presented here.

b) Photon assisted tunneling in an SIS diode

The sharp onset of current at a bias voltage of $V_b = V_{gap} \equiv 2\Delta/e$ is due to the high density of states of electrons in the right superconductor matching the high density of holes in the left superconductor. Below that bias voltage no current can flow (ignoring thermal excitations). However, if radiation of frequency ω is applied to the junction photons of energy $\hbar\omega$ can be absorbed by an electron like quasi–particle. Tunneling is energetically then possible if $V_b + \hbar\omega/e \geq V_{gap}$. The absorption of more than one photon will also occur, but with decreasing probability as

the number of absorbed photons per electron increases. These photon assisted tunneling events lead to increases in current, *i.e.*, current steps, below the gap voltage when $V_b + n\hbar\omega/e = V_{gap}$. This can be seen in Figure 1.11, which has been calculated using an equation derived in the next section.

The strong non–linearity of SIS's current voltage characteristic on the scale of a photon energy in the millimeter and submillimeter frequency range leads to several advantages over Schottky diodes when used in a mixer. As discussed earlier the conversion gain will increase as the non–linearity increases. Furthermore, the fact that this non–linearity is on the scale of the photon energy has two additional advantages. First, the required power levels for the local oscillator in a mixer can be much lower than those in a Schottky diode mixer. This can be important at high frequencies, where the generation of LO power is difficult. Second, the onset of quantum mechanical effects, such as gain in a mixer rather than conversion loss as classical mixers have, requires that the non–linearities be strong on the scale of the photon energy. These effects will be discussed in more detail in the following section.

1.4 Quantum Theory of Mixing with SIS Devices

SIS detectors exhibit quantum mechanical phenomena that dominate their characteristics when used in a mixer. It is therefore necessary to treat SIS detectors quantum mechanically if one wants to understand their behavior and predict their performance. The mathematical derivations presented here are based on the 1979 article by Tucker [1.9] and work by Tien and Gordon [1.10]. A useful review article is that by Tucker and Feldman [1.11]. The full quantum mechanical treatment of Tucker for photon assisted quasi–particle tunneling is of considerable complexity and an abbreviated version has been adopted here allowing the same physical insight for understanding the performance of an SIS mixer.

a) Current–voltage characteristic of an SIS diode with applied RF power

As in the previous section, the right–hand electrode of the SIS diode is assumed to be grounded

so that the reduced energies of this electrode's levels are simply $\hat{E}_i$. The Schrödinger equation,

$$\frac{\partial \psi_i^R(\vec{r},t)}{\partial t} = \frac{-i}{\hbar}\hat{E}_i\ \psi_i^R(\vec{r},t) \tag{1.23}$$

for the right–hand superconductors energy states can be separated and solved for the time dependent part

$$\begin{aligned}\psi_i^R(\vec{r},t) &= \psi_i^R(\vec{r})\ e^{\frac{-i}{\hbar}\int_0^t \hat{E}_i\ dt'} \\ &= \psi_i^R(\vec{r})\ e^{\frac{-i}{\hbar}\hat{E}_i\ t}\quad .\end{aligned} \tag{1.24}$$

The left–hand side superconductor has a time dependent potential $eV(t)$ relative to the right hand superconductor composed of the bias voltage V_b and the local oscillator voltage V_ω,

$$eV(t) = eV_b + eV_\omega \cos(\omega t)\quad , \tag{1.25}$$

which yields for the left hand side wave function

$$\begin{aligned}\psi_i^L(\vec{r},t) &= \psi_i^L(\vec{r})\ e^{\frac{-i}{\hbar}\int_0^t [\hat{E}_i\ + eV_b + eV_\omega \cos(\omega t')]\ dt'} \\ &= \psi_i^L(\vec{r})\ e^{\frac{-i\,t}{\hbar}[\hat{E}_i\ +\ eV_b]}\ e^{\frac{-i}{\hbar\omega}eV_\omega \sin(\omega t)}\quad .\end{aligned} \tag{1.26}$$

Using the relation (Abramovitz and Stegun [1.12])

$$e^{-ix\sin(\omega t)} = \sum_{n=-\infty}^{\infty} J_n(x)\ e^{-in\omega t} \tag{1.27}$$

$$\text{yields}\qquad \psi_i^L(\vec{r},t) = \psi_i^L(\vec{r})\ e^{\frac{-i\,t}{\hbar}[\hat{E}_i\ + eV_b]}\ \sum_{n=-\infty}^{\infty} J_n\left(\frac{eV_\omega}{\hbar\omega}\right)\ e^{-in\omega t}\quad . \tag{1.28}$$

Comparing (1.28) with (1.24) allows for the interpretation of the left–hand superconductor having additional virtual states with energies $n\,\hbar\omega$ at probability amplitudes $J_n\left(\frac{eV_\omega}{\hbar\omega}\right)$. Note that the formalism is similar to the analysis of a frequency modulation of a high frequency carrier signal, *i.e.*, FM radio.

The virtual shifts in energy levels are equivalent to shifts of the DC voltage to $V_b + n\hbar\omega/e$ with a probability of $J_n^2\left(\frac{eV_\omega}{\hbar\omega}\right)$. Using the current voltage characteristic without an RF signal

$I_{DC}(V_b)$ as derived in (1.22) then yields the DC current $I_b(V_b, V_\omega)$ with an RF signal voltage V_ω at frequency ω applied,

$$I_b(V_b, V_\omega) = \sum_{n=-\infty}^{\infty} J_n^2\left(\frac{eV_\omega}{\hbar\omega}\right) I_{DC}\left(V_b + \frac{n\hbar\omega}{e}\right) \quad . \tag{1.29}$$

The probability weights $J_n^2\left(\frac{eV_\omega}{\hbar\omega}\right)$ depend on the rate of photons absorbed by the SIS junction and for a rate of zero the current is just $I_b(V_b, V_\omega = 0) = I_{DC}(V_b)$. Figure 1.11 shows an I–V characteristic of an SIS diode with different amounts of LO power applied. Two steps are visible ($n = 1$ and $n = 2$), which correspond to one and two photons absorbed by each tunneling quasi–particle. The presence of these quasi–particle photon steps was first discovered by Dayem and Martin [1.13] in 1962.

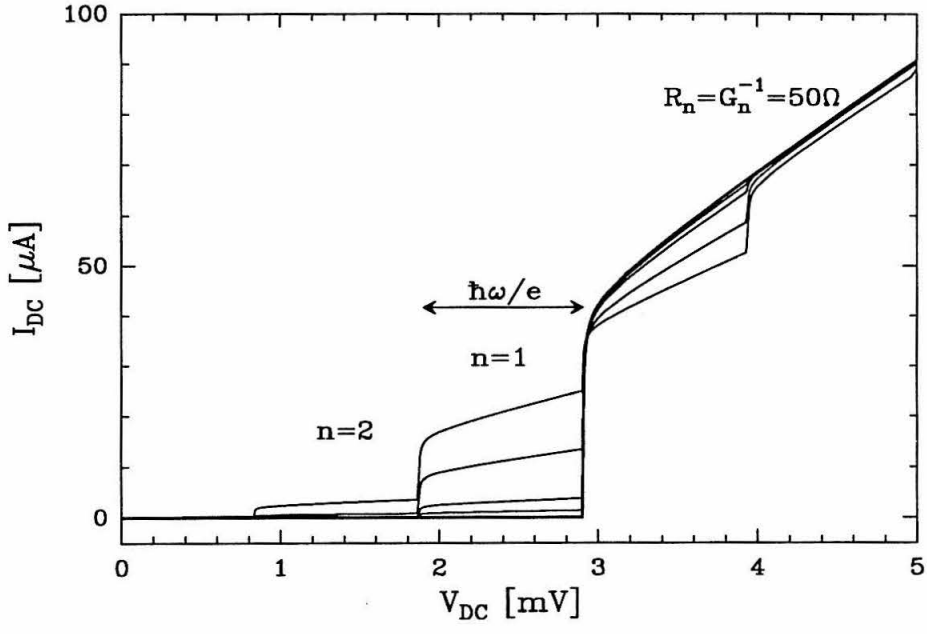


Figure 1.11 Current–voltage characteristic of an SIS diode with different LO power levels applied as calculated from equations (1.22) and (1.29). The LO frequency is 500 GHz and the SIS junction has a gap voltage of $V_{gap} = 2.9mV$, like NB/AlO_x/Nb SIS junctions. The heavy drawn line is for no LO power and the lines above it are for P_{LO} =7, 10, 20 and 30 mW, respectively.

b) Assumptions for the mixer calculations

To calculate the mixer gain and input and output impedances the following seven assumptions are made:

1) The IF is so low that the mixing element sees the same impedance at the LO frequency and both sidebands.

2) The mixer sees the same impedance at the IF and at DC so that the IF signal can be treated as small changes in the bias voltage.

3) The DC bias is applied by an ideal voltage source with an output load conductance G_L^{IF}.

4) No harmonic frequencies appear at the RF terminals of the mixer and the equivalent embedding circuits are purely resistive.

5) The currents and voltages across the mixing element are in phase.

6) The signal will only produce an amplitude modulation of the LO signal, *i.e.*, no phase variation of the LO.

7) Only devices where the current (for no local oscillator power) increases monotonically with the voltage will be considered.

Assumptions 1) through 4) are usually fulfilled in a mixer, especially the one described in chapter III of this thesis. The IF of 1.5 GHz is much lower than typical signal frequencies in the submillimeter band. The large capacitance of an SIS junction shorts the harmonic frequencies and the matching circuit as described in chapter III tunes out the large capacitance at the LO and both sideband frequencies (quasi–optical receivers with broad band matching circuits are truly double sideband receivers). A SSB receiver would clearly violate assumption 1). Assumption 5) is actually not correct due to quantum mechanical reactances. However, as shown by Feldman [1.14], these reactances are rather small and can be ignored. Assumption 6) is somewhat artificial and certainly not realistic for any receiver since it requires signals in both sidebands set up so that their phase contribution to the LO cancels. However, this assumption greatly simplifies the formalism and yields the same results as a rigorous treatment of the theory of mixing by Torrey and Whitmer [1.15]. Assumption 7) is usually fulfilled for SIS diodes in the bias voltage range of interest $V_{gap}/2 < V_b < V_{gap}$. Note that the assumption is for the unpumped I–V characteristic only, *i.e.*, without LO power applied. When LO power is applied the I–V characteristic of an SIS diode can develop areas of negative differential resistance indicating quantum behavior, which is within the following theoretical treatment.

c) The dissipative RF currents through the SIS diode

Before calculations of mixer conversion gain, input and output impedances can be made the dissipative, *i.e.*, in phase, currents induced by the RF signals (mostly from the local oscillator) need to be known. From equation (1.29) one can set up a photon rate equation as follows. For each energy level n the number of electron like quasi–particles that tunnel per second due to photon assisted tunneling are $J_n^2\left(\frac{eV_\omega}{\hbar\omega}\right)\ I_{\rm DC}\left(V_b + n\hbar\omega/e\right)$. Each electron requires n photons so that the total rate N is the sum for all energy levels,

$$N = \frac{1}{e}\sum_{n=-\infty}^{\infty} nJ_n^2\left(\frac{eV_\omega}{\hbar\omega}\right)\ I_{\rm DC}\left(V_b + \frac{n\hbar\omega}{e}\right)\ . \tag{1.30}$$

The total power absorbed by the SIS junction is $P = V_\omega I_\omega/2$ yielding a photon rate of $N = V_\omega I_\omega/2\hbar\omega$. From this the dissipative current follows to be

$$I_\omega(V_b, V_\omega) = \frac{2\hbar\omega}{eV_\omega}\sum_{n=-\infty}^{\infty} nJ_n^2\left(\frac{eV_\omega}{\hbar\omega}\right)\ I_{\rm DC}\left(V_b + \frac{n\hbar\omega}{e}\right)\ . \tag{1.31}$$

Using the recurrence relation for Bessel functions (Bronstein and Semendjajew [1.16])

$$\frac{2n}{x}J_n(x) = J_{n-1}(x) + J_{n+1}(x) \qquad \text{yields} \tag{1.32}$$

$$I_\omega(V_b, V_\omega) = \sum_{n=-\infty}^{\infty} J_n\left(\frac{eV_\omega}{\hbar\omega}\right)\ \left[J_{n-1}\left(\frac{eV_\omega}{\hbar\omega}\right) + J_{n+1}\left(\frac{eV_\omega}{\hbar\omega}\right)\right] I_{\rm DC}\left(V_b + \frac{n\hbar\omega}{e}\right)\ . \tag{1.32}$$

Equation (1.32) is identical to the one derived more rigorously in Tucker and Feldman [1.11].

d) Gain, input and output conductances of the SIS mixer

All required relations that contain the physics of the SIS tunnel junction are now available and the mixer's characteristics can be calculated. The DC current $I_b(V_b, V_\omega)$ and the dissipative component of the RF current $I_\omega(V_b, V_\omega)$ as given in (1.29) and (1.32), respectively, are treated as dependent variables. It will prove useful to define the following conductances:

$$G_{bb} \equiv \frac{\partial I_b}{\partial V_b} \qquad G_{\omega\omega} \equiv \frac{\partial I_\omega}{\partial V_\omega} \qquad G_{b\omega} \equiv \frac{\partial I_b}{\partial V_\omega} \qquad G_{\omega b} \equiv \frac{\partial I_\omega}{\partial V_b}\ . \tag{1.33}$$

The G_{bb} and $G_{\omega\omega}$ can be thought of as the low (DC and IF) and high (RF) frequency conductances of the diode, respectively, when conversion effects are ignored. The $G_{b\omega}$ is the conductance for

the conversion from high to low frequencies and $G_{\omega b}$ the reverse. Small changes in the junction's current can then be written as

$$dI_b = G_{bb}dV_b + G_{b\omega}dV_\omega \tag{1.34}$$

$$dI_\omega = G_{\omega b}dV_b + G_{\omega\omega}dV_\omega \quad . \tag{1.35}$$

To derive the differential low frequency output conductance of the mixer G_M^{IF} from (1.34) at constant LO power

$$G_M^{IF} \equiv \left(\frac{dI_b}{dV_b}\right)_{P_{LO}=const} = G_{bb} + G_{b\omega}\left(\frac{dV_\omega}{dV_b}\right)_{P_{LO}=const} \quad , \tag{1.36}$$

the quantity $\frac{dV_\omega}{dV_b}$ at constant LO power needs to be determined. The LO source current in the RF circuit is the sum of the LO current through the source conductance G_S^{RF} and the dissipative current I_ω through the SIS junction

$$I_{LO} = G_S^{RF}V_\omega + I_\omega \quad . \tag{1.37}$$

Holding the LO power constant, *i.e.*, $dI_{LO} = 0$ yields

$$0 = dI_{LO} = G_S^{RF}dV_\omega + G_{\omega\omega}dV_\omega + G_{\omega b}dV_b \tag{1.38}$$

$$\left(\frac{dV_\omega}{dV_b}\right)_{P_{LO}=const} = -\frac{G_{\omega b}}{G_S^{RF} + G_{\omega\omega}} \quad , \tag{1.39}$$

so that (1.39) inserted in (1.36) is

$$G_M^{IF} = G_{bb} - \frac{G_{\omega b}\,G_{b\omega}}{G_S^{RF} + G_{\omega\omega}} \quad . \tag{1.40}$$

The conversion efficiency is measured and calculated with a constant supply voltage, V_{DC}. Note, that the bias voltage of the SIS diode V_b contains the fixed DC bias V_{DC} and a time dependent component V_{IF}. The IF current then is $I_{IF} = dI_b$ at constant supply voltage V_{DC}. Assumption 6) stated that the RF signal is like an amplitude modulation of the LO signal, *i.e.*, $I_{signal} = dI_{LO}$ at constant DC supply voltage V_{DC}. Using equation (1.38) for dI_{LO},

$$I_{signal} = dI_{LO}|_{V_{DC}=const} = (G_S^{RF} + G_{\omega\omega})dV_\omega + G_{\omega b}\left(\frac{dV_b}{dV_\omega}\right)_{V_{DC}=const} dV_\omega \quad . \tag{1.41}$$

The low frequency voltage across the SIS junction is $V_b = V_{DC} - I_b/G_L^{IF}$, with G_L^{IF} the IF circuits load conductance. Differentiating this and using equation (1.34) yields

$$\left(\frac{dV_b}{dV_\omega}\right)|_{V_{DC}=\text{const}} = -\frac{G_{b\omega}}{G_L^{IF} + G_{bb}} \quad . \tag{1.42}$$

The signal current then finally is

$$I_{\text{signal}} = (G_S^{RF} + G_{\omega\omega} - \frac{G_{b\omega}G_{\omega b}}{G_L^{IF} + G_{bb}})dV_\omega \quad . \tag{1.43}$$

As stated above, the IF current is

$$\begin{aligned} I_{IF} = dI_b|_{V_{DC}=\text{const}} &= G_{bb}\left(-\frac{G_{b\omega}}{G_L^{IF} + G_{bb}}\right)dV_\omega \\ &= \frac{G_{b\omega}G_L^{IF}}{G_L^{IF} + G_{bb}} \end{aligned} \quad , \tag{1.44}$$

where equations (1.34) and (1.42) were used.

The equations derived so far allow us to calculate the conversion gain of the mixer. For most mixers, the conversion gain is less than one, as is unavoidable for classical mixers. However, SIS diodes can yield conversion gain larger than unity due to the quantum mechanical nature of the mixing process invoking photon assisted tunneling. The conversion gain C is defined as the ratio of the IF power delivered into the load resistor to the available power in the RF signal.

$$\begin{aligned} C(V_{DC}, P_{LO}, G_S^{RF}, G_L^{IF}) &\equiv \frac{I_{IF}^2}{(I_{\text{signal}}/2)^2}\,\frac{G_S^{RF}}{G_L^{IF}} \\ &= \frac{G_S^{RF}G_{b\omega}^2}{(G_{\omega\omega} + G_S^{RF})^2}\,\frac{4G_L^{IF}}{(G_L^{IF} + G_M^{IF})^2} \end{aligned} \quad . \tag{1.45}$$

The factor of 2 in $I_{\text{signal}}/2$ arises from assumption 6) and is included here to make the final results in accordance with more detailed calculations that assume the signal to be in only one sideband rather than both. The following dimensionless quantities can be defined,

$$C_o \equiv \frac{G_{b\omega}}{G_{\omega b}} \qquad \eta \equiv \frac{G_{\omega b}G_{b\omega}}{G_{bb}G_{\omega\omega}} \qquad g_S^{RF} \equiv \frac{G_S^{RF}}{G_{\omega\omega}} \qquad g_L^{IF} \equiv \frac{G_L^{IF}}{G_{bb}} \qquad g_M^{IF} \equiv \frac{G_M^{IF}}{G_{bb}} \; . \tag{1.46}$$

As will be shown later, the quantity C_o, which is the ratio of the down to up conversion conductances, is constant at a value of 0.5 for a classical mixer. η is the normalized product of the up and down conversion conductances and can thus be thought of as a measure of intrinsic

conversion efficiency. The lower case g's are simply normalized IF and RF conductances. Now the conversion gain can be rewritten in a clearer way,

$$C(V_{DC}, P_{LO}, G_S^{RF}, G_L^{IF}) = C_o\,\eta \frac{g_S^{RF}}{(1+g_S^{RF})^2} \frac{4 g_L^{IF}}{(g_L^{IF}+g_M^{IF})^2} \quad . \tag{1.47}$$

The conversion gain is thus the product of the intrinsic conversion efficiency, η, the constant (at least for a classical mixer) C_o, a fraction that depends on the RF match and one that depends on the IF match. However, some interaction between all terms exists. For the IF side of the mixer, the maximum available conversion gain C_{max} is available when $g_L^{IF} = |g_M^{IF}|$,

$$C_{max}(V_{DC}, P_{LO}, G_S^{RF}) = C_o \eta \frac{g_S^{RF}}{(1+g_S^{RF})(1+g_S^{RF}-\eta)} \quad \text{with} \quad g_M^{IF} > 0 \quad , \tag{1.48}$$

i.e., the IF side of the mixer is power matched. However, the maximum available conversion gain C_{max} of equation (1.48) is **infinite** if $g_M^{IF} = -g_L^{IF}$, *i.e.*, $g_M^{IF} < 0$. It will be shown later, that infinite conversion gain is a possible solution for a non–classical mixer.

e) Conversion gain of a classical mixer

For the classical mixer, *i.e.*, a resistor with a non–linear current–voltage characteristic, the current will follow the applied voltage instantaneously

$$I(t) = I_{DC}\left(V_b + V_\omega \cos(\omega t)\right) \quad . \tag{1.49}$$

The Fourier components of the low frequency current component I_b and the high frequency dissipative current component I_ω then are

$$\begin{aligned} I_b(V_b, V_\omega) &= \frac{1}{\pi}\int_0^\pi d(\omega t) I_{DC}\left(V_b + V_\omega \cos(\omega t)\right) \\ I_\omega(V_b, V_\omega) &= \frac{2}{\pi}\int_0^\pi d(\omega t)\cos(\omega t)\; I_{DC}\left(V_b + V_\omega \cos(\omega t)\right) \quad . \end{aligned} \tag{1.50}$$

The conductances as defined in (1.33) then are (with $\phi \equiv \omega t$),

$$\begin{aligned} G_{bb} &= \frac{1}{\pi}\int_0^\pi d\phi \frac{\partial I_{DC}(V_b + V_\omega \cos\phi)}{\partial V_b} \\ G_{\omega\omega} &= \frac{2}{\pi}\int_0^\pi d\phi \cos^2\phi \frac{\partial I_{DC}(V_b + V_\omega \cos\phi)}{\partial V_b} \\ G_{\omega b} = 2G_{b\omega} &= \frac{2}{\pi}\int_0^\pi d\phi \cos\phi \frac{\partial I_{DC}(V_b + V_\omega \cos\phi)}{\partial V_b} \quad . \end{aligned} \tag{1.51}$$

Recall from equation (1.40) that the sign of G_M^{IF} is determined by the sign of $G_{bb}G_{\omega\omega} - G_{b\omega}G_{\omega b}$, which is equivalent to η being larger than unity for a negative G_M^{IF} and η between zero and one for a positive G_M^{IF}. For a classical mixer

$$G_{bb}G\omega\omega - G_{b\omega}G_{\omega b} = \tag{1.52}$$

$$\frac{1}{\pi^2}\int_0^{\pi} d\phi_1 \int_0^{\pi} d\phi_2 (\cos\phi_1 - \cos\phi_2)^2 \frac{\partial I_{DC}(V_b + V_\omega \cos\phi_1)}{\partial V_b} \frac{\partial I_{DC}(V_b + V_\omega \cos\phi_2)}{\partial V_b} \quad ,$$

which is, of course, always positive. Thus $0 \leq \eta \leq 1$ and $G_M^{IF} \geq 0$ and therefore from inspection of equation (1.48) one can conclude that a classical mixer can not have a conversion gain larger than unity, *i.e.*, it will always have a conversion loss. From equations (1.51) it follows that $C_o = 2$ for a classical mixer and thus for the best possible case† of $\eta = 1$ the conversion gain can not exceed $C = 0.5$, *i.e.*, a 3dB conversion loss per sideband of a classical double sideband mixer. The only assumption for the classical mixer that forces this result was that the current instantaneously follows the applied voltage for the non–linear resistive mixer as stated in equation (1.49).

f) Conversion gain of an SIS diode mixer

The current response for an SIS diode is explicitly given in equations (1.29) and (1.32). The conductances as defined in (1.33) can be calculated

$$G_{bb} = \sum_{n=-\infty}^{\infty} J_n^2(v_\omega) \frac{dI_{DC}\left(V_b + \frac{n\hbar\omega}{e}\right)}{dV_b} \qquad \text{with } v_\omega \equiv \left(\frac{eV_\omega}{\hbar\omega}\right)$$

$$G_{\omega b} = \sum_{n=-\infty}^{\infty} J_n(v_\omega)\ \left[J_{n-1}(v_\omega)\ + J_{n+1}(v_\omega)\ \right] \frac{dI_{DC}\left(V_b + \frac{n\hbar\omega}{e}\right)}{dV_b} \tag{1.53}$$

$$G_{b\omega} = \frac{v_\omega}{V_\omega} \sum_{n=-\infty}^{\infty} J_n(v_\omega)\ J_{n+1}(v_\omega)\ \left[I_{DC}\left(V_b + \frac{(n+1)\hbar\omega}{e}\right) - I_{DC}\left(V_b + \frac{n\hbar\omega}{e}\right)\right]$$

$$G_{\omega\omega} = \frac{v_\omega}{2V_\omega} \sum_{n=-\infty}^{\infty} \left[J_n^2(v_\omega) + J_{n-1}(v_\omega)\ J_{n+1}(v_\omega)\right] \left[I_{DC}\left(V_b + \frac{(n+1)\hbar\omega}{e}\right) - I_{DC}\left(V_b + \frac{n\hbar\omega}{e}\right)\right].$$

Tucker [1.18] has shown that these expressions for the conductances reduce to those for a classical mixer given in (1.51) for photon energies (converted to voltages) small compared to the voltage

† McColl [1.17] has shown that a Schottky diode mixer can indeed theoretically approach the limit $\eta = 1$ for LO amplitudes V_ω large compared to the DC non linearity scale of the Schottky diode.

scale of the non–linearity of a mixer diode. The quantity $C_o = G_{b\omega}/G_{\omega b}$, which is 0.5 for the classical mixer, can now be much larger than 0.5. Comparing the terms of $G_{b\omega}$ and $G_{\omega b}$ of equation (1.53) for each n, it can be seen that if for the dominating n the following condition holds,

$$\frac{2e}{\hbar\omega}\left[I_{\rm DC}\left(V_b+\frac{(n+1)\hbar\omega}{e}\right)-I_{\rm DC}\left(V_b+\frac{n\hbar\omega}{e}\right)\right] > \left[\frac{dI_{\rm DC}\left(V_b+\frac{(n+1)\hbar\omega}{e}\right)}{dV_b}+\frac{dI_{\rm DC}\left(V_b+\frac{n\hbar\omega}{e}\right)}{dV_b}\right] \tag{1.54}$$

C_o will be larger than 0.5. Figure 1.12 shows for $n = 0$, which dominates the Bessel function series, how the left–hand side of equation (1.54) can be larger than the right–hand side. The left–hand side, for $n = 0$, is the slope of a line connecting the two points $I_{\rm DC}(V_b)$ and $I_{\rm DC}(V_b + \frac{\hbar\omega}{e})$. The right–hand side of (1.54) is the average of the DC I–V curve's slope at those points. The average of the slopes at the two points is much lower than the slope in between the two points since the photons *stepped* over the region of strong non–linearity. For a classical device, *i.e.*, one where the photon energy is small compared to the non–linearity of the I–V characteristic, the slopes of the dashed line and the average at the two points are very similar and the inequality in (1.54) becomes an equality.

Computer simulations [1.18] show that C_o can be much larger than unity. Furthermore, and not immediately obvious from equations (1.53), those computer simulations showed that the quantity η can exceed unity, which is equivalent to the mixers IF output conductance $G_{\rm M}^{\rm IF}$ being negative. The LO power can thus induce negative differential conductances on the I–V characteristic of an SIS diode. This then yields theoretically infinite available conversion gain or, in practice, conversion gain of larger than unity. This was first demonstrated in experiments by Kerr *et al.* [1.19].

g) Design considerations for SIS mixers

Concluding this section are some comments on practical design considerations for SIS mixers. The input circuit, *i.e.*, RF signal side of the mixer, can be similar to a classical design. The IF side of an SIS mixer can be rather different from a classical one since the IF load conductance $G_{\rm L}^{\rm IF}$ presented to the mixer has to be matched to the mixers IF output conductance $G_{\rm M}^{\rm IF}$ of the SIS diode mixer, which can be very different from the SIS junction's unpumped DC I-V characteristic. The LO can induce regions of negative differential resistances on the I–V characteristics, which yield

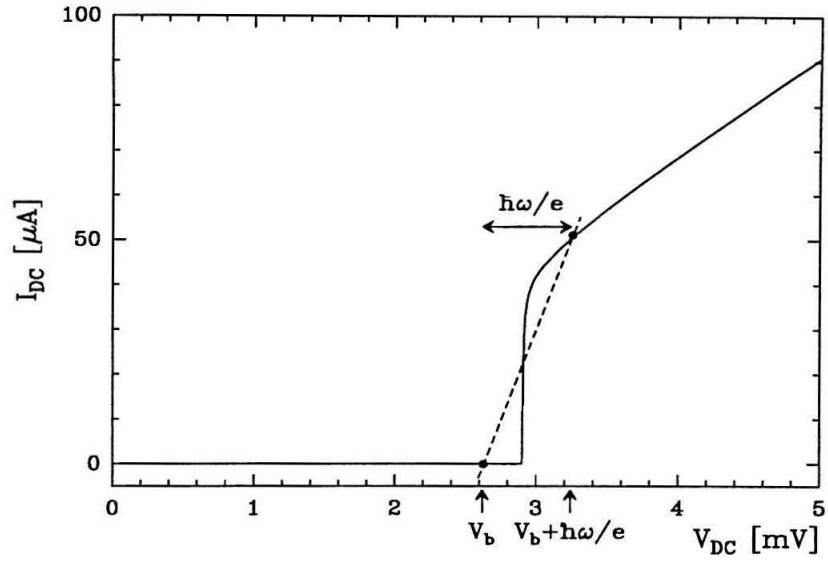


Figure 1.12 The lowest order term ($n = 0$) of equation (1.54) exhibiting strong quantum behavior for a device with a non–linearity of the current–voltage characteristic smaller than the photon energy. The dashed line connects the two points $I_{DC}(V_b)$ and $I_{DC}(V_b + \frac{\hbar\omega}{e})$. The slope of this is obviously greater than the average slope of the current–voltage characteristic at those two points, which yields the inequality of equation (1.54).

very high conversion gains and thus typically very high sensitivity. However, for a receiver in a radio astronomical application operating the mixer in a region of conversion gain much larger than unity is typically not practicable. To achieve conversion gain the IF impedance presented to the mixer would have to be much larger than the typical 50 Ω input impedance of an IF amplifier, thus requiring impedance transformer circuits. These transformers, depending on the transformation ratio, will significantly limit the IF bandwidth. Furthermore, even if the IF impedances can be matched over the desired bandwidth, the mixer will then easily be saturated. A saturated mixer will not yield a linear relation between the input and output power thus making calibration of data virtually impossible. The saturation stems from the IF component of the bias voltage V_b that will, if the mixer is saturated, have a voltage swing of larger than, or comparable to, a photon step. The power levels required to saturate an SIS mixer with gain and a reasonable bandwidth is typically already encountered from thermal radiation of the atmosphere in the millimeter and submillimeter wavelengths. It is therefore advisable to operate an SIS mixer with a conversion gain of unity or a loss of a few dB. This typically yields a mixer IF impedance of a few times the normal state resistance of the SIS diode.

1.5 Submillimeter SIS Receivers

Among the heterodyne receivers with large instantaneous bandwidths, those using superconductor-insulator-superconductor (SIS) tunneling junctions as the detector are the most sensitive in the millimeter wave band [1.20,1.21,1.22,1.23] and more recently also in the low–frequency–end of the submillimeter wave band [1.23,1.24,1.25,1.26,1.27]. The most common design in the past for millimeter wave heterodyne receivers used in radio astronomy is based upon waveguide structures which couple the radiation to the detector. Waveguide structures typically yield a tuning range of about 30% [1.28]. Using more than one tuning element, the range can be pushed to one octave [1.21,1.29]. These tuning elements are undesirable because they complicate the operation of the receiver and can suffer from irreproducible back lash and mechanical wear. To reduce these problems for waveguide receivers with mechanical tuners, designs using SIS detectors with integrated tuning structures are used in the millimeter wave band. A good review can be found in Kerr *et al.* [1.30] and more recent results in [1.31,1.32,1.33]. However, in those designs it is still necessary to machine a costly waveguide mixer block including waveguide horn antennas that do not easily allow for the construction of array receivers.

An alternative to the waveguide structure is to mount the detector at the center of a planar microantenna or at the end of an impedance matching circuit fed by the antenna. The planar antenna is then quasi–optically coupled to the telescope. This avoids the problems of tuners and high frequency waveguide component fabrication, and offers the potential of high-performance operation over many octaves with a single receiver. Wengler *et al.* [1.5] built such a quasi–optical receiver† using a bow-tie antenna mounted on a hyperhemispherical lens [1.4] to couple radiation to the SIS junction. It was the first heterodyne receiver with a large instantaneous bandwidth, covering a frequency range of 2 octaves (116 to 466 GHz). Bow-tie antennas have a frequency independent impedance [1.34] and symmetric E- and H-plane response, as long as their linear dimensions are larger than a free space wavelength. However, in theory, their beam patterns show no single main beam in the desired direction, perpendicular to the antenna-plane, but instead show a complex large angle pattern [1.34]. In practice [1.5] the beam can be pulled forward by a lens

† Quasi–optical receivers are sometimes also referred to as open–structure receivers – a convention not used in this thesis.

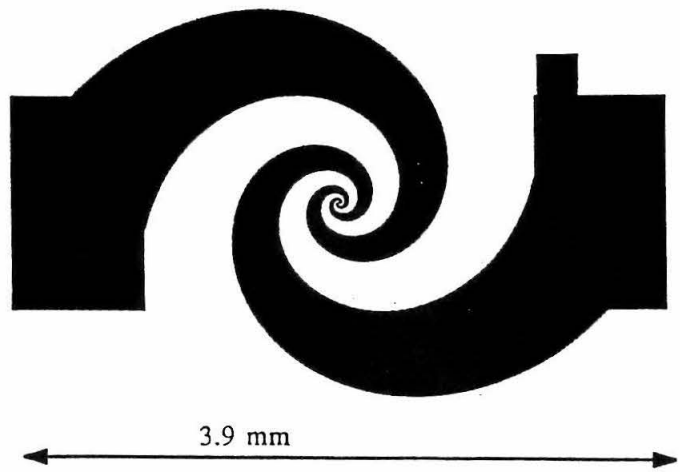


Figure 1.13 The planar two–arm logarithmic spiral antenna with the IF ports (contact pads to the left and right).

system. Wengler *et al.* observed noise temperatures almost as good as those for narrow band SIS waveguide receivers, which was very encouraging. This led to the development of an improved quasi–optical receiver described in this section and in [1.23], using a planar two-arm logarithmic spiral antenna (Figure 1.13) rather than a bow-tie antenna. In addition to frequency independent impedance, and nearly symmetric E- and H-plane patterns, planar spiral antennas have frequency independent beam patterns, with a main beam perpendicular to the antenna-plane [1.35]. This receiver was essentially as sensitive as the best SIS waveguide receivers in the millimeter band, and showed superior performance in the submillimeter band in its time (1988). Even today it is still very competitive with the best receivers reported in the literature. Its design frequency range was 100 to 1000 GHz, and noise temperature measurements were done between 115 and 761.4 GHz. In addition to the laboratory measurements, this receiver has been tested at 115, 230, 345 and 492 GHz at the Caltech Submillimeter Observatory (CSO) on Mauna Kea, Hawaii. This receiver constituted the first SIS receiver used for submillimeter wave astronomy. A more detailed description of this early quasi–optical receiver, serving as an example of a typical layout

for a quasi–optical receiver, will be given in the following section. Improvements to this receiver, the design of a new antenna system (the hybrid antenna see [1.36]), and a broadband matching structure (the end–loaded stub see [1.24]) to eliminate the impedance mismatch problems of quasi–optical SIS receivers are described in chapters II and III, respectively.

1.6 Quasi–Optical Receiver Description: The Spiral Antenna Receiver

Figure 1.14 shows the overall layout of the receiver. The cryostat used for this receiver is purely passive with a liquid helium reservoir cooling the innermost part (cold plate) to 4.2 K where the mixer block and low noise preamplifier are mounted. The helium boil–off is used to cool a radiation shield to about 80 K. A hold time of more than 24 hours was achieved with all electronics operating. The SIS detector, spiral antenna, RF optics and IF chain will now be described in detail.

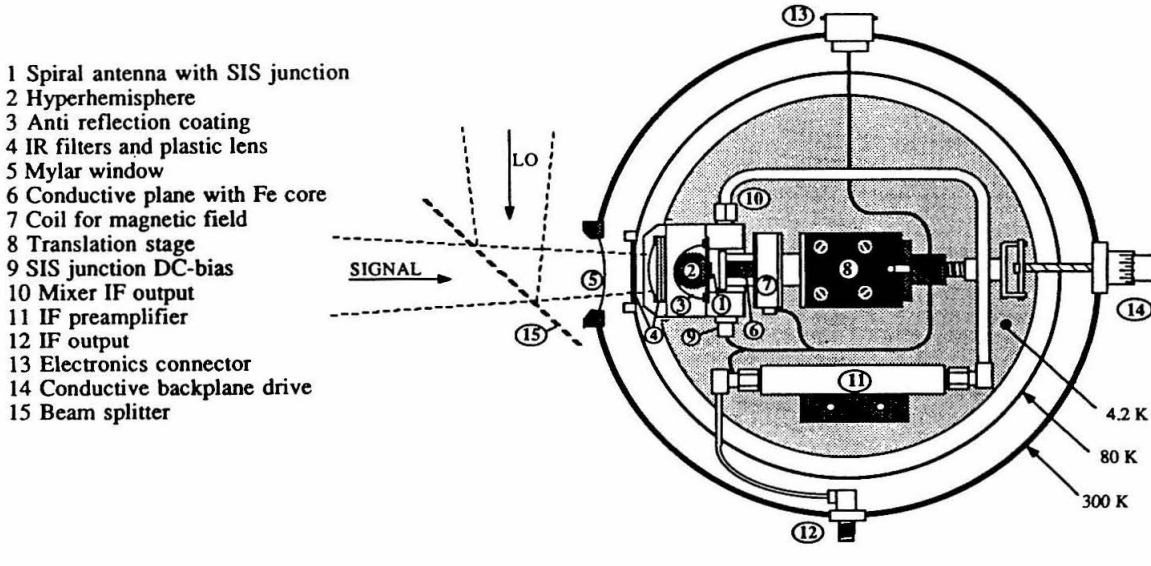


Figure 1.14 Receiver layout in the cryostat.

a) The SIS Detector

The PbInAu SIS tunnel junctions used for this receiver were produced in collaboration with Ron Miller at AT&T Bell Labs. Standard electron beam lithography for the masks and the tri-level photoresist stencil technique [1.37,1.38] were used for the fabrication of the devices [1.39] (Figure 1.15). A scanning electron micrograph of a junction is shown in Figure 1.16. With the SIS junction at a temperature of 4.2 K mounted in the receiver, a gap voltage of 2.40 mV and a critical current density of 7000 A/cm^2 were measured. The current–voltage characteristic of a typical junction is shown in Figure 1.17. The junction overlap area is about 0.5 μm^2, which yields a capacity of about 10 fF. With a normal state resistance of 50Ω the roll-off frequency is thus about 300 GHz. The two electrodes from the SIS junction extend out to the two arms of the spiral antenna. Hence the antenna and the SIS junction in its center are manufactured simultaneously from the same material (PbInAu) on a single crystal quartz substrate, 4 mm square by 0.1 mm thick.

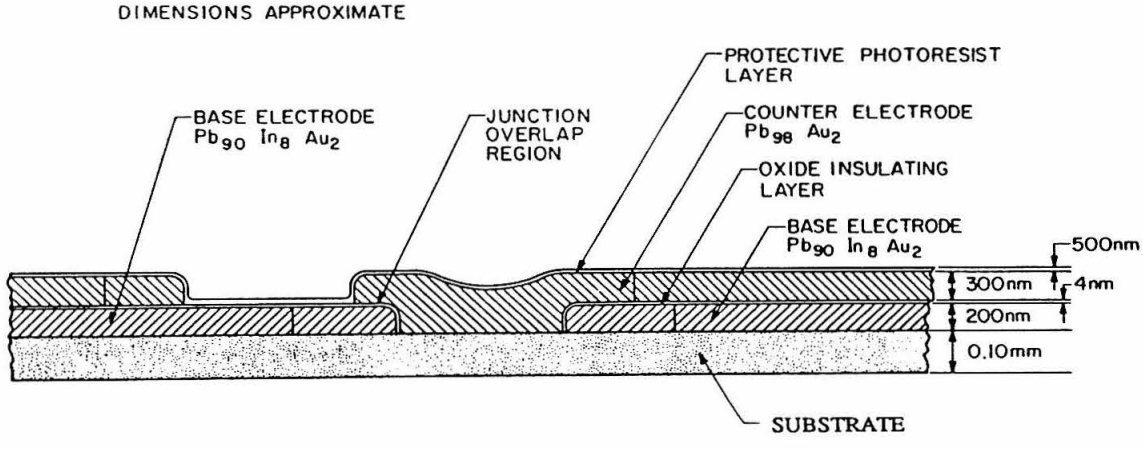


Figure 1.15 Section through SIS junction produced with the trilevel photoresist stencil technique.

b) The Spiral Antenna

The planar two-arm logarithmic spiral antenna (Figure 1.13) belongs to a family of frequency independent antennas, for which characteristics like impedance, beam pattern *etc.* do not depend

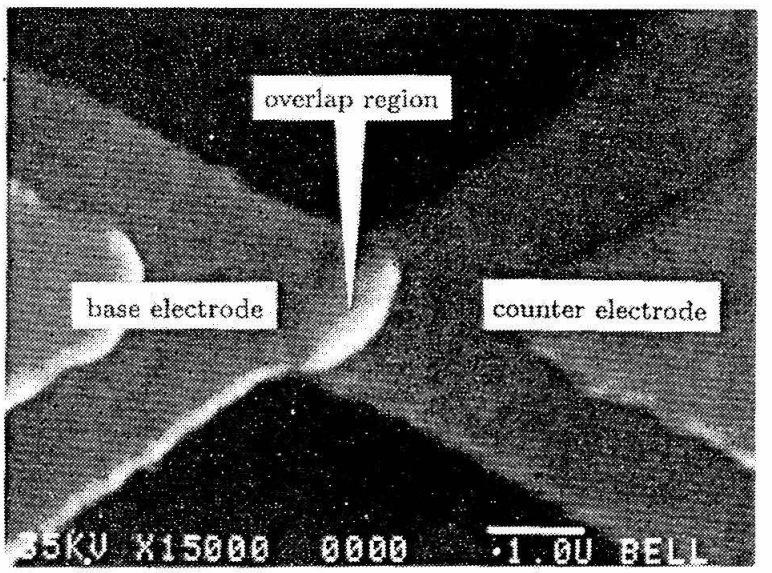


Figure 1.16 Scanning electron micrograph of an SIS junction. The overlap area is about $0.5\mu m^2$ in size.

on frequency over several octaves. Rumsey [1.40] proposed that this can be achieved when the antenna shape is described without a characteristic length scale, in terms of ratios (logperiodic antennas) or angles (spiral antennas). The shape for a single arm of a planar spiral antenna is given by

$$r = r_o e^{c\theta} \tag{1.55}$$

with (r, θ) polar coordinates and c a dimensionless constant (r_o = "one length unit"). Scaling this with wavelength

$$\frac{r}{\lambda} = \frac{r_o e^{c\theta}}{\lambda} = e^{c(\theta - \theta_\lambda)} \tag{1.56}$$

shows that changing the wavelength results in a rotation of the antenna due to a change of θ_λ. However, since the spiral antenna has circular polarization this is of no concern. In order to have a frequency independent beam pattern the effective aperture must increase with wavelength. This has been experimentally verified by Dyson [1.41], who showed that the fields decay by about 20 dB in the first wavelength along the spiral arm. Since the fields decay rapidly, the spiral can be

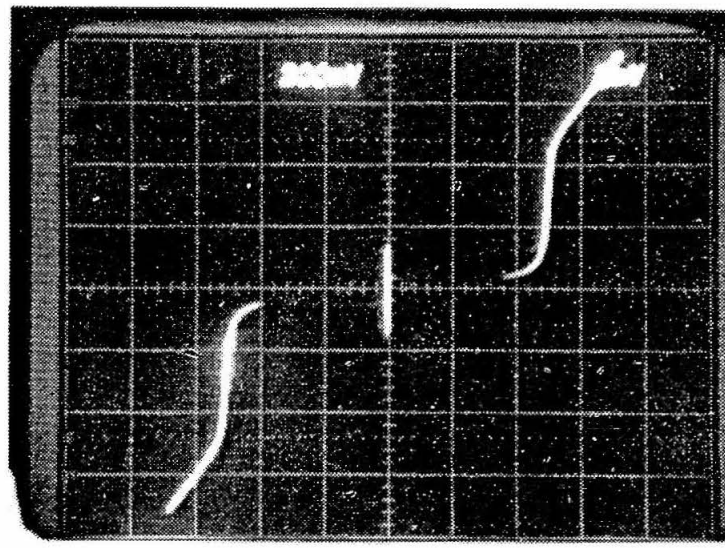


Figure 1.17 Current–voltage characteristic of the SIS junction as used in this receiver measured with the junction mounted in the mixer block at 4.2 K. The gap voltage is about 2.4 mV. The vertical scale is 20 μA/div and the horizontal 1 mV/div.

truncated at a radius R without affecting the antenna characteristics for $\lambda \leq R$. A radius of R = 1.5 mm was chosen, yielding an upper wavelength limit outside the dielectric of about 3 mm. An area with about 15 μm radius is needed for the SIS junction with its leads connecting to the spiral, yielding a lower wavelength limit outside the dielectric of about 300 μm. This shortest operating wavelength is about 10 times larger than the scale on which the antenna deviates from the ideal spiral shape.

A self-Babinet-complementary antenna structure was chosen, because all such structures with two ports have a constant impedance of $Z_o = 60\pi\Omega \simeq 188\Omega$. Mounting such an antenna on to a half–space of dielectric (approximated by the hyperhemisphere) yields an antenna impedance of

$$Z_{ant} = \frac{Z_o}{\sqrt{(1+n_h^2)/2}} \simeq 114\Omega \tag{1.57}$$

where $n_h = 2.11$, the refractive index of single crystal quartz in the submillimeter region [1.42].

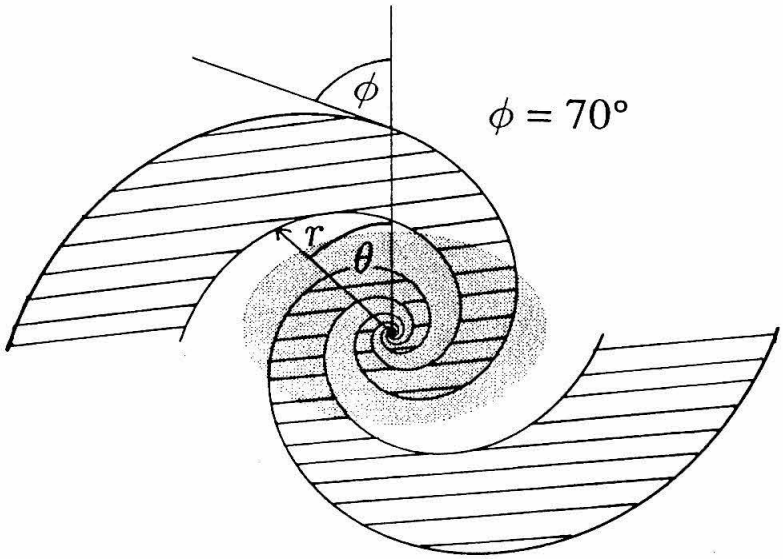


Figure 1.18 A spiral arm intersecting with a radial line from the center of the antenna at angle ϕ. The active region of the spiral antenna (about one wavelength in dimension) is shaded, showing the eccentricity of the antenna's effective aperture. An increase of ϕ will decrease the eccentricity, but also decrease the effective aperture thus widening the beam pattern. The antenna is a self–Babinet–complimentary structure since a rotation of 90° will move the metalized area (hashed) exactly onto the non–metalized area.

The dimensionless constant c in equation (1.55) and (1.56) determines the angle ϕ under which a radial line from the center of the antenna intersects with any boundary of a spiral arm† (Figure 1.18). They are related through

$$c = \cot \phi \ . \tag{1.58}$$

For the antenna used in this receiver, ϕ = 70° was chosen yielding c = 0.364. A more tightly wound spiral, with larger ϕ, will yield a more symmetric beam pattern, since the asymmetry is caused by the rapidly decaying fields (Figure 1.18). However, the effective aperture will decrease, resulting in a wider beam. This causes more problems in the RF optics which match the beam

† The two–arm logarithmic spiral antenna is thus also called a two–arm equiangular spiral antenna

from the telescope to the spiral antenna. The choice for ϕ represents a compromise between these effects, yielding an $f/0.87$ beam at -10 dB relative to the peak, with an eccentricity of less than 1.3 .

c) RF optics

The SIS device with the planar antenna structure sits on a dielectric (crystal quartz) substrate which is mounted on the flat side of a hyperhemispherical lens made out of the same material. This produces an asymmetric beam pattern with respect to the antenna plane due to the different dielectric media on each side of the antenna. For crystal quartz one gets a beam-coupling ratio of about 7 dB in favor of the quartz side [1.4]. To further increase this ratio, the antenna is backed by a conductive plane which can be moved perpendicular to the plane of the antenna according to the receiving frequency. The positioning of this back–plane is very uncritical – unlike back shorts in waveguide receivers – and can be left unchanged for frequency changes of up to a few times 10 GHz. A hyperhemispherically shaped dielectric lens, of radius r = 6.35 mm, converts the $f/0.87$ beam from the spiral antenna to an $f/2$ beam. In this early design, the antenna was placed at $d_f = r/n_h$ = 3.01 mm from the center of the lens, which is the aplanatic focus of the hyperhemispherical lens. This position is later modified and described in more detail in chapter II on the hybrid antenna.

Kasilingam and Rutledge [1.43] showed that for hyperhemispherical lenses with a diameter of more than two free space wavelengths, the focusing gain in the focal plane decreases by about a factor of two for a distance of 0.15 λ off–axis due to abberations. The focusing gain on axis at d_f is ϵ_h^2 = 19.8. Hence, the alignment of the center of the planar antenna with the optical axis of the hyperhemispherical lens should be within at least $\lambda/20$.

The $f/2$ beam from the hyperhemisphere is finally matched to an $f/4$ beam from the telescope optics with a plastic lens on the 4 K stage. Note that in the later design of the quasi–optical receiver employing a hybrid antenna (see chapter II) the f–numbers increase significantly thus making the plastic lens obsolete.

Radiative heating of the SIS junction, mainly from infrared radiation, can significantly decrease the mixing performance of the SIS junction. A higher junction temperature will decrease the gap voltage and reduce the non–linearity of the SIS device, resulting in more mixer noise

and less conversion gain [1.44]. A fused quartz filter, anti–reflection coated with polyethylene, on the 80 K stage reduces the thermal load for the helium stage. A series of fused quartz and fluorogold scattering filters, cooled to 4 K, were used to reject wavelengths shortward of 300 μm and mounted together with the plastic lens in the front part of the mixer block. To reduce reflections of the hyperhemispherical lens a quarter–wavelength thick foil of polyethylene was stretched over the hyperhemispherical lens as an anti–reflection coating.

d) IF circuit

The submillimeter signal is converted to an intermediate center–frequency (IF) of 1.5 GHz with a bandwidth of 500 MHz. No RF rejection filters at the terminals of the spiral antenna are necessary, since the RF-fields on the antenna decay rapidly. One arm of the spiral antenna is grounded, and the other one leads to a low pass filter. This filter uses the leads to the junction as inductors and the mount of the insulated lead as a capacitor. The cut-off has been set with a network analyzer in such a way as to short all frequencies above the IF band. The IF signal is amplified by a three stage liquid-helium-cooled preamplifier especially constructed for this receiver. The design is similar to that described by Weinreb [1.45], but with a high-electron mobility transistor (HEMT) in the first stage. An effective noise temperature of (2 ± 2) K averaged over the entire bandwidth was measured for this preamplifier. Since the amplifier is operated on the liquid helium stage its power consumption had to be low and a consumption for the complete amplifier of 30mW was achieved.

Table 1.1 Spiral antenna receiver noise temperatures

Frequency [GHz]	115	230	345	525	761
$T_{Receiver}$ (DSB) [K]	33	116	215	470	1100

e) Results

The results of this quasi–optical receiver are described in this paragraph. Table 1.1 shows the receiver's noise performance averaged over a 500 MHz bandwidth. These measurements were made using hot (295 K) and cold (78 K) loads as described in section 1.2b. In order to verify the results at 115, 230 and 345 GHz, the receiver was tested at the Caltech Submillimeter Observatory (CSO) on Mauna Kea, Hawaii. A Gunn oscillator and a Schottky diode multiplier [1.46] were used to supply local oscillator (LO) power at these frequencies. At the higher frequencies measurements were made in the laboratory using a far-infrared laser as the LO power source. The Josephson currents had to be suppressed with a magnetic field for frequencies above 350 GHz [1.47]. A conversion loss for the mixer could not be measured precisely, since the matching between the mixer and the IF amplifier and the losses in the optics are not well known. However, at 345 GHz the conversion loss is estimated to be about 11 ± 2 dB. At 115 GHz the mixer showed a saturation of about 10 %, but at higher frequencies no saturation was seen, probably due to higher conversion losses.

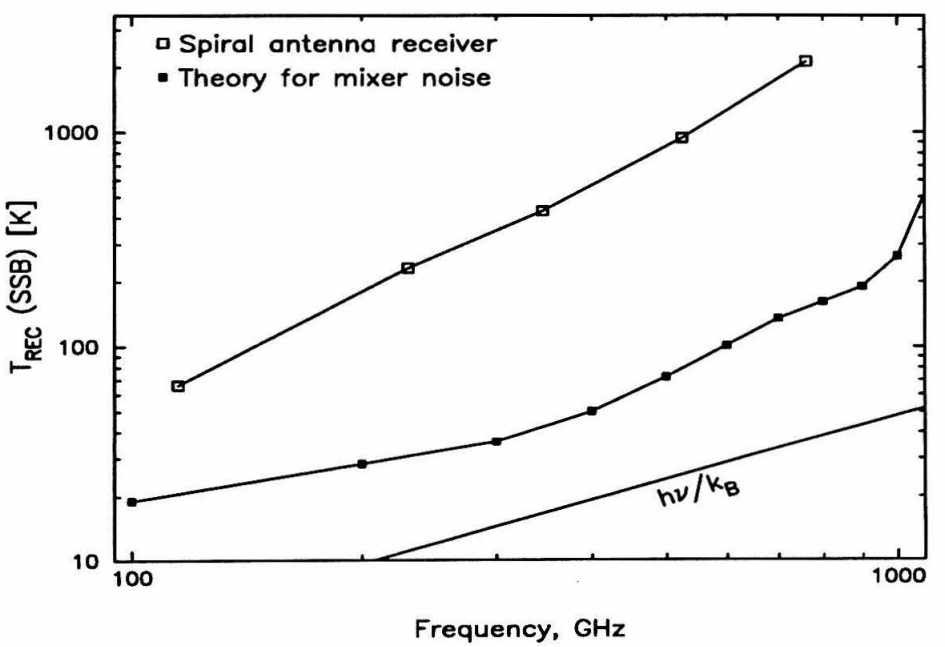


Figure 1.19 Comparison of a theoretical prediction for noise temperature, based on the I–V characteristic of the SIS junction used in the receiver, with the measured noise temperature of the receiver. The quantum limit is also shown.

A theoretical treatment of noise temperature vs. frequency [1.44] is compared to the receiver's performance in Figure 1.19. For reference the quantum limit is also shown. The current–voltage characteristic of the SIS junction, which enters the theoretical prediction of the mixing performance, was shown in Figure 1.16. The curves for the spiral antenna receiver's performance and the theoretical mixer noise are similar, but the measured points are shifted upward to higher noise temperatures with respect to the theoretical prediction. The shift is attributed to losses in the optics, additional noise from the preamplifiers and mismatch of the SIS junction to the antenna, which is due to the junction's capacitance. The curve for the theoretical mixer noise is a prediction, assuming optimum impedance matching conditions for the SIS device, *i.e.*, an ideal mixer noise temperature whereas the measured points are receiver noise temperatures.

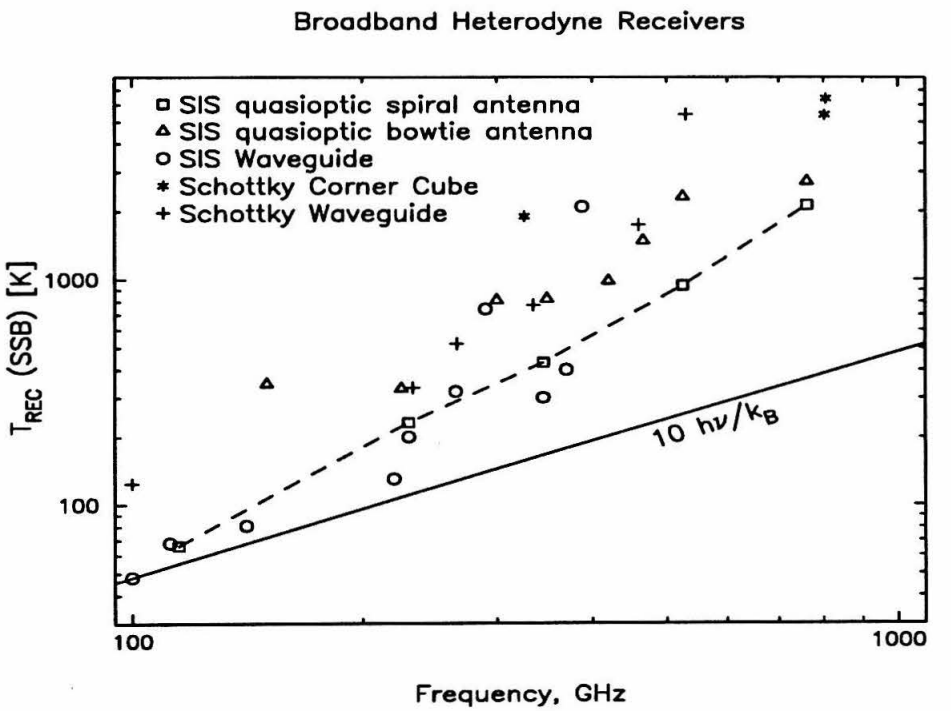


Figure 1.20 Comparison of T_{Rec} of the best SIS and Schottky receivers reported in the literature (as of 1988) with the spiral antenna receiver. For better comparison all double sideband (DSB) noise temperatures have been converted to single sideband noise temperatures. The $10h\nu/k_B$ line corresponds to ten times the quantum noise limit, which is about the best currently achievable for these receiver systems. [1.21,1.5,1.39,1.48–1.56]

Figure 1.20 shows a comparison of the spiral antenna receiver with a bow-tie antenna re-

ceiver, state of the art SIS waveguide and Schottky receivers as of 1988. During the first set of astronomical measurements the main beam efficiency at 345 GHz was about 30 %, which is lower by a factor of two than expected. This was attributed to a focusing problem inside the mixer block, which has now been rectified with the introduction of the hybrid antenna discussed in detail in chapter II. Spectra with the rectified RF optics have been taken, at 345 GHz and 492 GHz and the coupling efficiencies found to be similar to that of waveguide receivers. Again, more on the efficiencies can be found in chapter II.

Figure 1.21 shows a spectrum taken with the receiver described here, which was the first submillimeter spectrum taken with an SIS receiver. The observed position is in the Orion molecular cloud region (OMC1) with the LO frequency set at 341.54 GHz. The RF optics used was as described in this section. The Acousto Optical Spectrometer (AOS) has a center–frequency of 1.4 GHz, and the receiver responds to both sidebands, which are centered at 342.94 Ghz and 340.14 GHz. The integration time was 3.5 hours. The spectrum is confusion limited, *i.e.*, the background of astronomical lines limits the sensitivity. The displayed spectrum contains no baseline correction, showing the excellent stability of the system.

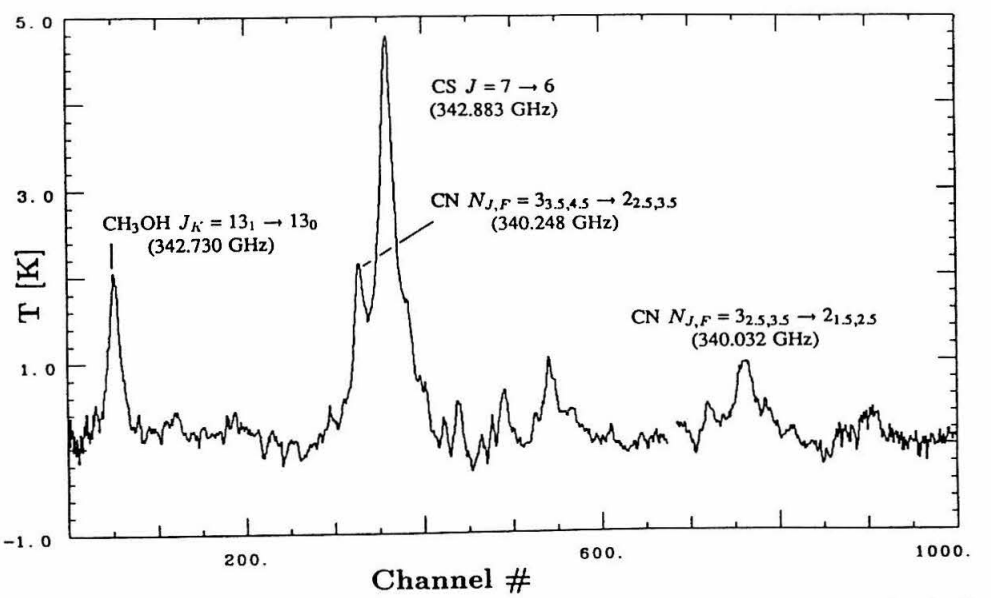


Figure 1.21 341.54 GHz spectrum of the core of OMC1. Response in both sidebands (centered at 342.94 GHz and 340.14 GHz) is included. The total integration time was 3.5 hours and the spectrum is confusion limited, *i.e.*, essentially all features with $T > 0.1K$ are real.

1.7 References Chapter I

[1.1] Phillips, T.G. and Keene, J., "Submillimeter Astronomy," *IEEE Proceedings Special Issue on Terrahertz Technology*, September 1992.

[1.2] Groesbeck, T.D., Blake, G.A., and Phillips, T.G., "A 325 to 360 GHz Emission Line Survey of the Orion KL Region," in preparation, 1993.

[1.3] Grossman, E., "AT – Atmospheric Transmission Software," 1989, Airhead Software, 2069 Bluff St., Boulder, CO 80304.

[1.4] Rutledge, D.B., Neikirk, D.P., and Kasilingam, D.P., "Integrated–Circuit Antennas" in *Infrared an Millimeter Waves*, K. J. Button, Ed., vol. 10, pp. 1 – 90, New York: Academic Press, 1984.

[1.5] Wengler, M. J., Woody, D. P., Miller, R. E., and Phillips, T. G., "A Low Noise Receiver for Millimeter and Submillimeter Wavelengths," *Int. J. of IR and MM Waves*, Vol. 6, pp. 697 – 706, 1985.

[1.6] Caves, C.M., "Quantum Limits on Noise in Linear Amplifiers," *Phys. Rev. D*, vol. 26, pp. 1817 – 1839, 1982.

[1.7] Bardeen, J., Cooper, L.N., and Schrieffer, J.R., "Theory of Superconductivity," *Phys. Rev.*, vol. 108, pp. 1175 – 1204, 1957.

[1.8] Josephson, B.D., "Supercurrents Through Barriers," *Advances in Physics*, vol. 14, pp. 419 – 451, 1965.

[1.9] Tucker, J.R., "Quantum Limited Detection in Tunnel Junction Mixers,"*IEEE J. Quantum Electronics*, vol. QE–15, pp. 1234 – 1258, 1979.

[1.10] Tien, P.K., and Gordon, J.P., "Multiphoton Process Observed in the Interaction of Microwave Fields with the Tunneling between Superconducting Films," *Phys. Rev.*, vol. 129, No. 2, pp. 647 – 651, Jan. 1963.

[1.11] Tucker, J.R., and Feldman, M.J., "Quantum Detection at Millimeter Wavelengths," *Phys. Rev.*, vol. 57, pp. 1055 – 1113, Oct. 1985.

[1.12] Abromovitz, M., and Stegun, I.A., "Handbook of Mathematical Functions," *Dover Publications, Inc. New York*, 10th edition, 1972.

[1.13] Dayem, A. H., and Martin, R. J., "Quantum Interaction of Microwave Radiation with Tunneling Between Superconductors," medit Phys. Rev. Lett., vol. 8, pp. 246, 1962.

[1.14] Feldman, M.J., "Some Analytical and Intuitive Results in the Quantum Theory of Mixing," *J. Appl. Phys.*, vol. 53, p. 584, 1982.

[1.15] Torrey, H.C., and Whitmer, C.A., "Crystal Rectifiers," *MIT Rad. Lab Series, McGraw Hill, New York*, vol. 15, 1948.

[1.16] Bronstein, I.N., and Semendjajew, K.A., "Taschenbuch der Mathematik," *Verlag Harri Deutsch,* 21st edition, 1984.

[1.17] McColl, M., Millea, M.F., Silver, A.H., Bottjer, M.F., Pedersen, R.J., and Vernon Jr., F.L., "The Super–Schottky Microwave Mixer," *IEEE Trans. Magn.*, vol. MAG–13, p. 221, 1977.

[1.18] Tucker, J.R., "Predicted Conversion Gain in Superconductor– Insulator– Superconductor Quasiparticle Mixers," *Appl. Phys. Lett.*, vol. 36, p. 477, 1980.

[1.19] Kerr, A.R., Pan, S.–K., Feldman, M.J., and Davidson, A., "Infinite Available Gain in a 115 GHz SIS Mixer," *Physica*, vol. 108B, p. 1369, 1981.

[1.20] D'Addario, L. R.,"An SIS Mixer for 90 – 120 GHz with Gain and Wide Bandwidth," *Int. J. of IR and MM Waves*, Vol. 5, No. 11, pp. 1419 – 1433, 1984.

[1.21] Ellison, B. N., Miller, R. E., "A Low Noise 230 GHz SIS Receiver," *Int. J. of IR and MM Waves*, Vol. 8, No. 6, pp. 608 – 625, 1987.

[1.22] Räisänen, A. V., Crété, D. G., Richards, P. L., and Lloyd, F. L., "Wide–Band Low Noise MM–Wave SIS Mixers with a Single Tuning Element," *Int. J. of IR and MM Waves*, Vol. 7, No. 12, pp. 1835 – 1852, 1986.

[1.23] Büttgenbach, T.H., Miller, R.E., Wengler, M.J., Watson, D.M., and Phillips, T.G., "A Broad–Band Low–Noise SIS Receiver for Submillimeter Astronomy," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT–36, No. 12, pp. 1720 – 1726, 1988.

[1.24] Büttgenbach, T.H., LeDuc, H.G., Maker, P.D., and Phillips, T.G., "A Fixed Tuned Broadband Matching Structure For Submillimeter SIS Receivers," *IEEE Trans. Applied Supercond.*, Vol. 2, No. 2, September 1992.

[1.25] Walker, C.K., Kooi, J.W., Chan, M., LeDuc, H.G., Carlstrom, J.E., and Phillips, T.G., "A Low–Noise 492 GHz SIS Waveguide Receiver," *Int. J. of IR and MM Waves*, vol. 13, No. 6, pp. 785 – 798, 1992.

[1.26] Zmuidzinas, J., LeDuc, H.G., "Quasi Optical Slot Antenna SIS Mixers," *IEEE Trans. Microwave Theory Tech.*, Vol. 40, No. 9, 1992.

[1.27] de Lange, G., Honingh, C.E., Dierichs, Schaeffer, H.H.A., M.M.T.M., Panhuyzen, R.A., Klapwijk, T.M., van de Stadt, H., and de Graauw, H.H.A., "A Low Noise 410 – 495 GHz $Nb/Al_2O_3/Nb$ SIS WAveguide Mixer," *Appl. Supercond. Conference Proceedings, Chicago*, 1992.

[1.28] Räisänen, A. V., McGrath, W. R., Crété, D. G., and Richards, P. L., "Scaled Model Measurements of Embedding Impedances for SIS Waveguide Mixers," *Int. J. of IR and MM Waves*, Vol. 6, pp. 1169 – 1189, 1985.

[1.29] Büttgenbach, T. H., Groesbeck, T. D., and Ellison, B. N., "A Scale Mixer Model for SIS Waveguide Receivers," *Int. J. of IR and MM Waves*, Vol. 11, No. 1, pp. 1 – 20, 1990.

[1.30] Kerr, A.R., and Pan, S.–K.,"Integrated Tuning Elements for SIS Mixers," *Int. J. of IR and MM Waves*, Vol. 9, No. 2, pp. 203 – 212, 1988.

[1.31] Ermakov, An.B., Koshelets, V.P., Kovtonyuk, S.A., and Shitov, S.V., "Parallel Biased SIS–Arrays for MM Wave Mixers: Main Ideas and Experimental Verification," *IEEE Trans. Magn.*, Vol. 27, No. 2, pp. 2642 – 2645, 1991.

[1.32] Carpenter, J.A., Smith, A.D., Arambula, E.R., Lee, L.P.S., Nelson, T. and Yujiri, L., "100 GHz SIS Mixer With Improved RF Matching," *IEEE Trans. Magn.*, Vol. 27, No. 2, pp. 2654 – 2657, 1991.

[1.33] Kerr, A.R., Pan, S.–K., Lichtenberger, A.W., and Lea, D.M., "Progress on Tunerless SIS Mixers for the 200 – 300 GHz Band," *IEEE Microwave and Guided Wave Lett.*, Vol. 2, No. 11, Nov 1992.

[1.34] Compton, R. C., McPhedran, R. C., Popović, Z. P., Rebeiz, G. M., Tong, P. P., Rutledge, D. B., "Bow–Tie Antennas on a Dielectric Half–Space: Theory and Experiment," *IEEE*

Trans. Antennas Propag. AP–35, 622 (1987).

[1.35] Cheo, B. R.–S., Rumsey, V. H., Welch, W. J., " A Solution to the Frequency–Independent Antenna Problem," *IRE Trans. Antennas Propag.* AP–9, no.6, Nov. 1961.

[1.36] Büttgenbach, T.H., "An Improved Solution For Integrated Array Optics In Quasi Optical MM and SUBMM Receivers: The Hybrid Antenna," *IEEE MTT* , to appear in October 1993.

[1.37] Dunkleberger, L. N., "Stencil Technique for the Preparation of Thin–Film Josephson Devices," *J. Vac. Sci. Tech.*, vol. 15, pp. 88 – 90, Jan. 1978.

[1.38] Dolan, G. J., "Offset Masks for Lift–Off Photoprocessing," *Appl. Phys. Lett.*, vol. 31, pp. 337 – 339, Sept. 1977.

[1.39] Woody, D. P., Miller, R. E., and Wengler, M. J., "85 – 115 GHz Receivers for Radio Astronomy," *IEEE Trans. Microwave Theory Tech.*, vol. MTT–33, pp. 90 – 95, Feb. 1985.

[1.40] Rumsey, V. H., "Frequency Independent Antennas," 1957 *IRE National Convention Record*, pt. 1, pp. 119 – 128.

[1.41] Dyson, J. D., "The Equiangular Spiral Antenna," *IRE Trans. Antennas Propag.*, vol. AP–7, pp. 181 – 187, April 1959.

[1.42] Loewenstein, E. V., Smith D. R., and Morgan, R. L., *Appl. Opt.* 12, 398 (1973).

[1.43] Kasilingam D., and Rutledge, D. B., "Focusing Properties of Small Lenses," *Int. J. of IR and MM Waves*, Vol. 7, No. 10, pp. 1631 – 1647, 1986.

[1.44] Wengler, M. J., and Woody, D. P., "Quantum Noise In Heterodyne Detection," *IEEE J. Quantum Electron.*, vol. QE–23, pp. 613 – 622, May 1987.

[1.45] Weinreb, S., Fenstermacher, D.L., and Harris, R. W., "Ultra–low–noise 1.2 to 1.7 GHz Cooled GaAs–Fet Amplifiers," *IEEE Trans. Microwave Theory Tech.*, vol. MTT–30, pp. 849 – 853, June 1982.

[1.46] Rothermel, H., Phillips, T. G., and Keene, J. B., " A Solid–state Frequency Source for Radio Astronomy in the 100 to 1000 GHz Range," *Int. J. of IR and MM Waves*, Vol. 10, No. 1, pp. 83 – 100, 1989.

[1.47] Dolan, G. J., Phillips, T. G., and Woody, D. P., "Low–noise 115 GHz Mixing in Superconducting Oxide–barrier Tunnel Junctions," *Appl. Phys. Lett.*, vol. 34, pp. 347 – 349, Mar. 1979.

[1.48] Predmore, C. R., Räisänen, A. V., Erickson, N. R., Goldsmith, P. F., and Marrero, J. L. R., "A broad–band, Ultra–low–noise Schottky Diode Mixer Receiver from 80 to 115 GHz," *IEEE Trans. Microwave Theory Tech.*, vol. MTT–32, pp. 498 – 506, May 1984.

[1.49] Erickson, N. R., "A Very Low Noise Single Sideband Receiver for 200 – 260 GHz," *Preprint #260 of Five College Radio Astronomy Observatory*, 1985.

[1.50] Röser, H. P., Durwen, E. J., Wattenbach, R., and Schultz, G. V., "Investigation of a Heterodyne Receiver with Open Structure Mixer at 324 GHz and 693 GHz," *Intl. J. of IR and Millimeter Waves*, vol. 5, pp. 301 – 314, March 1984.

[1.51] Röser, H. P., Wattenbach, R., and van der Wal, P., "Tunable Heterodyne Receiver from 100 μm to 1000 μm for Airborne Observations," *Airborne Astronomy Symposium*, NASA Conference Publication 2353, pp. 330 – 334, 1984.

[1.52] Pan, S.–K., Feldman, M. J., Kerr, A. R., and Timbie, P., " Low–noise 115 GHz Receiver Using Superconducting Tunnel Junctions," *Appl. Phys. Lett.*, vol. 43, pp. 786 – 788, Oct. 1983.

[1.53] Sutton, E. C., "A Superconducting Tunnel Junction Receiver for 230 GHz," *IEEE Trans. Microwave Theory Tech.*, vol. MTT–31, pp. 589 – 592, July 1983.

[1.54] Harris, A. I., Jaffe, D. T., Stutzki, J., and Genzel, R., "The UCB/MPE Cassegrain Submillimeter Heterodyne Spectrometer," *Intl. J. of IR and Millimeter Waves*, vol. 8, No. 8, pp. 857 – 883, 1987.

[1.55] Woestenburg, E. E. M., Nieuwenhuis, L., "Low–Noise 320 – 360 GHz Cryogenically Cooled Waveguide Schottky Diode Mixer," *Electron. Lett.*, 1986, 22, pp. 1305 – 1307.

[1.56] Keene, N. J., Mischerikow, K. D., Ediss, G. A., Perchtold, E., "Low–Noise 460 GHz Waveguide Schottky Mixer Radiometer For Radioastronomy," *Electron. Lett.*, 1986, 22, pp. 353 – 355.

CHAPTER II.

AN IMPROVED SOLUTION TO INTEGRATED ARRAY OPTICS: THE HYBRID ANTENNA

2.1 Motivation for the Development of a New Antenna System

Remote sensing in the millimeter and submillimeter wavelength bands requires sensitive detectors and well defined antenna beam properties. In this chapter an antenna system that provides such a well defined beam pattern will be described. In the introductory chapter (section 1.5 and [2.1]) a quasi–optical receiver based on a planar logarithmic spiral antenna was presented as an example of a quasi–optical receiver layout. It was noted, however, that the coupling of the receiver to the telescope optics was poor, as it had been with all quasi–optical receivers based on planar antennas in the past. This chapter describes a new antenna – named the hybrid antenna (published in [2.2]) – that eliminates the coupling problems of previous designs. The hybrid antenna has proven to couple as well to the telescope optics as waveguide horn antenna based receivers. Quasi–optical receivers based on planar antennas are desirable since they are much easier to manufacture and are thus preferred candidates for imaging arrays. A more detailed discussion on the differences of waveguide techniques to quasi–optical techniques will follow below. The issue of sensitivity of the detector will only be addressed in this chapter in so far as it depends on the coupling of the detector to the radiation field through the detector's receiving antenna. A specific application of remote sensing in the millimeter and submillimeter wavelength bands – heterodyne spectroscopy in radio astronomy – will be emphasized since the instrumentation developed was aimed at this application. Of course, the same basic principles apply to most other applications for coherent detection of radiation.

Radio astronomy uses large aperture antennas to focus the incoming radiation onto a second, much smaller antenna, which feeds the received power to a detector, either directly or via an impedance matching circuit (see figure 2.1). The properties of the second antenna, *i.e.*, the receiver antenna, and its coupling to the primary antenna, *i.e.*, the radio telescope, will be discussed.

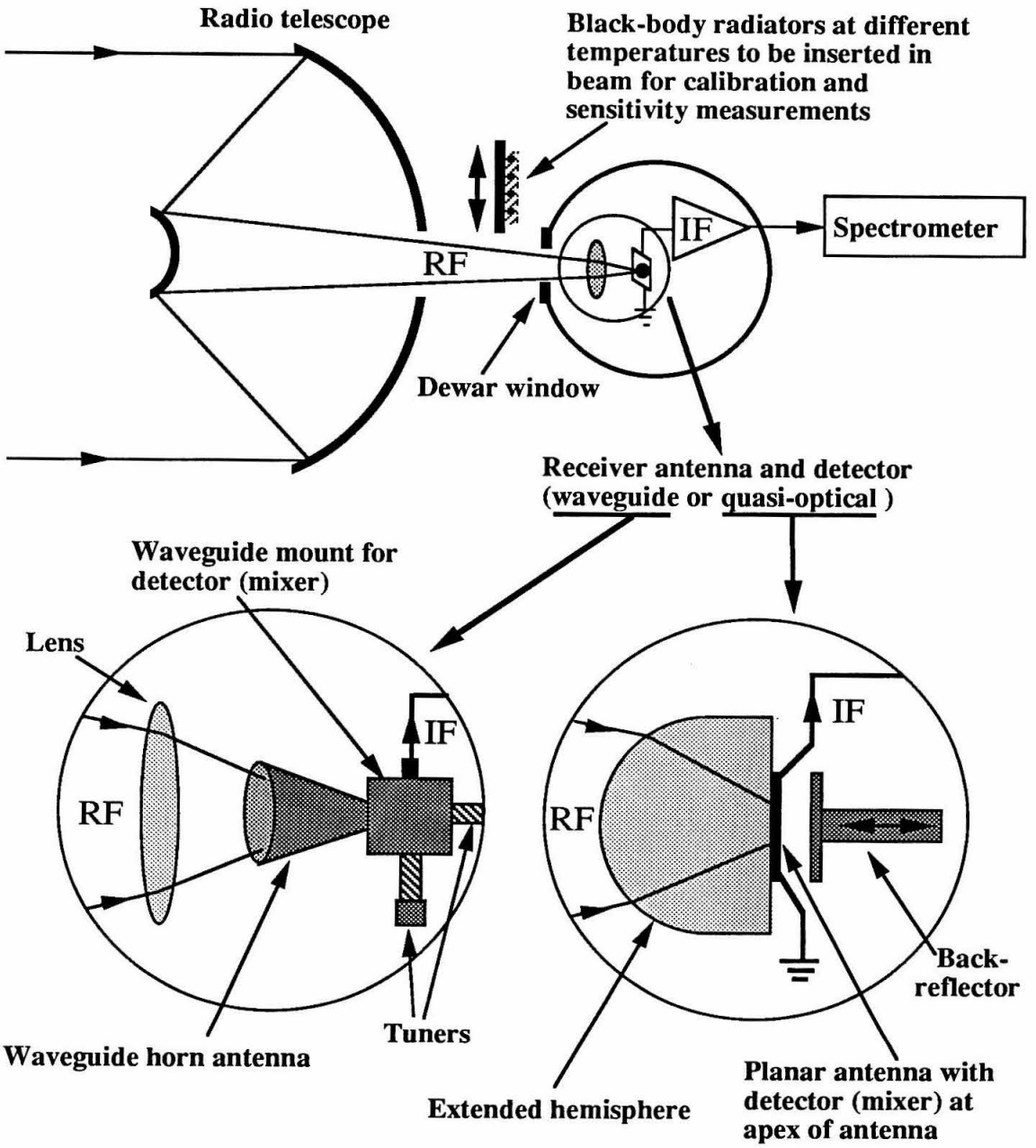


Figure 2.1 Optics layout for a radio telescope. The radiation from the telescope is focused onto the detector via a lens and a waveguide horn or a quasi–optical lens system. Loads of different temperatures are inserted in the optical path between the primary dish and the receiver dewar to measure the sensitivity of the system.

Traditionally, the receiver antennas used are waveguide horn antennas that transform the free space TEM mode coming from the telescope into a waveguide mode where the radiation is detected in a non–linear element suspended across the waveguide (see for example [2.3]). However, in the submillimeter band these waveguide structures become expensive and difficult to manufacture due to their small size. Since the skin depth gets smaller at shorter wavelengths the surface roughness of the walls of the waveguide structures becomes increasingly more important and thus losses will increase. Waveguide horn antennas with the associated metal waveguide structures are also not well suited for array applications since they are traditionally manufactured by machining the individual waveguide components. However, modern approaches use semiconductor lithographic techniques to manufacture parts of the horn antennas [2.4,2.5,2.6,2.7].

An alternate approach to waveguide techniques is to use quasi–optical coupling where the waveguide horn antenna is replaced by a planar antenna on a thick dielectric substrate that supports the antenna (see review by Rutledge [2.8] and references therein). The thick dielectric substrate simulates a semi–infinite dielectric half–space thus preventing the propagation of surface modes. Since broadside planar antennas like the bow–tie antenna [2.9], the logarithmic spiral antenna [2.10,2.11] the double dipole [2.12] or twin slot antennas [2.13,2.14,2.15] have very broad radiation patterns (typically $f/0.5$*) the dielectric substrate is shaped to be a hyperhemispherical lens to reduce the beam pattern's width by n [2.16], where n is the refractive index yielding $f/1$ to $f/2$ depending on the dielectric used. The hyperhemispherical lens uses the aplanatic focus of a sphere at a distance $d = r/n$ from the center of the sphere where r is the radius of the sphere [2.17]. The detector, or an impedance matching circuit feeding the detector, receives the power from the apex of the planar antenna.

The advantages of planar–structure antennas compared to waveguide systems are their low cost of manufacture, ease of installation, applicability to mass production using photo–lithographic techniques [2.8], and lower losses at high frequencies. The down side to this approach is that, so far, lower Gaussian coupling efficiencies have been found [2.18,2.19,2.1], when compared to waveguide horns. Also the beam launched by the hyperhemispherical lens and planar antenna combination is so broad [2.15,2.18,2.1] that additional optics are required to match it to a typical

* f–numbers in this paper are defined through the FWHP angle — see equation (2.1)

beam ($f/6$ to $f/20$) of a Cassegrain–focus radio telescope. However, earlier work with planar antennas on hyperhemispherical lenses like the bow–tie antenna [2.18] or logarithmic spiral antenna [2.1] yielded high receiver sensitivities. This effect of poor coupling but high receiver sensitivity can be understood when the different ways of applying input radiation to the detector are considered. In the case of the coupling efficiency measurements, a single–mode Gaussian beam from the telescope has to be coupled to the receiver antenna, thus the amplitude and phase properties of the receiving antenna's beam are important. For sensitivity measurements, blackbodies are used as sources of radiation in front of the receiver, *i.e.*, between the receiver and the telescope (see figure 2.1). Blackbodies are multi–modal sources, thus all components of the receiver antenna's beam pattern that are not blocked by apertures between the receiver antenna and the outside of the receiver (such as dewar windows, etc.) will receive power from the blackbodies. Therefore these measurements are insensitive to the beam pattern quality and the phase of the receiver's antenna. The beam pattern quality only contributes to the loss of sensitivity through that fraction that is blocked by apertures between the antenna and the blackbodies. Since the log periodic spiral antenna has superior amplitude beam patterns compared to the bow–tie antenna [2.9,2.11], the receiver based on the log periodic spiral antenna [2.1] described in chapter I naturally showed higher sensitivities than one based on the bow–tie antenna [2.18]. However, both receiver systems showed relatively poor coupling to a single mode Gaussian beam from a telescope when compared to receiver systems with waveguide horn antennas. This was probably due to problems in the optics resulting from the challenge in matching the very broad beam launched by the planar antenna ($f/0.5$) to the telescope optics (typically $f/6 - 10$). A more detailed discussion of the reasons for the low Gaussian coupling efficiencies of those previous quasi–optical systems can be found in section 2.3b.

The goal here was to develop a quasi–optical antenna system that would allow the receiver to couple to the telescope optics without any degradation, *i.e.*, antennas with high quality beam patterns, high f–numbers and high Gaussian coupling efficiencies. Ideally, no additional optics should be required between the antenna that launches the beam and the telescope optics. The hybrid antenna, introduced in this chapter, is such an antenna. The new hybrid antenna adds the properties of excellent radiation patterns and high aperture efficiencies to the several existing

advantages of planar antennas. The size of the beam can be designed to match the requirements (*e.g.*, $f/4$ to $f/20$) without any additional optics between the receiver antenna and the telescope. Furthermore the dielectric antenna is very space efficient, *i.e.*, it has a high aperture efficiency (76 %), thus being well suited for heterodyne receiver focal plane array applications. Beam pattern and aperture efficiency measurements of hybrid antennas as well as a theoretical analysis based on geometric ray calculations including diffraction limit effects are discussed in section 2.2. An application of a single hybrid antenna in an SIS receiver and considerations for array optics in section 2.5.

Throughout this chapter the following four nomenclatures will be used :

1) A dielectric sphere of radius r and refractive index n that is cut off at a plane a distance d from its center is called an extended hemispherical lens of extension length d (see figure 2.2).

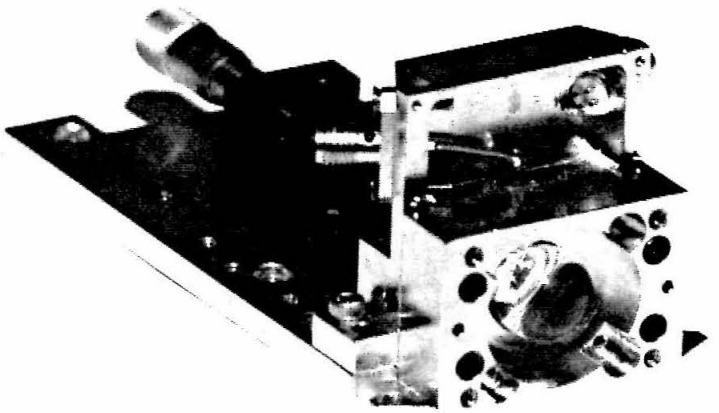
2) If that extension length is at $d = r/n$, *i.e.*, the lens is aplanatic [2.17], it is termed a hyperhemispherical lens.

3) When the extension length d is increased beyond the aplanatic point (for a lens with radius larger than a wavelength) the magnification will increase until the diffraction limit of the lens is reached. At that extension length ($d = d_{opt}$) the diffraction limited lens acts as a lens–antenna and the combination of it with the planar feed antenna mounted on its flat surface is named a *hybrid antenna*. The planar antenna is then called the feed antenna of the hybrid antenna. The reason for choosing the name *hybrid antenna* is that it is made of two antennas, a dielectric lens–antenna that defines the beam pattern of the hybrid antenna through the diffraction limit given by the lens–antenna's radius and a planar antenna that defines the polarization properties of the hybrid antenna. Both of the component antennas contribute to the overall radiation properties of the final antenna. A more detailed discussion of the operating principles of the hybrid antenna will follow later.

4) An elliptical lens–antenna is an ellipsoid cut perpendicular to its major axis at its second geometric focus, with a planar antenna mounted on the flat surface.

Also, throughout this work the following definition for the f–number of an optical system will

a)



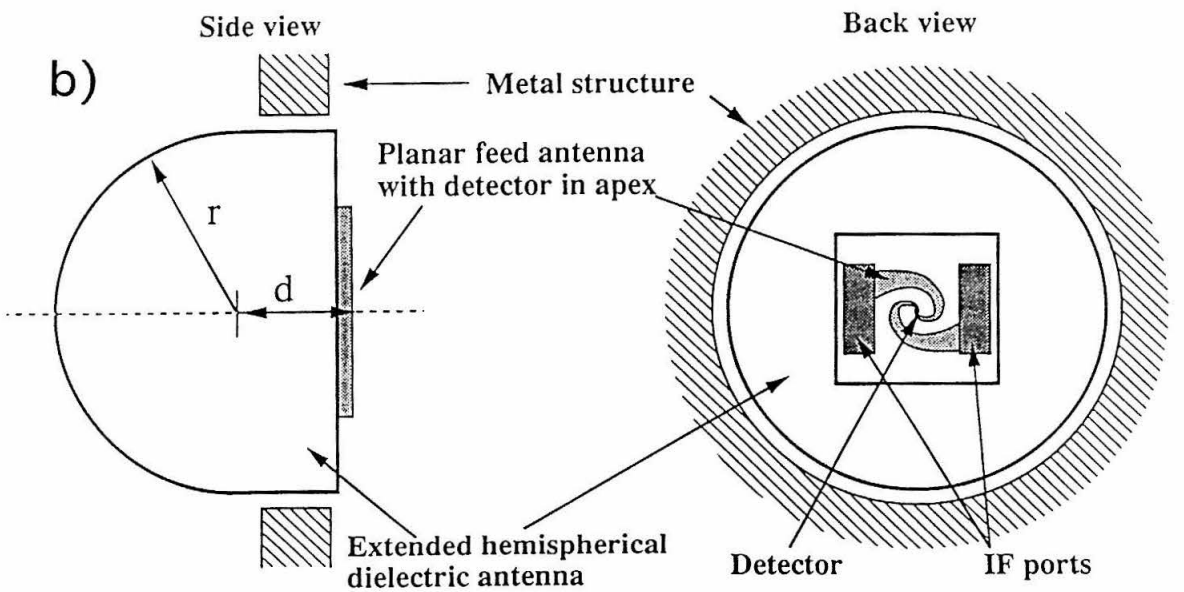


Figure 2.2 a) Photograph of a hybrid antenna in a mixer block. The anti–reflection coating has been removed to show the dielectric lens antenna of the hybrid antenna in the center of the mixer block. The block is mounted on a base plate that holds the micrometer stage, visible in the background, used to adjust the back plane position. b) Schematic of a hybrid antenna in a metal structure. The extension length d is defined from the center of the hemispherical dielectric to the metallization of the planar antenna.

be used:

$$f - \text{number} \simeq \frac{1}{2\tan(\theta_{\text{FWHP}}/2)} \quad , \tag{2.1}$$

with θ_{FWHP} the full width at half power (FWHP) angle of the beam.

Previous measurements of elliptical lens–antennas showed good beam patterns [2.20,2.21] requiring no further optics to couple to a typical Gaussian beam from a telescope, but the important question of Gaussian coupling or aperture efficiency was not addressed. Adding to the known advantages of the elliptical lens–antenna, the hybrid antenna is less expensive to manufacture. The important figure of merit for the hybrid antenna then is the Gaussian coupling efficiency of the system or the aperture efficiency when used in an imaging array receiver. Aperture efficiency measurements of the hybrid antenna, based on total power measurements, and a discussion of Gaussian coupling efficiencies based on calculations and measurements by Filipovic *et al.* [2.22] will be presented in the next section.

In summary, the new hybrid antenna, introduced in this chapter, allows the addition of the properties of excellent radiation patterns and high aperture efficiencies to the advantages of planar antennas. The size of the beam can be designed to match the requirements (*e.g.*, $f/4$ to $f/30$) without any additional optics between the receiver antenna and the telescope. Furthermore the dielectric antenna is very space efficient, *i.e.*, it has a high aperture efficiency, thus being well suited for heterodyne receiver focal plane array applications.

2.2 Hybrid Antenna Concept, Theory and Properties

Two issues, those of quality of beam patterns and aperture efficiency of the antenna, have to be addressed for an antenna in a quasi–optical imaging array receiver. For a single element receiver the Gaussian coupling efficiency is most important, but in an array receiver the aperture efficiency of the individual antennas is important, since it is a measure of the efficiency in the use of focal plane space with which the antennas sample the incoming radiation. In general, these properties are, of course, related. However, for simplicity they will be treated separately.

a) Physical description of the hybrid antenna

Figure 2.2 shows a photograph and a schematic diagram of the hybrid antenna with all accessories

as discussed below. Thinking of the antenna as a transmitter, the radiation is fed into the system by a planar antenna that uses the extended dielectric hemisphere as a substrate. Planar antennas suffer from power loss to substrate modes when the dielectric substrate is of comparable thickness to a wavelength, but mounting the planar feed antenna on a substrate lens antenna eliminates this problem by simulating a semi–infinite half–space of dielectric for the planar feed antenna. This also causes the feed antenna to radiate preferentially in the direction of the dielectric. For a dielectric constant of $\epsilon_r = 3.8$ and a spiral feed antenna, the ratio of power radiated into the dielectric to that radiated to the opposite face was found to be about 7 dB. This ratio depends on the beam width of the planar antenna and will increase for wider beams and higher dielectric constants ϵ_r [2.8]. The planar spiral antenna used throughout this work is identical to the one described in chapter I, which is a two–turn, self–Babinet–complimentary structure, with a diameter of about $3mm$ and an opening angle of 30°. A metal back plane on the free space side of the feed antenna was used to reflect forward that power which would otherwise be lost from the beam. To verify that the back reflector does not impact the beam patterns but acts only to recover the power otherwise lost, it was replaced with an absorber. When the back reflector was positioned for peak response, *i.e.*, about $\frac{3}{8}\lambda$ away from the planar feed antenna, the patterns were identical to those measured with an absorber. The back reflector can be eliminated by using a dielectric substrate of very high dielectric constant, such as high resistivity silicon ($\epsilon_r = 11.7$), since the power radiated into the free space direction is then negligible. However, the transition from the front surface of the dielectric lens–antenna to free space is then more critical, requiring the use of an anti–reflection coating.

b) Beam pattern measurements and concept of the hybrid antenna

The beam pattern measurements were performed using a computer controlled full two–dimensional angular far–field scanning antenna range in a microwave absorbing chamber. The source for the 115 GHz measurement was a Gunn oscillator, and for frequencies up to 500 GHz Gunn oscillators followed by multipliers were used. The measurement at 584 GHz used a far infrared laser system for the source. The distance between the source and the hybrid antenna was about 1m. The sources were all linearly polarized and modulated with a chopper wheel. The detector for the power received by the antenna was a bismuth bolometer placed at the apex of the planar antenna

[2.23]. The bolometer was DC–biased and the chopped signal amplified with a lock–in amplifier. The dynamic range of the set–up was about 25 dB. To get a better dynamic range than the one achieved here with room temperature techniques would require the use of different detectors such as Schottky diodes. However, bolometers were chosen since they could be manufactured lithographically in situ with the antenna structure rather than having to mount a separate detector in the apex of the antenna. The size of the bolometers is about $1\mu m$, enabling the antenna measurements to be performed in the submillimeter band without having the size of the detector affect the characteristics of the antenna system.

The concept of the hybrid antenna is to use the dielectric substrate lens itself as a radiating aperture antenna by choosing the extension length d of the extended hemispherical lens to be large enough so that the f–number of the lens is increased to the point where, for its particular diameter, the diffraction limit is reached. The beam launched from that position will approximate a wave with constant phase outside the dielectric antenna. Figure 2.3 shows measurements of beam patterns as a function of the extension length d of an extended hemispherical lens performed at 115 GHz. The extension length d was increased in the measurements by adding quartz slabs of $0.254mm$ thickness between the flat surface of the dielectric lens and the substrate of the planar antenna. The radius r of the dielectric lens antenna used was $r = 6.35mm$ and the refractive index of the fused quartz dielectric is $n = 1.95$. The quality of the patterns increases when the distance d is increased from the hyperhemispherical case of $d = r/n = 3.25mm$ up to the point $d = d_{opt}^{meas} = 4.27mm$ where the beam is diffraction limited and the sidelobes are at a minimum. A further increase of d then raises the sidelobe level while the main beam remains diffraction limited.

Figure 2.4 shows the excellent beam pattern quality of a hybrid antenna in a two–dimensional linear scale depiction (2.4a) and two perpendicular cuts in a logarithmic scale depiction (2.4b,c). The pattern was taken at 115 GHz with a 12.7 mm diameter fused quartz lens of dielectric constant $\epsilon_r = 3.8$ with the hybrid antenna mounted in a metal mixer block (see figure 2.2) like the one used in the SIS receiver with a back reflector, as described later. The measurements were performed in a metal mixer block, as encountered in most applications, so as not to exclude the possibility of problems of distortions of the beam pattern arising from the proximity of conducting surfaces to

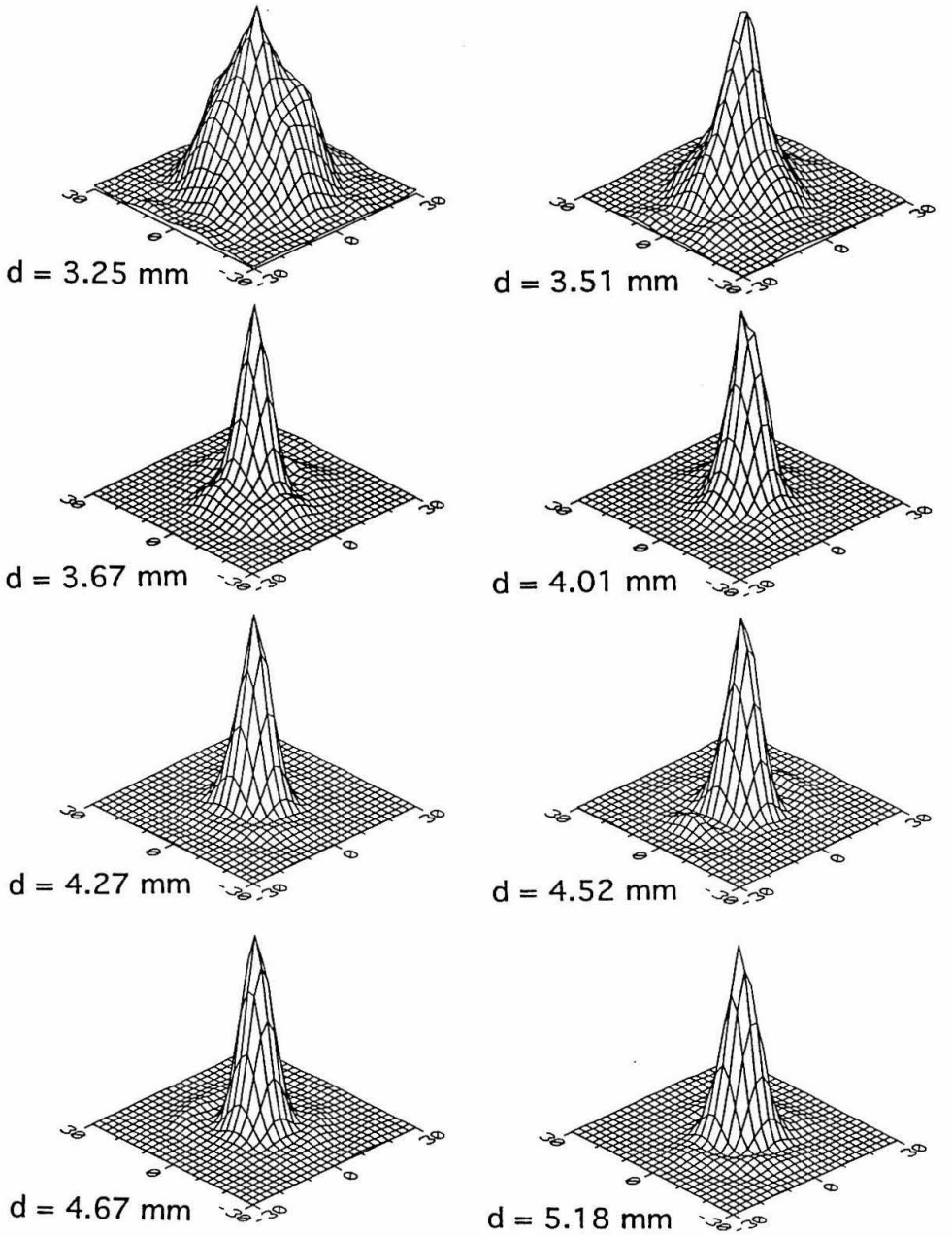


Figure 2.3 Beam pattern measurements in linear intensity depiction for different distance parameters d. The radius of the hybrid antenna was $6.35mm$, the dielectric constant ϵ_r = 3.8 and the frequency 115 GHz (λ = $2.6mm$). The position of d = $3.25mm$ corresponds to the aplanatic case (*i.e.*, a hyperhemispherical lens) where $d = r/n$. The measured beam for this position is that of the planar spiral antenna reduced in width by n by the lens. At $d = 4.27mm$ the beam pattern is diffraction limited and a further increase of d only increases the sidelobe levels. This position is the experimentally determined optimum position $d_{opt}^{meas}(115GHz)$. The beam pattern is now defined by the radius of the lens rather than the beam properties of the planar feed antenna.

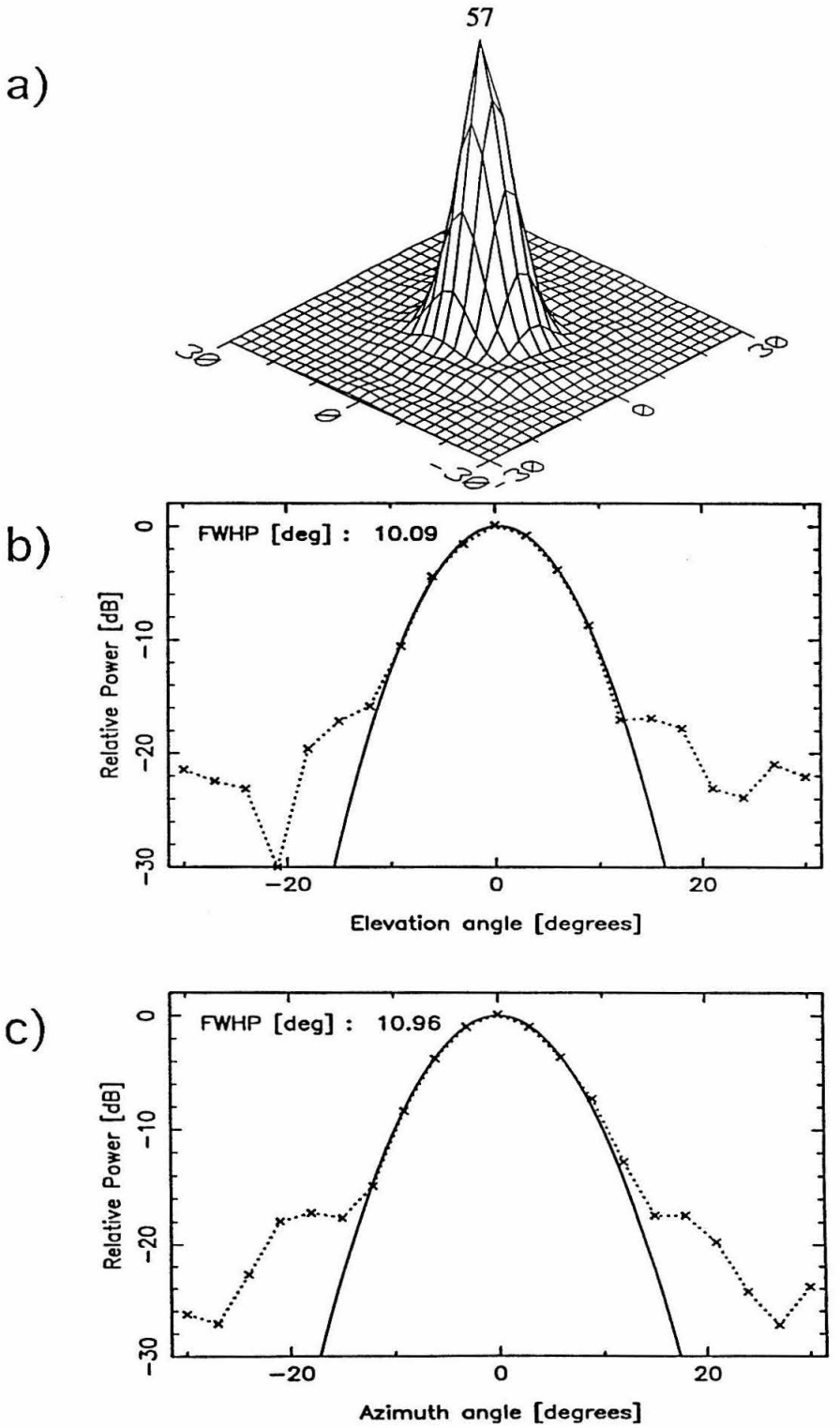


Figure 2.4 Beam pattern of a hybrid antenna with $\epsilon_r = 3.8$, $r = 6.35mm$, $d = d_{opt}^{meas} = 4.27mm$ at 115 GHz. a) the full two–dimensional pattern on a linear scale. b) and c) two perpendicular cuts with a logarithmic intensity scale. The solid line is a best fit Gaussian profile and matches the measured data very well down to about – 17 dB.

the hybrid antenna. The metal of the mixer block in the configuration used is concentric around the hybrid antenna in the same plane as the planar antenna with a distance from the apex of the planar antenna equal to the radius of the extended hemisphere (see figure 2.2).

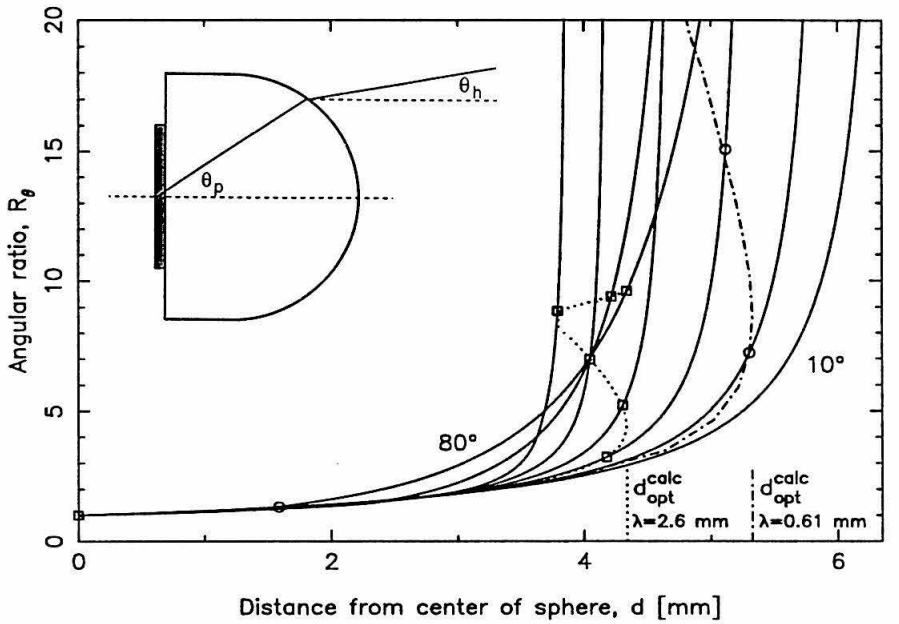


Figure 2.5 Geometric ray calculations including the diffraction limit for hybrid antennas. The angular ratio is the ratio of an angle θ_p of a ray as launched by the planar antenna inside the dielectric to the angle θ_h of the ray launched by the hybrid antenna (see insert in upper left corner). This ratio increases monotonically with increasing distance d of the position of the planar antenna to the center of the extended hemispherical lens. Rays are shown (solid lines) starting at an angle $\theta_p = 10^o$ with increments of 10^o up to 80^o. The intersection of the dotted line ($\lambda = 2.6mm$ or 115 GHz) and dash–dotted line ($\lambda = 0.61mm$ or 492 GHz) with the solid lines indicate where the angular ratio is large enough to have the ray, as launched by the hybrid antenna, to be within the diffraction limit of the dielectric lens antenna. For a distance $d = d_{opt}$ (indicated by the dotted and dashed–dotted vertical lines) all rays launched by the planar antenna are within the diffraction limit. This position depends on the wavelength through the diffraction limit. Beam pattern measurements, as shown in figure 2.3 for 115 GHz agree very well to the d_{opt}^{calc}(F) predicted by the calculations shown in this figure.

To investigate the optimum extension length d_{opt} dependance on wavelength, measurements at 492 GHz were carried out the same way as those described above at 115 GHz shown

in figure 2.3. A new optimum position $d_{opt}^{meas}(492\,\mathrm{GHz})$ was found with $d_{opt}^{meas}(492\,\mathrm{GHz}) > d_{opt}^{meas}(115\,\mathrm{GHz})$. This effect can be understood from figure 2.5, which is a geometric ray calculation including the effects of the diffraction limit for a lens of same refractive index and radius as those of the measurements. Defining the angular ratio R_θ as the ratio of the angle of a ray launched by the planar antenna to that of the ray as it leaves the extended hemispherical lens, figure 2.5 shows R_θ as a function of extension length d for a set of different rays launched by the planar antenna. As d increases from zero, *i.e.*, a hemispherical lens, towards the length where the extended hemispherical lens approximates an elliptical lens at its second geometric ray focus, R_θ increases from unity to infinity. However, for wavelengths λ comparable to the radius r of the lens, geometric ray optics alone is not a good approximation anymore, but requires modification due to the diffraction limit of the lens, which is governed by the radius of the lens. The full width at half power (FWHP) diffraction angle θ_{FWHP} is given by [2.17]

$$\theta_{FWHP} \simeq \frac{1.2\lambda}{2r}. \tag{2.2}$$

In figure 2.5 the positions where this diffraction angle is reached for the ray leaving the extended hemisphere are denoted by boxes for 115 GHz (2.6 mm) and circles for 492 GHz (0.61 mm). The dotted and dash–dotted line connect those points and were calculated using a larger sample of rays launched by the planar antenna. The important feature to note is that there is an extension length d_{opt}^{calc} beyond which it is not necessary to increase d since every ray launched by the feed antenna is either already within the diffraction limit (given by (2.2)) of the beam leaving the hybrid antenna or is refracted into it by the extended hemispherical lens, as demonstrated in figure 2.5.

The calculated optimum extension length d_{opt}^{calc} can be determined numerically from graphs as the ones shown in figure 2.5. Figure 2.6 shows a plot of the ratio d_{opt}^{calc}/r versus the radius of the lens r measured in units of wavelength. Comparing the calculated and measured optimum positions one finds them to agree very well and the results are listed in table 2.1.

Table 2.1. Comparison of calculated to measured optimum extension length d_{opt}.

frequency / Wavelength	115 GHz / 2.6 mm	492 GHz / 0.61 mm
calculated: d_{opt}^{calc}	4.33 mm	5.34 mm
measured: d_{opt}^{meas}	4.3 ± 0.2 mm	5.4 ± 0.2 mm

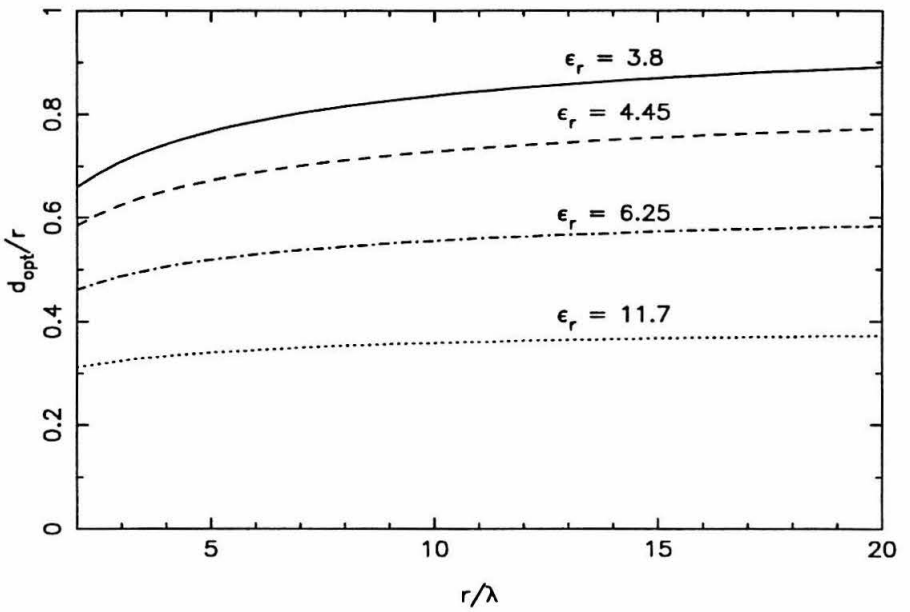


Figure 2.6 The optimum extension length as a fraction of the radius of the lens, d_{opt}^{calc}/r, as a function of radius of the lens measured in units of wavelength, r/λ, for different dielectric constants (ϵ_r = 3.8 : fused quartz, ϵ_r = 4.45 : single crystal quartz, ϵ_r = 11.7 : high resistivity silicon). The geometric ray method of figure 2.5 was used to generate the results shown here. A hyperhemispherical lens would yield a flat line at $1/\sqrt{\epsilon_r}$.

A hybrid antenna that is to cover a wide range of frequencies must use a $d_{opt}(\mathrm{F}_h)$ as determined for the highest frequency F_h. As shown later (see table 2.4) the aperture efficiency at the lowest frequency F_l will then be slightly lower than the optimum attainable for that frequency, since $d_{opt}(\mathrm{F}_h) > d_{opt}(\mathrm{F}_l)$. However, unless the operating range is more than an octave, the reduction in aperture efficiency is typically well under 10%.

Figure 2.7 shows beam pattern measurements at 214, 321, 492 and 584 GHz for hybrid antennas with $d(\mathrm{F}) \geq d_{opt}^{calc}(\mathrm{F})$. The 214 and 321 GHz measurements used low efficiency multipliers to generate the transmitter signal, thus the lower signal to noise levels. The 492 GHz and 584 GHz measurements yielded signal to noise ratios as good as in the 115 GHz measurements due to narrower beams and higher available transmitter power using a high efficiency Gunn multiplier chain [2.24] and a far infrared laser system, respectively.

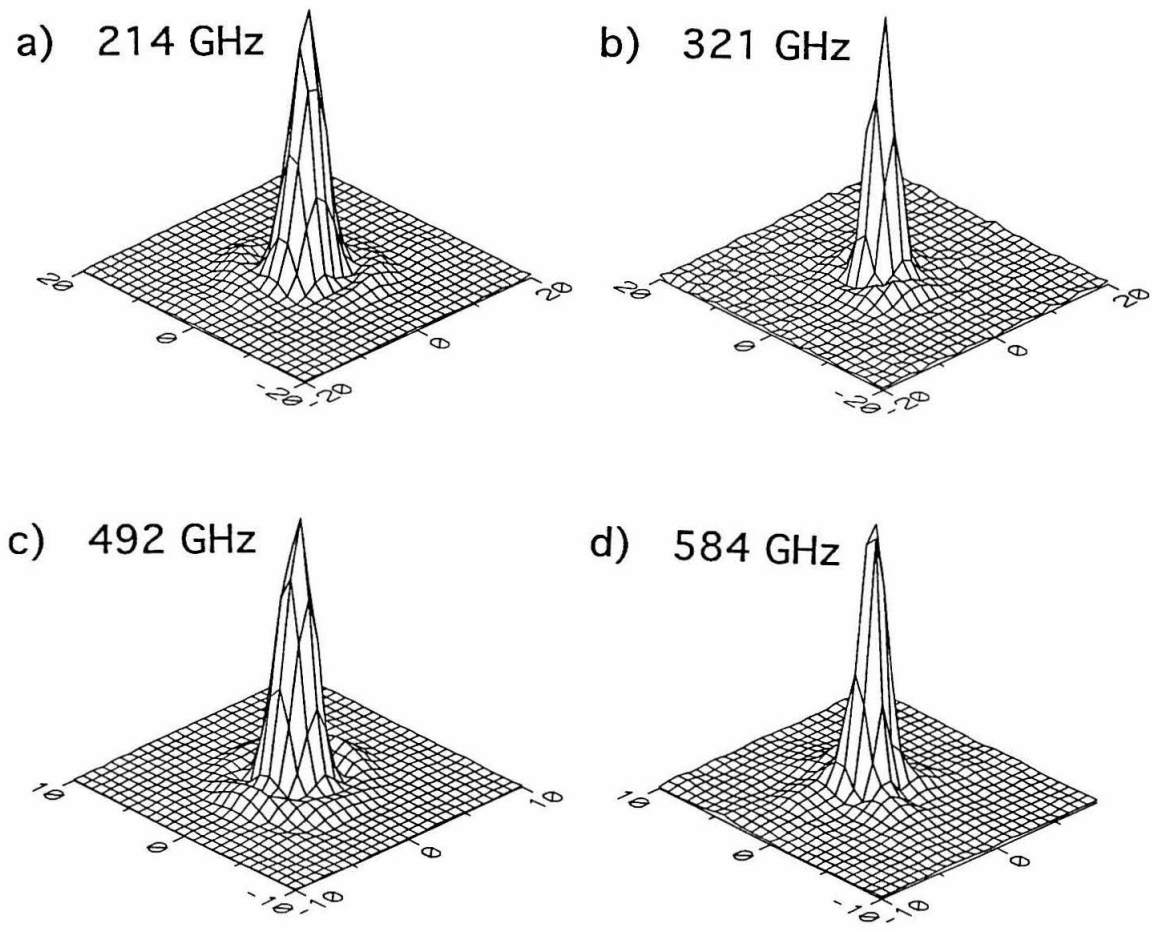


Figure 2.7 Beam patterns of hybrid antennas at different frequencies. a) 214 GHz, ϵ_r = 3.8, r = 6.35mm, d = 5.18mm (d_{opt}^{calc} = 4.80) b) 321 GHz, ϵ_r = 3.8, r = 6.35mm, d = 5.18mm (d_{opt}^{calc} = 5.07) c) 492 GHz, ϵ_r = 3.8, r = 6.35mm, d = 5.44mm (d_{opt}^{calc} = 5.33) d) 584 GHz, ϵ_r = 4.45, r = 10.0mm, d = 8.01mm (d_{opt}^{calc} = 7.71). The extension length d was not optimized in these measurements, but set to have d(F) $\geq$ d_{opt}^{calc}(F).

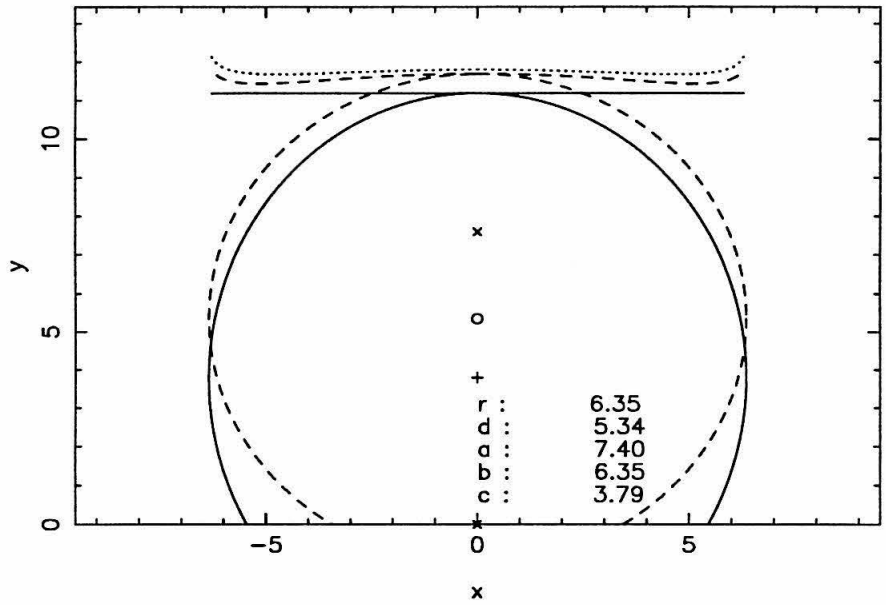


Figure 2.8 Truncated ellipsoidal lens with wavefront (solid lines) and extended hemispherical lens with wavefront (dashed line) from geometric ray optics calculations. The dotted line is the difference between the wavefront of the ellipsoidal lens and the extended hemispherical lens after a quadratic term for refocusing was removed. The radius of the extended hemispherical lens is $r = 6.35mm$, $\epsilon_r = 3.8$ and $d = d_{opt}^{calc} = 5.34mm$ as determined from figure 2.5. The elliptical lens has the same length of its minor axis b as the radius r of the hemisphere. The (x) symbols denote the foci of the ellipse, which are a distance $c = a/n$ from the center (+) of the ellipse. The major axis a of the ellipse is then determined from $a^2 = c^2 + b^2$. The circle (o) denotes the center of the spherical lens.

c) Calculated reduction of aperture efficiency from phase errors

In the limit of very high frequencies and large lens sizes, *i.e.*, the geometric ray approximation, the hybrid antenna would approach that equivalent to the second geometric ray focus of an ellipsoid of revolution. For these conditions it would be advantageous to actually use an elliptical lens rather than an extended hemispherical lens. In the geometric ray approximation an elliptical lens focuses parallel light to a single point within the lens, *i.e.*, with no abberations, whereas an extended hemispherical lens will have some abberations (see figures 2.8 and 2.9). However, in those cases where the lens size is not much larger than the operating wavelength, an elliptical lens can be approximated with a much lower cost extended hemispherical lens, which was done

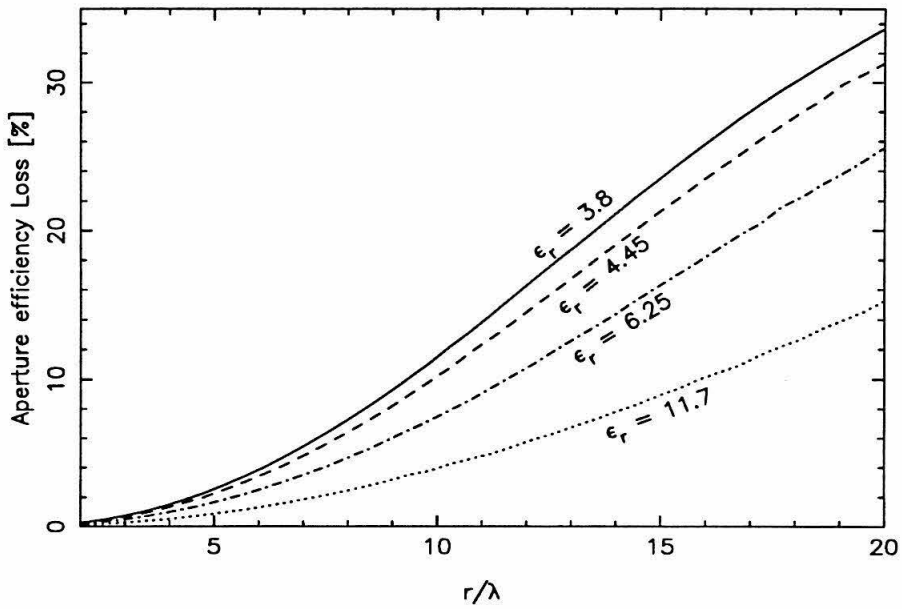


Figure 2.9 Calculations of loss of aperture efficiency due to phase errors of an extended hemispherical lens as a function of the ratio of the radius of the lens to the wavelength, r/λ, for different dielectric constants (ϵ_r = 3.8 : fused quartz, ϵ_r = 4.45 : single crystal quartz, ϵ_r = 11.7 : high resistivity silicon). A uniform amplitude in the aperture plane was chosen.

in all measurements of this thesis. The reduction of aperture efficiency, due to phase errors in the aperture plane, for an extended hemisphere versus a truncated elliptical lens, depends on the diameter of the lens, the wavelength and the refractive index of the lens material. Figure 2.8 shows the phase fronts as calculated with geometric ray optics for two different 12.7 mm diameter lenses with a refractive index of n = 1.95 at 500 GHz. One lens is a truncated ellipsoidal lens where the wavefront and lens are shown by solid lines and the other lens is an extended hemispherical lens with wavefront shown as dashed lines. The dotted line shown in figure 2.8 is the difference between the wavefront of the ellipsoidal lens and the extended hemispherical lens after a quadratic term for refocusing was removed. The parameter d of the extended hemispherical lens is $d = d_{opt}^{calc} = 5.34mm$ as determined from figure 2.5. The elliptical lens has the same length of its minor axis b as the radius r of the hemisphere. The (x) symbols denote the foci of the ellipse, which are a distance $c = a/n$ from the center (+) of the ellipse. The major axis a of the

ellipse is then determined from $a^2 = c^2 + b^2$. The circle (o) denotes the center of the spherical lens.

The aperture efficiency loss due to the phase error $\chi(\rho, \phi)$ is calculated from

$$A_{loss}^{phase} = 1 - |< e^{i\chi} >|^2 = 1 - \frac{|\int_0^{2\pi}\int_0^r e^{i\chi(\rho,\phi)}\rho d\rho\ d\phi|^2}{[\int_0^{2\pi}\int_0^r \rho d\rho\ d\phi]^2} \tag{2.3}$$

and is about 10%. The electric field is assumed to be constant in amplitude across the aperture. The onset of sidelobe shoulders at about – 17 dB, as shown in figure 2.4b and c, is a typical signature of an Airy pattern from the constant illumination in phase and amplitude [2.17] of a circular aperture and are consistent with the above assumption. However, it is important to stress that there can be very different illumination functions that will still produce beams with sidelobes at – 17 dB. Most of the phase errors occur at the edges of the aperture as can be seen in figure 2.8. Since different planar antennas will yield different illumination functions, especially at the edges, a constant amplitude in the aperture plane was chosen. This was done to simplify comparison of the different parameters of figure 2.9. A constant amplitude in the aperture plane corresponds to a FWHP beam angle for the planar antenna inside the dielectric of about $f/0.5$. Again, a quadratic term in the phase front was removed for figure 2.9 since this reflects only a different focusing position. The ratio of the radius r of the lens to the wavelength λ is proportional to the f–number of the hybrid antenna, since from (2.1) and (2.2)

$$f - \text{number} \simeq \frac{1}{2\tan(\frac{0.3\lambda}{r})} \quad . \tag{2.4}$$

Figure 2.9 shows the calculated reduction in aperture efficiency as a function of r/λ for different refractive indices. The reduction in aperture efficiency increases with increasing size of the lens since the approximation of an extended hemispherical lens to an elliptical lens becomes worse, which will increase abberations. A higher dielectric lens will have less abberations since an elliptical lens of that material will be closer to a sphere. The extended hemispherical lens is thus a better approximation to the elliptical lens.

Table 2.2. Beam pattern measurements summary

diam. [mm]	freq. [GHz]	FWHP(E)	FWHP(H)	$f\#$	$f\# \cdot \lambda$ [mm]
6.35	115	20.3	17.8	3.0	7.8
6.35	208	11.6	11.3	5.0	7.2
6.35	492	4.86	5.39	11.2	6.8
12.7	115	10.9	10.2	5.4	14.1
12.7	208	5.4	6.3	9.8	14.2
12.7	214	4.9	6.1	10.4	14.6
12.7	321	4.0	4.1	14.2	13.3
12.7	428	2.85	2.91	19.9	13.9
12.7	492	2.92	2.52	21.1	12.8

Table 2.2 summarizes beam pattern measurements performed between 115 GHz and 492 GHz with dielectric antennas of two different diameters: 6.35 mm and 12.7 mm. Measurements done at 584 GHz are not listed since the lens used for those measurements (20 mm diameter) was from a different supplier and had significant surface errors resulting in abberations which dominated the measurement results. The beam size is given as the full width at half power (FWHP) in the E– and H–plane of the transmitting horn antennas. The $f\#$ is calculated from the geometric mean FWHP angle θ_{FWHP} via equations (2.1) and (2.2). The product, $f\# \cdot \lambda$, yields the spot size in the image plane and should correspond to the diameter of the dielectric lens–antenna, if the antenna behaves as a diffraction limited, uniformly illuminated aperture. As shown in table 2.2, this is approximately the case for all the measurements. However, table 2.2 shows a general trend for $f\# \cdot \lambda$ to decrease with increasing frequency. By inspection, it can easily be verified that this is not due to the tan–function affecting the low f–number results more than the high f–number results. The systematic decrease of $f\# \cdot \lambda$ is attributed to an increase in the measured beam width due to increasing phase errors. In addition to the phase errors discussed previously, there could be phase errors from surface inaccuracies of the lens. The lenses used have a surface accuracy of better than $2\mu m$. The loss of coupling efficiency L can be estimated from the Ruze [2.25] formula for telescopes, modified for a lens with refractive index n :

$$L = 1 - e^{-(2\pi(n-1)E_{RMS}/\lambda)^2}, \tag{2.5}$$

which is negligible at submillimeter wavelengths for $E_{RMS} \approx 2\mu m$. It is thus concluded, that the increased beam sizes are caused by phase errors from abberations as discussed in the text accompanying figures 2.8 and 2.9.

2.3 Aperture Efficiency Measurements and Comments on Gaussian Coupling

a) Aperture efficiency measurements

Beam pattern measurements can usually be performed rather easily whereas aperture and Gaussian coupling efficiency measurements require absolute power calibration, which can be difficult at millimeter and submillimeter wavelengths. Laboratory measurements at 115 GHz with a planar–logarithmic–spiral–structure as the feed antenna of a hybrid antenna were performed and an aperture efficiency of $(76 \pm 6)\,\%$ was obtained. These measurements were performed at room temperature with a bismuth bolometer at the apex of the planar feed antenna. The manufacture of the bismuth bolometers and their responsivity calibration have been described by Neikirk *et al.* [2.23]. The measured aperture efficiency depends on absolute power measurements done with the bolometer, which was thermally calibrated with direct currents provided through the bias circuit. Figure 2.10 shows a measurement of this crucial responsivity calibration for the bolometer.

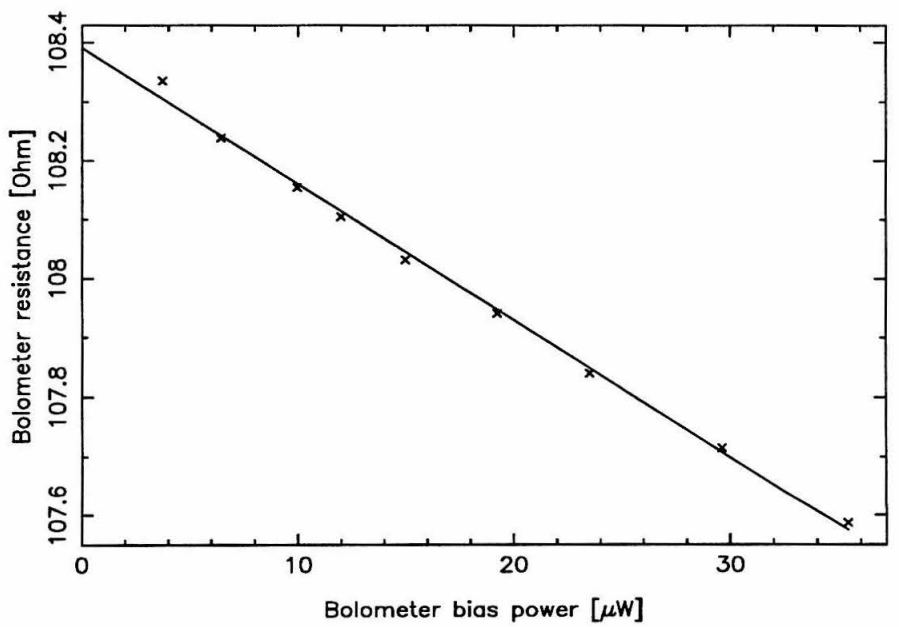


Figure 2.10 Total power bolometer calibration for efficiency measurements. The bolometer is calibrated by changing the DC–bias of the bolometer and measuring its change in resistance. The responsivity of the bolometer is the slope of the curve shown multiplied with the bias current of the bolometer and is 23 V/W at 1 mA bias current.

For the RF measurements the extended hemisphere was covered with a quarter–wave anti–reflection coating of polyethylene to avoid reflection from the dielectric surface, and the back reflector was positioned for maximum response. In the design presented here, the hybrid antenna is fed by a planar logarithmic spiral antenna, which accepts elliptical polarization [2.1]. The polarization of the hybrid antenna is therefore elliptical too. The transmitter used a standard gain horn with linear polarization. Two measurements with the transmitter horn rotated by 90° were performed and the received power from both measurements was added together. The difference between the received power for the two perpendicular linear polarizations of the transmitter measurements was less than 10% showing that the hybrid antenna with a logarithmic spiral antenna is nearly circularly polarized, *i.e.*, the eccentricity of the elliptical polarization is small. By adding the power of the two polarization measurements together the hybrid antenna's circular co–polarized component is added to the circular cross–polarized component. In millimeter and submillimeter wavelengths radio astronomy the signal is typically randomly polarized so that the addition correctly yields the aperture efficiency as applicable for a radio–astronomical receiver. However, some receivers – like Schottky diode receivers, which require high local oscillator power levels – may have polarizing optics in front of the mixer eliminating one polarisation of the signal.

No correction was made for any mismatch between the antenna impedance and the bolometer, since the resistance of the feed antenna's arm material was not well known and the bolometer's resistance could not be measured without the feed antenna in series. The thickness of the antenna arms was approximately $0.2\mu m$ and RF losses due to the surface resistance of the antenna arms were also not taken into account. The actual efficiency will therefore be higher than quoted here. However, these effects are estimated to be less than 5%.

Subsequent to the measurements discussed here, efficiency and beam pattern measurements using planar Schottky diodes soldered into the apex of a logarithmic periodic antenna at 90 GHz and 180 GHz (G. Rebeiz, private communication, 1991) and with a double slot antenna at 246 GHz [2.22] were performed. They confirmed the measurements of this work with higher signal to noise levels for the pattern measurements and calculated similar aperture efficiencies from the pattern measurements.

To calculate the aperture efficiency from the received and transmitted power, P_r and P_t

respectively, Friis' transmission formula [2.26] is solved for the effective aperture of the hybrid antenna

$$A_e = \frac{P_r}{P_t} \frac{\ell^2 \lambda^2}{A_{et}}, \tag{2.6}$$

with ℓ the distance between the transmitting antenna and the receiving hybrid antenna, and A_{et} the effective aperture of the transmitting antenna. The physical aperture of the hybrid antenna with a lens radius of $r = 6.35mm$ is $A_p = \pi r^2 = 127mm^2$. The effective area of the transmitting antenna, a standard gain horn (Alpha Ind., model F861 – 33), was calculated [2.27] and also measured in a symmetric (transmit/receive) setup using two identical standard gain horns. The effective area of the horn was found to be $A_e(horn) = (142 \pm 9)mm^2$. The effective area of the hybrid antenna is

$$A_e(hybrid) = (95 \pm 7)mm^2 \tag{2.7}$$

and thus for the aperture efficiency

$$\eta_a = 0.76 \pm 0.06. \tag{2.8}$$

The error in the measurement is mostly due to the uncertainty in the measurement of the effective area of the transmitting horn antenna (1σ: 6%) and the absolute power calibration of the bolometer (1σ: 5%). Also note, that all measurements were made in a realistic environment for the hybrid antenna, *i.e.*, in a metal mixer block rather than idealized conditions.

For applications requiring only one polarization, the cross polarized power would have to be subtracted, reducing the aperture efficiency by that fraction. Using a linearly polarized planar logarithmic periodic antenna as the feed antenna for the hybrid antenna, a maximum cross–polarized beam of – 7 dB relative to the co–polarized beam was found. However, the cross–polarized component of a log–periodic antenna has been found to vary with frequency [2.28] and lies between – 5 and – 15 dB. The cross–polarized beam pattern followed the co–polarized pattern so that it only reduces the aperture efficiency for applications with a singly polarized source. If linear polarization is a requirement for a particular application but multi–octave bandwidth can be sacrificed, work by Rogers and Neikirk [2.14] and Zmuidzinas and LeDuc [2.15] with a double slot antenna suggests that this planar antenna is a good choice as a feed antenna for a hybrid antenna. This was recently verified by Filipovic *et al.* [2.22], who made beam pattern measurements of

a hybrid antenna with high dynamic range (40 dB) thus allowing them to calculate the aperture efficiency. The calculation yielded an aperture efficiency of $(73 \pm 5)\%$ for a hybrid antenna with a twin slot feed antenna at 246 GHz, which is in good agreement with the results of this work.

b) Comments on the Gaussian coupling efficiency

From theory, using ray–tracing inside the dielectric lens and electric and magnetic field integration on the spherical surface of the lens, Filipovic et al. [2.22] find the Gaussian coupling efficiency (GCE) of the hybrid antenna reduced by about 8% compared to a hyperhemispherical lens (the aplanatic, thus aberration–free case, of an extended hemispherical lens) system, which they calculate to have a GCE of 97%. However, they were unable to experimentally verify the higher GCE for the aplanatic optics but rather measured a lower GCE (87%) for the aplanatic case compared to the hybrid antenna case (GCE$\simeq$ 89%). The reasons are probably the same as in the case of earlier receivers with aplanatic hyperhemispherical lenses which showed poor coupling to a single Gaussian mode of a telescope, as mentioned above. In the case of receivers this effect was even stronger than for the measurements of Filipovic *et al.*. Filipovic *et al.* used room temperature optics that were optimized, yielding focusing parameters different from the ones predicted by Gaussian optics calculations, whereas the receiver optics with cryogenically cooled SIS detectors allow for much less optimization thus possibly yielding much worse results. The discrepancy between the optimized positions of components in the experiment and their calculated positions using Gaussian optics are most likely due to the Gaussian optics formalism breaking down for very low f–numbers as are encountered in the optics of the aplanatic hyperhemispherical lens. In addition to optimizing the position of optical components the shape of the lenses used to couple the power from the transmitter to the hyperhemispherical lens would also have to be optimized, which was not done, causing lower GCE. Note that all above GCE values are only relative since Filipovic *et al.* did not carry out any total power measurements but normalized the measured GCE to calculated ones for an extension length they call the simulated elliptical lens position, which they derive from geometric ray considerations not including wavelengths effects.

In conclusion, the GCE of the hybrid antenna is high with about 89% (ignoring reflections of the surface of the lens) and possibly close to the best achievable for an extended hemispherical lens system without complex optics following it. This allows good coupling of a hybrid antenna

based receiver to a single mode beam from a radio astronomical telescope, as verified at the Caltech Submillimeter Observatory (CSO) at 345 and 492 GHz (see chapter IV or [2.29,2.30]).

2.4 Application of a Single Hybrid Antenna in an SIS Receiver

A single hybrid antenna was successfully tested at the Caltech Submillimeter Observatory (CSO), a 10.4 m diameter submillimeter telescope on Mauna Kea, Hawaii, in an application with a superconducting insulator superconductor (SIS) detector in heterodyne mode and an RF matching circuit integrated on the arms of a planar logarithmic spiral feed antenna (see chapter III or [2.31]). Aperture, main beam and forward efficiencies of the radio telescope with a hybrid antenna based receiver [2.29,2.30] (also see chapter IV) and scalar–feed horn waveguide receiver systems [2.3,2.32] were measured at 345 GHz and 492 GHz. When the respective efficiencies were compared between the hybrid antenna based receiver and the waveguide horn based receiver, they were found to be identical within the measurement uncertainties ($< 10\%$). These efficiencies include the coupling efficiency between the telescope and the receiver, besides other factors, which depend on the performance of the telescope, that are constant at each frequency when one receiver is replaced with another one. Since the telescope performance, independent of the receiver coupling, is not known well enough, an absolute efficiency for the receiver coupling to the telescope can not be deduced. Thus, the only conclusion that can be drawn is that the coupling of the hybrid antenna based receiver to a single mode Gaussian beam from a telescope is thus about the same as that of a scalar–feed horn waveguide receiver. This is the first quasi–optical receiver tested on a radio astronomical telescope to achieve such good performance. Figure 2.11 shows a double sideband spectrum taken in the core of the Orion molecular cloud (OMC–1) with the two sidebands centered at 492.16 GHz and 494.96 GHz [2.30]. Note that the good coupling between the hybrid antenna and the telescope optics is due to the high quality beam patterns of the hybrid antenna and is not necessarily a statement about the intrinsic efficiency of the hybrid antenna itself. The hybrid antenna's coupling efficiency affects the sensitivity of the receiver. Table 2.3 shows the sensitivities obtained with the receiver system, expressed in double sideband noise temperatures. The sensitivities obtained are very high and approach those of the best waveguide receivers [2.3,2.32]. The increase of noise temperature at 492 GHz is due to the fact that the

lithographic matching circuit, which is designed to tune out the SIS junction capacitance and will be described in the next chapter, rolls off at about 475 GHz [2.31].

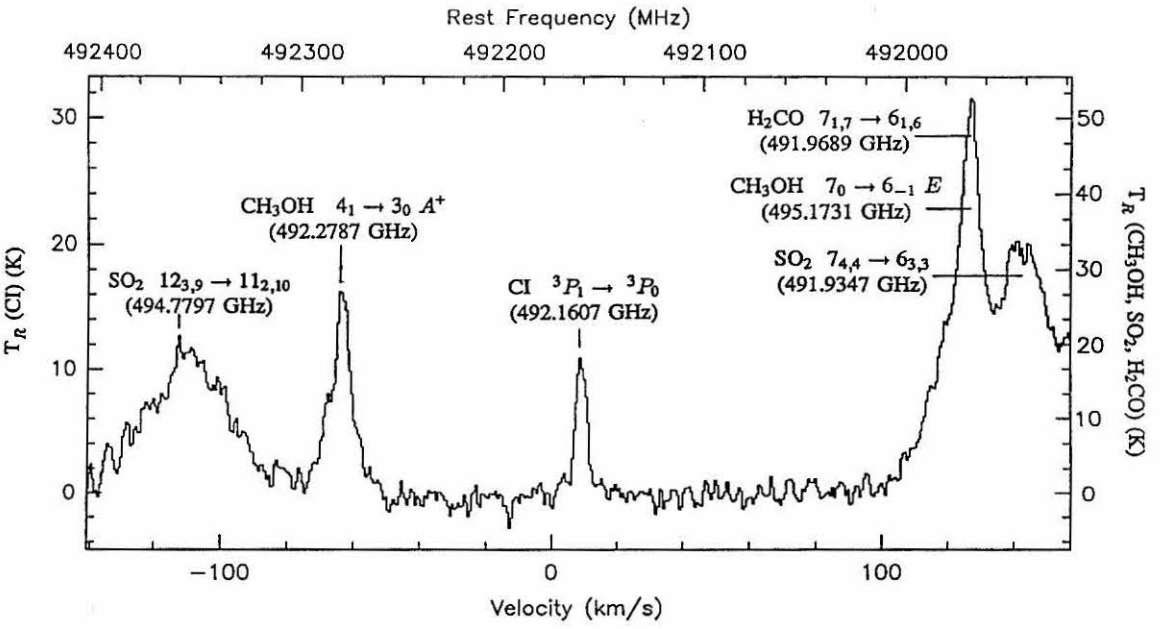


Figure 2.11 Submillimeter spectrum of the core of the Orion molecular cloud towards IRc2 taken with a Superconducting–Insulator–Superconductor (SIS) receiver utilizing a hybrid antenna. The data were taken with the 10.4 m aperture telescope of the Caltech Submillimeter Observatory (CSO), Hawaii, in 1990. The response of two sidebands, centered at 492.16 GHz and 494.96 GHz, is included. The hybrid antenna receiver was compared to waveguide based receivers and yielded similar results for coupling to the telescope and overall telescope efficiencies.

Table 2.3. Receiver noise temperatures

Frequency [GHz]	318	395	426	492
T_{Rx} (DSB) [K]	200	230	220	500

The very high sensitivities obtained with the receiver are an indication that the intrinsic coupling efficiency of the hybrid antenna is high. However, it was not possible, as is usually the case, to quantify the coupling efficiency of the hybrid antenna from the noise temperature

measurements. The coupling efficiency is just one of many parameters that determine the receivers sensitivity, most of which are not easily measured to better than 10%.

2.5 Considerations for Array Optics

An antenna that is to be used as an element in a heterodyne array receiver must have several features in addition to being a good single element. Its aperture efficiency has to be high to efficiently sample the image plane, the beam width should be narrow and preferably matched to the telescope optics without further optics, and finally, the cost and ease of manufacture has to be reasonable if large arrays are anticipated.

Table 2.4. Aperture efficiencies η_A for different lens extension length d measured at 115 GHz

Extension d [mm]	3.25	3.51	3.76	4.01	4.27	4.52	4.67	5.18
Mean FWHP [°]	24±3	17.2±1.7	10.8±0.5	11.2±0.5	10.5±0.5	10.2±0.5	10.5±0.5	10.0±0.5
η_A [%]	18±2	29±3	58±7	65±7	76±6	67±7	66±6	71±7

Table 2.4 shows the aperture efficiencies as determined from total power measurements and the beam widths as determined from the pattern measurements as a function of extension length d. The aperture efficiency peaks at the optimum extension length d_{opt}, which is determined experimentally (figure 2.3) and theoretically (figures 2.5 and 2.6). The lower aperture efficiency, for extension lengths d smaller than d_{opt}, are due to the increase in beam size (*i.e.*, lower directivity), whereas from theory the Gaussian coupling efficiency is expected to increase towards the aplanatic case ($d = r/n = 3.25mm$), due to smaller aberrations [2.17]. However, as discussed earlier, Gaussian coupling efficiencies are experimentally typically found to be lower for the aplanatic case [2.18,2.1,2.22].

The hybrid antenna in a fly's–eye configuration (see figure 2.12) is considered a good candidate for a single element of an array. Hybrid antennas have high aperture efficiencies and diffraction limited beams, thus can sample the image plane at a spatial frequency of half the Nyquist sampling rate, *i.e.*, undersampled by a factor of two. Planar antennas, which are the feed antennas for hybrid antennas, are inexpensive and easy to manufacture lithographically. The extended hemispherical or elliptical lenses can be manufactured from a mold since the surface accuracy requirements in the millimeter and submillimeter wavelength ranges do not require optical

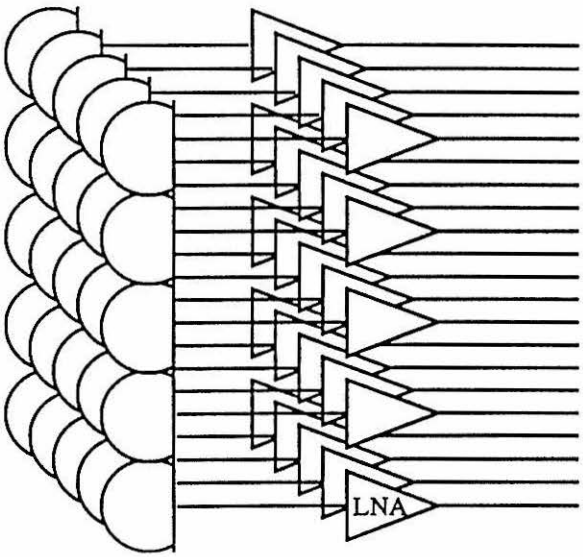


Figure 2.12 "Fly's–eye" configuration of a 5 by 5 focal plane array of hybrid antennas. The array of antennas is shown feeding an array of intermediate frequency (IF) low noise amplifiers (LNA). The shown array would be about 35 mm on a side for a telescope with an $f/13$ beam at 492 GHz.

quality finish. To keep the power loss due to surface inaccuracies below 1%, the RMS surface error, as determined from (2.5), has to be better than $\lambda/20\pi(n-1)$, which is about $10\mu m$ at 500 GHz for a quartz lens.

It is important to note that if the receiver is operated in a total power mode, the image plane has to be sampled at twice the rate (for each linear dimension) compared to a mode where the electric field with its phase is measured (see for example [2.8]). Radio astronomical receivers used for single telescope observations are typically operated in a total power mode (*e.g.*, autocorrelator spectrometers produce power spectra) , despite the fact that, in principle, they are heterodyne receivers and measure amplitude and phase, *i.e.*, they are field sensitive. The image plane of a given optical system contains Fourier components of the electric field up to a cutoff frequency f_c^E, which determines the maximum spatial resolution of the source obtainable with the particular optical system. For power measurements there are Fourier components up to *twice* the cutoff frequency for the electric field components due to squaring of the fields, *i.e.*, $f_c = f_c^P = 2f_c^E$. Nyquist sampling then requires twice the spatial cutoff frequency f_c^P. This implies that a two–

dimensional array receiver in power detection mode requires four times as many detectors as one that preserves the electric field with the phase information until the image is reconstructed.

At considerable reduction in Gaussian coupling efficiency, the size of the receiving antenna could be made half the linear size of the diffraction limit for field detection, or one–quarter the linear size for power detection, to allow for Nyquist sampling. This is often done for optical systems that are background noise limited. However, in broadband (IF) millimeter and submillimeter wavelength heterodyne receivers, the detector's sensitivity typically determines the overall system sensitivity. Reducing the size of the antenna would reduce the amount of power received by it. Since the noise power produced by the detector stays constant, the signal to noise ratio will suffer. The quadratic relation between the integration time required to achieve a certain signal to noise ratio and the system's sensitivity thus rules out this approach as long as the system's sensitivity remains detector limited.

In this paragraph the reason for suggesting the fly's–eye configuration over a single lens system will be discussed. Measurements of individual planar feed antennas on one big hyper-hemispherical lens, *i.e.*, in the aplanatic focus position, showed poor beam patterns for the off–axis elements [2.33]. A lens with 4λ diameter showed significant distortions of the main beam when operated $\frac{1}{4}\lambda$ off axis and sidelobe levels as high as -4 dB were present when operated $\frac{1}{2}\lambda$ off axis. However, Gaussian coupling efficiencies could still be reasonably high despite some distortions of the main beam. Measurements of an array of feed antennas on an extended hemispherical or elliptical lens, *i.e.*, as a hybrid antenna with an array of feed antennas, were not performed. Since the required size of the lens to accommodate an array with low distortions would produce beams too narrow to match directly to typical f–numbers of a telescope this approach was not chosen. However, for arrays with few elements, feeding telescopes with relatively high f–numbers, this would be a possible configuration (see for example [2.21]). In the opinion of the author, the fly's–eye technique is more versatile since it does not restrict the number of elements in the array (the feed antennas are usually fairly big due to the IF and DC connection pads), allowing for the size of the beam to be designed to directly match the beam from a telescope, and allowing all elements of the array to perform equally. Systems that do not provide for a direct match to the telescope optics may suffer from losses introduced from the additional optics required to match

the beams. The hybrid antenna in the fly's–eye configuration avoids these problems. Additionally, the size of the feed antenna is much smaller than the size of the hybrid antenna, thus easily providing room for IF connections or circuits at each element of an array.

2.6 Chapter Summary

Beam pattern and aperture efficiency measurements of hybrid antennas were performed and hybrid antennas are found to be good candidates for focal plane imaging array receivers. Calculations based on geometric ray optics including diffraction limit effects were presented and showed excellent agreement with measurements thus providing all necessary parameters to design hybrid antennas. The manufacture of hybrid antennas is low cost and allows for mass production in arrays. Due to the hybrid antennas' diffraction limited performance they will allow half Nyquist sampling rate (undersampled by a factor of two) of the image plane for field detection or half that sampling rate for power detection. Depending on the application, the feed antenna can be chosen to be a broad band antenna (several octaves) like logarithmic spiral antennas with circular polarization, or a logarithmic periodic antenna with linear polarization. The f–number of the beam can be custom designed to match the optics of a telescope directly. The feed antenna is smaller than the hybrid antenna itself thus ample room for IF connections or circuitry is available at each array element.

Using a planar logarithmic spiral antenna for the feed of the hybrid antenna, an aperture efficiency of 76% was measured. The hybrid antenna was tested in an SIS receiver with a $Nb/AlO_x/Nb$ tunnel junction and a broad band matching circuit yielding coupling efficiencies to a telescope as high as those obtained with corrugated feed horn based receiver systems and sensitivities approaching those of the best waveguide receivers for submillimeter wavelengths.

2.7 References Chapter II

[2.1] Büttgenbach, T.H., Miller, R.E., Wengler, M.J., Watson, D.M., and Phillips, T.G., "A Broad–band Low–Noise SIS Receiver for Submillimeter Astronomy," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT–36, No. 12, pp. 1720 – 1726, 1988.

[2.2] Büttgenbach, T.H., "An Improved Solution For Integrated Array Optics In Quasi Optical MM and SUBMM Receivers: The Hybrid Antenna," *IEEE MTT* , to appear in October 1993.

[2.3] Ellison, B.N., Schaffer, P.L., Schaal, W., Vail, D., and Miller, R.E., "A 345 GHz SIS Receiver For Radio Astronomy," *Int. J. of IR and MM Waves*, Vol. 10, No. 8, 1989.

[2.4] Rebeiz, G.M., Kasilingam, D.P., Stimson, P.A., Guo, Y., and Rutledge, D.B., "Monolithic Millimeter–Wave Two–Dimensional Horn Imaging Arrays," *IEEE Trans. Antennas Propag.*, Vol. AP–28, pp. 1473 – 1482, 1991.

[2.5] Eleftheriades, G.V., Ali–Ahmad, W.Y., Katehi, L.P., and Rebeiz, G.M., "Millimeter–Wave Integrated–Horn Antennas, Part I: Theory," *IEEE Trans. Antennas Propag.*, Vol. AP–39, No. 11, pp. 1575 – 1581, 1991.

[2.6] Ali–Ahmad, W.Y., Eleftheriades, G.V., Katehi, L.P., and Rebeiz, G.M., "Millimeter–Wave Integrated–Horn Antennas, Part II: Experiment," *IEEE Trans. Antennas Propag.*, Vol. AP–39, No. 11, pp. 1575 – 1581, 1991.

[2.7] Walker, C.K., private communication 1991.

[2.8] Rutledge, D.P. Neikirk, D.P. Kasilingam, "Integrated Circuit Antennas," in *Infrared and Millimeter Waves*, Vol. 10, pp. 1 – 90, Ed. K.J. Button, New York, Academic Press, 1983.

[2.9] Compton, R.C., McPhedran, R.C., Popovic, Z., Rebeiz, G.M., Tong, P.P., and Rutledge, D.B., "Bow–Tie Antennas on a Dielectric Half–Space: Theory and Experiment," *IEEE Trans. Antennas Propagat.*, vol. AP–35, p. 622, 1987.

[2.10] Rumsey, V. H. "Frequency Independent Antennas," 1957 *IRE National Convention Record*, pt. 1, pp. 119 – 128.

[2.11] Dyson, J. D. "The Equiangular Spiral Antenna," *IRE Trans. Antennas Propag.*, vol. AP–7, pp. 181 – 187, April 1959.

[2.12] Filipovic, D.F., Aliahmad, W.Y., and Rebeiz, G.M., "Millimeter–Wave Double–Dipole Antennas for High Gain Integrated Reflector Illumination," *IEEE Trans. Microwave Theory Tech.*, Vol. 40, No. 5, pp. 962 – 967, 1992.

[2.13] Kerr, A.R., Siegel, P.H., and Mattauch, R.J., "A Simple Quasi–Optical Mixer for 100 – 120 GHz," *IEEE MTT–S Int. Symposium Digest*, p. 96, April 1977.

[2.14] Rogers, R.L., and Neikirk, D.P., "Use of Broadside Twin Element Antennas to Increase Efficiency on Electrically Thick Dielectric Substrates," *Int. J. of IR and MM Waves*, Vol. 10, pp. 697 – 728, 1988.

[2.15] Zmuidzinas, J., and LeDuc, H.G., "Quasi Optical Slot Antenna SIS Mixer," *IEEE Trans. Microwave Theory Tech.*, Vol. 40, No. 9, 1992.

[2.16] Kasilingam, D., and Rutledge, D. B., "Focusing Properties of Small Lenses," *Int. J. of IR and MM Waves*, Vol. 7, No. 10, pp. 1631 – 1647, 1986.

[2.17] Born, M., and Wolf, E., "Principles of Optics," Pergamon Press, New York, 2. ed., 1964.

[2.18] Wengler, M.J., Woody, D.P., Miller, R.E., and Phillips, T.G., "A Low Noise Receiver for Millimeter and Submillimeter Wavelength," *Int. J. of IR and MM Waves*, Vol. 6, pp. 697 – 706, 1985.

[2.19] Zah, C., Kasilingam, D., Smith, J. S., Rutledge, D. B., Wang, T.–C., and Schwarz, S. E., "Millimeter Wave Monolithic Schottky Diode Imaging Arrays," *Int. J. of IR and MM Waves*, Vol. 6, No. 10, pp. 981 – 997, 1985.

[2.20] Skalare, A., de Graauw, Th., and van de Stadt, H., "A Planar Dipole Array Antenna with an Elliptical Lens," *Microwave and Optical Tech. Lett.*, Vol. 4, No. 1, 1991.

[2.21] Alder, C.J., Brewitt–Taylor, C.R., Dixon, M., Hodges, R.D., Irving, L.D., and Rees, H.D., "Microwave and Millimeter–Wave Receivers with Integral Antenna," *IEEE Proc.–H*, Vol. 138, 1991.

[2.22] Filipovic, D.F., Gearhart, S.S., and Rebeiz, G.M., "Double Slot Antennas on Hyperhemispherical, Elliptical, and Extended Hemispherical Dielectric Lenses," *IEEE Microwave Theory and Tech.*,, to appear in October 1993.

[2.23] Neikirk, D.P., Lam, W.W., and Rutledge, D.B., "Far–Infrared Microbolometer Detectors," *Int. J. of IR and MM Waves*, Vol. 5, pp. 245 – 276, 1984.

[2.24] Radiometer Physics, Bergwiesenstr. 15, Meckenheim, FRG.

[2.25] Ruze, J., "Antenna Tolerance Theory – A Review," *Proc. IEEE* , Vol. 54, No. 4, p. 633, 1966.

[2.26] Friis, H.T., "A Note on a Simple Transmission Formula," *Proc. IRE*, 34, 254 – 256, 1946.

[2.27] Jakes, W.C., "Antenna Engineering Handbook," ed. H. Jasik, McGraw–Hill Book Co., chapter 10, 1961.

[2.28] Kormanyos, B.K., Ling, C.C., and Rebeiz, G.M., "A Planar Wideband Millimeter–Wave Subharmonic Receiver," *IEEE Microwave Symp. Digest*, pp. 213 – 216, 1991.

[2.29] Büttgenbach, T.H., Keene, J.B., Walker, C.K., and Phillips, T.G., "Submillimeter Detection Of Extragalactic C I Emission: IC342," *Astrophysical Journal Letters*, Vol. 397, L15 – L17, 1992.

[2.30] Büttgenbach, T.H., Keene, J.B., Young, K., Phillips, T.G., and Miller, R.E., "The Small–Scale Structure of C I in Orion," in preparation.

[2.31] Büttgenbach, T.H., LeDuc, H.G., Maker, P.D., and Phillips, T.G., "A Fixed Tuned Broadband Matching Structure For Submillimeter SIS Receivers," *IEEE Trans. Applied Supercond.*, Vol. 2, No. 2, September 1992.

[2.32] Walker, C.K., Kooi, J.W., Chan, M., LeDuc, H.G., Carlstrom, J.E., and Phillips, T.G., "A Low–Noise 492 GHz SIS Waveguide Receiver," *Int. J. of IR and MM Waves*, Vol. 13, No. 6, p. 785, 1992.

[2.33] Büttgenbach, T.H., and Rose, T.L., "Planar Antenna Structures for SIS Receivers," in preparation.

CHAPTER III

RF IMPEDANCE MATCHING CIRCUITS: THE END–LOADED–STUB

3.1 The Need for RF Matching Circuits and Review of Previous Designs

Superconductor Insulator Superconductor (SIS) tunnel junction mixers are gradually approaching the quantum noise limit in the millimeter band [3.1,3.2,3.3,3.4]. As discussed in section 1.2 of this thesis, high speed, low sub–gap leakage current and very sharp non–linearity in the current–voltage characteristic of the mixer's SIS diode are required to yield high receiver sensitivities. Much of the recent progress is due to the use of $Nb/AlO_x/Nb$ SIS junctions. Even though $Nb/AlO_x/Nb$ SIS junctions have higher specific capacitances ($\approx 50-90 fF/\mu m^2$ [3.5]) relative to Pb devices used in previous SIS mixer designs ($\approx$ 20*$-40 fF/\mu m^2$ [3.6]), their speed has increased to $\approx$ 120 GHz (ωRC = 1) while maintaining low sub–gap leakage currents and very sharp non–linearities in the current–voltage characteristic. The increases in speed were possible due to improvements of the manufacturing processes.

Waveguide based mixers with one or two tuning elements are able to match out the large reactive part of the junction impedance and thus achieve very good noise performance [3.1,3.2,3.3,3.4, 3.7]. However, such mixers are not very well suited for large focal plane receiver arrays for several reasons. Current manufacturing techniques would make an array excessively expensive. Tuning each individual element would make frequency changing very time consuming, and the packing density in the focal plane would be low, making the array physically large. The large size of an array poses problems for the cryogenic system and the telescope optics. However, quasi–optical mixers usually avoid most of the drawbacks of waveguide receivers mentioned above, but in the past have lacked the capability for tuning out the capacitance over a wide bandwidth, thereby

* This number is from experiments where the capacitance was determined from the roll–off frequency in FTS measurements performed in 1988 with SIS junctions made by R. Miller at Bell Labs.

failing to attain the same sensitivity as waveguide mixers. The new matching circuit design introduced in this chapter has been published in [3.8].

Review of previous designs.

The coupling efficiency t from a driving impedance Z_D into the junction impedance $Z_j = R_j + X_j$ is given by

$$t = 1 - \left| \frac{Z_D - Z_j^*}{Z_D + Z_j^*} \right|^2 \quad . \tag{3.1}$$

Note that R_j is the real part of the SIS junction *RF* impedance, which is not necessarily the junction normal state resistance R_n. However, for the frequency range of this design, typical biasing conditions and LO power levels, $R_n \approx R_j$ [3.9]. Furthermore the imaginary part of the junction RF impedance in principle has components of quantum mechanical origin [3.10 and 3.11], but in the case of devices as used here it is dominated by the parallel capacitance of the device. Z_j is therefore assumed to be

$$Z_j = (R_n^{-1} + i\omega C)^{-1}. \tag{3.2}$$

Several possibilities exist to attain the resonant condition $Z_D = Z_j^*$, with perfect coupling ($t = 1$). Designs using an open-ended inductive stub [3.12] to tune out the SIS junction capacitance have a much narrower bandwidth compared with other designs as discussed later. This limits their usefulness for many applications and puts very tight constraints on the manufacturing parameters, which are often not well known or reproducible and will reduce the yield when attempting to make these devices for specific frequencies. Other designs use series arrays of SIS junctions that are tuned out by an inductor parallel to the array [3.13,3.14]. These designs are somewhat more wide band than open-ended inductive stubs but suffer from a parasitic series inductance limiting the maximum operating frequency to about 100 GHz. A design used by Kerr *et al.* [3.14] uses an array of SIS junctions with each individual junction tuned out by a parallel inductor at 75 to 110 GHz. This approach should be useful up to a few hundred GHz and has very recently been applied at 215 to 280 GHz, yielding excellent sensitivity in a waveguide receiver without back shorts [3.15]. A review of many of the previously mentioned designs can be found in Kerr *et al.* [3.16]. An elegant approach to an individually tuned series array of SIS junctions has recently

been presented by Ermakov *et al.* [3.17]. In their design, the SIS junctions are in series as seen by the RF but are parallel DC–biased through the same inductors that provide the tuning for the SIS junctions capacitance. The advantages are i) all SIS junctions have the same bias voltage, thus reducing the uniformity constraints on the manufacturing tolerances for the array, and ii) the IF impedance of the mixer, just as for the DC–bias, is a parallel circuit of the individual IF impedances of the SIS junctions, thus making it possible to design the array to have an IF output impedance of about 50 Ω. Excellent mixer performance was demonstrated at 48 GHz. The bandwidth of the circuit was not well defined in the data presented due to several very narrow resonances.

If no attempt is made to tune out the capacitance (Z_D real, *i.e.*, $Z_D = R_D$) the transmission is

$$t = \frac{4\ R_n/R_D}{(1 + R_n/R_D)^2 + (\omega R_n C)^2} \quad . \tag{3.3}$$

For a typical case of $R_n \approx R_D$ and $\omega R_n C \gg 1$

$$t \approx \frac{4}{(\omega R_n C)^2} \quad . \tag{3.4}$$

For $\omega R_n C = 5$ at 500 GHz (3.3) will only yield $t = 0.14 = -8.6dB$. Keeping Z_D real, the best possible coupling for (3.3) can be achieved by making it the same as the absolute value of the junction's impedance, *i.e.*,

$$R_D = |Z_j| = \frac{R_n}{\sqrt{1 + (\omega R_n C)^2}}. \tag{3.5}$$

The transmission is then given by

$$t = \frac{2}{\sqrt{1 + (\omega R_n C)^2}\ + 1} \approx \frac{2}{\omega R_n C}\ , \tag{3.6}$$

which yields $t = 0.33 = -4.8dB$ for the case given above. The required driving impedance is $R_D = 10\Omega$ for a junction with $R_n = 50\Omega$, since it is dominated by the low impedance of the capacitance. This makes it necessary to transform the antenna impedance down to achieve such a low driving impedance.

Zmuidzinas and LeDuc [3.18] have built a receiver for 500 GHz using a double slot dipole antenna with a continuous real impedance transformer to nearly realize the requirement of equation

(3.5). Their specific design yielded a theoretical coupling of $t = 0.23 = -6.3dB$. The results obtained with this design were very encouraging. The work presented here goes one step further and provides a full match to the complex impedance of the SIS tunnel junction. The obvious addition to the design by Zmuidzinas and LeDuc to obtain a full match would be to add an open-ended tuning stub to the SIS junction to tune out the junction's capacitance. However, if an open–ended inductive stub is used to tune out the SIS junction's capacitance, the bandwidth of the circuit is very small. This can be seen from the following calculation.

3.2 Bandwidth Calculations for Different Circuits

To get a comparison for the bandwidth of different tuning circuits the change of admittance with frequency scaled by the real part of the admittance at resonance frequency will be considered. This quantity is inversely proportional to the bandwidth of the circuit.

a) The open–ended stub

The simplest way to tune out the capacitance of an SIS device is with open–ended stubs. However, they are known to yield much smaller bandwidths than, for example, designs with lumped parallel inductors. The reason for this is that the inductance of the open–ended stub itself is a function of frequency making its admittance a very strong function of frequency. Generally, the admittance Y of a transmission line of length ℓ, propagation velocity v, characteristic admittance $Y_c = Z_o^{-1}$, and load admittance Y_L is

$$Y = Y_c \frac{Y_L + iY_c \tan(\omega\ell/v)}{Y_c + iY_L \tan(\omega\ell/v)} \quad . \tag{3.7}$$

First consider an open–ended stub. It has a load admittance $Y_L = 0$ so that

$$Y^{stub} = Z_o^{-1} i \tan(\omega\ell/v) \quad . \tag{3.8}$$

The SIS junction is parallel to the open–ended stub so that the total admittance of the circuit is

$$Y_{total}^{stub} = Z_o^{-1} i \tan(\omega\ell/v) + i\omega C + R_n^{-1} \quad . \tag{3.9}$$

To calculate the bandwidth of the circuit the change of admittance with frequency will be considered.

$$\frac{dY_{total}^{stub}}{d\omega} = \frac{i\ell}{Z_o v}\left(1 + \tan^2(\omega\ell/v)\right) + iC \tag{3.10}$$

The resonant condition for the stub to tune out the SIS junction's capacitance C is for $\mathrm{Im}(Y_{total}^{stub}) = 0$, *i.e.*,

$$Z_o^{-1}\tan(\omega\ell/v) + \omega C = 0 \quad . \tag{3.11}$$

Using equation (3.11) in (3.10) yields

$$\frac{dY_{total}^{stub}}{d\omega} = \frac{i\ell}{vZ_o}\left(1 + (\omega Z_o C)^2\right) + iC \quad . \tag{3.12}$$

To find the optimum characteristic impedance Z_o for the open–ended stub the value for Z_o has to be found where the change of admittance with frequency in (3.10) has a minimum, *i.e.*,

$$\frac{d}{dZ_o}\left(\frac{dY_{total}^{stub}}{d\omega}\right) = 0 \quad . \tag{3.13}$$

Solving this equation requires solving a transcendental equation due to ℓ's dependence on Z_o. With $Q \equiv \omega R_n C$ and $Q^2 \gg 1$ one gets

$$Z_o \simeq \frac{R_n}{Q} = \frac{1}{\omega C} \quad . \tag{3.14}$$

The condition in (3.11) then simply reduces to

$$\tan(\omega\ell/v) = -1 \quad \text{or} \quad \omega\ell/v = \pi - \arctan(-1) = \frac{3\pi}{4} \quad , \tag{3.15}$$

and so the length is $\frac{3}{8}\lambda$, which is the standard result. The fractional bandwidth B of a circuit is defined by

$$B \equiv \frac{2\Delta\omega}{\omega} \quad . \tag{3.16}$$

With $\Delta\omega$ approximated by the change of admittance with frequency scaled by the real part of the admittance at resonance frequency,

$$\Delta\omega \approx \left| \frac{\mathrm{Re}(Y_{total}^{stub})}{\left(\frac{dY_{total}^{stub}}{d\omega}\right)} \right| \tag{3.17}$$

so that the fractional bandwidth for the open–ended stub with (3.9),(3.12),(3.14) and (3.17) is

$$B^{stub} = \frac{1}{\left(\frac{3\pi}{4} + \frac{1}{2}\right) Q} \simeq \frac{1}{3\,Q} \quad . \tag{3.18}$$

b) A lumped element parallel inductor

If a lumped element inductor were used to tune out the SIS junctions' capacitance rather than an open-ended stub, the analysis described above yields the well known relation for the fractional bandwidth of such a circuit

$$B^{ind} = \frac{1}{Q} , \tag{3.19}$$

justifying equation (3.17). This makes the circuit utilizing a parallel inductor a factor of 3 higher bandwidth than that of an open–ended tuning stub. However, a parallel inductor would short–circuit the DC bias voltage and the IF output of the SIS junction. Thus, an RF short–circuit which is open–circuit at DC and IF frequencies must be placed in series with the inductor. This RF short–circuit itself is usually frequency dependent and limits the bandwidth of the design. Recent work by Carpenter *et al.* [3.19] uses a quarter wave transformer to lower the RF impedance thus increasing the RF saturation power of the mixer before they feed the SIS junction. The capacitance of the SIS junction is tuned out with an inductor that is RF short circuited with another quarter wave section yielding a bandwidth of about 10% at 90 GHz.

c) The end–loaded–stub

In the new design presented in this thesis a different way of tuning out the capacitance was chosen. The SIS junction is put at the **end** of a short transmission line of length ℓ henceforth called an end–loaded–stub. Unlike the parallel inductor, the end–loaded–stub does not require any additional circuitry for the DC bias of the SIS junction since the SIS junction itself terminates both the RF and DC–bias currents supplied via the transmission lines of the matching circuit.

Using equation (3.7), but now with the load admittance that of the SIS junction $Y_L = R_n^{-1} + i\omega C$ yields

$$Y_{total}^{tl} = \frac{1 + i(Q + \frac{R_n}{Z_o}\tan(\alpha))}{R_n + (i - Q)Z_o \tan(\alpha)} \qquad \text{with} \quad \alpha \equiv \omega\ell/v \ . \tag{3.20}$$

Again, the same analysis method to calculate the fractional bandwidth as for the open–ended stub will be followed. Note that this method only yields accurate results for matching circuits with small fractional bandwidths. As will be shown, the designed circuit has a fractional bandwidth large enough to violate that assumption. Fortunately, as seen by the results from computer simulations

and measurements of the device, the fractional bandwidth is actually larger than calculated here. Furthermore, the main interest here is in a comparison between different techniques to show which circuits are favorable.

The change of the circuit's admittance with frequency is

$$\frac{dY^{tl}_{total}}{d\omega} = \frac{iR_nC + i\frac{R_n\ell}{Z_o v}(\tan^2(\alpha)+1)}{R_n - Z_oQ\tan(\alpha) + iZ_o\tan(\alpha)} -$$

$$\frac{\left[1 + i(Q + \frac{R_n}{Z_o}\tan(\alpha))\right]\left[(i-Q)Z_o\frac{\ell}{v}(\tan^2(\alpha)+1) - R_nCZ_o\tan(\alpha)\right]}{(R_n - Z_oQ\tan(\alpha) + iZ_o\tan(\alpha))^2} \quad . \tag{3.21}$$

The resonant condition for the transmission line from $Im(Y^{tl}_{total}) = 0$ in (3.20) is

$$\tan(\omega\ell/v) = -\frac{1}{2}\left(\frac{Z_oQ}{R_n} - \frac{R_n}{Z_oQ} + \frac{Z_o}{R_nQ}\right) + \sqrt{\frac{1}{4}\left(\frac{Z_oQ}{R_n} - \frac{R_n}{Z_oQ} + \frac{Z_o}{R_nQ}\right)^2 + 1} \quad . \tag{3.22}$$

Solving (3.13) with (3.21) the optimum characteristic impedance Z_o is obtained. It turns out that within the limits

$$R_n > Z_o \geq \frac{R_n}{Q} \qquad \text{for } Q^2 \gg 1 \tag{3.23}$$

the solution shows no strong dependence of Z_o. This condition makes the short section of transmission line look like a series inductor. With the choice of $Z_o = R_n/Q$ for further calculations, which reduces (3.22) to

$$\tan(\alpha) = \frac{-1}{2Q^2} + \sqrt{\frac{1}{4Q^4} + 1} \approx 1 \qquad \text{or} \qquad \alpha = \omega\ell/v = \frac{\pi}{4}\ , \quad i.e., \quad \ell = \frac{1}{8}\lambda. \tag{3.24}$$

Using the same definition as in (3.17) for $\Delta\omega$ one gets

$$\Delta\omega = \left|-\frac{\pi R_nC}{2Q} + iR_nC\left(\frac{\pi}{4Q^2} - 1 - \frac{\pi}{2}\right)\right|^{-1} \tag{3.25}$$

and thus for the fractional bandwidth

$$B^{tl} \approx \frac{1}{Q}. \tag{3.26}$$

Comparing (3.18) with (3.19) and (3.26)

$$B^{tl} \approx B^{ind} \approx 3\ B^{stub}$$

shows that the design with the SIS junction at the end of the transmission line yields a much broader match than one with open–ended stubs by a factor of ≈ 3 and about the same bandwidth as parallel inductor circuits but avoiding the need for an RF short – DC and IF open circuit.

The short transmission line that is used to tune out the capacitance of the junction (section 2 in table 3.2) also transforms the real part of the SIS junction's admittance to a higher conductance as can be seen from (3.20) with (3.24)

$$\mathrm{Re}(Y^{tl}_{total}) = R_n^{-1}\ 2Q^2. \tag{3.27}$$

Note that this impedance is typically too low to be driven by an antenna directly. A two–section quarter wave transformer is used (sections 3 and 4 in table 3.2) to transform the low real impedance up to the antenna's impedance (see figures 3.1 and 3.2).

The overall design allows the operation of a quasi–optical receiver system requiring no adjustable tuners over an octave of bandwidth well into the submillimeter band. The size of the tunnel junctions used is not critical and can be larger than 1 μm^2. The optimum size for the design used would have been about $1.5\mu m^2$. We had designs with that size but none of them made contact due to an alignment problem in the manufacturing as will be discussed later. Even though electron beam lithography was used to define the junction area, it is possible to manufacture junctions of this size with optical lithography [3.18]. Additionally, the large size of the junctions eliminates saturation problems, makes it easy to suppress the Josephson oscillations with a magnetic field and makes the device more robust against static discharges.

The matching circuit developed here for the quasi–optical receivers can, of course, also be used in waveguide mixers, reducing the amount of tuning that the waveguide tuners have to provide, thus reducing losses, or even allowing for designs without any adjustable tuners. An investigation of this technique is currently underway at Caltech.

3.3 Quasi–Optical Receiver Design Employing an End–Loaded–Stub Matching Circuit

a) Receiver design: Hybrid antenna and IF circuit

The basic mixer design, as far as coupling the radiation into the antenna and processing the IF, is similar to the one described in chapter I part 1.6 and [3.20], with modified optics for the hyperhemisphere as described in chapter II and [3.21].

The extended hemisphere of the hybrid antenna is made from single crystal quartz, as is the substrate of the spiral feed antenna, to obtain a high thermal conductivity. The parameters used to bring the extended hemisphere and spiral feed antenna into the regime of a hybrid antenna structure over the frequency range of the matching circuit are: radius of the hemisphere $r_h = 6.35mm$ with a refractive index of $n_h = 2.11$, and distance of the antenna to the center of the hemisphere $d_a = 5.33mm$. Note that the optics contain no further refracting elements since the Gaussian beam launched by the hybrid antenna structure is between 10 and 5 degrees FWHP for the range of 200 to 475 GHz. This matches the beam of the Caltech Submillimeter Observatory (CSO) 10.4 meter antenna in the Cassegrain focus with slight over–illumination at 200 GHz and slight under–illumination at 475 GHz. Recalling from chapter I, the spiral feed antenna is a self–Babinet–complementary antenna structure with a frequency independent free space impedance of $Z_o = 60\pi\Omega \simeq 188\Omega$. Mounting such an antenna onto a half–space of dielectric (approximated by the extended hemisphere) yields an antenna impedance of about

$$Z_{ant} = \frac{Z_o}{\sqrt{(1+n_h^2)/2}} \simeq 114\Omega. \tag{3.28}$$

The schematic layout of the matching design and an optical picture of it is shown in figure 3.1. The transmission lines and their ground planes are made from Nb with an SiO dielectric. A symmetric design is used which will halve the driving impedance as seen by the circuit, *i.e.*, the effective driving impedance of the antenna seen by each half of the circuit is

$$Z_{ant,eff} = \frac{Z_{ant}}{2} = 57\Omega \quad . \tag{3.29}$$

The advantage of this design is twofold: The maximum impedance that can be realized for a transmission line is determined by the minimum width of the line ($\approx 2.5\mu m$) and the maximum

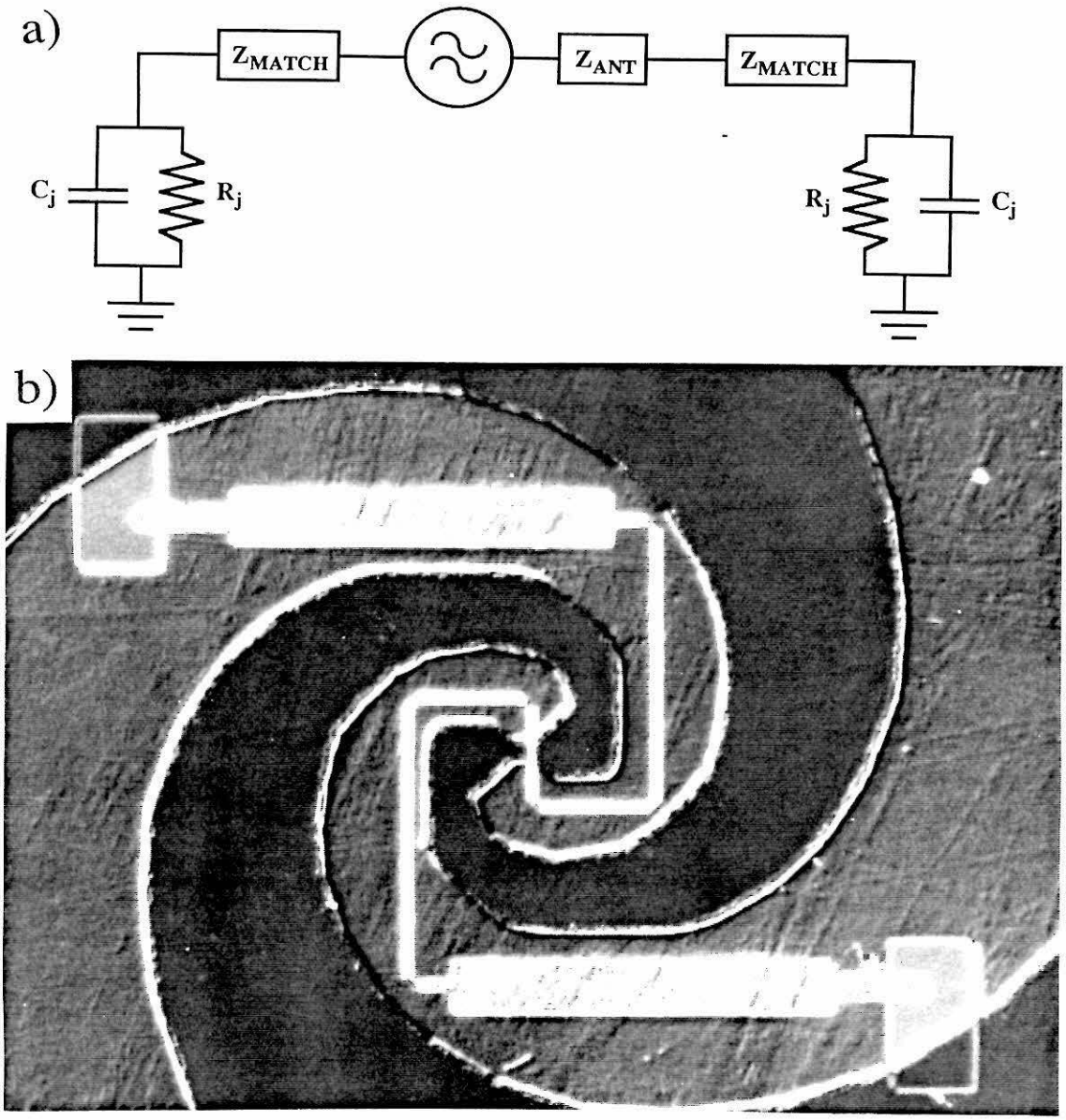


Figure 3.1 a) Schematic of the symmetric matching circuit. b) Picture of the apex of the device through an optical microscope. The width of the transmission line in the center is 2.5 μm. The SIS junctions are visible at the ends of the matching circuit as very small rectangular boxes with 1 μm on the side. The $0.45\mu m$ thick dielectric between the transmission lines of the matching circuit and the spiral arms acting as the ground plane extends over the whole area of the picture shown, except for two windows (13 by 25 μm^2) around the end of the last part of transmission line where the thickness of the dielectric layer is only $0.15\mu m$. The reduced thickness of the SiO is necessary for the lift–off technique when making the actual SIS tunnel junction. Note that the position of the SIS junctions is asymmetric relative to the transmission line due to an alignment problem of the transmission line layer.

thickness of the insulating layer ($\approx 0.5\mu m$), yielding about 26Ω. The lower driving impedance of the antenna makes a broad–band transformer design easily possible with this limitation. The optimum normal state impedance R_n of the tunnel junction with a two–section real transformer and an end–loaded–stub is then determined to be about R_n = 13Ω. Secondly, the IF signals produced by the tunnel junctions are in series, equivalent to a single junction with R_n^e = 26Ω. The IF impedance of an SIS tunnel junction is typically 2 – 3 times R_n^e [3.9] since it is just the dynamic DC impedance of the SIS with LO power applied at the bias voltage of operation. In principle, this makes a matching circuit to a standard 50Ω low noise amplifier input unnecessary. However, since the actual device tested had R_n^e = 36 for the two junctions in series, this condition was not quite fulfilled in practice and an IF matching circuit would further improve the results obtained. A small capacitor from the IF terminals of the SIS junction to ground is necessary to terminate the IF at frequencies well above the IF amplifiers' passband to avoid saturation of the SIS junction [3.22].

b) Design and numerical predictions of end–loaded–stub matching circuit

Figure 3.2 shows the equivalent non–symmetric circuit of figure 3.1 with the corresponding Smith chart plot of the RF impedances. The short transformer section next to the SIS junction transforms the capacitive part of the SIS junction to a real impedance at the center frequency. The following two–section transformer (approximately a quarter wavelength at center frequency) then transforms the impedance up to match the high effective impedance of the antenna. The impedance, group velocity and losses for the transmission lines were calculated according to Zmuidzinas and LeDuc [3.18]. The parameters assumed for the calculations are based on results by Stern [3.5] and are given in table 3.1.

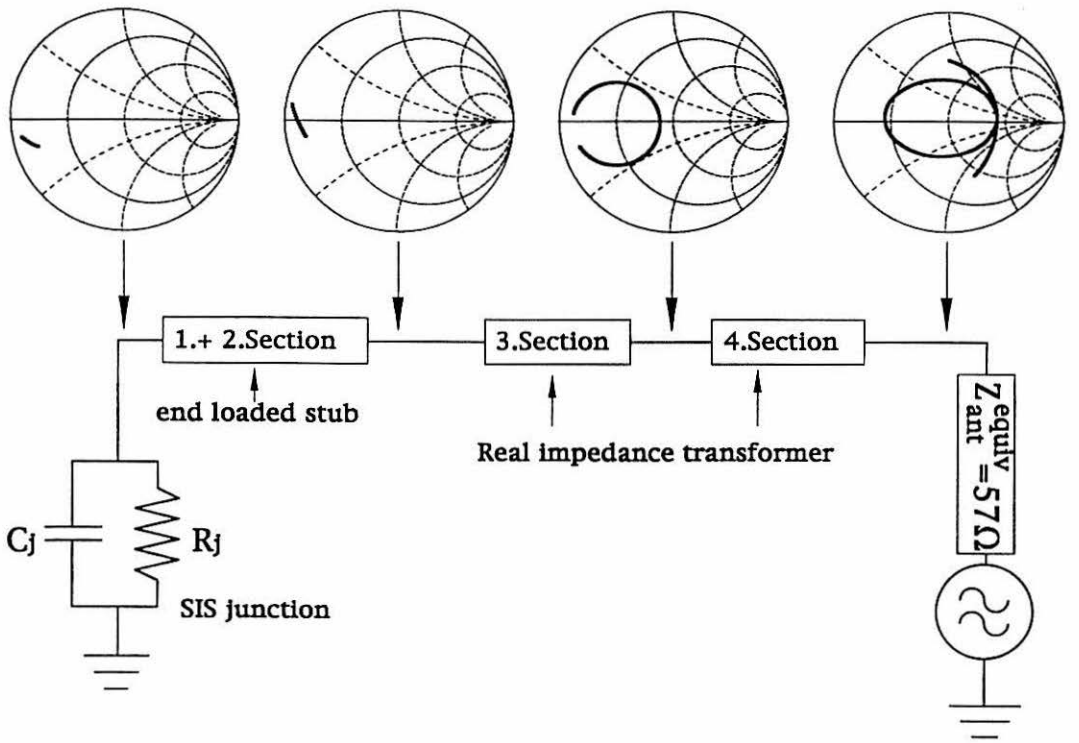


Figure 3.2 The equivalent non–symmetric matching circuit with Smith charts showing the impedance at each position of the circuit from 200 GHz to 460 GHz. The Smith charts are normalized to $Z_{ant,equiv.} = 57\Omega$, *i.e.*, the driving impedance of the circuit. The impedances shown in the Smith charts are towards the SIS junction with the circuit towards the antenna disconnected.

Table 3.1 Assumed parameters to calculate the transmission lines

junction specific capacitance	$C_s = 70\ fF/\mu m^2$
center frequency	$f = 345\ GHz$
superconducting film thickness	$d = 250\ nm$
normal state surface resistance just above T_c	$\rho_N = 5\ \mu\Omega\ cm$
physical temperature	$T = 4.2\ K$
critical temperature of the Nb film	$T_c = 9.2\ K$
gap voltage of the Nb film at $T = 0K$	$V_{gap} = 2.9\ mV$
dielectric constant of SiO	$\epsilon = 5.8\epsilon_o$

Variations with frequency of the calculated transmission line parameters were checked over the frequency range of the matching circuit and found to be negligible, which is generally true as long as the frequencies are well below the gap frequency of the superconductor used ($f_{gap}(Nb) =$ 700 GHz). The calculated losses over the frequency range of interest were less than 1% per wavelength and were ignored, as were losses in the SiO dielectric layer. The part of the $2.5\mu m$

wide line in the apex of the antenna connecting the two circuits together is $2\mu m$ long and was included as a transmission line with varying distance to its ground plane. However, its contribution is very small and could be ignored. The calculated transmission line parameters for the different sections of the matching circuit of figure 3.2 are given in table 3.2. Section 1 is the part of the line that is necessary to make contact to the SIS junction on the thin ($0.15\mu m$) SiO dielectric layer and forms a box of $5x5\mu m^2$ with the SIS junction assumed centered. It is not part of the matching circuit but was included in the computer simulations.

Table 3.2 Results of the transmission line calculations

	$width[\mu m]$	$height[\mu m]$	$Z_o[\Omega]$	v/c	$h_{equiv.}[\mu m]$	$\epsilon_{equiv.}$
Section 1	5	0.15	6.23	0.298	0.352	13.2 ϵ_o
Section 2	5	0.45	13.6	0.381	0.683	8.55 ϵ_o
Section 3	10	0.45	7.35	0.372	0.648	8.23 ϵ_o
Section 4	2.5	0.45	24.0	0.391	0.759	9.05 ϵ_o

The goal of the design was to have a bandpass that ranges from 230 GHz to 492 GHz and initially the matching circuit was calculated based on standard quarter–wave transformer designs [3.23]. Subsequently, *Touchstone* software by EEsof [3.24] was used to optimize this circuit for maximum coupling in that range. However, *Touchstone* is unable to use superconducting transmission line parameters. The calculated superconducting transmission line parameters were therefore converted into pseudo normal metal transmission line parameters by keeping the width of the lines and their propagation speed constant, but varying the height and dielectric constant to yield the same impedances for the lines (see table 3.2). This method was used to keep the geometry for the computer simulation as close as possible to that of the actual device, so that the program would more correctly account for discontinuities due to changes in width, bends, etc. The new set of matching circuit parameters as obtained from the optimization were then recalculated for superconducting lines and then applied for a new run of the optimization program. The convergence for this method was very fast, requiring typically only 2–3 iterations. The resulting transmission line parameters were thus in agreement with superconducting microstrip line properties while *Touchstone* provided a convenient method of optimizing the design. The dashed line in figure 3.3 shows the predicted matching performance of the circuit.

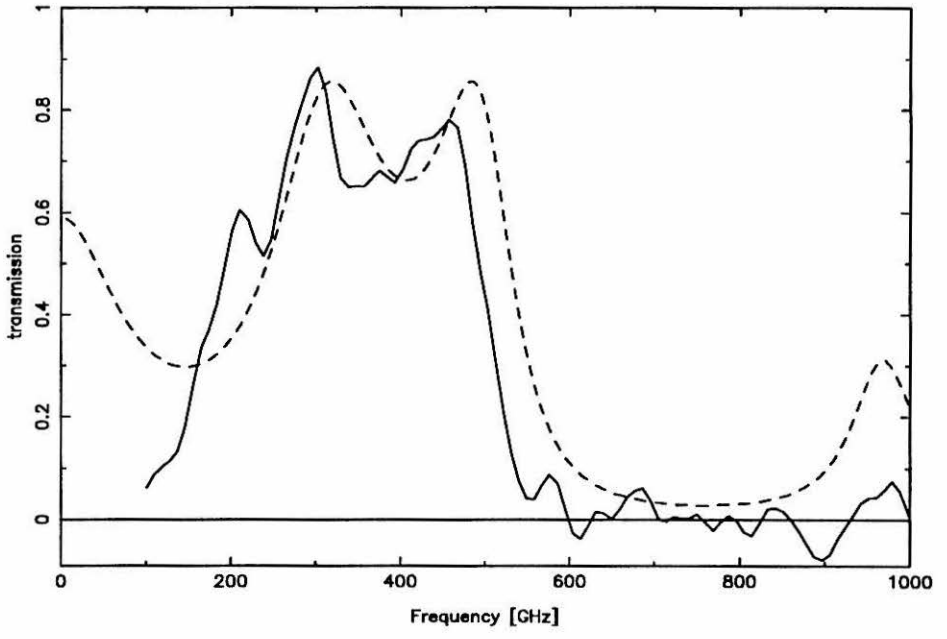


Figure 3.3 The measured response of the receiver in direct detection mode corrected for the calculated beam splitter efficiency and responsivity of the SIS direct detector (solid line). The correction for the responsivity, which has a $1/f$ dependency for the frequency range of interest (see Figure 3.5), is necessary since the SIS direct detector is a photon counting device rather than a frequency independent power detector. The throughput of the FTS is assumed to be a slow function of frequency, therefore not altering the measured characteristic frequencies of the matching circuit. The FTS is not usable below 150 GHz due to the limiting size of the optics. The data depicted has been scaled to the calculated transmission of the matching circuit (dashed line) since our FTS measurements have no absolute calibration. The predicted characteristic frequencies are shifted by only 8% from the measurement.

The ground plane for the transmission lines are the arms of the spiral antenna. Since the electric fields are concentrated in the thin dielectric layer between the transmission line and the ground plane, the ground plane can be assumed to be of infinite size for this calculation.

The calculated transmission of the matching circuit within its passband is typically $t \geq 0.75 = -1.2dB$ as shown in figure 3.3. A higher value for the transmission, if desired, is easily attainable by trading for bandwidth. The choice for the frequencies of the bandpass was based on the astronomically important transition frequencies of $CO(2 \rightarrow 1)$ at 230 GHz, $CO(3 \rightarrow 2)$ at

345 GHz and the $^3P_1 \rightarrow {}^3P_0$ transition of C I at 492 GHz in interstellar molecular clouds [3.25].

Finally, the attained bandwidth is compared to the maximum possible as given by the *Bode–Fano* criteria [3.26,3.27]. The assumptions for the *Bode–Fano* criteria are for the circuit to be causal and lossless. For a load consisting of a resistor R_n parallel to a capacitor C at the input plane of the matching circuit it states

$$\int_0^\infty \ln \frac{1}{|\Gamma(\omega)|} d\omega \leq \frac{\pi}{R_n C}, \tag{3.30}$$

with Γ the voltage reflection coefficient at the output plane of the matching circuit. For a matching circuit with a power transmission of $t = 0.75$ and $Q = 3$, as approximately given for this circuit, the maximum attainable fractional bandwidth is $B \leq 1.49$. The end–loaded–stub circuit yields a fractional bandwidth of about half of that. Using more than 2 sections for the transformer to transform the real part of the admittance it was possible to increase the fractional bandwidth to about 1, at least in computer simulations as described earlier. However, no device with more than a two–section real impedance transformer could be fabricated because of an alignment problem in the manufacturing process, as discussed below. On the other hand, a significant increase in bandwidth relative to the one tested is not necessarily useful for this application due to constraints from the optics, radiation filtering and saturation problems for the SIS junction. However, the increased bandwidth could be traded for a better match, which is always desirable.

c) Superconducting device manufacture: SIS tunnel junctions and and transmission lines

The tunnel junctions were manufactured at the Microdevices Lab at JPL, in a process similar to that described by LeDuc *et al.* [3.28], and subsequently tested at Caltech. The critical current density was 12 kA/cm^2 ($I_c = \frac{\pi}{4} V_{gap}/R_n$) as required for the proper operation of the matching circuit. The junctions were fabricated using a standard self–aligned lift–off trilayer process. The trilayer was deposited in–situ in a high vacuum deposition system with a base pressure of $4 \cdot 10^{-9}$ Torr onto a 0.25 mm thick single crystal quartz disc of 50 mm diameter. Using standard optical lithography, the shape of the ground plane structure, *i.e.*, the arms of the spiral antenna, were defined and then etched. The SIS tunnel junction was defined with direct electron beam writing on a $0.12\mu m$ thick PMMA layer followed by evaporation of a $0.05\mu m$ chrome layer and subsequent lift–off. The combined chrome photoresist mask was then used to protect the junction in a parallel

plate reactive ion etcher. Even though the size of the junction tested does not require electron beam lithography, all junctions were made with electron beam lithography, since some of the devices on the same substrate are as small as $0.25\mu m^2$.

The dielectric for the microstrip transmission line is made out of SiO with a total thickness of $0.45\mu m$. This film is deposited by thermal evaporation of SiO in two steps. The first step produces a layer that is $0.15\mu m$ thick and is used as the insulator for the lift–off of the tunnel device. The second layer is evaporated on top of the first except in the immediate vicinity of the tunnel junction and has a thickness of $0.3\mu m$. This two–step method is necessary since the lift–off for the tunnel junction with a film thickness of more than $\approx 0.2\mu m$ causes problems with the self aligned lift–off technique. However, since the high impedance transmission lines could not be made narrower than about $2.5\mu m$ a film thickness of $0.45\mu m$ was needed to achieve an impedance of Z_o = 24Ω. The transmission lines were then defined with standard optical lithography and were made out of a $0.2\mu m$ thick Nb layer. Finally the contact pads for the IF connectors and DC bias supply were covered with gold for easy contacting. Due to an alignment problem of the transmission line layer the yield of the batch was low. Figure 3.1b shows the position of the two tunnel junctions relative to the transmission line, for a successful device, one being several microns from the end, and the other one just barely making contact.

Figure 3.4 shows the current voltage characteristic of the device. Note that the gap voltage appears as 5.8 mV due to the two tunnel junctions being in series as seen by the DC bias circuit. The area of an individual SIS junction is $1\mu m^2$ with a normal state resistance of R_n = 18Ω.

3.4 Heterodyne and Direct Detection Results

Two types of high frequency tests were performed: Heterodyne testing, determining the sensitivity of the receiver as it will be used for radio astronomical observations at specific frequencies and, secondly, direct detection response measurements, performed with a Fourier Transform Spectrometer (FTS) over the range of 100 to 1000 GHz with a resolution of 20 GHz [3.29]. Note that all results measured with the FTS are convolved with the resolution of the spectrometer. The radiation source in the FTS is a matched load (Eccosorb) at liquid Nitrogen temperatures (78 K) chopped against a room temperature (295 K) load. The chopped signal is detected as an increase

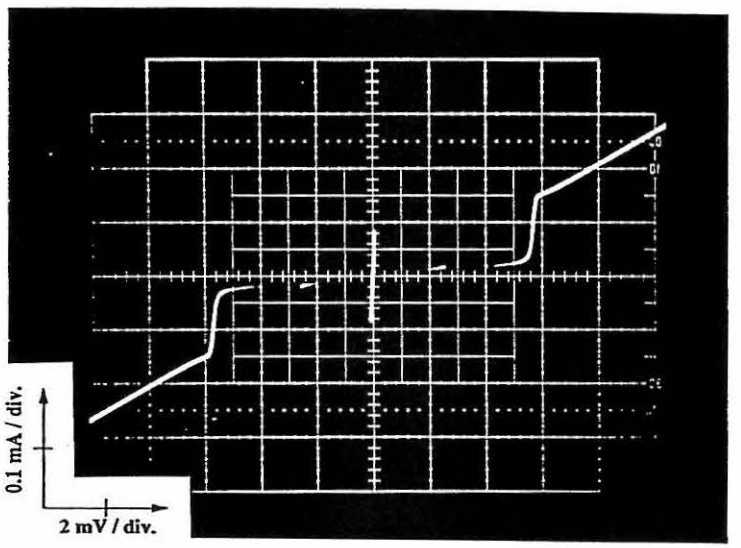


Figure 3.4 I–V characteristic of the $Nb/AlO_x/Nb$ SIS tunnel diode. The horizontal scale is 2mV per division and the vertical scale 0.1 mA per division. Using $I_c = \frac{\pi}{4} V_{gap}/R_n$ yields a critical current density of 12 kA/cm^2. The gap voltage of an individual tunnel junction is 2.9 mV yielding 5.8 mV for the two devices in series, as shown here.

in current in the SIS tunnel junction biased at 2.52 mV. The receiver's direct detection response, depicted by the solid line in figure 3.3, has been derived from the measured data by dividing by the calculated beam splitter efficiency and the calculated responsivity of the SIS junction in direct detection mode. The beam splitter is made from $76\mu m$–thick mylar foil with a dielectric constant of $\epsilon = 2.99\epsilon_o$. The thickness of the beam splitter was chosen to give optimum efficiency at 500 GHz. The responsivity of the SIS direct detector is now calculated :

The DC tunneling current was derived in chapter I equation (1.29),

$$I_o(V_b, V_\omega) = \sum_{n=-\infty}^{\infty} J_n^2(eV_\omega/\hbar\omega)\ I_{\rm DC}(V_b + n\hbar\omega/e) \tag{3.31}$$

with $I_{\rm DC}(V_b)$ the current of the unmodulated SIS junction at bias voltage V_b, and V_ω the applied RF voltage amplitude at frequency ω. For a direct detector the incoming signal is small ($eV_\omega \ll \hbar\omega$)

so that only single photon processes are relevant and (3.31) reduces to the direct current increase under chopping as given in [3.10].

$$\Delta I_{DC}(V_b) = \frac{1}{4} V_\omega^2 \; \frac{I_{DC}(V_b + \hbar\omega/e) - 2I_{DC}(V_b) + I_{DC}(V_b - \hbar\omega/e)}{(\hbar\omega/e)^2} \tag{3.32}$$

The RF impedance Z_{RF} of the tunnel junction can be calculated from the dissipative currents given by Tucker and Feldman [3.10] and the applied RF voltage V_ω. As discussed earlier one can neglect the quantum susceptance and have

$$Z_{RF}(V_b, V_\omega) = V_\omega \left(\sum_{n=-\infty}^{\infty} J_n(eV_\omega/\hbar\omega) \; [J_{n+1}(eV_\omega/\hbar\omega) + J_{n-1}(eV_\omega/\hbar\omega)] \; I_{DC}(V_b + n\hbar\omega/e) \right)^{-1} \tag{3.33}$$

which reduces to

$$Z_{RF} = \frac{2\hbar\omega/e}{I_{DC}(V_b + \hbar\omega/e) - I_{DC}(V_b - \hbar\omega/e)}. \tag{3.34}$$

Equations (3.33) and (3.34) are valid if the SIS junctions' capacitance terminates all harmonic frequencies, which is not entirely satisfied in this case. However, the effects of this are small enough for the qualitative analysis of interest here. The responsivity $\Re$ is the ratio of the current increase to the power absorbed by the junction

$$\begin{aligned} \Re &= \frac{\Delta I_{DC}}{(V_\omega^2/2Z_{RF})} \\ &= \frac{e}{\hbar\omega} \left(\frac{I_{DC}(V_b + \hbar\omega/e) - 2I_{DC}(V_b) + I_{DC}(V_b - \hbar\omega/e)}{I_{DC}(V_b + \hbar\omega/e) - I_{DC}(V_b - \hbar\omega/e)} \right) \\ &= \frac{e}{\hbar\omega} \, \eta \end{aligned} \tag{3.35}$$

with η the quantum efficiency of the direct detector in analogy to a photo–diode detector [3.30, 3.31]. The responsivity of the SIS direct detector as a function of frequency at a bias voltage of 2.52 mV was calculated using a digitized DC I–V curve of the SIS junction shown in figure 3.4. The result is shown in figure 3.5 with the dashed curve showing the responsivity of the theoretical device with a perfectly sharp I–V curve and zero sub–gap leakage currents but with otherwise identical parameters to the device of figure 3.4 (*i.e.*, $\eta = 1$ between f_c and $2f_{gap} - f_c$).

The drop off below $f_c \simeq 100$ GHz is due to the choice of the DC bias voltage $V_b = 2.52$ mV since the photon energy for frequencies $f < f_c$ is then too small for photon assisted tunneling,

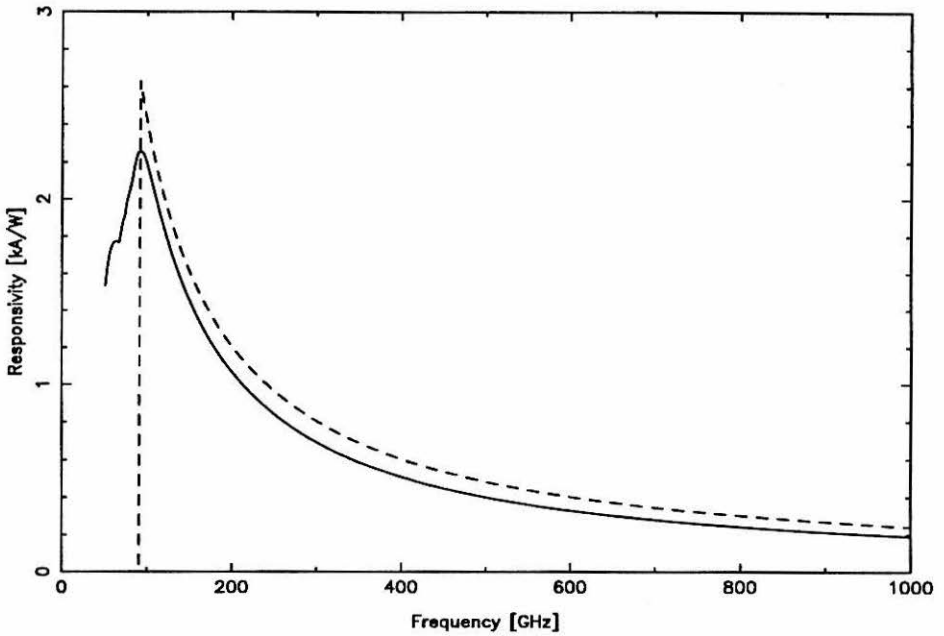


Figure 3.5 Calculated responsivity of the SIS diode in direct detection mode. Data below 50 GHz has been omitted since the photon energy is smaller than our voltage sampling of the I–V curve. This effect is also responsible for the small bump at 70 GHz. The dashed line is for a theoretical SIS direct detector with a perfectly sharp I–V curve and zero sub–gap leakage current, but otherwise identical to the one from figure 3.4.

i.e., $\hbar\omega < e(V_{gap} - V_b)$. It is interesting to note that the response of the ideal detector and the SIS direct detector are nearly identical above f_c, which is, of course, due to the sharp non–linearity and low dark current of the device. The frequency response of a quantum limited detector above the cut–off frequency is simply f^{-1} up to the gap frequency.

In the Rayleigh–Jeans limit ($h\nu \ll kT$) a single mode detector, like the SIS junction in a hybrid antenna, receives power independent of frequency, $P = kT\Delta\nu$, from a thermal source at temperature T. The throughput of the FTS is assumed to be independent of frequency. This assumption has been verified to within $\approx$ 20% at frequencies above 150 GHz. At frequencies below 150 GHz the throughput of the FTS decreases rapidly due to the limited size of the optics. The direct detection response measured with the FTS is a combination of the coupling efficiency

of the receiver optics to the FTS, the frequency response of the matching circuit and the detector, and the beam splitter efficiency. The measured data shown in figure 3.3 is corrected for the beam splitter efficiency and the responsivity of the SIS junction. The remaining corrections are assumed to have much slower variations with frequency than the matching circuit so that the characteristic frequencies of the matching circuit can be determined rather accurately.

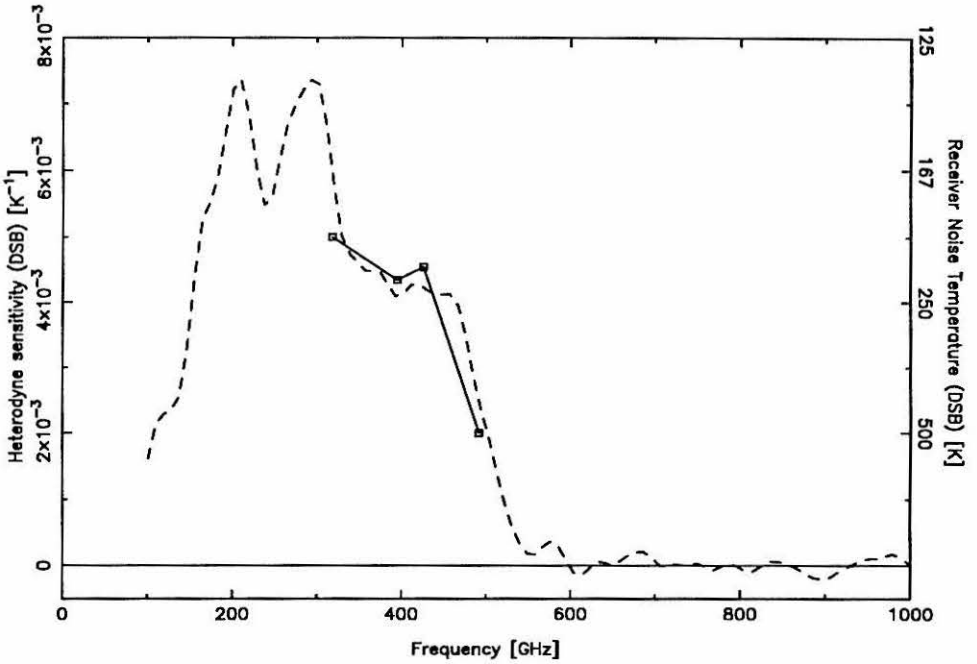


Figure 3.6 Comparison between the heterodyne (solid line) and direct detection response (dashed line). The FTS results are corrected for the beam splitter efficiency only. This plot demonstrates the usefulness for direct detection measurements done with an FTS to predict the relative frequency dependence of the heterodyne sensitivity of a receiver. The direct detection measurement has an arbitrary scale. Note, that contrary to figure 3.3 the direct detection measurement depicted here has not been corrected for the responsivity of the SIS junction, since the heterodyne sensitivity measurement is a responsivity measurement itself.

The measured bandwidth of the receiver in direct detection agrees very well with the predicted response of the matching circuit. The characteristic frequencies show only an 8% difference, approximately, which can easily be corrected in future experiments by changing the thickness of

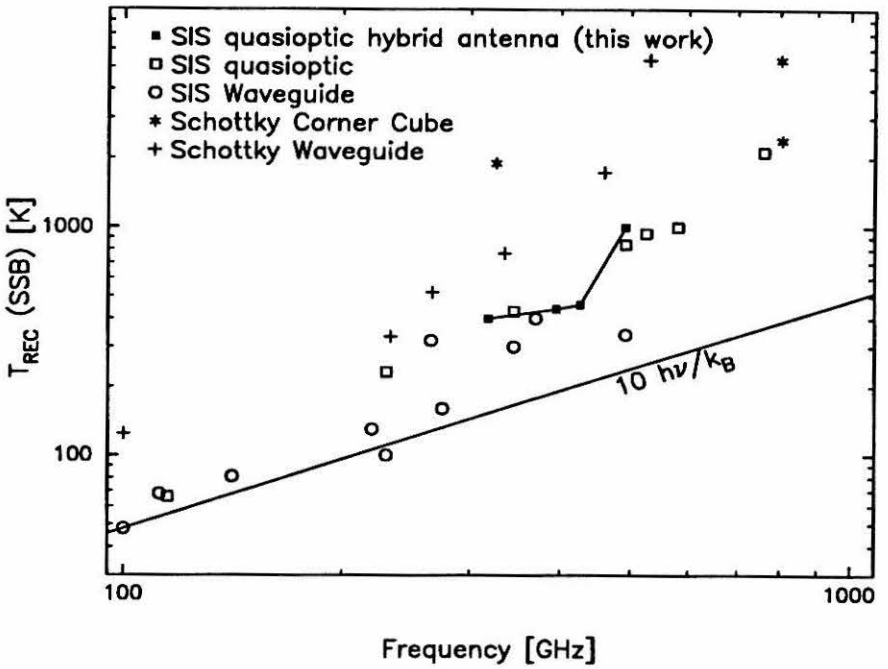


Figure 3.7 Single sideband noise temperature comparison for SIS receivers reported in the literature. The $10h\nu/k_B$ line corresponds to ten times the quantum noise limit, which is about the best currently achievable for these receiver systems.

the dielectric insulator in the transformers. This deviation is smaller than expected for a first attempt, since many of the parameters are not well known, like the penetration depth of the fields into the superconducting film, the exact thickness of the dielectric insulator between the superconducting lines, the specific capacitance, and the dielectric properties of thin film SiO. However, it cannot be concluded that the parameters are now known to within 8%, since errors of the individual parameters may cancel. Additionally, the accuracy of the numerical simulations of step discontinuities in the transmission line calculations are not known since *Touchstone* uses algorithms optimized for transmission line height to width ratios near 1, whereas the geometry used here has height to width ratios much smaller than 1.

Table 3.3 Receiver noise temperatures

Frequency [GHz]	318	395	426	492
T_{sys} (DSB) [K]	200	230	220	500

Heterodyne measurements were performed at 318 GHz, 395 GHz, 426 GHz, and 492 GHz. The double sideband noise temperatures for those frequencies are given in table 3.3. The increase to 500 K at 492 GHz is at least in part due to the fact that the impedance transformer rolls off at about 475 GHz as seen with FTS measurements. Figure 3.6 is a comparison between the direct detection response (corrected for the beam splitter efficiency only as measured by the FTS) and the heterodyne response showing that the receiver's heterodyne frequency range can be predicted with FTS direct detection measurements. The sensitivity of the receiver compares very well to the most sensitive submillimeter receivers reported in the literature (see figure 3.7) to date (*e.g.*, Walker *et al.* 1992 [3.7]) , most being based on tunable waveguide mounts.

3.5 Chapter Summary and Technological Outlook

A receiver for millimeter and submillimeter wavelengths has been demonstrated successfully using $Nb/AlO_x/Nb$ tunnel junctions and a fixed–tuned broadband matching circuit made from Nb with SiO for the dielectric. This design combines high sensitivity with the ease of fabrication and operation of quasi– optical mixers. Using a two–section real impedance transformer in conjunction with an end–loaded–stub, the matching circuit covers over an octave of bandwidth, from 200 GHz to 475 GHz, which is higher than that of the best waveguide mixers, without having to use external tuners like most waveguide mounts. The heterodyne receiver sensitivity is comparable to the best numbers reported in the literature although the best waveguide receivers have slightly better performances. The quasi–optical mixer is manufactured as an integrated circuit and is the best candidate for imaging array receiver systems. The matching circuit can also be used in waveguide mixers, reducing the losses from mechanical tuners or even making them unnecessary.

3.6 References Chapter III

[3.1] Ogawa, H., Mizuno, A., Hoko, H., Ishikawa, H., and Fukui, Y., "A 110 GHz SIS Receiver For Radio Astronomy," *Int. J. of IR and MM Waves*, Vol. 11, No. 6, 1990.

[3.2] Kooi, J.W., Chan, M., Phillips, T.G., Bumble, B., and LeDuc, H.G., "A Low Noise 230 GHz Heterodyne Receiver Employing $.25\mu m^2$ Area $Nb/AlO_x/Nb$ Tunnel Junctions," *IEEE Microwave Theory and Tech.*,, submitted.

[3.3] Ellison, B.N., Schaffer, P.L., Schaal, W., Vail, D., and Miller, R.E., "A 345 GHz SIS Receiver For Radio Astronomy," *Int. J. of IR and MM Waves*, Vol. 10, No. 8, 1989.

[3.4] Carter, M.C., Navarro, S., Karpov, A., Billon–Pierron, D., Lehnert, T., Gundlach, K.H., "Receiver Development with Nb/Al–Oxide/Nb SIS Mixers in the Frequency Ranges of (210 – 270) and (320 – 370) GHz," *Proc. 16th Int. Conf. Infrared and Millimeter Waves*, eds. M.R. Siegrist, M.Q. Tran, & T.M. Tran, Ecole Polytechnique Fédérale de Lausanne, p. 218, 1991.

[3.5] Stern, J., private communication.

[3.6] Kircher, C.J., and Murakami, M., "Josephson Tunnel–Junction Electrode Materials," *Science*, vol. 208, pp. 944 – 950, May 1980.

[3.7] Walker, C.K., Kooi, J.W., Chan, M., LeDuc, H.G., Carlstrom, J.E., and Phillips, T.G., "A Low–Noise 492 GHz SIS Waveguide Receiver," *Int. J. of IR and MM Waves*, Vol. 13, No. 6, pp. 785 – 798, 1992.

[3.8] Büttgenbach, T.H., LeDuc, H.G., Maker, P.D., Phillips, T.G., "A Fixed Tuned Broadband Matching Structure For Submillimeter SIS Receivers," *IEEE Trans. Applied Supercond.*, Vol. 2, No. 2, September 1992.

[3.9] Smith, A. D., Richards, P. L., "Analytic Solution to Superconductor-Insulator-Superconductor Quantum Mixer Theory," *J. Appl. Physics*, Vol.53, pp. 3806 – 3812, May 1982.

[3.10] Tucker, J. R., Feldman, M.J., "Quantum Detection at Millimeter Wavelength," *Rev. Mod. Physics*, Vol.57, pp. 1055 – 1113, 1985.

[3.11] Worsham, A.H., Ugras, N.G., Winkler, D., Prober, D.E., Erickson, N. R., and Goldsmith, P.

F., "Quantum Tunneling Currents in a Superconducting Junction," *Phys. Rev. Letters*, vol. 67, pp. 3034 – 3037, Nov. 1991.

[3.12] Räisänen, A. V., McGrath, W. R., Richards, P. L. and Lloyd, F. L., "Broadband RF Match to a Millimeter–Wave SIS Quasi–Particle Mixer," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT–33, No. 12, pp. 1495 – 1500, 1985.

[3.13] D'Addario, L. R., "An SIS Mixer for 90 – 120 GHz with Gain and Wide Bandwidth," *Int. J. of IR and MM Waves*, Vol. 5, No. 11, pp. 1419 – 1433, 1984.

[3.14] Kerr, A.R., Pan, S.–K., Whiteley, S., Radparvar, M., and Faris, S., "A Fully Integrated SIS Mixer for 75 to 110 GHz," *IEEE International Microwave Symposium*, May 1990.

[3.15] Kerr, A.R., Pan, S.–K., Lichtenberger, A.W., and Lea, D.M., "Progress on Tunerless SIS Mixers for the 200 – 300 GHz Band," *IEEE Microwave and Guided Wave Lett.*, Vol. 2, No. 11, Nov 1992.

[3.16] Kerr, A.R., and Pan, S.–K., "Integrated Tuning Elements for SIS Mixers," *Int. J. of IR and MM Waves*, Vol. 9, No. 2, pp. 203 – 212, 1988.

[3.17] Ermakov, An.B., Koshelets, V.P., Kovtonyuk, S.A., and Shitov, S.V., "Parallel Biased SIS–Arrays for MM Wave Mixers: Main Ideas and Experimental Verification," *IEEE Trans. Magn.*, Vol. 27, No. 2, pp. 2642 – 2645, 1991.

[3.18] Zmuidzinas, J., LeDuc, H.G., "Quasi Optical Slot Antenna SIS Mixers," *IEEE Trans. Microwave Theory Tech.*, Vol. 40, No. 9, 1992.

[3.19] Carpenter, J.A., Smith, A.D., Arambula, E.R., Lee, L.P.S., Nelson, T., and Yujiri, L., "100 GHz SIS Mixer With Improved RF Matching," *IEEE Trans. Magn.*, Vol. 27, No. 2, pp. 2654 – 2657, 1991.

[3.20] Büttgenbach, T.H., Miller, R.E., Wengler, M.J. Watson, D.M., Phillips, T.G., "A Broad–band Low–Noise SIS Receiver for Submillimeter Astronomy," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT–36, No. 12, pp. 1720 – 1726, 1988.

[3.21] Büttgenbach, T.H., "An Improved Solution For Integrated Array Optics In Quasi Optical MM and SUBMM Receivers: The Hybrid Antenna," *IEEE MTT* , to appear in October 1993.

[3.22] Feldman, M.J. and D'Addario, L.R., "Saturation of the SIS direct detector and the SIS Mixer," *IEEE Trans. Magnetics*, Vol. MAG–23, 1987.

[3.23] Young, L., "Tables for Cascaded Homogeneous Quarter–Wave Transformers," *IRE Trans. Microwave Theory Tech.*, vol. MTT–7,No 2, pp. 233 – 237, 1959.

[3.24] Touchstone CAD Software, EEsof, Westlake Village, CA.

[3.25] Büttgenbach, T.H. , Keene, J.B., Walker, C.K., and Phillips, T.G., "Submillimeter Detection Of Extragalactic C I Emission: IC342," *Astrophysical Journal Letters*, 397, L15 – L17, September 1992.

[3.26] Bode, H.W. "Network Analysis and Feedback Amplifier Design," *Van Nostrand*, N.Y., 1945.

[3.27] Fano, R.M, "Theoretical Limitations on the Broad–Band Matching of Arbitrary Impedances," *J. of the Franklin Institute*, vol. 249, pp. 57 – 83, Jan. 1950, and, pp. 139 – 154, Feb. 1950.

[3.28] LeDuc, H.G., Stern, J.A., Thakoor, S., Khanna, S., "All Refractory NbN/MgO/NbN SIS Tunnel Junctions," *IEEE Trans. Magn.*, vol. MAG–23, pp. 863 – 865, 1987.

[3.29] Hu, Qing, Mears, C. A., and Richards, P. L., "Measurements of Integrated Tuning Elements For SIS Mixers With A Fourier Transform Spectrometer," *Int. J. of IR and MM Waves*, Vol. 9, No. 4, pp. 303 – 320, 1988.

[3.30] Wengler, M.J., and Woody, D.P., "Quantum Noise In Heterodyne Detection," *IEEE J. Quantum Electron.*, Vol. QE–23, pp. 613 – 622, 1987.

[3.31] Wengler, M.J., "Submillimeter Wave Detection with Superconducting Tunnel Diodes," Proc. IEEE , Vol. 80, No. 11, pp. 1811 – 1826, 1992.

CHAPTER IV

ASTROPHYSICAL OBSERVATIONS OF C I IN THE INTERSTELLAR MEDIUM

This chapter concludes this thesis with a presentation of astrophysical observations and interpretation of the distribution of neutral carbon in the interstellar medium. Most of the observational data presented here were taken with quasi–optical SIS receivers as described in the previous chapters. Since the observations were carried out in parallel with the development of the quasi–optical receiver system, the actual quasi–optical receivers used for the different observing runs represent different stages of the quasi–optical receiver development. To compare the performance of the quasi–optical receiver with a conventional waveguide system some of the 492 GHz measurements were repeated with a conventional SIS diode waveguide receiver that became available as a facility instrument on the Caltech Submillimeter Telescope in the fall of 1991.

4.1 Presence of Neutral Carbon in the ISM

Carbon is the fourth most abundant element in the universe after hydrogen, helium and oxygen, with an abundance relative to hydrogen of 3.3×10^{-4} (Allen 1973, see table 4.1). It is an important element in molecular chemical processes and is a constituent of many commonly observed interstellar molecules (CO, CS, CH, HCN, H_2CO, etc.), and, in fact, of all observed interstellar molecules larger than 4 atoms, except SiH_4 (Verschuur 1992). However, its interstellar distribution, especially in the neutral atomic and ionic states, is not well known. The observations made for this thesis represent the first extragalactic submillimeter detection of neutral carbon and some of the first high spatial resolution studies of neutral carbon in the interstellar medium (ISM).

Atomic carbon plays an important role in the cooling of the dense interstellar medium because of its large abundance, widespread distribution, the small splitting of its ground state fine–structure levels ($E(^3P_1)/k = 23.6$ K and $E(^3P_2)/k = 62.5$ K – see figure 4.1), and the moderate H_2 density

required for collisions to dominate the excitation of these levels ($n_{crit}(^3P_1) \approx 10^3$ cm^{-3} and $n_{crit}(^3P_2) \approx 2 \times 10^3$ cm^{-3}; Schröder *et al.* 1991). C I and CO provide the bulk of the gas cooling for the cool component of the interstellar medium (ISM) of our Galaxy. As shown by observations with the Cosmic Background Explorer satellite (COBE; Wright *et al.* 1991), C I is a more important coolant for our Galactic ISM than CO. The strongest CO line ($J = 4 \rightarrow 3$) detected by COBE is not as intense as either C I line, and the sum of all observed CO lines is not as intense as the sum of the two C I lines. However, the emission from both C I and CO is overpowered by the extremely bright C II emission which arises primarily from the warm photo-dissociated regions (PDRs) at the surfaces of dense clouds.

Table 4.1 Cosmic Abundances

Element	Abundance	Ionization Threshold (eV)
H	1.0	13.60
He	8.5×10^{-2}	24.59
O	6.6×10^{-4}	13.62
C	3.3×10^{-4}	11.26
N	9.1×10^{-5}	14.53
Ne	8.3×10^{-5}	21.56
Fe	4.0×10^{-5}	7.87
Si	3.3×10^{-5}	8.15
Mg	2.6×10^{-5}	7.65
S	1.6×10^{-5}	10.36
Molecule		**Dissociation Threshold (eV)**
CO		11.1

The neutral carbon (CI) ground state fine structure line ($^3P_1 \rightarrow {}^3P_0$) at 492 GHz (or 609 μm) has been observed over the past decade within our Galaxy (*e.g.*, Phillips and Huggins 1981; Keene *et al.* 1985; Frerking *et al.* 1989; White and Padman 1991) and found to be bright, widespread, and bearing a strong resemblance in antenna temperature, line shape, and spatial distribution to the low J rotational lines of ^{13}CO. This resemblance has been attributed to the similar opacities and excitation requirements of C I and ^{13}CO and shows that, to some degree, the species must coexist.

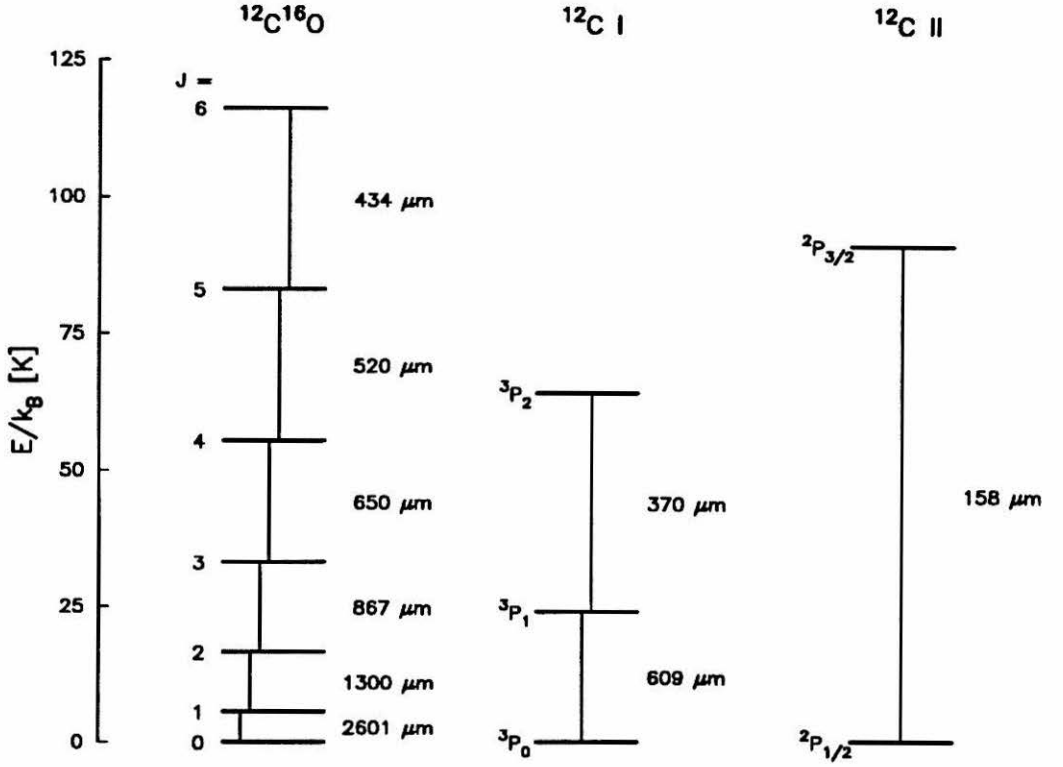


Figure 4.1 Energy level diagrams of CO, C I, and C II (from Keene 1990). These three species are the most common gas phase species including carbon. With the exception of the two lowest rotational transitions of CO, all transitions are in the submillimeter band. The temperature scale shows the good energetic match to typical molecular clouds in the ISM with $T = 10 - 100K$.

The most prominent theory of carbon gas phase abundances is that of ion–chemistry models which include photo-dissociation processes (Langer 1976; Tielens and Hollenbach 1985; van Dishoeck and Black 1988). In this theory, henceforth simply referred to as PDR model, the significant neutral carbon abundance requires the presence of UV radiation creating a photo–dissociation region (PDR) on the surface of a cloud that is externally illuminated by UV radiation. UV radiation of energy higher than 13.6 eV, *i.e.*, radiation that is capable of ionizing hydrogen, typically gets absorbed close to its own source so that it does not penetrate the ISM significantly. This leaves hydrogen in PDRs in the neutral atomic or molecular form. The part of the UV radiation that can penetrate the ISM, *i.e.*, with energy lower than 13.6 eV, is called near–ultraviolet (NUV) radiation. Since CO is a very stable molecule and the cosmic abundance of oxygen is

twice as high as that for carbon, previous PDR models find that all the carbon is tied up in CO inside a molecular cloud where it is shielded from NUV radiation. According to these PDR models, only on the surface of the molecular cloud where the NUV flux is sufficiently high can neutral carbon exist through photo–dissociation of CO. The photo–dissociation threshold for CO is 11.1 eV, which is very close to the ionization threshold of carbon at 11.26 eV (see table 4.1), so that carbon, once set free, will quickly be ionized. The surface layer in which neutral carbon can exist according to the PDR models is thus thin, bounded towards the UV source by C II and towards the center of the molecular cloud by CO. It only extends over a few A_V at an optical depth of about $A_V = 8 - 10$ into the cloud for an incident NUV flux of $10^2 - 10^4 F_o$, where F_o is our local Galactic radiation field, 1.6×10^{-3} erg s^{-1} cm^{-2} in the 6–13.6 eV energy range (Habing 1968).

If PDRs are the dominant source of the emission for C I, C II, and CO lines, then, indeed, a correlation between the emission strengths of the species is to be expected. Such a correlation has been established by Crawford *et al.* (1985) and Stacey *et al.* (1991) for C II vs. CO ($J = 1 \rightarrow 0$) in star-forming molecular clouds and starburst galaxies.

However, work by Keene *et al.* (1985) showed that neutral carbon can be widespread within molecular clouds and that the column density peaks even further into the cloud (away from the UV source) than that of CO, in contradiction to the PDR models. The goals of the investigations described in this chapter were to find out what the distribution of neutral carbon in the ISM is, and how it relates to the well studied distribution of CO. As will be discussed in section 4.6, the PDR models can be modified to raise the C I abundance inside molecular clouds using a revised ion chemistry.

To investigate the distribution and abundance of neutral carbon at large scales a nearby galaxy was chosen: IC 342. The results of these first extragalactic submillimeter observations of C I (Büttgenbach *et al.* 1992a), and a comparison to COBE data of the Milky Way, are presented in the following two sections. The small scale structure of neutral carbon in the ISM is studied with observations of the Orion Molecular Cloud (OMC) in the vicinity of the embedded infrared source IRc2 and the ionizing stars of the Trapezium (Büttgenbach *et al.* 1993). Those findings will be presented in sections 4.4 and 4.5. Theoretical calculations of ion chemistry in the interstellar

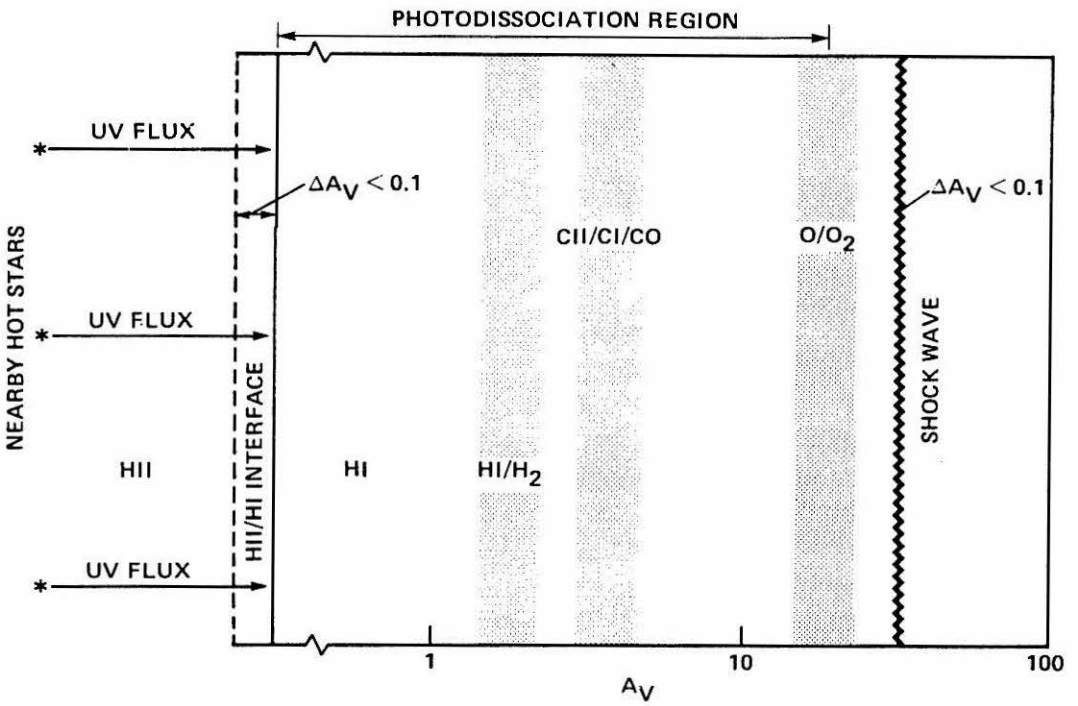


Figure 4.2 Photo–dissociation region (PDR) model of the transition from ionized medium to molecular medium (from Tielens and Hollenbach 1985). In the PDR models neutral carbon is embraced in a thin layer between ionized carbon close to the ionizing source and CO inside the molecular cloud that ties up all the gas phase carbon.

medium were performed for this thesis, using code by Le Bourlot *et al.* (1992), which allow for an explanation of the observed relation of neutral carbon to CO. These results will be discussed in section 4.6 as a modification to previous PDR models. The chapter is concluded with a summary describing a unified model of the observed large and small scale distribution and abundance of neutral carbon.

4.2 First Submillimeter Detection of Extragalactic C I: IC342

Within the Milky Way, previous studies have shown that there is a general correlation of C I and CO emission. The first extragalactic detection of CO dates back to 1975 by Rickard *et al.* However, until now, no extragalactic observations of C I have been possible. This section reports the first observation of CI in IC 342 ($\alpha(1950) = 3^h\ 41^m\ 57^s$, $\delta(1950) = 67°\ 56'\ 29''$), a nearby Scd type galaxy. It has a low inclination angle ($i = 25°$, *i.e.*, nearly face on), resulting in relatively narrow emission lines. Its optical size is $> 10'$. The nuclear bar ($\approx$ 700 pc) shows strong infrared emission with an infrared luminosity in the 3 to 1000μm wavelength range of $4 \times 10^9 L_\odot$ (Becklin *et al.* 1980). The high IR luminosity indicates active star formation in the central region of the galaxy.

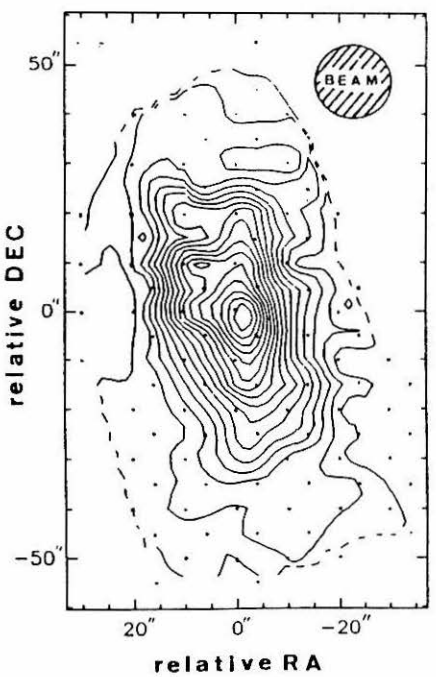


Figure 4.3 Integrated line intensity of ^{12}CO ($2 \rightarrow 1$) emission in the nuclear bar of IC 342 (from Eckart *et al.* 1990). Contour intervals of 20 K km s^{-1} with a peak of 324 K km s^{-1}. The dots mark positions where data were taken.

IC 342 was first detected and partially mapped in CO by Morris and Lo (1978) and subsequently mapped by Rickard and Palmer (1981), Young and Scoville (1982), Lo *et al.* (1984),

Eckart *et al.* (1990), and Ishizuki *et al.* (1990). Figure 4.3 shows a map of the nuclear bar of IC 342 of ^{12}CO (2 → 1) emission (Eckart *et al.* 1990). As observed by Eckart *et al.* (1990) the main beam temperatures of the ^{13}CO (1 → 0) and (2 → 1) transitions in IC 342 are high for extragalactic sources, about 0.5 K in a 21″ and 14″ beam, respectively.

a) Observations

The $^3P_1 \rightarrow {}^3P_0$ fine structure transition of C I lies at the frequency of 492.1607 GHz (Frerking *et al.* 1989; Yamamoto and Saito 1991). Most of the previous astronomical data (*e.g.*, Phillips and Huggins 1981; Keene *et al.* 1985) were taken from the Kuiper Airborne Observatory with a beam size of 2′.8. Although the atmospheric transparency at airplane altitudes (12.5 km) is nearly 100%, thus allowing high sensitivity, the instantaneous bandwidth of the InSb hot–electron bolometer receiver in use at that time was too small to detect the broad lines of extragalactic sources. The work reported here was possible due to the wide instantaneous bandwidth of the new SIS receivers. These receivers have been developed over the past few years using quasi–optical techniques (see chapters I through III or Büttgenbach *et al.* 1988, 1992b; Büttgenbach 1993) and waveguide techniques (Ellison *et al.* 1989; Walker *et al.* 1992) for use at the Caltech Submillimeter Observatory (CSO), a 10.4 m diameter submillimeter telescope on Mauna Kea. The large atmospheric opacity at 492 GHz requires excellent observing conditions even at a high site like Mauna Kea (4100 m). On the best night for these observations the atmospheric transmission at a typical zenith angle for this source (50°) was found to be about 25%, which corresponds to 0.8 mm of precipitable water vapor (Grossman 1989). With a distance to IC 342 of 1.8 Mpc (McCall 1989) the 15″ beam of the CSO at the line frequency of C I (492 GHz) yields a linear spatial resolution of 130 pc.

IC 342 was detected in three separate observing runs at the CSO. During the first, in October 1990, a quasi–optical receiver similar to that described in chapter I and by Büttgenbach *et al.* (1988), but improved by utilizing a hybrid antenna fed by a planar logarithmic–spiral structure (see chapter II and Büttgenbach 1993), was used. The measured coupling efficiency at 492 GHz between the receiver and the telescope was about the same as for that of the waveguide based system (Walker *et al.* 1992) used for the September and October 1991 observations. The measured beam efficiencies (η_B) of both receivers at 492 GHz on the Moon was found to have a value of

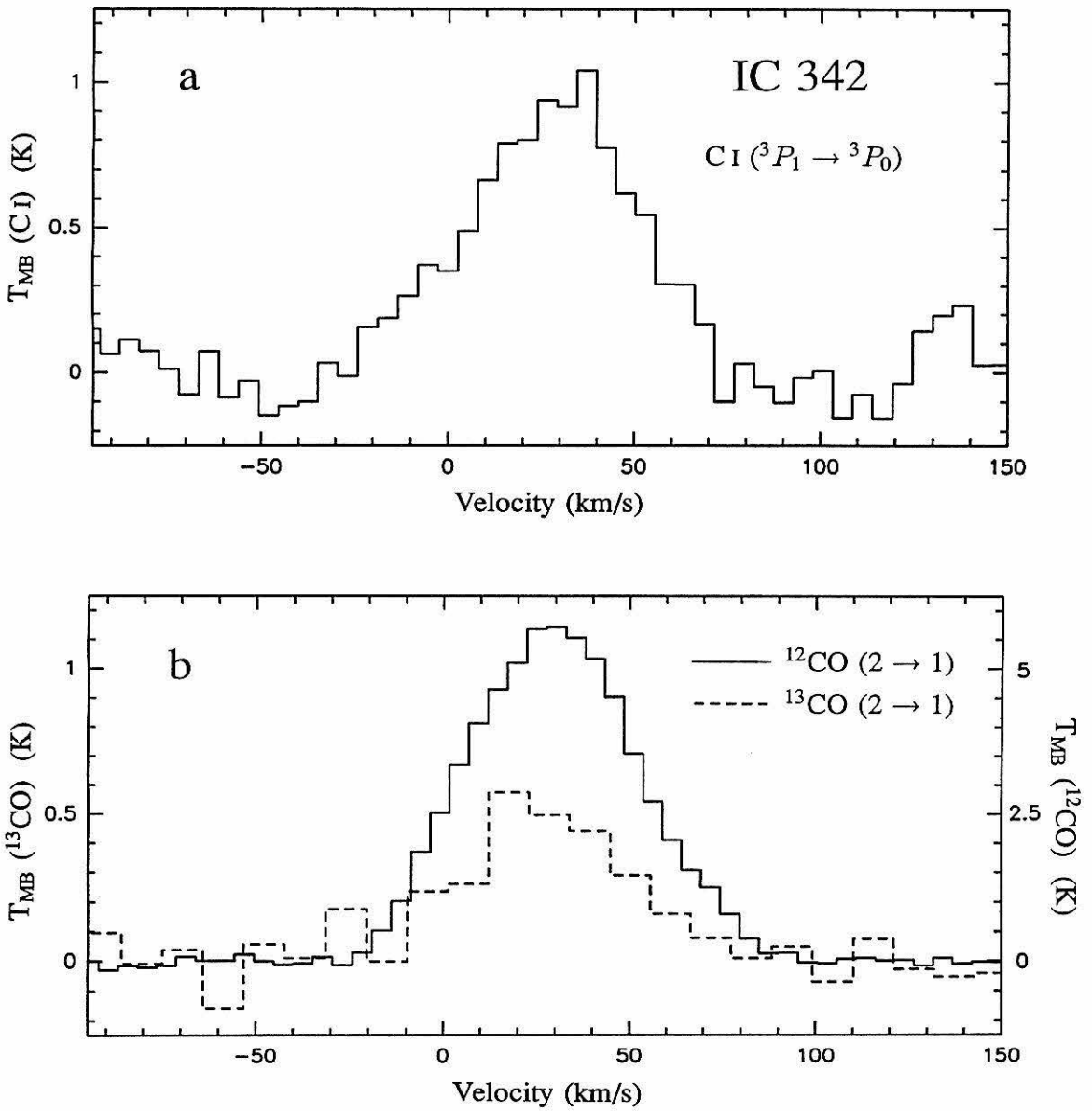


Figure 4.4 *a*) C I ($^3P_1 \rightarrow {}^3P_0$) spectrum of the central 15″ of IC 342. The C I spectrum is the weighted sum of all three individual observations taken at the CSO in the period of October 1990 to October 1991 with a linear baseline removed. *b*) ^{12}CO and ^{13}CO(2 → 1) spectra (solid and dashed lines, respectively) from Eckart *et al.* (1990) with 14″ beam size. The ^{12}CO spectrum has been divided by 5 for ease of comparison.

0.68 ± 0.03 assuming a Moon temperature of 395 K and an emissivity of 0.97 (Linsky 1973). The main beam efficiency η_{MB} as measured on Jupiter (33.″9 mean diameter) was 0.38 ± 0.02 assuming a brightness temperature of $T_A = 163K$ (Hildebrand *et al.* 1985). By mapping Mars the beam size was measured to be 15″. The pointing accuracy was approximately 5″.

All spectra were calibrated with the standard hot-sky chopper method and divided by the main beam efficiency. This method of calibration is appropriate for a source the size of IC 342's nuclear bar (approximately 10″ EW ×40″ NS in CO; Eckart 1990), as shown in figure 4.3 in the ^{12}CO (2 → 1) transition, and for comparison with the ^{13}CO(2 → 1) and ^{12}CO(2 → 1) spectra of Eckart *et al.* (1990) since they used the same method and also had essentially the same beam size (14″). To confirm that the extent of C I emission in IC 342 is the same as the 10″ × 40″ found for CO three additional positions were observed in September of 1992. A cut from the center towards the south was chosen to minimize the pointing accuracy requirements due to the four times larger size of IC 342's nuclear bar in that direction (see figure 4.3).

b) Results

Figure 4.4a shows the average of the spectra of the central position of IC 342 from all three observing runs of the $^3P_1 \rightarrow {}^3P_0$ fine structure line of C I. The C I line has a peak main beam temperature of 1.0 ± 0.2 K and an integrated line temperature of 51 ± 10 K km s^{-1}, or equivalently an intensity of $(6.2 \pm 1.2) \times 10^{-6}$ erg s^{-1} cm^{-2} sr^{-1}. For comparison, figure 4.4b shows the ^{12}CO(2 → 1) and ^{13}CO(2 → 1) spectra measured by Eckart *et al.* (1990) with the IRAM 30 m telescope. The ^{12}CO(2 → 1) spectrum shown has been divided by a factor of five. The C I and CO lineshapes are very similar suggesting that the emission comes from the same regions. The ^{12}CO line is shown to be optically thick for the center of IC 342 by Eckart *et al.* (1990) and Wall and Jaffe (1990), while the ^{13}CO line's opacity is at most moderately thick ($\tau \leq 1$). The difference in intensity of the ^{12}CO and ^{13}CO line could arise mainly from the different filling factors of the emission from the different isotopes. Since the main beam temperature of the C I line is a little greater than that of the ^{13}CO(2 → 1) line, it would follow that the C I emission has an opacity similar to that of ^{13}CO(2 → 1) and that the area filling factor lies between that of the ^{12}CO and ^{13}CO sources. However, as will be discussed in section 4.4, it is also possible that the ^{13}CO emission is from regions of colder gas than the ^{12}CO. The strength of the C I emission

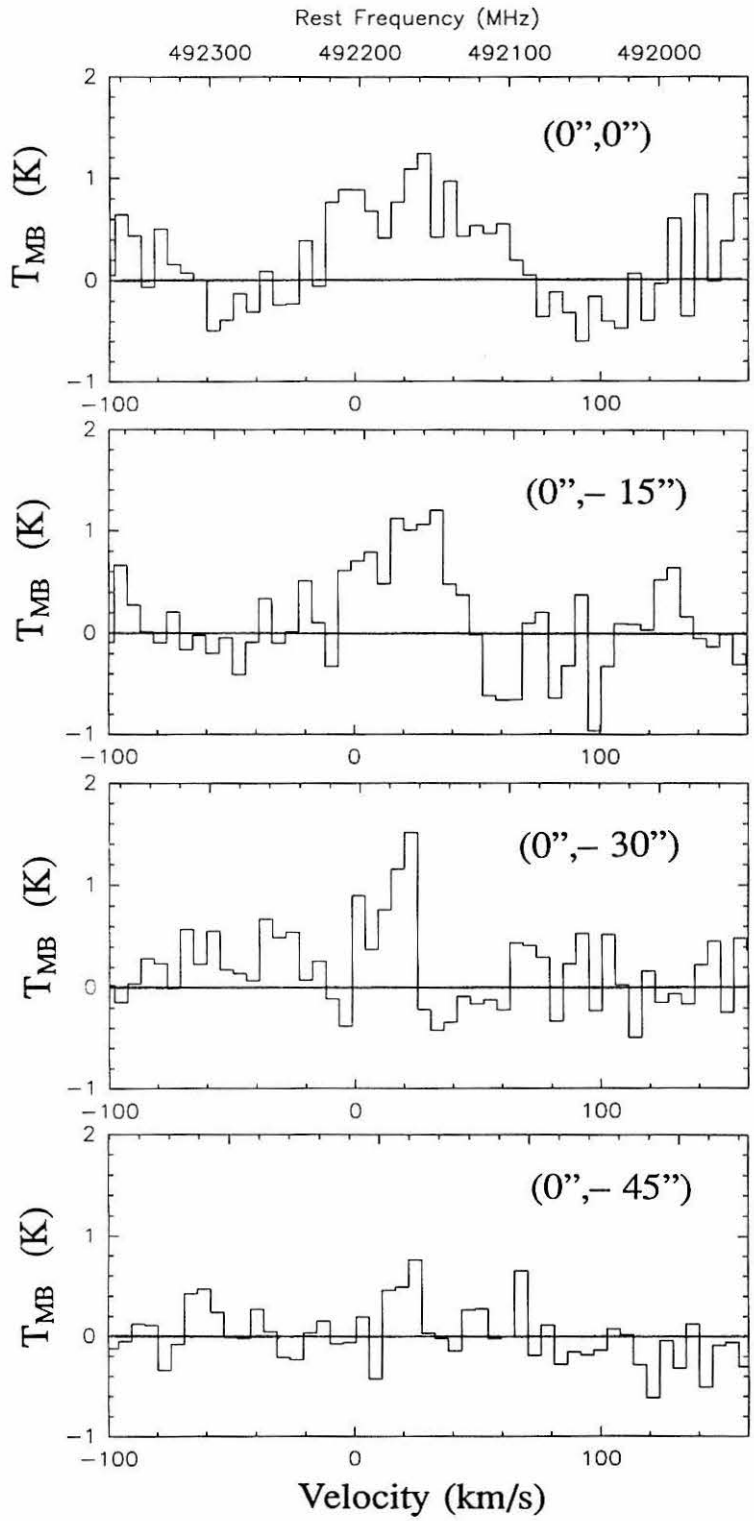


Figure 4.5 Cut through IC 342 in C I emission from the center towards the south in 15″ steps.

(stronger than ^{13}CO but weaker than ^{12}CO) could then be due to a combination of C I having a somewhat lower opacity than ^{13}CO, *i.e.*, $\tau < 1$, but emanating mostly from the same warm regions as ^{12}CO. Observations of the C I $^3P_2 \rightarrow {}^3P_1$ transition at 809 GHz (or 371μm) would be necessary to confirm this.

To determine the column density of neutral carbon it is assumed that the C I emission is optically thin or at most of moderate opacity. The low opacity assumption, if anything, leads to an underestimation of the actual column density. The formula to calculate the total neutral carbon column density is derived in the appendix, section 4.8,

$$N_{\rm CI}^T = 1.9 \times 10^{15} \left[e^{\frac{23.6}{T_{ex}}} + 3 + 5e^{\frac{-38.9}{T_{ex}}} \right] \frac{\tau_o}{1 - e^{-\tau_o}} \int \hat{T}_R^* \, dV \quad , \tag{4.1}$$

where $\hat{T}_R^*$ is the observed Rayleigh–Jeans† line temperature, corrected for telescope losses and beam efficiency. The opacity correction term, $\tau_o/1 - e^{-\tau_o}$, can be ignored due to the low opacity assumption. As shown in the appendix, the excitation temperature, T_{ex}, is uncritical for neutral carbon as long as it is above 15 K, but will be more critical for the CO column density derivation used for comparison with the C I data. The lowest number for the gas temperature of IC 342 found in the literature is by Eckart *et al.* (1990), who derive a lower limit of 20 K. An upper limit can be derived from measurements by Ho, Turner and Martin (1987) of the 1.2 cm inversion lines of NH_3, which emanate mostly from dense hot cores. They derive an NH_3 excitation temperature of about 70 K. An average from the lower and upper limit agrees well with the dust temperatures of 42 K derived by Becklin *et al.* (1980) with a 1′ beam. Güsten *et al.* (1993) recently observed the CO(4 → 3) transition at the CSO, with about the same beam size as the C I data, and fitted a two component model to their data deriving $T_1 = 33K$ and $T_2 = 50K$ for the two components, respectively. The average of their two numbers incidentally is also 42 K, which will be adopted here as the excitation temperature for C I and CO. With the low opacity assumption, this yields $N_{\rm CI}^T \geq 7 \times 10^{17}{\rm cm}^{-2}$.

It is interesting to compare the C column density to the total CO column density as derived by Eckart *et al.* (1990) of $N_{\rm CO}^T = 4 \times 10^{18}{\rm cm}^{-2}$, using an excitation temperature of $T_{ex} = 20K$.

† Rayleigh–Jeans temperatures are denoted with a hat–symbol throughout this thesis – see the appendix for conversion between R.J. temperatures, $\hat{T}$, and kinetic temperatures, T.

This implies that the ratio of $N_{\rm CI}^T/N_{\rm CO}^T$ is about 18%, a rather high number for PDR models. PDR models usually will yield a maximum of 10% for a narrow range of optical depth into the cloud (see figure 4.2) and about 0.1% deeper into the cloud (Hollenbach, Takahashi and Tielens 1991). Using the observed $C^{18}O$ intensity of Eckart *et al.* (1990) with an average excitation temperature of 42 K from Güsten *et al.* (1993) yields a somewhat higher CO column density (see the appendix for a derivation of CO column densities based on $C^{18}O$ measurements) of $N_{\rm CO}^T = 5.4 \times 10^{18}{\rm cm}^{-2}$, which is consistent with $N_{\rm CO}^T \simeq 5 \times 10^{18}{\rm cm}^{-2}$ derived by Güsten *et al.* 1993. This yields a carbon to CO ratio of $N_{\rm CI}^T/N_{\rm CO}^T \simeq 0.13$, *i.e.*, $\simeq$ 13%. Using the N_{H_2} column density of Eckart *et al.* (1990), but corrected for the different excitation temperature of CO from Güsten *et al.* (1993), yields a total hydrogen nucleus column density of $N_{\rm H}^T \simeq 10^{23}{\rm cm}^{-2}$. The cosmic abundance of carbon relative to hydrogen is 3.3×10^{-4} (Allen 1973), which implies that about 2% of all carbon in the ISM is in the form of neutral carbon. Assuming that most of the carbon in the gas phase is tied up in CO, yields that of all the carbon in the gas phase more than 10% is in form of neutral atomic carbon. Despite all uncertainties, it is clear that a huge amount of neutral carbon is observed in the ISM.

Table 4.2 $I({\rm C\,I}\ (1 \rightarrow 0))/I({\rm CO}(2 \rightarrow 1))$ Ratios for different positions in IC 342

Position relative to center $(\Delta\alpha, \Delta\delta)$	(0″,0″)	(0″,– 15″)	(0″,– 30″)	(0″,– 45″)
C I : $\int T_{\rm MB} dv$ $[K\ km\ s^{-1}]$	51 ± 10	31 ± 13	15 ± 8	9 ± 6
$I({\rm C\,I}\ (1 \rightarrow 0))/I({\rm CO}(2 \rightarrow 1))$	1.4 ± 0.4	1.3 ± 0.6	1.5 ± 0.7	0.7 ± 0.8

Figure 4.5 shows four spectra of IC 342 from the center towards the south in 15″ steps taken in September 1992. Table 4.2 summarizes the integrated line intensities and the ratio of the integrated line intensity of C I to CO(2 → 1). The ratios are constant for the first three positions and only deviate for the southern most position where the signal to noise ratio of the C I measurement is poor. The central line velocities of the C I spectra also agree well with those of the CO(2 → 1) from Eckart *et al.* (1990). It can thus be concluded that there is a strong correlation on large scales between C I and CO emission.

4.3 Comparison with COBE Data

The measurements of IC 342 can be compared with our Galaxy by making use of data from the FIRAS instrument on COBE. In this important work, Wright *et al.* (1991) have generated a FIR intensity weighted average spectrum of the Milky Way and have derived line intensities for several lines. The useful FIRAS frequency coverage is from about 200 GHz to 3 THz, with a resolution of 20 GHz. The transition lines that were detected by COBE are summarized in table 4.3 (from Wright *et al.* 1991).

An interesting comparison is that $I(\mathrm{C\,I}\ (1 \rightarrow 0))/I(\mathrm{CO}(2 \rightarrow 1)) = 1.4 \pm 0.4$ in IC 342 and 2.3 ± 0.6 in the Milky Way. In the CO component of this ratio for IC 342 the line intensities of the ^{12}CO, ^{13}CO, and C^{18}O ($2 \rightarrow 1$) transitions have been included, which is what is measured by COBE for the Milky Way. Also, the ^{13}C isotopic line of C I lies within the linewidth of the ^{12}C I line (Cooksy *et al.* 1986) and is therefore included in the C I measurement but, because of the, at most, moderate opacity of the ^{12}C I line, the ^{13}C I is thought to be unimportant.

Table 4.3 Molecular and Atomic Transition Lines Detected by COBE in the Milky Way

Species	Transition	Frequency GHz	Wavelength μm	Intensity erg s^{-1} cm^{-2} sr^{-1}
CO	$v = 0$: $J = 2 \rightarrow 1$	230.54	1302	$(2.8 \pm 0.4) \times 10^{-8}$
CO	$v = 0$: $J = 3 \rightarrow 2$	345.80	867.2	$(3.9 \pm 0.9) \times 10^{-8}$
CO	$v = 0$: $J = 4 \rightarrow 3$	461.04	650.4	$(4.2 \pm 0.9) \times 10^{-8}$
CO	$v = 0$: $J = 5 \rightarrow 4$	576.23	519.8	$(3.1 \pm 0.9) \times 10^{-8}$
C I	$2p^2$: $^3P_1 \rightarrow {}^3P_0$	492.16	609.1	$(6.4 \pm 1.3) \times 10^{-8}$
C I	$2p^2$: $^3P_2 \rightarrow {}^3P_1$	809.35	370.4	$(1.16 \pm 0.16) \times 10^{-7}$
C II	$2p$: $^2P_{3/2} \rightarrow {}^2P_{1/2}$	1900.7	157.7	$(1.73 \pm 0.07) \times 10^{-5}$
N II	$2p^2$: $^3P_1 \rightarrow {}^3P_0$	1460.3	205.3	$(1.66 \pm 0.09) \times 10^{-6}$
N II	$2p^2$: $^3P_2 \rightarrow {}^3P_1$	2459.3	121.9	$(2.61 \pm 0.36) \times 10^{-6}$

First, it is remarkable that the intensity ratios $I(\mathrm{C\,I}\ (1 \rightarrow 0))/I(\mathrm{CO}(2 \rightarrow 1))$ between IC 342 and the Milky Way do not differ by more than a factor of about two, despite the fact that the ratio for IC 342 is from the nucleus of the galaxy and the ratio from the COBE data is from an average over the entire Milky Way. Second, the ratio is actually somewhat smaller for IC 342 than for the Milky Way, unless both ratios are taken at the limits of their error bars. The uncertainties

of the ratios have been derived from the absolute uncertainties of the individual measurements, including calibration. This probably overestimates the uncertainty of the ratios since the same calibration technique was used in each of the two measurements forming each ratio. Therefore, the difference in the ratios is believed to be real.

Despite the requirement of UV–flux to produce C I in PDRs, recent calculations by Hollenbach, Takahashi and Tielens (1991) show that the ratio of C I to ^{12}CO ($1 \rightarrow 0$) emission is relatively independent of ultraviolet flux, but is more strongly dependent on density. In IC 342 this ratio is about 12, or equivalently 1.5 for the C I to ^{12}CO($2 \rightarrow 1$) ratio. (The factor scales with ν^3 for optically thick CO lines, as verified by observation [Eckart 1990; Wall and Jaffe 1990].) This requires a model density of $(5.2 \pm 1.4) \times 10^3$ cm^{-3}. The relative intensities of the C I, CO, C II (Crawford *et al.* 1985; Stacey *et al.* 1991), and O I lines (Eckart *et al.* 1990) and the FIR continuum (Becklin *et al.* 1980; Rickard and Harvey 1984), are consistent with this density and an average incident far-ultraviolet radiation field flux $F_{UV} \approx 10^3 F_o$.

To find a model density for the Milky Way based on the C I ($1 \rightarrow 0$) to CO($1 \rightarrow 0$) ratio the models of Hollenbach, Takahashi, and Tielens (1991) are used. It is assumed that the CO($1 \rightarrow 0$) to CO($2 \rightarrow 1$) intensity ratio is 1/8, as in IC 342. The assumption for the CO lines is necessary because COBE was unable to measure the CO ($1 \rightarrow 0$) line. The resulting density estimate of $(3.5 \pm 0.8) \times 10^3$ cm^{-3} is relatively insensitive to the UV field strength. With $F_{UV} = 10^3 F_o$ the other line ratios, C I, C II, and O I, are also well matched. This estimate is consistent with that of Wright *et al.* (1991) who used the high flux UV field PDR models ($F_{UV} = 10^6 F_o$) of Tielens and Hollenbach (1985) and the COBE line ratios for C I, C II, and O I, to derive a density of 2.5×10^3 cm^{-3}.

The uncertainties in the densities above were derived from the uncertainties in the line ratios, excluding any possible contributions from the models. Any model uncertainties could easily dominate, thus the absolute densities may be less certain than indicated. The difference of derived densities may largely be due to the dissimilar nature of the regions observed in the two cases. The COBE spectrum is an average over the entire Milky Way, biased particularly towards our local Galactic neighborhood, while, in contrast, the C I data presented here come from the central 130 pc region of IC 342.

In summary, C I emission from IC 342 has been observed, and found to be comparable in intensity to $^{12}CO(2 \rightarrow 1)$. This result is quite similar to the Milky Way as measured by COBE. On a large scale it can thus be concluded that there is, indeed, a strong correlation of C I and CO emission. Overall, there is good agreement between PDR models and observation. It can not be said, however, if types of regions other than PDRs might be contributing to the C I and CO emission. Although the C I data do not constrain the average UV flux, PDR models provide a constraint for the average density of the ISM that is responsible for the bulk of the observed C I and CO emission. However, as discussed later, the PDR models used here assume no depletion of heavy elements (C,N,O,S,...) in the gas phase of the ISM, which is found to be unrealistic. Section 4.6 of this thesis will give an explanation of the high C to CO and C to H abundance ratios observed, based on a revised ion–chemistry model of PDR's by Pineau des Forêts, Roueff and Flower (1992) that do not avoid the depletion of heavy elements.

> Looking at Orion is like looking at a bowl of porridge and trying to figure out why there is a carrot over here and a bean over there.
>
> – Raoul Taco Machilvich

4.4 The Small–Scale Structure of Neutral Carbon in Orion

To investigate the distribution of neutral carbon within molecular clouds high spatial resolution observations are required. As mentioned in section 4.2, most of the previous data (*e.g.*, Phillips and Huggins 1981; Keene *et al.* 1985) were taken from the Kuiper Airborne Observatory, which in addition to the bandwidth limitations due to the InSb detector used for those observations, only has a 0.9 meter aperture telescope, yielding an angular resolution of 2.8′. This resolution is too poor to see detailed structure in maps of molecular clouds. For mapping small structures only ground-based telescopes have sufficiently large apertures at present. The Caltech Submillimeter Observatory (CSO) on Mauna Kea with a 10.4 m aperture provides a 15″ beam at 492 GHz. Orion was chosen as a source since its relative closeness of about 450 pc yields a spatial resolution of 0.033 pc and many maps in various transitions of other atomic and molecular species are available in the literature. However, despite the fact that ^{12}CO and ^{13}CO molecules are generally the most favored tracers of molecular gas, only rather old and uncertain or low resolution data were available in the literature. Thus, maps of those two isotopes in the $J = 2 \rightarrow 1$ rotational transition were taken at the CSO.

a) Introduction to the Orion Molecular Cloud

The Orion molecular cloud complex is one of the best studied molecular clouds. It shows recent and probably ongoing OB star formation and extends over approximately 20°. A recent review of the Orion molecular cloud complex is given by Genzel and Stutzki (1989). The molecular cloud complex was mapped by Tucker, Kutner and Thaddeus (1973) in the ^{12}CO($1 \rightarrow 0$) transition. Molecular line emission from the two main components, Orion A and Orion B, correlates well

with infrared data from the Infrared Astronomical Satellite (IRAS), both of which show a ridge with a sharp edge to the west (Beichman 1988, Robinson 1984), as seen for Orion A in figure 4.6 . Within the molecular ridge, higher density condensations ($n(H_2) > 10^4 cm^{-3}$) are visible in molecular tracers such as OH and H_2CO (Cohen *et al.* 1983). The dense molecular clumps are locations of recent and ongoing star formation and also show up as prominent features in the IRAS data. Figure 4.6 shows a $^{12}CO(1 \rightarrow 0)$ intensity contour map of the Orion A molecular cloud with an angular resolution of 8.7′ (Maddalena *et al.* 1986). Orion A covers about a 3° by 10° area, corresponding to about 24 by 80 pc^2, and contains approximately $10^5 M_\odot$ (Maddalena *et al.* 1986).

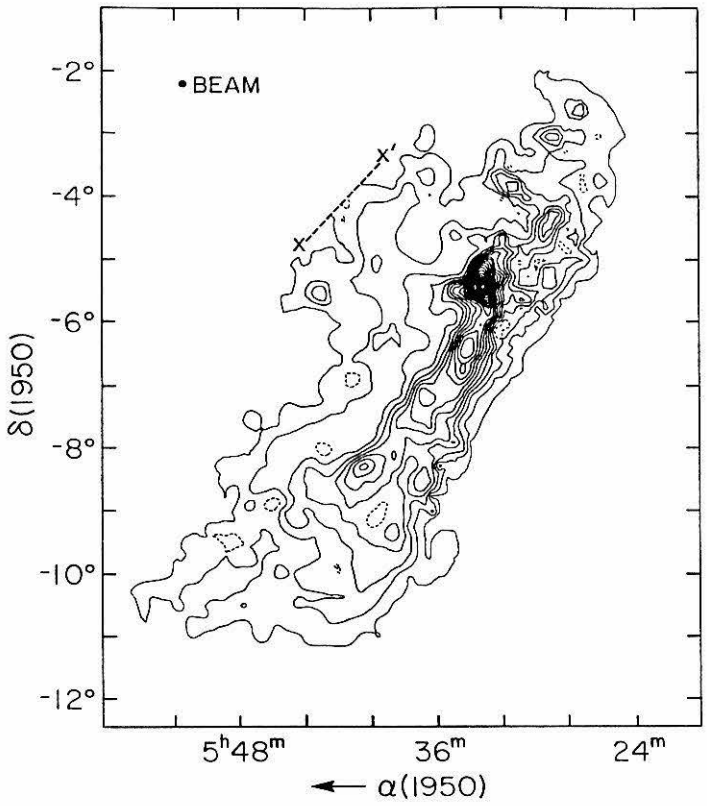


Figure 4.6 $CO(1 \rightarrow 0)$ intensity map of the Orion A cloud (from Maddalena *et al.* 1986). The dashed line connecting X and X' is where the Orion A cloud connects to Orion B.

Bally *et al.* (1987) have mapped the Orion A molecular ridge (40 by 2 pc^2) in the ^{13}CO $(1 \rightarrow 0)$ transition and found over 100 condensations ($\approx$ 1 pc lengthscale) in the ridge. Higher

resolution studies of NH_3 (Ziurys *et al.* 1981; Batrla *et al.* 1983) showed these clumps to be about 0.1pc or less in size. Using interferometric techniques, the cores of these clumps were resolved to be typically 0.05 pc in size (Harris *et al.* 1983; Mundy *et al.* 1986; Wilson and Johnston 1989).

The most prominent and best studied area of condensations within the molecular ridge of Orion A is a region referred to as OMC–1, visible as the strongest peak in figure 4.6. It is centered on the infrared source IRc2 ($\alpha(1950) = 5^h\ 32^m\ 47^s$, $\delta(1950) = -5°\ 24'\ 24''$) and the Becklin–Neugebauer (BN) object (Becklin and Neugebauer 1967), about (–5″, 7″) relative to IRc2. Several prominent features are related to the OMC–1 region, two of which will be mentioned here. The "hot core," with kinetic gas temperatures in excess of 200 K and densities of a few times $10^7\,cm^{-3}$ (Goldsmith *et al.* 1983) located 3″ south of IRc2 and about 5–10″ in size, and the "plateau," a high velocity bipolar flow (P.A. = – 45°) of molecular gas observed in CO (about 30″ FWHP, *e.g.*, Wilson, Serabyn and Henkel 1986) with velocities as high as 100 $km\,s^{-1}$.

About 90″ south of OMC–1 is another condensation found by Keene, Hildebrand and Whitcomb (1982) denoted OMC–1 south, or the "southern source." The C I data taken for this thesis covers both condensations, OMC–1 and the southern source.

b) Observations

The receiver used an SIS mixer quasi–optically coupled to the telescope with a logarithmic spiral antenna and is similar to the one described in chapter I and Büttgenbach *et al.* (1988) with the quasi–optics modified as described in chapter II and Büttgenbach (1993). The coupling efficiency between the quasi–optical receiver and the telescope optics was compared with a waveguide receiver (Ellison *et al.* 1989) in the same position at the telescope. Both receivers showed nearly identical performance at 345 GHz. At the time of these measurements no waveguide receiver existed at 492 GHz for comparison.

The beam efficiency at 492 GHz on the Moon was found to be $\eta_B = 0.68 \pm 0.03$ assuming a Moon temperature of 395 K and an emissivity of 0.97 (Linsky 1973). By mapping Mars the beam size was measured to be approximately 15″. The main beam efficiency (η_{MB}) was 0.35 ± 0.05 measured on Mars (14.″2 diameter) or 0.39 ± 0.02 measured on Jupiter (33.″9 mean diameter) assuming brightness temperatures of 225 and 163 K respectively (Neugebauer *et al.* 1971; Ulich

1981; Hildebrand *et al.* 1985). The higher uncertainty for the efficiency on Mars is due to its smaller size making the result more sensitive to uncertainty in the measured beam diameter. The double sideband receiver noise temperature was 1700 K and 1100 K for the 1990 October and 1991 January observing runs, respectively. The systems poorer performance compared to the numbers quoted in chapter I was due to a lack of high quality SIS diodes at that time. Due to bad weather in January 1991, most of the data were taken in October of 1990. The IF bandwidth was 500 MHz allowing observations of several lines simultaneously.

Due to the high atmospheric opacity at 492 GHz the initial pointing of the telescope can be difficult. Planets could only be detected in very good weather conditions. However, since the quasi–optical receiver has a frequency operating range from 100 to 700 GHz without any changes that would affect the pointing, it was possible to point with high accuracy on the 345 GHz CO(3 $\rightarrow$ 2) line in CRL 2688. Throughout the observing period, weather allowing, the pointing at 492 GHz was verified on Mars and Jupiter and found to be better than 5″. While mapping Orion, the IRc2 position ($\alpha(1950) = 5^h\ 32^m\ 47^s$, $\delta(1950) = -5°\ 24'\ 24''$) was frequently re–observed. Using the methanol $4_1 \rightarrow 3_0\ A^+$ line at 492.2787 GHz and the SO_2 ($12_{3,9} \rightarrow 11_{2,10}$) line in the upper sideband at 494.7797 GHz (see figure 4.7) the pointing could be verified in any weather conditions allowing observations. The atmospheric transmission in the two sidebands is different by only about 3% and the receiver is inherently double sideband so the response in both sidebands is equal.

With a grid spacing of 20″ the map was undersampled. The integration time was 20 seconds per point on source. However, the 1′.7 EW ×2′.3 NS region centered 20″ south of IRc2 was mapped twice and thus has higher S/N than the rest of the map, with typical RMS noise of 1.5 K per MHz.

c) Data reduction

It was not immediately obvious what coupling efficiency should be assumed to calibrate the C I data. The C I emission is very widely spread and in fact does not disappear at the edges of the map. Clearly, using a main beam efficiency as measured on a planet like Jupiter will overestimate the line intensities. The Moon is perhaps a more appropriate calibration object due to the extended nature of the C I emission.

It was thus decided to calculate the source coupling efficiency by estimating the shape of the 492 GHz beam from a high SNR beam map of the CSO generated from holographic telescope surface measurements at 300 GHz (Serabyn, Phillips and Masson 1991). The measured beam at 300 GHz was Fourier transformed to yield the surface error of the telescope, and then back transformed to the far field at 492 GHz (Serabyn, private communication, 1992). The resulting beam map is nearly 4′ by 4′ in size, approximately the same size as the Orion map. As a first approximation, the integrated beam map was normalized to the measured moon efficiency by assuming that there is no contribution to the beam power from the area of the Moon not covered by the calculated beam, *i.e.*, between 2′ and 15′ radius. This assumption was checked by comparing the measurement of Jupiter with the coupling efficiency this method predicts. The measured coupling efficiency for Jupiter was $\eta_m(Jupiter) \equiv T_A^*(Jupiter)/T(Jupiter) = 0.38$ (note that this is *not* the same as the main beam efficiency discussed above) whereas the calculation predicts that $\eta_c(Jupiter) = 0.41$. Doing the same for Mars the measured ratio $\eta_m(Mars) \equiv T_A^*(Mars)/T(Mars) = 0.16$ while the predicted value is $\eta_c(Mars) = 0.20$. For both planets, the agreement between the measured and predicted values is very good, within the errors of the measurements, justifying the calculations and assumptions. The discrepancies are likely due to the amount of radiation falling between 2′ and 15′ radius on the Moon. The bigger discrepancy for the Mars comparison is due to two additional effects: pointing, focusing and collimation errors taken out in the calculated beam, but presumably affecting the measurements, especially of small objects, plus any differences in the illumination patterns of the bolometer used to measure the 300 GHz beam pattern and the 492 GHz receiver with its quasi–optically coupled spiral antenna. No attempt was made to correct for any of these effects, because the large extent of the C I emission made them irrelevant.

The observed source map was used to derive a source coupling efficiency of η_c(C I) = 0.65 for a single spectral resolution element at the peak of the C I line at IRc2. For the other lines in the spectrum (CH_3OH, SO_2 and H_2CO) the source coupling efficiency was calculated to be $\eta_c(CH_3OH) = 0.39$, *i.e.*, about the same as the point source coupling efficiency as measured on Jupiter, using the Methanol integrated line map (figure 4.8b) as a source model. All spectra and maps have been calibrated with the standard hot–sky chopper method and divided by the

appropriate source coupling efficiency.

The 500 MHz bandwidth AOS that was used in the 1990 October observing run suffered from gain compression resulting in measured line temperatures that were too high by a then unknown factor. The calibration was corrected afterwards by observing the IRc2 position again with a 50 MHz bandwidth AOS that has been measured to be linear in power response. All data presented here have this calibration correction included.

The maps of ^{12}CO and ^{13}CO in the $J = 2 \rightarrow 1$ rotational transition were taken at the CSO with a resolution of 30″. To obtain a resolution comparable to that of the C I map these very high SNR maps were deconvolved to 15″ using an error beam map of the CSO measured at 230 GHz by Serabyn, Phillips and Masson (1991).

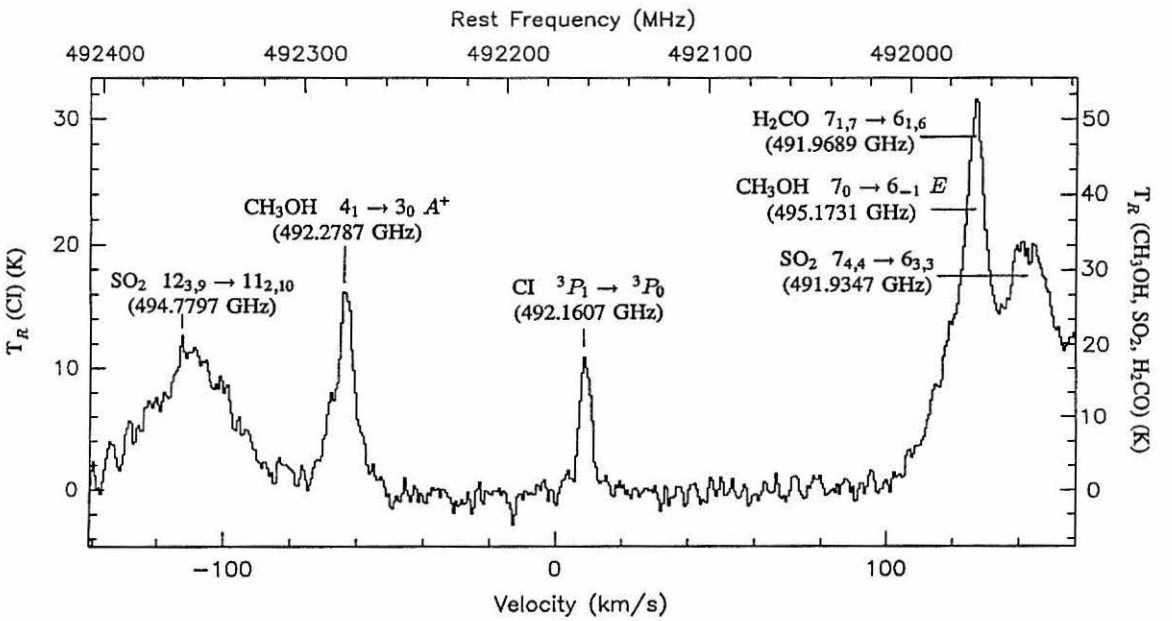


Figure 4.7 492.1607 GHz (LSB) spectrum of IRc2 ($\alpha(1950) = 5^h\ 32^m\ 47^s$, $\delta(1950) = -5°\ 24'\ 24''$) in Orion. Response from the upper sideband (USB) is also included with a center frequency of about 2.8 GHz higher than the LSB. The total integration time was 6 minutes. The C I line is the weakest line detectable in this spectrum but is fortuitously located in the center of an apparently blank portion of an otherwise crowded spectrum. Because the source coupling coefficients are so different for the widespread C I line as compared to the nearly unresolved methanol, formaldehyde and SO_2 lines, the spectrum is labeled with both temperature scales.

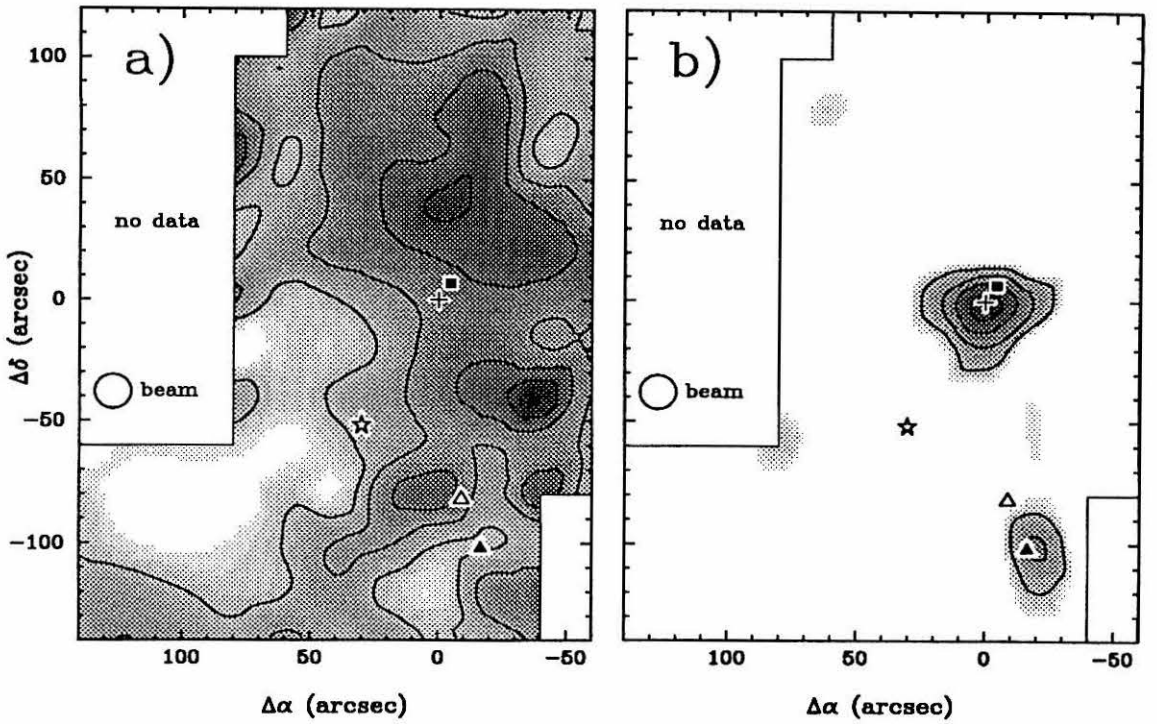


Figure 4.8 *a)* C I Integrated line intensity map centered at IRc2 (plus–symbol; $\alpha(1950) = 5^h\ 32^m\ 47^s$, $\delta(1950) = -5°\ 24'\ 24''$). Other symbols are : BN (square), the southern source of Keene, Hildebrand and Whitcomb (1982) which was resolved into two individual sources, FIR 3 (open triangle) and FIR 4 (filled triangle), by Mezger, Wink and Zylka (1990) and the main source of ionization, $\theta^1 C$ (star). The minimum contour is 25 $K\ km\ s^{-1}$ with intervals of 15 $K\ km\ s^{-1}$. The first grey–scale level and approximately 1σ level is about 8 $K\ km\ s^{-1}$.
b) The methanol $4_1 \rightarrow 3_0$ A^+ transition at 492.2787 GHz integrated line intensity map centered at IRc2. The minimum contour is 60 $K\ km\ s^{-1}$ with intervals of 40 $K\ km\ s^{-1}$. The 1σ level is about 20 $K\ km\ s^{-1}$ and the first grey level starts at 2σ.

d) Results

The spectrum of Orion IRc2, centered at 492.1607 GHz for the lower sideband (LSB), is shown in figure 4.7. Spectral lines from the upper sideband are marked with their respective frequencies. The C I line is the weakest line detectable in the spectrum but is fortuitously located in the center of an apparently blank portion of an otherwise crowded spectrum. The difference between C I and the molecular lines is apparent from a comparison between the map of the C I integrated line intensity shown in figure 4.8a and the methanol integrated line intensity in figure 4.8b. Whereas the C I line is widespread and shows no peak at the IRc2 position, the methanol line (and all the

other lines in the spectrum) is strongly peaked at the IRc2 position (to within ±5″) and is nearly unresolved in this map. Because the source coupling coefficients are so different for the C I and methanol lines figure 4.7 was labeled with both temperature scales.

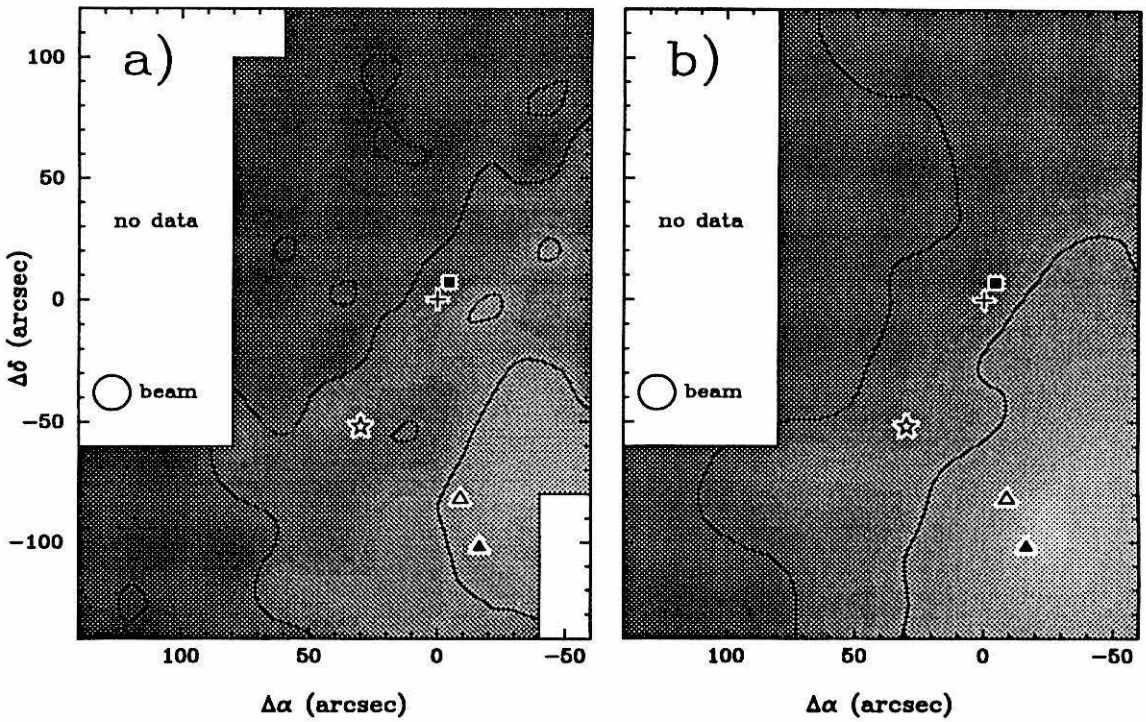


Figure 4.9 *a)* C I line velocity map. The minimum line velocity contour of 8.5 $km\,s^{-1}$ in the south west increasing at 1 $km\,s^{-1}$ intervals to the north east. *b)* ^{13}CO ($2 \rightarrow 1$) line velocity map with the same contour and grey–scale levels as for the C I map.

The peak integrated line intensity of 79 $K\,km\,s^{-1}$ in the C I intensity map (figure 4.8a) is located at (– 40″,– 40″) while at the IRc2 position C I has an integrated line intensity of only 45 $K\,km\,s^{-1}$. A line of intensity peaks that form a circle around the ionizing stars of the Trapezium cluster ($\theta^1 C$ is marked on the maps) is visible with the peaks at (– 40″,– 80″), (– 40″,– 40″), (0″, 40″) and (80″, 60″). This feature will become more clear when the data is compared to C II data below.

The methanol peak of figure 4.8b is located at IRc2 (± 5″) with an integrated line intensity of 192 $K\,km\,s^{-1}$. The peak at (– 20″,– 100″), close by FIR 4 (filled triangle), has an integrated line intensity of 105 $K\,km\,s^{-1}$. Formaldehyde emission with integrated line intensity approximately

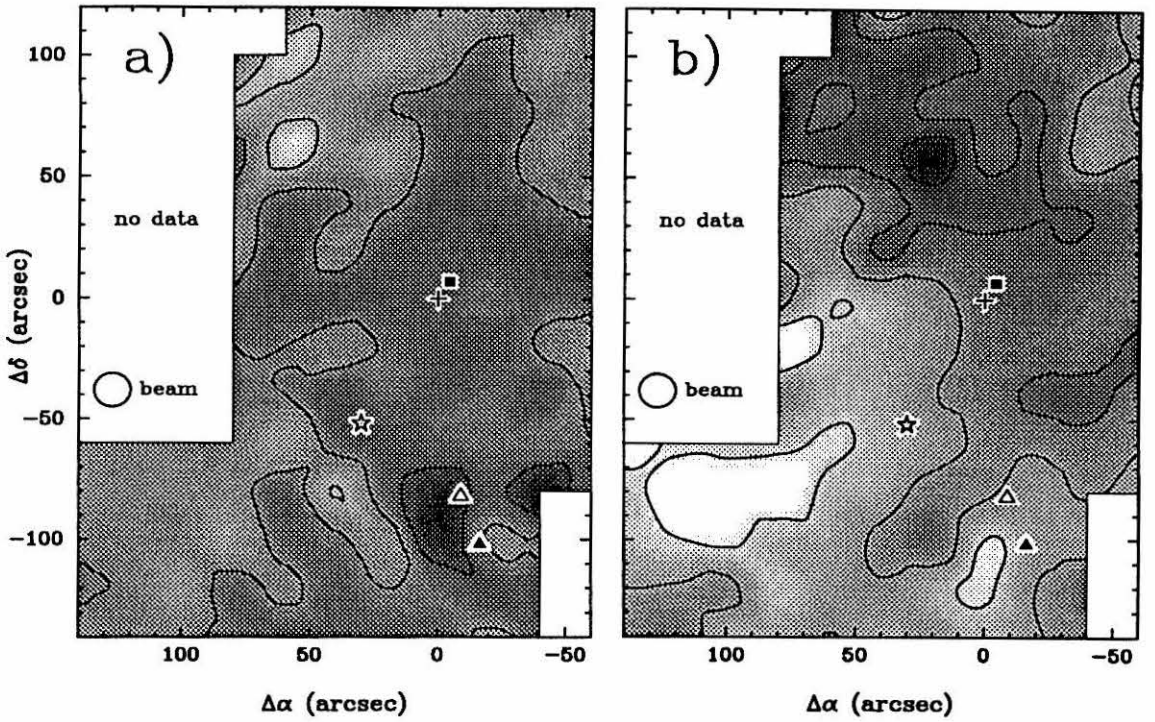


Figure 4.10 *a)* C I line dispersion. The minimum contour is 2 $km\,s^{-1}$ with intervals of 1.5 $km\,s^{-1}$. *b)* C I peak line temperature map centered at IRc2. The minimum contour is 4 K with intervals of 5 K. The peak temperature of 22 K in this map is located at (20″,60″) while the temperature at the IRc2 position is only 11 K.

105 Kkm s^{-1} was also found at the (–20″,–100″) position. SO_2 was possibly detected there as well. This position corresponds to the southern source of OMC-1 as found in dust emission at 400 μm by Keene, Hildebrand and Whitcomb (1982) and at 1.3 mm by Mezger, Wink and Zylka (1990) who denoted its position as FIR 4 (filled triangle in the maps of this work).

An examination of the C I LSR line velocities show that they vary smoothly from 7.5 km s^{-1} in the SW corner of the map to 10.2 km s^{-1} in the NE corner (see figure 4.9a). For comparison a ^{13}CO (2 → 1) LSR line velocity map is shown in figure 4.9b with the same contour and grey–scale levels as for the C I LSR line velocity map of figure 4.9a. The very similar LSR line velocity structure suggest that the same volume of gas is observed for both species. The C I FWHP line widths are typically 3.5 to 4.5 km s^{-1} in the middle and SW portion of the map narrowing to 2 km s^{-1} in the NE (see figure 4.10a). The line dispersion of ^{13}CO (2 → 1) follows that of C I with the exception of the IRc2 position (0″, 0″). At IRc2 the ^{13}CO line can be fitted with a narrow

and a wide component, 4km s^{-1} and 20km s^{-1}wide, respectively. The narrow component has the same width as the C I line. Even though the signal to noise is insufficient for a definite detection, the C I line as shown in figure 4.7 may also show the wide line component (the "plateau") with a FWHP of approximately 20km s^{-1}.

Throughout most of the Orion ridge, as observed in CS (*e.g.*, CS(2 $\rightarrow$ 1) Mundy *et al.* 1986), the C I line dispersion is constant at about 4km s^{-1}. The line peak temperatures are also fairly constant in that region with values varying from 9 to 13 K as shown in figure 4.10b. In the NE, where the line dispersion reduces, the C I peak line temperatures increase resulting in approximately constant C I intensities throughout the ridge. There is only one significant peak in the line temperature at (20″, 60″) where the line is 22 K. However, the integrated line intensity does not peak there since the line width in that position is narrower than usual at about 2.5 kms^{-1}.

There are two peaks in the C I intensity and velocity dispersion (figures 4.8a and 4.10a) located at the southern source at (0″,–80″) and (–40″, –80″). However, they lie to the east and west of FIR 3 or to the NE and NW of the methanol peak (FIR 4) by approximately one map pixel or 20″. They have velocity dispersions of 6 km s^{-1}, the highest in the map, while the ^{13}CO (2 $\rightarrow$ 1) lines do not show any significant increase in width at those positions. The C I dispersion peak at (0″, –80″) possibly corresponds to the blue lobe of the outflow source from FIR 4 (P.A.$\approx$ –30°) found by Schmid–Burgk *et al.* (1990) in ^{12}CO. The other C I dispersion peak (–40″, –80″) may correspond to a NW extension of the generally SW oriented red lobe of the same outflow. However, higher signal to noise ratios for the carbon data would be necessary to confirm this correlation. At the position of the methanol peak (FIR 4) the C I line is relatively weak with no increase in dispersion.

e) C I column densities

To calculate the C I column densities the opacity of the C I line needs to be estimated. The relatively low observed peak temperatures of the C I line (typically $\hat{T}_R^* = 10$ to 15 K or $T_R^* \geq 20$ to 25 K using equation (4.11b) of the appendix) can have several explanations: (1) it could be optically thin (2) it could have a beam filling factor of significantly less than unity, (3) the excitation temperature is actually as low as the observed Planck corrected line temperature or (4) it could be subthermally excited. Since the critical density of the lower C I line is very low

($n_{crit} \approx 10^3 \mathrm{cm}^{-3}$) the $^3P_1 \rightarrow {}^3P_0$ C I transition should be thermalized. If the PDR models are correct then the C I emission should only emanate from the surface of molecular clouds or of the clumps they are made up from, *i.e.*, outside the region where ^{12}CO becomes optically thick or roughly from the same region (see figure 4.2). The beam filling factor thus should be greater or equal to that of ^{12}CO, *i.e.*, about unity. Also, the excitation temperature should be similar to that of ^{12}CO. The remaining choice is an opacity of less than or about unity. In this case, the opacity correction term of equation (4.1), $\tau_o/1 - e^{-\tau_o}$, can thus be ignored. Observations of the J =2 $\rightarrow$ 1 transition of C I at 809 GHz by Zmuidzinas *et al.* (1988) lead to the same conclusion.

A choice of 90 K for the excitation temperature (later shown to be a typical value for ^{12}CO in the vicinity of IRc2) yields for the total C I column density

$$N_{\mathrm{CI}}^T = 1.4 \times 10^{16} \int \hat{T}_R^* \, dV \quad . \tag{4.2}$$

For the C I intensity peak position at (– 40″, – 40″), $N_{\mathrm{CI}}^T \geq 1.1 \times 10^{18}$, for an average position, *e.g.*, (0″, – 40″), $N_{\mathrm{CI}}^T \geq 7.1 \times 10^{17}$, which is the same as for the IRc2 position. The values quoted here for N_{CI}^T are lower limits because of the low opacity assumption. If assumptions (2), (3) or (4) are partially applicable the value for N_{CI}^T could increase.

4.5 Comparison of Neutral Carbon with other species in the ISM

The IRc2 position is featureless in the C I maps presented here, while at its position most other species show extremely strong features. Due to IRc2's complex and non–representative structure – an internally infrared heated source with external UV illumination – its position will not be used for derivation of typical parameters of giant molecular clouds (GMCs). A position 40″ south of IRc2 is chosen instead as a "typical" reference position.

a) CO and the C I to CO abundance ratio

To derive an abundance ratio for C I to CO the total CO column density N_{CO}^T is required in addition to the above derived neutral carbon column density, N_{CI}^T. Since both the ^{13}CO and ^{12}CO low J rotational transitions are optically thick in Orion (Castets *et al.* 1990), the C^{18}O isotopic emission is required. Wilson *et al.* (1986) mapped an area covering a part of the Orion A ridge in

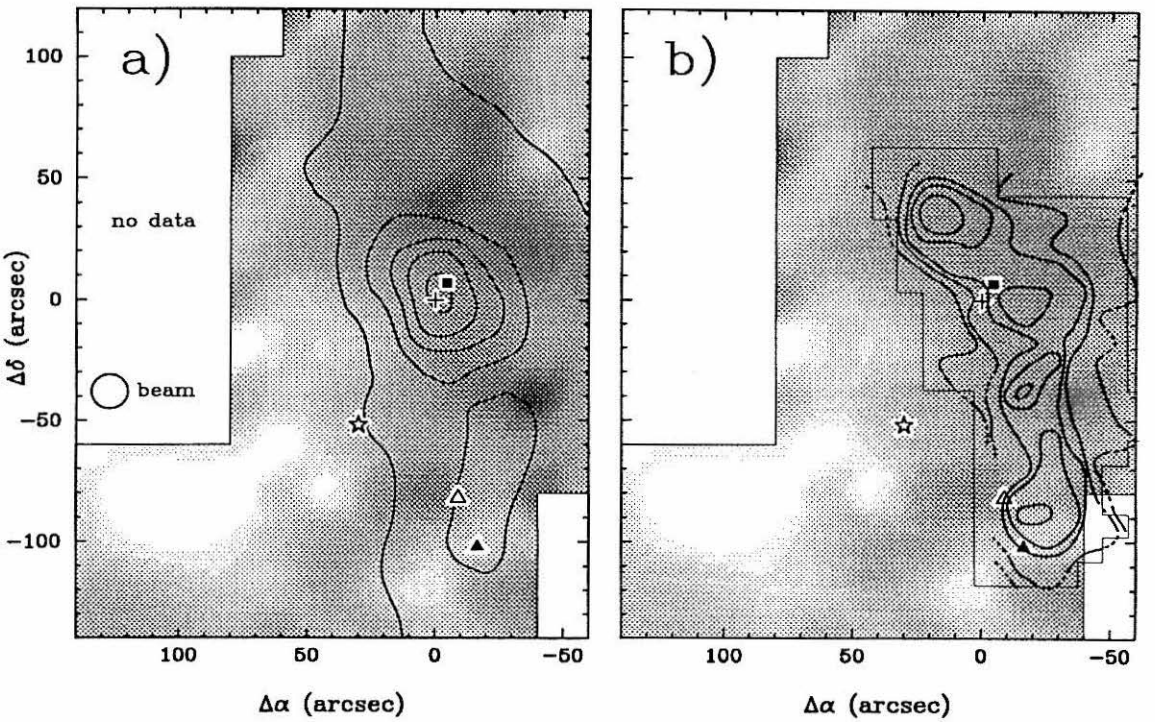


Figure 4.11 *a)* ^{13}CO (2 → 1) intensity map contours taken at the CSO with a beam of 30″ and then deconvolved to 15″ superposed on a C I intensity grey–scale map as in figure 4.8a. The minimum contour of the ^{13}CO map is 90 K km s^{-1} with steps of 90 K km s^{-1}. The ^{13}CO (2 → 1) peak intensity is at IRc2 with 470 K km s^{-1}. Except for a different intensity scale, a ^{12}CO(2 → 1) map taken at the CSO looked very similar to the ^{13}CO (2 → 1) map displayed here. *b)* C^{18}O (1 → 0) intensity contour map by Wilson *et al.* (1986) superimposed on a C I intensity grey–scale map. The numbers labeling the contours correspond to the intensity of the contours in K km s^{-1}, starting at 4K km s^{-1}and increasing in 2K km s^{-1}steps to 12K km s^{-1}. Note that there is no C^{18}O peak at the position of IRc2 (0″, 0″).

the vicinity of OMC–1 in the C^{18}O(1 → 0) transition. Their intensity contour map is superposed on the C I intensity grey–scale map in figure 4.11b. Peak line temperatures are typically 3 to 6 K ensuring that the line is optically thin since the beam filling factor should be about the same as that for the other CO isotope transitions. The CO column density can then be derived similarly to that for neutral carbon and a derivation can be found in the appendix. Unlike for the column density derivation of C I, the column density of CO depends roughly linearly on the excitation temperature, T_{ex}, through the partition function. This makes the correct choice of T_{ex} much more critical for the determination of CO column densities.

Wilson *et al.* (1986) used a kinetic gas temperature of 100 K as found in CO observations in a 35″ beam by Schloerb and Loren (1982) and derived a CO column density (assuming LTE) towards BN of $N_{CO}^{T} = 1.6 \times 10^{19}\mathrm{cm}^{-2}$ using the terrestrial ratio of $^{16}O/^{18}O = 490$. From the intensity map of Wilson *et al.* (1986) a CO column density of $N_{CO}^{T} = 1.3 \times 10^{19}\mathrm{cm}^{-2}$ would follow for a typical position (0″, – 40″). This yields a typical C to CO abundance ratio of 0.06. At the edge of the $C^{18}O$ map the intensity drops to about half of that found towards BN yielding $N_{CO}^{T} = 8 \times 10^{18}\mathrm{cm}^{-2}$ resulting in a C to CO ratio of about 0.1. However, all CO column density values quoted from Wilson *et al.* are believed to be an overestimation for reasons discussed now.

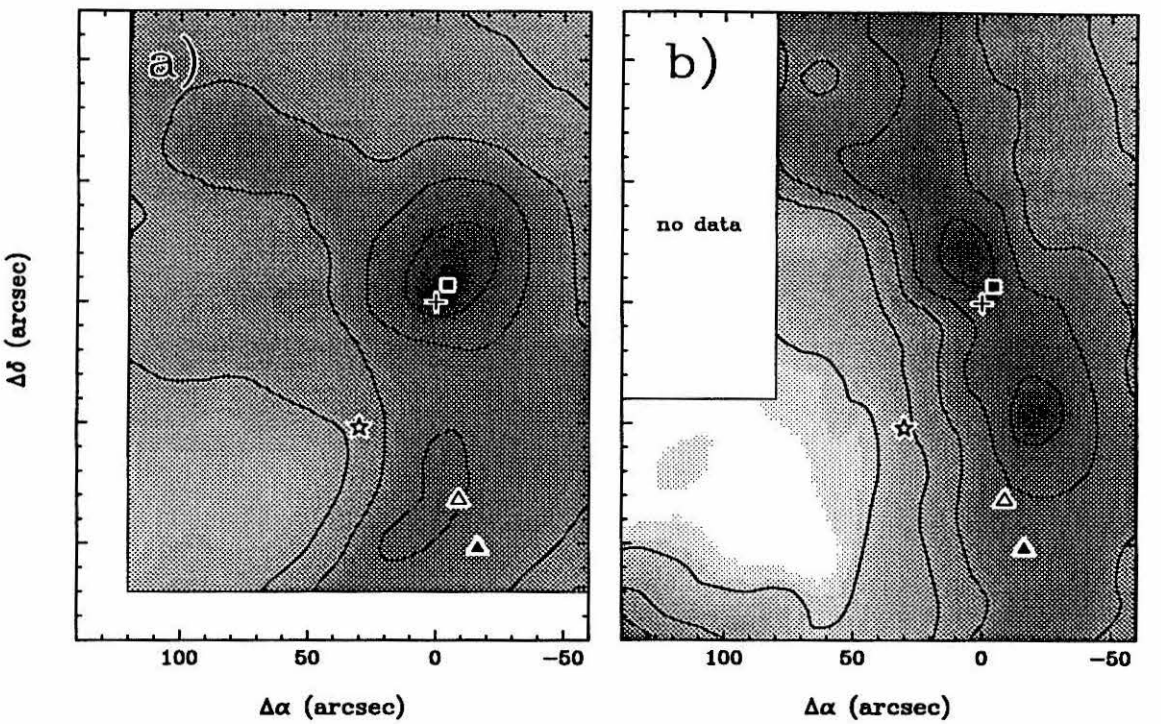


Figure 4.12 $^{12}CO(2 \rightarrow 1)$ *[b)* ^{13}CO $(2 \rightarrow 1)$] peak line temperature map taken at the CSO with a beam of 30″ deconvolved to 15″. The minimum contour is 60 K [10 K] with steps of 30 K [10 K]. The peak temperature is 165 K [55 K].

To find the excitation temperature for the $C^{18}O$ line, the $^{12}CO(2 \rightarrow 1)$ and ^{13}CO $(2 \rightarrow 1)$ maps of peak line temperatures, shown in figures 4.12a,b, are now investigated. While a typical Rayleigh–Jeans temperature for the ^{12}CO transition is indeed about 100 K, with a peak value of 165 K at IRc2, the ^{13}CO peak line temperature only reaches a maximum of 55 K and a more typical value of 40 K. The ^{12}CO line is thus about 2–3 times hotter than the ^{13}CO, which could

indicate the ^{13}CO line is optically thin. This is not supported by observations of the $J = 1 \rightarrow 0$ transitions, which clearly show that the ^{13}CO line is optically thick, too (Castets *et al.* 1990). An explanation to this discrepancy was first offered by Young and Scoville (1984), who studied M82 and suggested a decrease of kinetic gas temperature from the outside to the inside of a molecular cloud to obtain a physically reasonable model. A multi–transition study of CO in Orion A by Castets *et al.* (1990) also required such a temperature gradient. With the exception of the very luminous internally heated regions around IRc2 and possibly the southern source (FIR 4), most clumps of molecular gas are thought to be mainly heated externally by NUV radiation from the ionizing stars of the Trapezium cluster (most notably $\theta^1 C$) in OMC–1. ^{12}CO should be optically thick on the surface of those clumps while ^{13}CO requires greater depth into the cloud before reaching $\tau \approx 1$. Gierens, Stutzki and Winnewisser (1992) modeled the intensities and line ratios of low J CO transitions in Orion, as observed by Castets *et al.*, by implementing a temperature and density gradient in a radiative transfer code. They found that a different beam filling factor due to the smaller area where ^{13}CO is optically thick did not significantly contribute to the different observed peak temperatures, but that the main effect is indeed from colder gas in the core of clumps.

If ^{13}CO lines do indeed emanate mainly from regions of colder gas, as compared to the surface of clouds where ^{12}CO lines are optically thick, then there is no reason to use the ^{12}CO peak line temperature for the kinetic gas temperature of the C^{18}O lines. Instead, the ^{13}CO line peak temperature is more appropriate as an upper limit. This may not be true for those clumps that are heated internally by strong IR sources like the clump at the IRc2 position. However, these positions are not representative for a large cloud like OMC–1. Using an excitation temperature of 40 K, a typical column density of $N_{CO}^{T} = 5.8 \times 10^{18}\mathrm{cm}^{-2}$ is obtained at a position of $(0'', -40'')$ and thus a C to CO abundance ratio of 0.12. For the position of the carbon intensity peak $(-40'', -40'')$ the C to CO abundance ratio reaches 0.33 (using the scaled CO column density for that position), whereas for the areas of CO column density peaks the abundance ratio decreases to 0.05.

b) Comparison of C I with dust.

Keene, Hildebrand and Whitcomb (1982; hereafter KHW) made a 400μm continuum dust map,

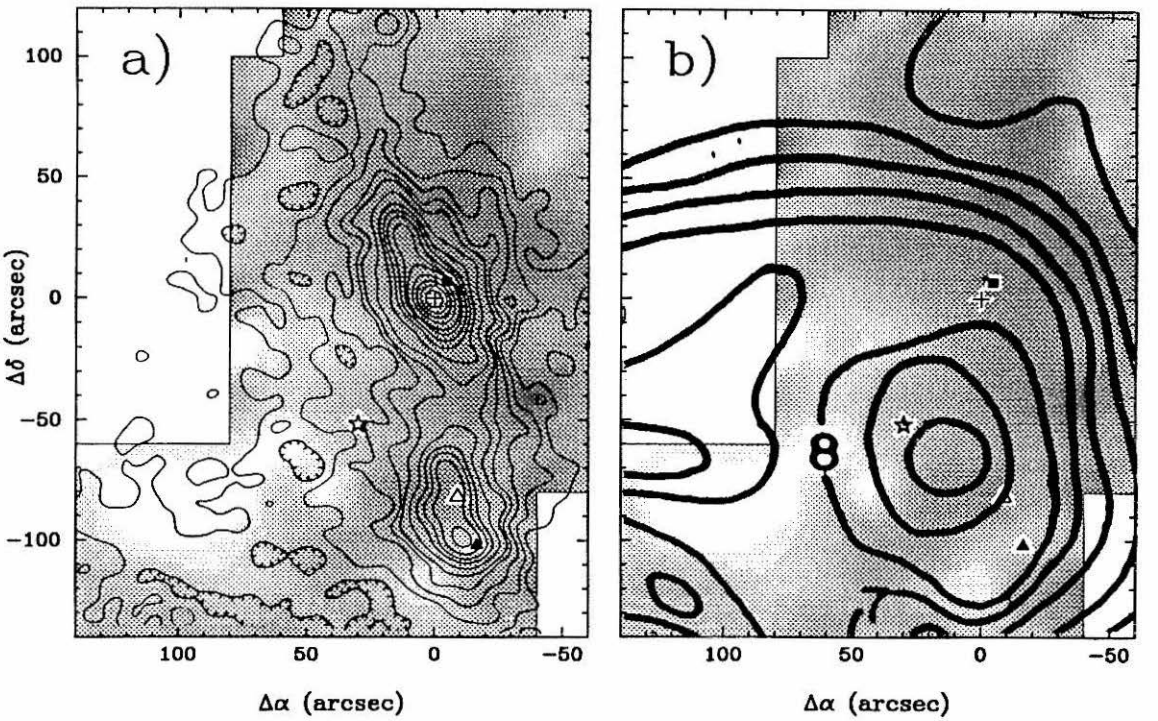


Figure 4.13 *a)* 1.3 mm continuum observation (contours) by Mezger, Wink and Zylka (1990) with 11″ resolution overlayed on the C I grey–scale map. *b)* C II intensity map (contours) by Stacey *et al.* (1993) with a 55″ beam overlayed on the C I grey–scale map. The two contours, labeled with a "7" and "8," correspond to 70% and 80% of the peak C II integrated line intensity, respectively.

of 3′ by 5′ size, of the central region of OMC–1, with 35″ resolution. They found two distinct condensations, one at the position of IRc2 and a new one about 90″ south of IRc2, which they called the *southern source* (OMC–1 south). A higher resolution continuum map, taken at 1.3 mm with a resolution of 11″ (Mezger, Wink and Zylka 1990; hereafter MWZ), resolved each condensation seen by KHW into two separate ones giving a total of four sources, named, from north to south, FIR 1 to 4 by MWZ. A dust emission intensity contour map of the data by MWZ is shown in figure 4.13a overlayed on the intensity grey–scale map of C I emission. The strong peaks corresponding to the condensations show no resemblance to the C I intensity data, *i.e.*, there is no enhancement of carbon column density with dust column density. The dust emission measured by MWZ is optically thin, even for the cores of the condensations at 11″ resolution, thus the dust emission intensity is the product of the grain temperature, emissivity and column density. The peaks seen are thought to be temperature and column density peaks. For the condensations

FIR 1 – 4, MWZ derive hydrogen column densities, $N_H = N(H + 2H_2)$, of about 20, 10, 3 and 9 $\times 10^{24} cm^{-2}$, respectively. A more typical value, found by averaging a 3′ by 3′ area centered on the ridge, but subtracting out the contribution from FIR 1 – 4, yields $N_H = 6 \times 10^{23} cm^{-2}$. However, the derivation of N_H based on dust emission is uncertain to about a factor of 3, largely due to uncertainties in the dust grain emissivity. KHW derived a column density of $N_H = 1.6 \times 10^{23} cm^{-2}$ for the same region synthesizing a 3′ beam including the condensations FIR 1 – 4. An average value of $N_H = 3 \times 10^{23}$ is therefore adopted here for positions along the ridge but away from the condensations. The neutral carbon column density derived for a typical position in the molecular ridge is $N_{CI}^T \geq 7.1 \times 10^{17}$, which yields a neutral carbon abundance relative to hydrogen of 2.4×10^{-6}. For the C I peak ($N_{CI}^T \geq 1.1 \times 10^{18}$ at $(-40'', -40'')$) the C to H abundance ratio increases to about 3.7×10^{-6}.

The cosmic abundance of carbon relative to hydrogen (see table 4.1) is 3.3×10^{-4}, which implies that roughly 0.7% of the carbon in the ISM is in the form of neutral carbon for a typical position within the Orion ridge outside the dense dust condensations, rising to about 1.1% for the position of the carbon intensity peak, but drastically dropping to 0.02% at the strongest condensation (IRc2). A typical amount of carbon tied up in CO then is about 6%.

The important point here is, independent from the uncertainties of the individual results, that atomic carbon (C I) is typically found at the 1% level of total carbon in GMC's away from the condensations, with some slight local enhancements, but drops to basically zero within condensations (hot cores). The remaining atomic carbon seen towards the positions of the condensations is mostly from material not related to the condensations, since the absolute carbon abundance is the same for positions away from the condensations and the C I line width does not increase like the lines of species seen only in the condensations (CH_3OH, SO_2, H_2CO).

c) Comparison of C I with C II

PDR models, as discussed earlier, provide for little neutral carbon anywhere but in a small interface region sandwiched between the ionized phase of the ISM (*e.g.*, C II) around the ionizing stars and the shielded interior of the molecular cloud, where all the gas phase carbon supposedly is bound in CO. Figure 4.13b shows a C II intensity contour map by Stacey *et al.* (1993) superimposed on the grey–scale map of CI as the one shown in figure 4.8a. The CII emission peaks near $\theta^1 C$ where

the temperature and ionizing flux is high due to the proximity of the ionizing sources, *i.e.*, the Trapezium stars. It drops where it lacks sufficient UV flux to be ionized. The C I emission, like all comparison molecular lines, avoids the area to the east due to the strong UV flux dissociating molecules or ionizing C I. It then rises to a peak towards the west where the C II falls off most rapidly. Note, that all CI intensity peaks, with the exception of one at (0″, – 80″), fall on the sharp C II emission gradient indicated by the 40 – 70% C II intensity contours. This is consistent with previous PDR models (*e.g.*, Tielens and Hollenbach 1985), which argue that C I will peak where there is only just enough ionizing radiation to photo–dissociate CO to yield reasonable densities of C I. However, away from the PDR interface as outlined above (*e.g.*, in the NW corner of the map), the CI column density data displays about half of the peak value seen at the PDR interface, while geometrically simple PDR models predict the C I intensity vanishes. This underlying C I intensity seen far from the peak caused by the PDR interface has been explained in the past, either by geometric effects such as a non–edge–on ionization front or by a very clumpy gas structure as discussed below. In the following section a PDR model, based on revised ion–chemistry model calculations, will be discussed. It is interesting to note, that, on small scales, the C I to C II correlation established here is the only correlation found between the C I map and any other species investigated.

4.6 Modification to the Neutral Carbon Abundance from Ion Chemistry Models

The chemical composition of the gas phase in interstellar molecular clouds is not well known for several reasons. Some crucial chemical reaction rates in the gas phase are often not known to better than a few orders of magnitude (see for example Adams, Smith and Clary 1985) while reactions on dust grains that may play an important role are known to an even lesser extent. The assumption made in many chemistry models of the ISM is that, with the exception of pure hydrogen reactions, none of the reaction products on dust grains return to the gas phase. This leads to a depletion of heavy atoms (C,N,O,S,...) in the gas phase of typically a factor of 10 relative to the solar system abundance. As this may not be a good approximation for very hot regions such as the condensations in OMC–1 (Walmsley *et al.* 1987; Blake *et al.* 1987), attempts have been made to include grain chemistry into models (*e.g.*, Hasegawa, Herbst and Leung 1992). However, for most areas the depletion is necessary to match observed abundances. Some PDR models (*e.g.*, Tielens and Hollenbach 1985; Hollenbach, Takahashi and Tielens 1991) assume no depletion of heavy elements in the gas phase, which is clearly unrealistic. The data presented show a typical depletion for carbon of about 10 in IC 342 and Orion, since the CO column densities (most of the gas phase carbon is in CO) relative to hydrogen are about a factor of 10 lower than the cosmic abundance of carbon relative to hydrogen. In the gas phase, and all further discussion will be exclusively about this, ion chemistry plays an important role due to the higher cross sections of reactions. The following discussion of a modification to previous PDR models of the neutral carbon abundance in the ISM is due to ion chemistry models by Graedel, Langer and Frerking (1982), Pineau des Forêts, Roueff and Flower (1992) and Roueff (1993, private communication).

Ions are mostly produced by cosmic ray (CR) ionization in dark interstellar clouds through

$$H_2 + CR \rightarrow H_2^+ + e^- \qquad (4.3)$$

with an ionization rate of about $\xi = 10^{-17}s^{-1}$ (Black and Dalgarno 1977). Another possible outcome with a branching ratio of 5% (Shah and Gilbody 1982) is a dissociative process

$$H_2 + CR \rightarrow H + H^+ + e^- \quad . \qquad (4.4)$$

H_3^+, which drives the ion molecule chemistry, is rapidly formed after reaction (4.3) via

$$H_2^+ + H_2 \rightarrow H_3^+ + H \ . \tag{4.5}$$

Sequences of protonation reactions, which in their first step remove H_3^+ again, then yield the more complex carbon and oxygen bearing molecular ions that subsequently, through dissociative recombinations with electrons, yield neutral molecules. However, the competing route for the destruction of H_3^+ is dissociative recombination with electrons,

$$H_3^+ + e^- \rightarrow H_2 + H \ , \tag{4.6}$$

which has a rate for destruction of H_3^+ depending on the electron density and dissociative recombination rate. Since the ionization rate due to cosmic rays linearly depends on the density of the medium and on the cosmic ray ionization rate, ξ, but the recombination rate depends on the product of the densities of electrons and ions, a medium with lower density will have a higher relative ionization state (assuming CR ionization to be the dominant source of ionization, as is the case inside molecular clouds without embedded IR sources, where the interior is shielded from outside NUV heating). The dissociative destruction of H_3^+ thus increases with decreasing density. At a transition density, n_{tran}, the dissociative destruction of H_3^+ will be equal to the destruction of H_3^+ via the protonation reactions which yield oxygen and carbon bearing molecules. This transition density separates a *low–density* from a *high–density* regime (see figure 4.14). In the high–density regime the gas will thus be rich in O_2, H_2O, OH, HCO^+ etc. (see figure 4.14c), while in the low–density regime neutral and ionized atoms such as C, H^+, C^+, S^+, etc. are abundant (see figures 4.14a,b). Some molecules, such as CO for example, are abundant in both phases.

The transition from the low–density to the high–density regime in the model, as shown in figure 4.14a, is very sharp. Figure 4.14b shows the abundances relative to hydrogen of several key species for the low to high density transition. The reason for the sharp separation of the two regimes is that in the high–density regime, H_3^+, through a chain of protonation reactions, efficiently produces oxygen and carbon bearing molecules (see figure 4.14c) that through charge transfer reactions remove H^+ (*e.g.*, $H_2O + H^+ \rightarrow H_2O^+ + H$). Since the electron recombination rate for H^+ is much smaller than for large molecular ions, H^+ charge transfer reactions effectively reduce the electron abundance. Atomic ions, such as S^+ or C^+, are also destroyed through reactions

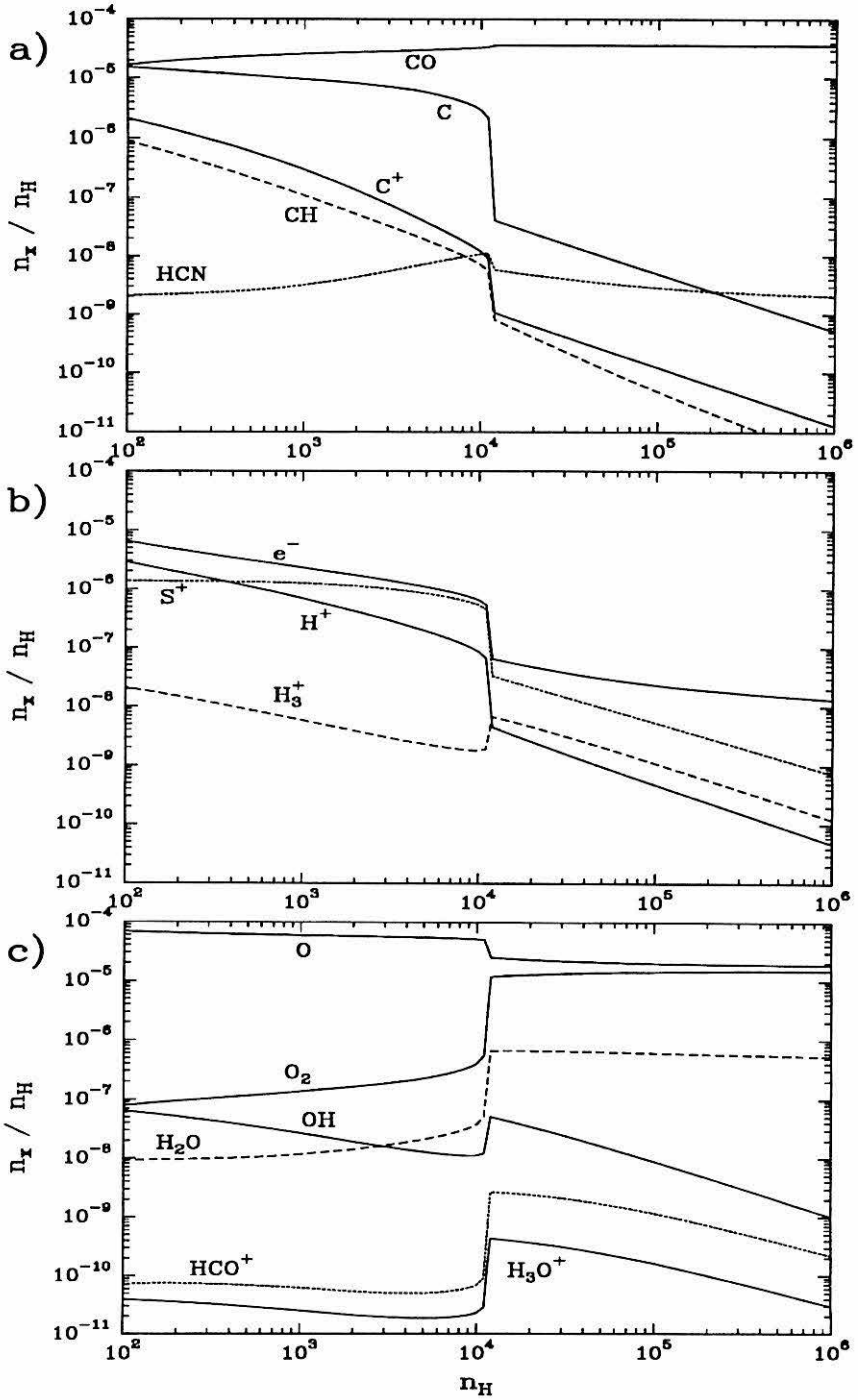


Figure 4.14 Chemical model calculations using a large dissociative recombination rate of H_3^+ yield a transition from the low– to high–density regime at about $n_{tran}n_H \simeq 10^4\,\mathrm{cm}^{-3}$. The abundances of the species shown are normalized to the hydrogen abundances and shown as a function of hydrogen density n_H. Relative abundance of *a)* carbon bearing species: C, C^+, CO, CH and HCN; *b)* species critical for the ion chemistry: electrons (e^-), H^+, $H_3^+$and S^+; *c)* oxygen bearing species: O, O_2, H_2O, OH, HCO^+ and H_3O^+.

with O_2 and H_2O thus decreasing the electron density further. Moving towards the low–density regime the electron abundance rises, thus the H_3^+ abundance drops and protonation reactions of H^+ start dominating the chemistry. The few oxygen bearing molecules, such as H_2O etc., that are now produced are efficiently removed through charge transfer reactions with the now more abundant H^+ as shown above. The high S^+, H^+ and C^+ abundances (they are not destroyed by O_2, H_2O, etc. anymore) further enhance the electron abundance due to their low recombination rate thus acting as a feedback in the system causing the chemical equilibrium to swing further into the state of a low–density regime. This feedback in the system causes the sharp transition.

In an early ion chemistry model (though not a PDR model) by Graedel, Langer and Frerking (1982) the phenomena of the transition density did occur at astrophysically relevant hydrogen densities. Their model discusses the different abundances of molecular species for low and high density clouds, but the full potential of this aspect was not exploited. The H_3^+ dissociative recombination rate of $k_e(H_3^+) \approx 3.3 \times 10^{-7}(T/300\,K)^{-0.5}\mathrm{cm}^3\,\mathrm{s}^{-1}$ used in their model was an average of measurements by Leu, Biondi and Johnson 1973, and McGowan *et al.* 1979. However, in the 1980's, when PDR models became the focus of attention, a much lower H_3^+ dissociative recombination rate was measured and used in the PDR models (*e.g.*, Adams and Smith 1988: $k_e(H_3^+) \approx 10^{-10}(T/300\,K)^{-0.5}\mathrm{cm}^3\,\mathrm{s}^{-1}$). This reduced the transition density for all the PDR models to values so low that it was no longer of astrophysical importance and observed molecular clouds were always thought to be in the high–density regime. With measurements of the H_3^+ dissociative recombination rate by Amano (1990) and Canosa *et al.* (1992) the card turned again in favor of a high H_3^+ dissociative recombination rate ($k_e(H_3^+) = 1.5\times10^{-7}\,(T/300\,K)^{-0.5}\mathrm{cm}^3\,\mathrm{s}^{-1}$). Recent chemical model calculations by Pineau des Forêts, Roueff and Flower (1992), hereafter PRF, using this newly established high rate, showed, just as the Graedel, Langer and Frerking model did, that the transition of the low– to high–density regime occurs at astrophysically relevant hydrogen densities, n_H[†]. The transition density found by PRF is rather typical for interstellar clouds ($n_H = 5.5 \times 10^3\,\mathrm{cm}^{-3}$). This has a significant impact on the abundances of some species

[†] Note that the number density of the medium is denoted n, while the number of H-nuclei is denoted $n_H = n(H) + 2n(H_2)$. Inside molecular clouds all hydrogen is molecular (see figure 4.15b) so that $n_H \simeq 2n$ (neglecting the He abundance).

when the molecular cloud observed is in the chemically low–density regime. The ratio of neutral carbon to other molecules, such as CO or H_2, is boosted up in their PDR model significantly when the medium is in the low–density regime. The C I to CO abundance ratio inside the cloud is about 10–20%, while it can exceed unity on the NUV photo–dissociated surface. This can explain why neutral carbon is seen at high levels throughout molecular clouds rather than **only** on their NUV illuminated surfaces as predicted by previous PDR models. In the dense cloud cores, such as IRc2, C I is found to be weak, in agreement with the models, which predict a very low C I abundance in the high–density regime.

For this thesis we performed ion chemistry model calculations of PDR's, based on code by Le Bourlot *et al.* (1992). The model uses a chemical network of 95 species (not including isotopes) and 750 gas–phase reactions of cosmic ray ionization, dissociative recombinations, ion–molecule reactions, neutral–neutral reactions, radiative recombinations, radiative associations, photo–dissociation, and secondary UV photons. One reaction on grains is also included : H + H (on grain) $\rightarrow$ H_2. Figure 4.14a shows calculations of abundances of C I, CO, C^+, CH and HCN relative to hydrogen as a function of hydrogen density for the center of a molecular cloud, *i.e.*, with no NUV flux, at $T \simeq 10K$, a cosmic ionization rate of $\xi = 2.5 \times 10^{-17}\mathrm{s}^{-1}$, and a depletion of heavy elements of 10. The transition from the low– to high–density regime for these model parameters occurs at a hydrogen density of $n_{tran} = n_H \simeq 10^4\,\mathrm{cm}^{-3}$. The C I to CO ratio is high, 0.3 to 0.1, for densities from $n_H = 10^3\,\mathrm{cm}^{-3}$ up to the transition density, $n_H \approx 10^4\,\mathrm{cm}^{-3}$, above which the C I abundance is suppressed.

When more free electrons exist, either through UV ionization, a higher metallicity, *i.e.*, lower depletion of heavier elements, or a higher cosmic ray ionization rate, the transition density will also be pushed up since the equilibrium for dissociative recombination of H_3^+ compared to the protonation of H_3^+ is pushed towards higher densities. Higher temperatures reduce the transition density roughly as $1/T$ because the H_3^+ dissociative recombination rate from Canosa *et al.* (1992) of $k_e(H_3^+) = 1.5 \times 10^{-7}\,(T/300\,K)^{-0.5}\mathrm{cm}^3\,\mathrm{s}^{-1}$ decreases for increasing temperature. However, at very high temperatures ($> 100K$), the depletion of heavy elements in the gas phase may be reduced thus offsetting the effect of a lower dissociative recombination rate. Due to the high electron abundance on the surface of a NUV illuminated cloud those areas are almost always in

the chemically low–density regime.

Figure 4.15a shows the CO, C, C^+ and electron abundance of an isobaric model as a function of optical depth into the cloud with a pressure of $nT = 5 \times 10^4\, K\,cm^{-3}$ for a molecular cloud in the low–density regime. The pressure of $nT = 5 \times 10^4\, K\,cm^{-3}$ is the average pressure of the gas in IC 342 as derived by Eckart *et al.* (1990). The cloud is externally heated by a strong NUV field of 1000 times the standard UV interstellar radiation intensity of Mathis, Mezger and Panagia (1983) of 1.2×10^5 photons $cm^{-2}\ s^{-1}\ Å^{-1}$. Integrated between 910 Å (13.6 eV) and 2100Å (6 eV) this yields the same standard local Galactic NUV radiation field F_o as used in previous sections of this work, $F_o = 1.6 \times 10^{-3}$ erg $s^{-1}\,cm^{-2}$ in the 6–13.6 eV energy range (Habing 1968). The C^+ abundance is highest for low optical depth, *i.e.*, on the very surface of the cloud and the C I abundance rises where the C^+ starts dropping, as is qualitatively seen in all PDR models. However, the C I abundance is very high, and only at an optical depth of ≥ 6, does the CO abundance dominate. Further into the cloud the C I to CO ratio is the same as shown in figure 4.14a in the low–density regime since the gas is now shielded from the NUV flux. The hydrogen density, $n_H = n(H) + 2n(H_2)$, inside the cloud is about $7 \times 10^3\,cm^{-3}$ and the temperature $T \simeq 12K$. Figure 4.15b shows the temperature profile as a function of τ. The carbon stays at a roughly 10–20% level of the CO abundance throughout the cloud with an enhancement at the interface region if the cloud is externally heated by NUV radiation. The temperature dip visible at $\tau \simeq 5$ in figure 4.15b is caused by C I cooling. Figure 4.15c shows the column density as a function of optical depth into the cloud. The measured C I column density for IC 342 of $N^T_{\rm C\,I} \geq 7 \times 10^{17} cm^{-2}$ corresponds to an average optical depth for a cloud observed in the nucleus of IC 342 of about $\tau \geq 65$ (assuming the beam filling factor to be about unity). At that optical depth the predicted CO abundance is $N^T_{\rm CO} \geq 4 \times 10^{18} cm^{-2}$, which is in good agreement with the measurements. The C II column density is independent of optical depth for $\tau \geq 3$ since its abundance is negligible inside the molecular cloud.

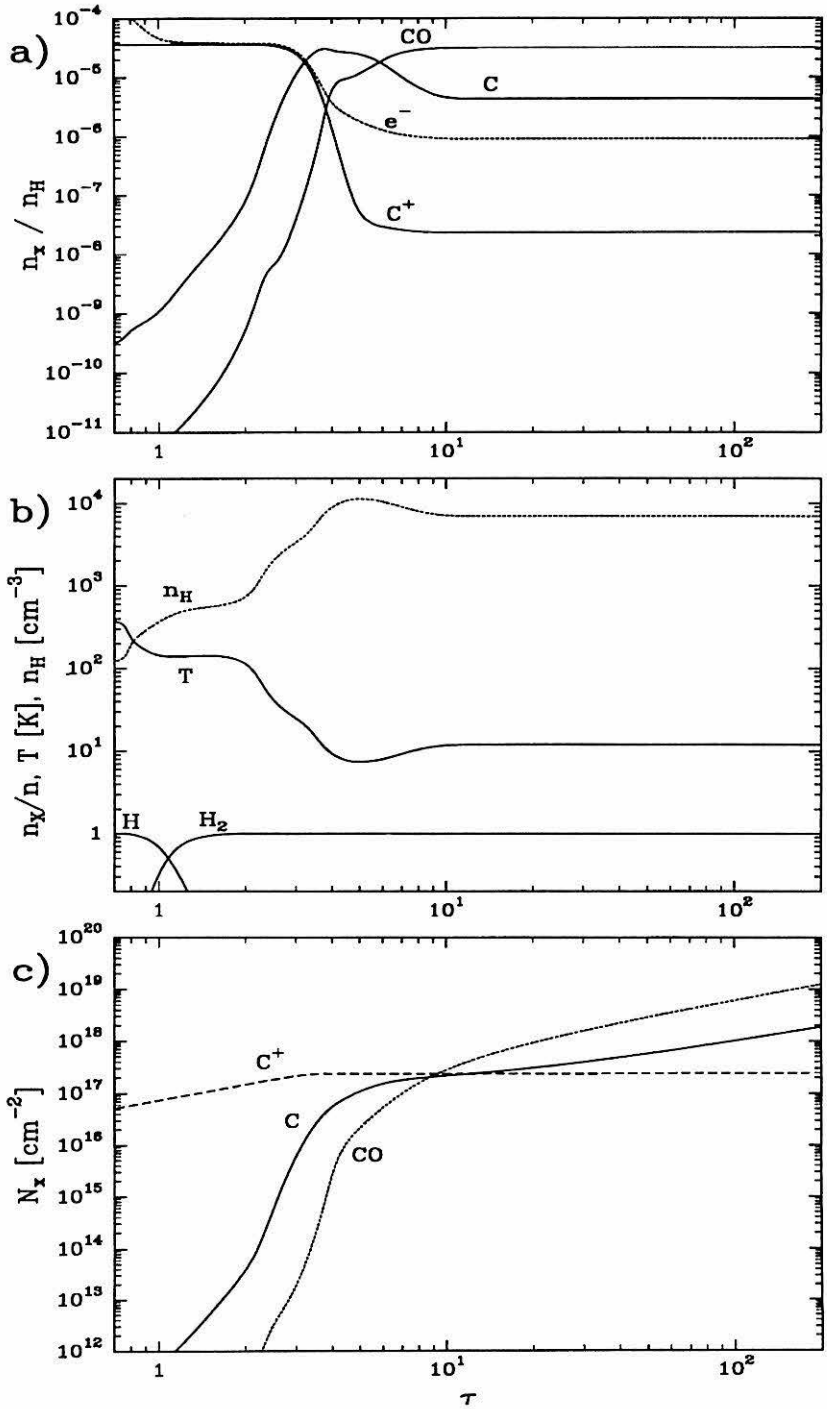


Figure 4.15 Isobaric model in the low–density regime ($nT = 5 \times 10^4\ K\,\mathrm{cm}^{-3}$) with a depletion of C,N,O and S of 10 for a cloud externally illuminated by a NUV source ($F_o = 1000$). *a)* Abundance ratios relative to hydrogen as a function of optical depth τ into the cloud, and *b)* temperature and density profile. The transition from atomic to molecular hydrogen at an optical depth of $\tau \simeq 1$ is also shown. *c)* Column density as a function of optical depth into the cloud.

4.7 Summary and Conclusions: A Unified Model of Neutral Carbon Abundance

Observations of the $^3P_1 \rightarrow {}^3P_0$ fine structure transition of C I were carried out on large and small scales to investigate the abundance and distribution of neutral carbon (C I) in the ISM. On large scales, as measured with the nuclear bar of the galaxy IC 342, neutral carbon shows a strong correlation with the low J rotational transitions of ^{13}CO. The C I ($J = 1 \rightarrow 0$) to CO ($J = 2 \rightarrow 1$) intensity ratio was found to be 1.4 for the central 130 pc of IC 342 and stayed constant for positions out to the radio edge of IC 342's nuclear bar (as measured in ^{12}CO by Eckart *et al.* 1990). The same ratio in the Milky Way, as measured by COBE (Wright *et al.* 1991), is 2.1. The similarity of this ratio was very surprising, especially when it was considered that the COBE result is an average over the entire Galaxy, while the IC 342 result is from the core of a galaxy and that the absolute intensities that entered into the ratios are higher by two orders of magnitude for the IC 342 results. The intensity of C I emission in the Orion ridge is identical to the intensity of the central 15″ of IC 342 despite the different excitation conditions apparent from the different CO($2 \rightarrow 1$) intensities. This yields rather different C I to CO($2 \rightarrow 1$) ratios. This ratio is, however, nearly identical for IC 342 and the COBE results, indicating that regions of condensations in Orion–like GMCs are not typical contributors to the average emission of galaxies such as the Milky Way. COBE also showed that the C I ($1 \rightarrow 0$) and the C I ($2 \rightarrow 1$) transition combined provide more cooling for the ISM than all observed CO rotational transitions together. This is most likely also true for IC 342.

An abundance ratio of C I to CO of 13% to 18% was found in the central 130 pc of IC 342. This is very similar to the roughly 10% to 15% C I to CO abundance ratio found for a typical position in the molecular ridge in the vicinity of OMC–1. The ratio of neutral carbon to hydrogen was found to be about $\approx 3 \times 10^{-6}$ and twice that in IC 342. Using the cosmic carbon abundance of 3.3×10^{-4} relative to hydrogen (Allen 1973), about 1% to 2% of the total carbon in the ISM is in the form of neutral carbon (at least in the regions where the excitation conditions are adequate). Assuming all the carbon in the gas phase to be either in CO, C I, or C II (which yields a gas phase depletion of $\approx$ 10 for carbon), yields about 10% to 15% of the carbon in the gas phase in form of neutral atomic carbon. Only in the condensations in OMC–1, like the clump of gas associated with IRc2, the total neutral carbon abundance sharply reduces by up to two orders of magnitude.

This raises two questions: (1) Why does the neutral carbon abundance drop sharply in the regions of dense cores and (2) why is carbon seen at about a 10% to 15% level throughout the cloud elsewhere?

To investigate the relation of C I to other atoms or molecules on small scales the C I intensity map was compared to ^{12}CO, ^{13}CO, $C^{18}O$, CS, dust and C II. Except for the void area east of θ^1C seen in all maps, the C I intensity map did **not** correlate with any of the CO or CS (Mundy *et al.* 1986) intensity maps, hydrogen column density maps as derived from the dust measurements (KHW, MWZ) or CO column density maps as measured with the $C^{18}O$ ($J = 1 \rightarrow 0$) transition (Wilson *et al.* 1986). The only correlation found was with an intensity map of C II (Stacey *et al.* 1993). This correlation showed that at the edge of the C II emission there were slight C I intensity peaks. Previous photo–dissociation region (PDR) models (*e.g.*, Langer 1976; Tielens and Hollenbach 1985; van Dishoeck and Black 1988) predict the neutral carbon abundance to peak in a thin layer sandwiched between a layer of ionized carbon and, deeper into the cloud, by molecular gas, where all the carbon is thought to be tied up in CO (see figure 4.2). According to those models, C I exists only in a thin PDR transition layer between the ionized and neutral molecular region, where the NUV flux is sufficient to photo–dissociate CO but not ionize all the carbon. The PDR models thus offer a natural explanation for the neutral carbon deficiency within the condensations since the interiors of the clumps are completely shielded from the NUV flux necessary to photo–dissociate CO. To explain the observed widespread distribution of C I many modifications to the existing PDR models have been brought forth (see discussions by Chiéze, Pineau des Forêts and Herbst 1989, and Genzel, Harris and Stutzki 1989). The most popular theory invokes a very clumpy structure of the ISM (Stutzki *et al.* 1988, Burton, Hollenbach and Tielens 1990) which would allow the NUV flux to penetrate space efficiently at a rather high clump to interclump density ratios of 1000 creating many PDR transition layers on the surfaces of those clumps (see review by Stutzki *et al.* 1991). The CI maps presented do not show enhancement at the surface of the clumps that are resolved by $C^{18}O$ or dust emission observations, which makes it unlikely that clumps at different scales, that most certainly exist (*e.g.*, Falgarone, Phillips and Walker 1991), are the main contributors to the observed underlying CI emission. On the contrary, observations in S140 by Keene *et al.* (1985) clearly showed an increase of C I emission deeper

into the cloud, *i.e.*, away from the NUV source, than the CO peak, which can not be explained with a clumpy structure anymore.

For the large–scale measurement of IC 342, the PDR model by Hollenbach, Takahashi and Tielens (1991) is able to reproduce the measured intensities for an average density of $3.5 \times 10^3\,\mathrm{cm}^{-3}$ and a NUV field flux of $10^3\,F_o$, where F_o is the local galactic NUV field. Two problems are seen to remain with this model. First, they use no depletion of heavy elements in the gas phase, due to molecules freezing out on grains, which probably was done to raise the predicted abundances of some species in the gas phase, such as neutral carbon, closer to the observed abundances. Second, the observed C to CO abundance ratio requires all observed CO emission to emanate from regions from which C I emission also emanates since the 10% measured abundance ratio is predicted by the PDR model to exist only in the thin PDR transition layer.

A solution to the afore mentioned problems is offered by a revised ion chemistry model of Pineau des Forêts, Roueff and Flower (1992), hereafter PRF, that includes photo–dissociation processes for up to moderately strong NUV flux ($F_o \leq 10^3$). Based on a much higher dissociative recombination rate of H_3^+ measured by Amano (1990) and Canosa *et al.* (1992), PRF show that the ISM is split into two different chemical regimes depending on the density. In the high–density regime, which is similar to traditional PDR model compositions of the ISM, the chemistry is dominated by protonation reactions with H_3^+. In the low–density regime, with typical hydrogen densities up to about $n_H \approx 10^4\,\mathrm{cm}^{-3}$, the chemical composition of the ISM is rather different from the traditionally (used in all the previous PDR models) predicted one. Charge transfer reactions with H^+ dominate. The medium has a much higher degree of ionization and atomic ions like H^+, C^+ and S^+ and neutral atoms like C are very abundant. C to CO ratios of 10% to 20% are typical throughout the cloud without irradiation by NUV. On the surface of the cloud, depending on the amount of NUV flux illuminating it, the C to CO ratio can exceed unity. Similar ratios were also found in an early ion chemistry model without NUV illumination by Graedel, Langer and Frerking (1982), which was also based on a high H_3^+ dissociative recombination rate, $k_e(H_3^+)$. In the time between the model by Graedel, Langer and Frerking (1982) and the model by PRF (1992), PDR models used much lower values for $k_e(H_3^+)$, thus having difficulty to explain the observed C I abundances.

The PDR model by PRF allows for a unified explanation of the observed phenomena. High C I to CO ratios exist in the interior of molecular clouds, where C I can now coexist with CO. The similarity for the ^{13}CO low J lines comes thus naturally, since the excitation conditions are similar. The measured neutral carbon abundance relative to hydrogen of $\approx 2.4 \times 10^{-6}$ matches the predicted one from the new chemical model calculations in the low–density regime very well. The high intensities for the CI line, as measured in IC 342 and by COBE, are not surprising either since the C I lines match the excitation conditions of the ISM as well as the CO lines, have very similar Einstein A–coefficients, but, in average, emit at a higher frequency thus more efficiently cool the medium. The C I intensity peaks seen in the Orion map that fall on the sharp gradient of reducing C II emission are, just as qualitatively with the old PDR models, due to the PDR transition layer enhancing the neutral carbon abundance. The fact that they are not very strong peaks (CI to CO ratios up to 0.33 were observed) relative to the average intensity is probably due to the small value of optical depth over which they emanate yielding low column densities for that range and small beam filling factors. Both effects will dilute the effect of a strong C I abundance increase. The fact that at the positions of the condensations (as measured by dust emission) the neutral carbon abundance vanishes is due to the local densities inside those clumps being much higher than $n_H \simeq 10^4\,\mathrm{cm}^{-3}$, leaving the medium in the high–density regime. Very little neutral carbon is predicted in the high–density chemistry regime, just as in the old PDR models.

In summary, it should be kept in mind, that the transition from the chemically low– to high–density regime sensitively depends on several parameters, such as the cosmic ray ionization rate and the H_3^+ dissociative recombination rate. The cosmic ray ionization rate is not better known than to a factor of a few, while the H_3^+ dissociative recombination rate has gone down and up again during the last decade by about 3 orders of magnitude. The exact value of the transition density thus remains uncertain. However, the model calculations, performed as part of this thesis, are able to predict the observed C I and CO column densities quite accurately, while other model predictions remain to be verified. The model is not able to predict very high NUV flux PDRs yet, since it was developed for clouds with little to moderate NUV illumination. Since the NUV flux only modifies the surface of the molecular cloud a higher NUV field should not change the C I and CO column densities significantly.

It is suggested, based on the observational data presented here and the good agreement of the observations with the ion–chemistry model, that the C I ($1 \rightarrow 0$) emission emanates from the bulk of gas in molecular clouds, *i.e.*, from regions similar to those emitting in the low J ^{13}CO lines. This requires the bulk of the observed medium to be in the chemically low–density regime. The fine structure emission lines from neutral carbon are thus a more important coolant for the ISM than the rotational transitions of CO for low to moderate hydrogen densities. This is in contrast to previous PDR models that only had significant neutral carbon abundances in a thin PDR transition layer on the surface of molecular clouds irradiated by NUV.

The model calculations shown here also predict very high C to CO ratios on the PDR surface of molecular clouds that, to some extent, have been observed in several positions in the OMC–1 complex.

4.8 Appendix: C and CO column density derivation

a) C I column density

The column density of C I can be calculated from the observed intensity I_ν^J under the assumption that the C I emission is optically thin or at most of moderate opacity. The neutral carbon atom splits its ^3P ground state into three fine structure levels due to spin–orbit coupling. The transition observed for this thesis is the $^3P_1 \rightarrow {}^3P_0$, *i.e.*, the lower of the two transitions (see figure 4.1). The number of carbon atoms emitting in the level with angular momentum J of the fine structure transition are

$$N_{\rm C\,I} = \frac{4\pi}{A_{1,0}h\nu} \int I_\nu^J \, d\nu \quad , \tag{4.7}$$

where $A_{1,0} = 7.9 \times 10^{-8} s^{-1}$ (Nussbaumer and Rusca 1979) is the Einstein A–coefficient for spontaneous emission from the $J = 1$ to the $J' = 0$ level. To calculate the total number of C I atoms the population of the three levels need to be known. In local thermodynamic equilibrium (LTE), *i.e.*, when the population of all levels can be characterized by one excitation temperature T_{ex}, the ratio of populations for levels with angular momentum J is given by the Boltzmann equation

$$\frac{n_J}{n_{J'}} = \frac{g_J}{g_{J'}} \exp\left(-\frac{E_J - E_{J'}}{k_B T_{ex}}\right) \quad , \tag{4.8}$$

with the statistical weights $g_J = 2J + 1$ and E_J the energy of the level for the angular momentum J. The total number of C I atoms then is

$$N_{\rm C\,I}^T = \frac{\sum_{J'=0}^{2} g_{J'} \exp\left(-\frac{E_{J'}}{k_B T_{ex}}\right)}{g_J \exp\left(-\frac{E_J}{k_B T_{ex}}\right)} \frac{4\pi}{A_{1,0}h\nu} \int I_\nu^J \, d_\nu \quad , \tag{4.9}$$

where the sum $f = \sum_{J'=0}^{2} g_{J'} \exp\left(-\frac{E_{J'}}{k_B T_{ex}}\right)$ is called the partition function. For the $J = 1 \rightarrow 0$ transition, with the energies of the levels of $E_0 = 0$, $E_1/k_B = 23.6K$ and $E_2/k_B = 62.5K$, this yields,

$$N_{\rm C\,I}^T = \left[e^{\frac{23.6}{T_{ex}}} + 3 + 5e^{\frac{-38.9}{T_{ex}}}\right] \frac{4\pi}{A_{1,0}h\nu} \int I_\nu^1 \, d_\nu \quad . \tag{4.10}$$

In millimeter and submillimeter astronomy intensities are often quoted in units of temperature, which, for an ideal telescope looking at a source much larger than the beam size at an optically thick line in LTE, would actually be the physical temperature of the gas in the Rayleigh–Jeans

approximation ($h\nu \ll k_B T$). Throughout this thesis any Rayleigh–Jeans temperature is denoted with a hat–symbol, $\hat{T}$. To obtain the true physical temperature of the emitting molecule of an optically thick and thermalized line where the source is larger than the beam of the telescope (*i.e.*, the beam filling factor is unity) the Rayleigh–Jeans temperature $\hat{T}$ has to be corrected with the Planck function†

$$T = \left(\frac{k_B}{h\nu} \ln \left(1 + \frac{h\nu}{k_B \hat{T}} \right) \right)^{-1} \tag{4.11a}$$

to obtain the kinetic gas temperature, T. If the observed line is optically thin or the beam filling factor less than unity, the observed Rayleigh–Jeans temperature can not be converted to a physical temperature. The only conclusion that can then be drawn for the physical temperature is a lower limit, *i.e.*,

$$T \geq \left(\frac{k_B}{h\nu} \ln \left(1 + \frac{h\nu}{k_B \hat{T}} \right) \right)^{-1} \quad . \tag{4.11b}$$

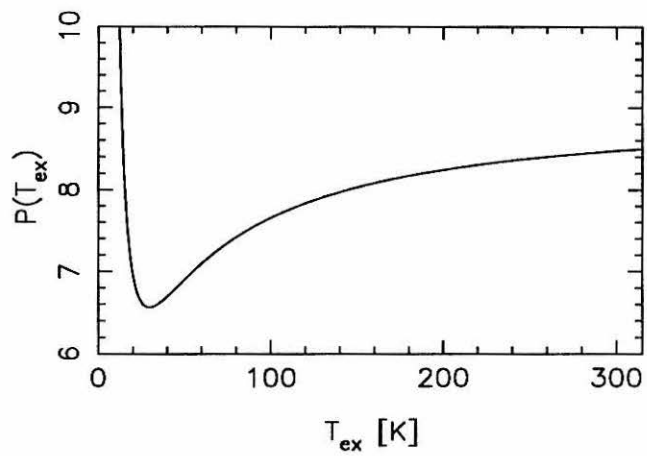


Figure 4.16 The partition function of neutral carbon's ^{3}P ground state divided by the relative abundance of the J=1 level of C I. This is the only term that directly depends on the excitation temperature T_{ex} in equation (4.4). It has values between 6.5 and 8.5 for excitation temperatures between 14.5 and 317 K.

Another convention in radio–astronomy is to use a velocity scale with respect to the local

† If the intensity was measured using a beam switching technique, a correction for the cosmic background radiation of 2.75 K has to be made. However, at the $^3P_1 \rightarrow {}^3P_0$ and $^3P_2 \rightarrow {}^3P_1$ transition frequencies of C I the contribution is 0.3 mK and 0.02 mK, respectively, and can thus be ignored.

standard of rest (LSR), *i.e.*, our local region of the Galaxy, rather than a frequency, thus the frequency interval $d\nu$ needs to be converted to a velocity interval dV. The *antenna temperature* $\hat{T}_R^*$, is corrected for all telescope losses and relates to the intensity via

$$k_B \hat{T}_R^* = \frac{c^2 I_\nu^J}{2\nu^2} \quad . \tag{4.12}$$

For an observed line with moderate opacity at line center $\tau_o \approx 1$ the column density needs to be corrected by a term $\tau_o / 1 - e^{-\tau_o}$, which is a photon escape probability (Goldreich and Kwan 1974). The total column density for the $^3P_1 \rightarrow {}^3P_0$ transition then is

$$N_{\rm CI}^T = 1.9 \times 10^{15} \left[e^{\frac{23.6}{T_{ex}}} + 3 + 5e^{\frac{-38.9}{T_{ex}}} \right] \frac{\tau_o}{1 - e^{-\tau_o}} \int \hat{T}_R^* \, dV \quad . \tag{4.13}$$

The intensity, $\int \hat{T}_R^* \, dV$, is a directly measured quantity, but the excitation temperature T_{ex} and the opacity τ_o at line center need to be determined by other means. Contrary to rotational transitions like those of CO, where the excitation temperature roughly linearly affects the column density as shown below, it is rather uncritical here, as long as $T \geq 15K$. Figure 4.16 shows the function $P(T_{ex})$, which is the partition function f divided by the relative population of the level with $E_1/k_B = 23.6K$,

$$P(T_{ex}) = e^{\frac{23.6}{T_{ex}}} + 3 + 5e^{\frac{-38.9}{T_{ex}}} \quad . \tag{4.14}$$

$P(T_{ex})$ includes all terms involving the excitation temperature T_{ex} of equation (4.13). For temperatures between 15 K and 300 K the excitation temperature does not affect the result significantly, that is $P(T_{ex})$ varies only between 6.5 and 8.5.

b) CO column density

The CO column density can be calculated in a similar way to the above derivation of the C column density. The CO partition function, f, now is

$$f = \sum_{J'=0}^{\infty} g_{J'} e^{-\frac{E_{J'}}{k_B T_{ex}}} \quad . \tag{4.15}$$

Townes and Schawlow (1955) showed that the partition function of a molecule with a $^1\sum$ electronic ground state at high excitation temperature T_{ex} can be approximated by $f = k_B T_{ex}/hB$, where B is the rotational constant, *i.e.*, half the frequency of the $J = 1 \rightarrow 0$ transition. The CO

column density thus roughly depends linearly on the assumed excitation temperature T_{ex} through the partition function. The Einstein A coefficient for a molecule with a $^1\sum$ electronic ground state is (Townes and Schawlow 1955)

$$A_{J,J-1} = \mu^2 \frac{J}{g_J} \frac{64\pi^4\nu^3}{3hc^3} \quad , \tag{4.16}$$

with $\mu = 0.11 Debyes$, the dipole moment of CO. The total CO column density, as derived from the $C^{18}O$ ($J = 1 \rightarrow 0$) transition thus is

$$N_{CO}^T = \frac{^{16}O}{^{18}O} 4.7 \times 10^{13} T_{ex} e^{\frac{5.3}{T_{ex}}} \int \hat{T}_R^* \, dV \quad , \tag{4.17}$$

where $E_1(C^{18}O)/k_B = 5.3K$.

4.9 References Chapter IV

Adams, N.G. Smith, D., & Clary, D. C. 1985, ApJ, 296, L31.

Adams, N.G., & Smith, D. 1988, in Millar T. J. Williams, D. A., eds., Rate Coefficients in Astrochemistry, Kluwer, Dordrecht, p. 173.

Allen, C. W. 1973, "Astrophysical Quantities," London, Athlone Press.

Amano, T. 1990, J. Chem Phys., 92, 6492.

Bally, J., Langer, W. D., Stark, A. A., & Wilson, R. W. 1987, ApJ, 312, L45.

Batrla, W., Wilson, T. L., Bastien, P., & Ruf, K. 1983, A&A, 128, 129.

Becklin, E. E., & Neugebauer, G., 1967, ApJ, 147, 799.

Becklin, E. E., Gatley, I., Matthews, K., Neugebauer, G., Sellgren, K., Werner, M. W., & Wynn-Williams, C. G. 1980, ApJ, 236, 441.

Beichman, C. A. 1988, *Astrophys. Lett. Comm., 27, 67.*

Black, J. H., & Dalgarno A. 1977, ApJS, 34, 405.

Blake, G. A., Sutton E. C., Masson, C. R. & Phillips, T. G. 1987, ApJ, 315, 621.

Burton, M. G., Hollenbach, D. J., & Tielens, A. G. G. M. 1990, ApJ, 365, 620.

Büttgenbach, T. H. 1993, to appear in IEEE Trans. Microwave Theo. and Tech,. October 1993.

Büttgenbach, T. H., Keene, J., Phillips, T. G., & Walker, C. K. 1992a, ApJ, 397, L15.

Büttgenbach, T. H., Keene, J., Phillips, T. G., & Young, K. 1993, in preparation.

Büttgenbach, T. H., LeDuc, H. G., Maker, P. D., & Phillips, T. G. 1992b, IEEE Trans. Appl. Superconductivity, 2, 165.

Büttgenbach, T. H., Miller, R. E., Wengler, M. J., Watson, D. M., & Phillips, T. G. 1988, IEEE Trans. Microwave Theo. and Tech., 36, 1720.

Canosa, A., Gomet, J. C., Rowe, B., Mitchell, J. B. A., & Queffelec, J. L., 1992, J. Chem. Phys., 97, 1028.

Castets, A., Duvert, G., Dutrey, A., Bally, J., Langer, W. D., & Wilson, R. W. 1990, A&A, 234, 469.

Chièze, J. P., Pineau des Forêts, G., & Herbst, E. 1991, ApJ, 373, 110.

Cohen, R. J., Matthews, N., Few, R. W., & Booth, R. S. 1983, *MNRAS, 203, 1123.*

Cooksy, A. L., Saykally, R. J., Brown, J. M., & Evenson, K. M. 1986, ApJ, 309, 828.

Crawford, M. K., Genzel, R., Townes, C. H., & Watson, D. M. 1985, ApJ, 291, 755.

Eckart, A., Downes, D., Genzel, R., Harris, A.I., Jaffe, D.T., & Wild, W. 1990, ApJ, 348, 434.

Ellison, B. N., Schaffer, P. L., Schaal, W., & Miller, R. E. 1989, Int. J. IR and MM Waves, 10, 937.

Falgarone, E., Phillips, T. G., & Walker, C. K. 1991, ApJ, 378, 186.

Frerking, M. A., Keene, J., Blake, G. A., & Phillips, T. G. 1989, ApJ, 344, 311.

Genzel, R., Harris, A. I., & Stutzki, J. 1989, in Kaldeich B. H., ed., Proc. 22nd ESLAB Symp., IR Spectr. in Astr., ESA–SP–290, p. 115.

Genzel, & Stutzki, J. 1989, Annu. Rev. Astron. Astrophys., 27, 41.

Gierens, K. M., Stutzki, J., & Winnewisser, G. 1992, A&A, 259, 271.

Goldreich, P. & Kwan J. 1974, ApJ, 189, 441.

Goldsmith, P. F., Krotkov, R., Snell, R. L., Brown, R. D., & Godfrey, P. 1983, ApJ, 274, 184.

Graedel, T. E., Langer, W. D., & Frerking, M. A. 1982, ApJS, 48, 321.

Grossman, E., 1989, Atmospheric transmission software, Airhead Software, 2069 Bluff St., Boulder, CO 80304.

Güsten, R., Serabyn, E., Kasemann, C., Schinkel, A., Schneider, G., Schulz, A. & Young, K. 1993, ApJ, 402, 537.

Habing, H. J. 1968, Bull. Astr. Inst. Netherlands, 19, 421.

Harris, A. I., Townes, C. H., Matsakis, D. N. & Plamer, P. 1983, ApJ, 265, L63.

Hasegawa, T.I., Herbst, E., & Leung, C. M. 1992, ApJS, 82, 167.

Hildebrand, R. H., Loewenstein, R. F., Harper, D. A., Orton, G. S., Keene, J., & Whitcomb, S. E. 1985, Icarus, 64, 64.

Ho, P. T. P., Turner, J. L., & Martin, R. N. 1987, A&A, 308, L7.

Hollenbach, D. J., Takahashi, T., & Tielens, A. G. G. M. 1991, ApJ, 377, 192.

Ishizuki, S., Kawabe, R., Ishiguro, M., Okumura, S. K., Morita, K.-I., Chikada, Y., & Kasuga, T. 1990, Nature, 344, 224.

Keene, J. 1990, Proc. of SETI Conf. *Carbon in the Galaxy: Studies from Earth and Space*, eds. Tarter, J., Chang, S. & DeFrees, D., NASA-CP 3016, p.181.

Keene, J., Blake, G. A., Phillips, T. G., Huggins, P. J., & Beichman, C. A. 1985, ApJ, 299, 967.

Keene, J., Hildebrand, & Whitcomb, S. E. 1982, ApJ, 252, L11.

Langer, W. D. 1976, ApJ, 206, 699.

Le Bourlot, J., Pineau des Forêts, G., Roueff, E., & Flower, D.R. 1992, A&A, 267, 233.

Leu, M. T., Biondi, M. A., & Johnson, R., 1973, Phys. Rev., A7, 292.

Linsky, J. L. 1973, ApJS, 25, 163.

Lo, K. Y. *et al.* 1984, ApJ, 282, L59.

Maddalena, R. J., Morris, M., Moscowitz, J., & Thaddeus, P. 1986, ApJ, 303, 375.

Mathis, J. S., Mezger, P. G. & Panagia, N. 1983, A&A, 128, 212.

McCall, M. L. 1989, AJ, 97, 1341.

McGowan, J. W., Mul, P. M., D'Angelo, V. S., Mitchell, J. B. A., Defrance, P., & Froelich, H. R. 1979, Phys. Rev. Lett., 42, 373.

Mezger, P. G., Wink, J. E., & Zylka, R. 1990, A&A, 228, 95.

Morris, M., & Lo, K. Y. 1978, ApJ, 223, 803.

Mundy, L. G., Scoville, N. Z., Baath, L. B., Masson, C. R. & Woody, D. P. 1986, ApJ, 304, L51.

Neugebauer, G., Münch, G., Kieffer, H., Chase, Jr. S. C. & Miner, E. 1971, AJ, 76, 719.

Nussbaumer, H., & Rusca, 1979, A&A, 72, 129.

Phillips, T. G., & Huggins, P. J. 1981, ApJ, 251, 533.

Pineau des Forêts, G., Roueff, E., & Flower, D.R. 1992, Mon. Not. R. Astron. Soc., Short Comm., 258 45.

Rickard, L. J, & Harvey, P. M. 1984, AJ, 89, 1520.

Rickard, L. J, & Palmer, P. 1981, A&A, 102, L13.

Rickard, L. J, Palmer, P., Morris, M., Zuckerman, B. & Turner B. E. 1975, ApJ, 199, L75.

Robinson, L. J. 1984, *Sky Telesc., 67, 4.*

Schloerb, F. P. & Loren, R. L. 1982, Symp. on the Orion Nebula to Honor Henry Draper, Ann. N.Y. Acad. Sci. (eds. A. E. Glassgold and P. J. Huggins), 395, 32.

Schmid–Burgk, J., Güsten, R., Mauersberger, R., Schulz, A., & Wilson, T. L. 1990, ApJ, 362, L25.

Schröder, K., Staemmler, V., Smith, M. D., Flower, D. R., & Jaquet, R. 1991, J. Phys. B: At. Mol. Opt. Phys., 24, 2487.

Serabyn, E., Phillips, T. G., & Masson, C. R. 1991, J. Appl. Optics, 30, 1227.

Shah, M.B., & Gilbody, H.B., 1982, J. Phys. B., 15, 3441.

Stacey, G. J., Geis, N., Genzel, R., Lugten, J. B., Poglitsch, A., Sternberg, A., & Townes, C. H. 1991, ApJ, 373, 423.

Stacey, G. J., Jaffe, D. T., Geis, N., Genzel, R., Harris, A. I., Poglitsch, A., Stutzki, J., & Townes, C. H. 1993, ApJ, 402, 173.

Stutzki, J., Genzel, R., Graf, U., Harris, A. I., Sternberg, A., & Güsten, R., 1991, IAU Symp. No. 147, "Fragmentation of Molecular Clouds and Star Formation," Falgarone, E., *et al.*, eds. pp. 235.

Stutzki, J., Stacey, G. J., Genzel, R., Harris, A. I., Jaffe, D. T., & Lugten, J. B., 1988, ApJ, 332, 379.

Tielens, A. G. G. M., & Hollenbach, D. 1985, ApJ, 291, 772.

Townes, C. H., & Schawlow, A. L. 1955, McGraw–Hill Book Co., Inc..

Tucker, K. D., Kutner, M. L., & Thaddeus, P., 1973, ApJ, 186, L13.

Ulich, B. L. 1981, AJ, 86, 1619.

van Dishoeck, E. F., & Black, J. H. 1988, ApJ, 334, 771.

Verschuur, G. L. 1992, Sky & Teles., 83, 379.

Walker, C. K., Kooi, J. W., Chan, M., LeDuc, H. G., Carlstrom, J. E., & Phillips, T. G. 1992, Int. J. IR and MM Waves, 13, 785.

Wall, W. F., & Jaffe, D. T. 1990, ApJ, 361, L45.

Walmsley, C. M., Hermsen, W., Henkel, C., Mauersberger, R. & Wilson, T. L. 1987, A&A, 172, 311.

White, G. J., & Padman, R. 1991, Nature, 354, 511.

Wilson, T. L. & Johnston, K. J. 1989, ApJ, 340, 894.

Wilson, T. L., Serabyn, E., Henkel, C., & Walmsley, C. M. 1986, A&A, 158, L1.

Wright, E. L. et al. 1991, ApJ, 381, 200.

Yamamoto, S., & Saito, S. 1991, ApJ, 370, L103.

Young, J. S., & Scoville, N. 1982, ApJ, 258, 467.

Young, J. S., & Scoville, N. 1984, ApJ, 287, 153.

Ziurys, L. M., Marin, R. N., Pauls, T. A., & Wilson, T. L. 1981, A&A, 104, 288.

Zmuidzinas, J., Betz, A. L., Boreiko, R. T., & Goldhaber, D. M. 1988, ApJ, 335, 774.