

**A FAR-INFRARED HETERODYNE SPECTROMETER
FOR AIRBORNE ASTRONOMY**

Thesis by
Erich N. Grossman

In Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy

California Institute of Technology
Pasadena, California
1988

submitted October 7, 1987

©1988

Erich N. Grossman

All Rights Reserved

Acknowledgements

My debts are many and varied. Firstly, I would like to express my deep gratitude to Tom Phillips, my faculty advisor. He has taught me a great deal of experimental physics, much of it not confined to the narrow sub-fields directly involved in my thesis project. Working together, I have often found myself taking an overly narrow view of some particular problem, while Tom, taking a much broader one, has been able to point out subtle analogies with other physical systems or alternative experimental strategies that make for progress. In short, he has taught me to generalize. The airborne spectrometer project, (in addition to being his idea originally), has also greatly profited from the presence of a viewpoint which was slightly removed from the furor of everyday details, which Tom provided.

Dan Watson has really been the prime mover of the project, making most of the basic design decisions, and supervising the details of the receiver's construction and testing. In particular, the far-infrared laser, without which none of the experiments described in this thesis would have been possible, was virtually entirely his work. He has also stepped in to act as my *de facto* advisor, explaining physics or suggesting the next thing for me to work on, whenever Tom was elusive or enigmatic. I have learned a vast amount from him, particularly the mundane but essential minutiae of cryogenics, mechanical engineering, and far-infrared optics, e.g. how to make polyethylene stick to quartz, when to use thread inserts, and the like. Also, I probably now face a long and weary period of sobriety, since, thanks to Dan, I have not had to buy my own drinks for several years now.

Several other individuals in the submillimeter group have also made important contributions to the receiver project. Pat Schaffer has built much of the electronics for the airborne instrument, and was a sorely needed source of fix-it skill, muscle power, and good humor on our trips to Mountain View. Jocelyn Keene built the mesh-on-dielectric low-pass filters used in most of our lab experiments and for the 150 μm receiver. She and Mike Wengler also did a considerable amount of work on upgrading the lab's Fourier transform spectrometer (originally constructed by Elliott Brown), which proved very useful in characterizing our stressed detector setup. She also adapted the software used for control of the InSb receiver on the KAO to our receiver, and assisted on the first KAO expedition, in which she remained an

level-headedness in an ocean of hysteria. Colin Masson was a valuable and dependable source of advice on YIG oscillator electronics, far-infrared optics, and other aspects of the receiver. Jeff Stern evaporated several planar mirrors and dichroics for the project, and has put up with my shortcomings as an office-mate for years longer than anyone should have to. Michael Wengler, in addition to his FTS work, has imparted a lot of insight to me in the course of many years of discussions about antennas, diodes, superconductors, quantum noise, and heterodyne detection generally. With Mike, however, such discussions rarely stop at the boundaries of experimental physics. At various times in the course of our long, and nearly coincident tenures at Caltech, he has gladly taken up the roles of colleague, confidante, motorcycle partner, paramedic, psychotherapist, literary critic, and procurer. How I ever shall adequately repay him I shudder to imagine.

Apart from our group at Caltech, the project owes a great deal to E. E. Haller and his colleagues at Lawrence Berkeley Laboratory, who have grown the germanium photoconductors used in our system as mixers. The bandwidth and gain characterization of our high compensation detectors, in particular, were done in collaboration with In-Shik Park, who grew a series of neutron-transmutation doped samples specially for these experiments. The Schottky diodes used in our sideband generator were fabricated by R. Mattauch's group at the University of Virginia. The latter also provided us with several important data about the diodes which were necessary in analyzing their performance. Finally, the physical construction of the receiver's components was no small part of the total work done on the project. For this I thank the staff of the Caltech Central Engineering Services, and Fernando Lopez and Larry Frazier of the Physics Shop for their workmanship and effort. Marty Gould, proprietor of the Zen Machine Shop, deserves special mention. He built our second-generation corner-cube and all our off-axis mirrors, and somehow managed to work within the often appalling deadlines we set him.

I am grateful to my fellow graduate students in Downs, past and present, including Elliott Brown, Geoff Blake, David Carico, Thomas Buettenbach, and Rick Edelson, for many interesting conversations on all manner of scientific subjects and for much simple good fellowship. Aside from my thesis project, my education in astrophysics has greatly profited from two other projects I worked on at Caltech. I am grateful to the Caltech Infrared Group, especially Gerry Neugebauer and Chas Beichmann, for whom

I worked my during my first two and a half years here, and who more or less introduced me modern observational astronomy. I also thank Nick Scoville, Anneilla Sargent, Colin Masson, and the staff of the Owens Valley Radio Observatory for tutelage in the practical aspects of millimeter interferometry, and for many discussions about the current observational understanding of star formation. The millimeter interferometrists contribute alot to the general élan with which research gets done in Downs. I also want to thank my friends in Bishop, Vicki and Justin Polcyn and the Woodys (all of 'em) for enlivining many a dull observing trip (or just random weekend) with days of windsurfing and skiing, mountainbiking and barbecues. Finally, my family, though geographically remote, has always been near in my thoughts, and their support and confidence in me has been a great comfort.

The long-suffering American taxpayers deserve a round of applause as well. On the airborne receiver project they have offered their support through NASA grant NAGW-107. During my first two years on the project, I was supported by a NASA graduate student researchers grant.

“Results ? Why man, I have gotten alot of results.
I know several thousand things that won't work.”

Thomas Alva Edison

Abstract

The design and construction of a novel heterodyne spectrometer for airborne astronomy in the $50 \mu\text{m} - 200 \mu\text{m}$ wavelength range is described, along with laboratory measurements of its performance. A bulk, extrinsic Ge:Ga photoconductor is used as the mixer. Its low bandwidth, determined by the hole recombination rate, necessitates the use of a continuously tunable local oscillator. This is provided by a far-infrared laser sideband generator, which is based on a GaAs Schottky diode mounted at the feed of a corner-cube antenna, the latter combination acting as a reflective FIR modulator.

The first chapter of this thesis describes the astronomical and technical context of the project – in particular, the constraints which the astronomical goals set on the instrument, and the advantages and drawbacks of each of the various broad instrumental strategies that are available for spectrometer design. The chapter's last section provides a very brief overview of our most successful laboratory results, which are described at greater length in chapters 2 – 4. In chapter two we describe the performance of Ge:Ga mixers as heterodyne mixers. We report on an extensive series of measurements of bandwidth, photoconductive gain, and direct detection responsivity for a series of highly compensated, NTD detectors grown specifically for this purpose. Chapter two also describes a number of experiments on FIR heterodyne performance, made using the direct, attenuated laser, rather than the output of the sideband generator, as the local oscillator. These confirm the expectation that germanium photoconductors are capable of quantum-limited noise performance with quantum efficiencies of $\sim 10\%$, at much lower LO powers than required for Schottky diodes. Our best achieved noise temperature is $T_N(DSB) = 655\text{K}$ at $P_{LO} = 1.6\mu\text{W}$, a factor of > 25 lower than the best reported corresponding figure for Schottky diodes.

Chapter 3 describes the operating principles and construction of our FIR laser, which formed a basic tool in nearly all our laboratory experiments. A brief discussion of the Lorenz instability in FIR lasers is also given, in connection with various observations we have made of spontaneous pulsations and excess low-frequency noise on the laser output, and which have recently been the subject of considerable study by other researchers. Chapter four describes FIR laser sideband generation using small-area Schottky diodes and corner-cube antennas. The construction and performance of our corner-cubes is outlined, including

the first direct measurement of the main beam efficiency of a corner-cube antenna in the FIR, and a comparison with theory. The construction and measured performance of the rest of the sideband generator is also described. A detailed, quantitative model has been developed for the conversion efficiency obtainable from Schottky diodes in this application. We find that the low conversion efficiency (-39 db) measured in our experiments, and comparable to that found by other researchers, is inherent in the diode and well predicted by the model. For our particular experiment, the model predicts -28 db loss due to the diode, plus approximately -10 db loss due to the antenna coupling efficiency. The dependence of conversion efficiency on diode parameters is studied and guidelines for future optimization derived. Unfortunately, the severe conversion loss we measure, combined with low FIR laser power and (somewhat less significantly) poor optics transmission, leads to our presently available LO power being inadequate to obtain astronomically useful sensitivity, by a large factor.

Table of Contents

Acknowledgements	iii
Abstract	vii
List of Figures	x
Chapter 1 – Introduction	1
1.1 Atmospheric transmission	2
1.2 Astronomical motivation	5
1.3 General instrumental strategies	11
1.4 Overall design of our instrument	19
1.5 Engineering formalism for description of Ge:Ga heterodyne performance	25
1.6 Highlights of our results	31
Chapter 2 – Ge:Ga Photoconductors as Heterodyne Mixers	37
2.1 Theory	38
2.2 Experimental details	54
2.3 Experimental results	72
Chapter 3 – Far-infrared laser	107
3.1 Basic theory	108
3.2 Construction details	115
3.3 FIR output : power and beam patterns	124
3.4 The Lorenz instability : spontaneous pulsations and deterministic chaos in FIR lasers	130
Chapter 4 – Sideband Generator	142
4.1 Corner-cube antennas	143
4.2 Theory of Schottky diodes as FIR modulators	157
4.3 Polarizing Michelson interferometer optics	183
4.4 Sideband generation results	190
References	200

List of Figures

1.1 Atmospheric FIR transmission	3
1.2 Functional block diagram of the receiver	20
1.3 Typical I-V curves of Ge:Ga photoconductors	27
1.4 Contributions to IF output power spectral density	30
1.5 Best reported heterodyne noise performance of various FIR and submillimeter mixer technologies .	36
2.1 Simplified band structure of germanium; schematic of extrinsic photoconduction	39
2.2 Theoretical and measured compensation dependence of mobility	44
2.3 Incorporation of Ge:Ga photoconductors into complete receiver system	55
2.4 DC bias circuit	56
2.5 Design and performance of cryogenic GaAsFET IF preamplifier	62
2.6 Mixer block designs for unstressed and stressed detectors	67
2.7 Photoconductivity spectra obtained with and without stress on detector LBL 82-4.6	69
2.8 Measured Ge:Ga detector noise spectra with laser illumination	80
2.9 Breakdown field versus compensation	82
2.10 Recombination bandwidth versus bias field, showing carrier heating	84
2.11 Recombination bandwidth versus minority impurity compensation	85
2.12 Direct detection responsivities of the 729-series detectors	87
2.13 Non-Lorentzian excess detector noise apparent at biases near breakdown	91
2.14 IF spectrum of detector 82-4.6, as used in the airborne system	94
2.15 G.R. noise spectrum of Eagle-Pitchard detector	96
2.16 Experimental setup for measurement of heterodyne noise temperature	98
2.17 Bandwidth determinations for detectors 712-7.2(NTD) and 108-17.7	100
2.18 Heterodyne tuning curves for detector 108-17.7, showing our best achieved noise temperature	104
3.1 FIR laser cycle	108
3.2 Gain spectrum of the laser transition	112
3.3 Lowest order modes of a highly overmoded waveguide	116

3.4	Polarization behavior of FIR laser output	119
3.5	Mechanical design of the fir-infrared cavity	120
3.6	Design and performance of the optical isolator	121
3.7	Optoacoustic laser lock	123
3.8	Physical layout of airborne laser system	125
3.9	Histogram of measured 119 μm laser power levels	128
3.10	Far-field intensity pattern of the laser at 119 μm	129
3.11	1-D scans of laser intensity pattern at various wavelengths	130
3.12	Regions of stability and instability of the Lorenz equations in the b-r plane	139
3.13	Spontaneous laser pulsations observed on various FIR laser lines	141
4.1	Corner-cube antenna (first version)	144
4.2	Geometry of Schottky diode chip and whisker	145
4.3	Zen corner-cube design	146
4.4	Schematic near, transition, and far-field antenna patterns, and definition of coordinate systems	148
4.5	Theoretical beam patterns for two different whisker geometries	150
4.6	Setup for corner-cube beam measurements	154
4.7	Theoretical and experimental beam patterns of the first corner-cube design at 214 μm	155
4.8	Main beam of the corner-cube at 119 μm	157
4.9	Energy-band structure of the Schottky diode	159
4.10	Theoretical bias dependence Schottky video responsivity at three different wavelengths	162
4.11	Laboratory setup for and results from experimental determination of V_{FB}	166
4.12	Equivalent circuit representation of the FIR modulation process	166
4.13	Reflectivity modulation in the low frequency limit	170
4.14	Reflectivity modulation for the baseline 1E12 diode at 2.52 THz	171
4.15	Reflectivity modulation for three different epilayer dopings	174
4.16	Reflectivity modulation as function of FIR frequency	175
4.17	Reflectivity modulation at three temperatures	177

4.18 Single-sideband diode conversion efficiency as a function of epilayer doping and diode radius	181
4.19 Optical configuration for the polarizing Michelson interferometer	184
4.20 First measurement of retardation “waveplating” due to wire grid polarizers	186
4.21 Second measurement of retardation due to wire grid polarizers	187
4.22 Scans of the corner-cube video signal as its orientation is scanned	190
4.23 Laboratory setup for sideband generation	192
4.24 First Fabry-Perot spectra taken of the sideband generator output, illustrating audio homodyne effect	194
4.25 FIR laser sidebands detected with ultra-high sensitivity detector LBL 583-4.6	195
4.26 Schottky bias dependence of sideband power	197

Chapter 1 – Introduction

Over the last several years, a number of efforts have been undertaken to extend the techniques of high resolution millimeter-wave spectroscopy to ever higher frequencies, motivated largely by the needs of the astronomical community. At the same time, interest among infrared astronomers in extending their spectroscopic techniques to longer wavelengths and higher spectral resolutions has also increased. However, the far-infrared (FIR) wavelength region (which we shall somewhat arbitrarily take to mean the range from $30\ \mu\text{m}$ to $350\ \mu\text{m}$.) presents a number of serious obstacles to spectroscopists, some fundamental and some practical. Both in terms of technological capabilities and in terms of astronomical exploitation, the far-infrared has largely remained *terra incognita* to this day.

The project on which this thesis is based had as its aims, firstly, the development of a new far-infrared receiver technology based on Ge:Ga photoconductive mixers and tunable laser sideband local oscillators, and secondly, the performance of preliminary astronomical observations with an airborne spectrometer which implemented this technology. Compared to competing techniques based on Schottky diode mixers, this technology is at a much earlier and more experimental stage of development. It is also considerably more complex. However, it has the potential for greatly improved noise performance. Indeed, one of the main results of our work has been the (first) demonstration of quantum-limited heterodyne reception in the far-infrared, with typical quantum efficiencies of $\sim 10\%$. This thesis reviews the laboratory work we have done and the improved current understanding that has been gained of the capabilities and limitations of this technology. It also describes the spectrometer we have built for astronomical observations from the Kuiper Airborne Observatory (KAO). Before addressing the scientific issues, however, it must be stated at the outset that the final aim of the project, namely spectroscopic observations of an astronomical source using the airborne receiver, has not been achieved. As will be made clear in this and succeeding chapters, the fundamental obstacle is production of sufficient local oscillator power by the sideband generator.

The purpose of this first chapter is to provide an introduction to the overall design of our receiver and to describe the astronomical, atmospheric, and instrumental considerations which motivated the design. It is also intended to provide an overview of the substantive results we have obtained in our laboratory

experiments, and which are described in detail in chapters 2 – 4. In order to properly describe these results, however, some background is necessary. We therefore devote a section of this introductory chapter to an exposition of the empirical, “engineering” formalism commonly used to describe photoconductor performance. The formalism used for the description of noise in heterodyne spectroscopy is also described. The last section then covers the highlights of our laboratory results, including extensive measurements of bandwidth, gain, and heterodyne noise performance of the Ge:Ga photoconductors, quantitative modeling of the performance of FIR laser sideband generators, and comparison with our detailed experimental results.

1.1 Atmospheric transmission

By far the most important constraint imposed on any instrument intended for far-infrared astronomical applications is due to the transmission of the atmosphere. When viewing from the ground, the Earth’s atmosphere is to all intents and purposes completely opaque between the “windows” at $30\ \mu\text{m}$ and $350\ \mu\text{m}$. (Indeed, this is the motivation for our definition of “far-infrared” as $30\ \mu\text{m} < \lambda < 350\ \mu\text{m}$.) Most of the absorption is due to transitions of water vapor, ozone, and molecular oxygen, and their isotopic variants. Transitions due to CH_4 , N_2O , CO_2 , CO , and a multitude of trace gases produce a very small (in the far-infrared) amount of absorption, (McClatchey *et al.* 1973, Traub and Stier, 1976). Water is heavily concentrated in the lower layers of the atmosphere, and therefore has the widest (pressure-broadened) absorption lines. For this reason, both mm-wave astronomers trying to push to shorter wavelengths and infrared astronomers trying to push to longer wavelengths go to great lengths to perform their observations from the highest and driest sites available.

The upper curve in figure 1.1 shows the calculated atmospheric transmission, under exceptionally good conditions, from the best easily accessible terrestrial site, the summit of Mauna Kea in Hawaii. The calculation was done using the line parameters of the standard AFCRL line compilation (Rothmann 1983, McClatchey *et al.* 1972), the same atmospheric parameters (e.g. effective temperature and pressure, and isotopic abundances for each species) as Traub and Stier (1976), and assuming Lorentzian lineshapes. Longward of $30\ \mu\text{m}$, Doppler broadening is negligible compared to collisional broadening for ground-based (and even airborne) observations, so the approximation of Voigt profiles by Lorentzians is very good. (At

Atmospheric Transmission at Zenith

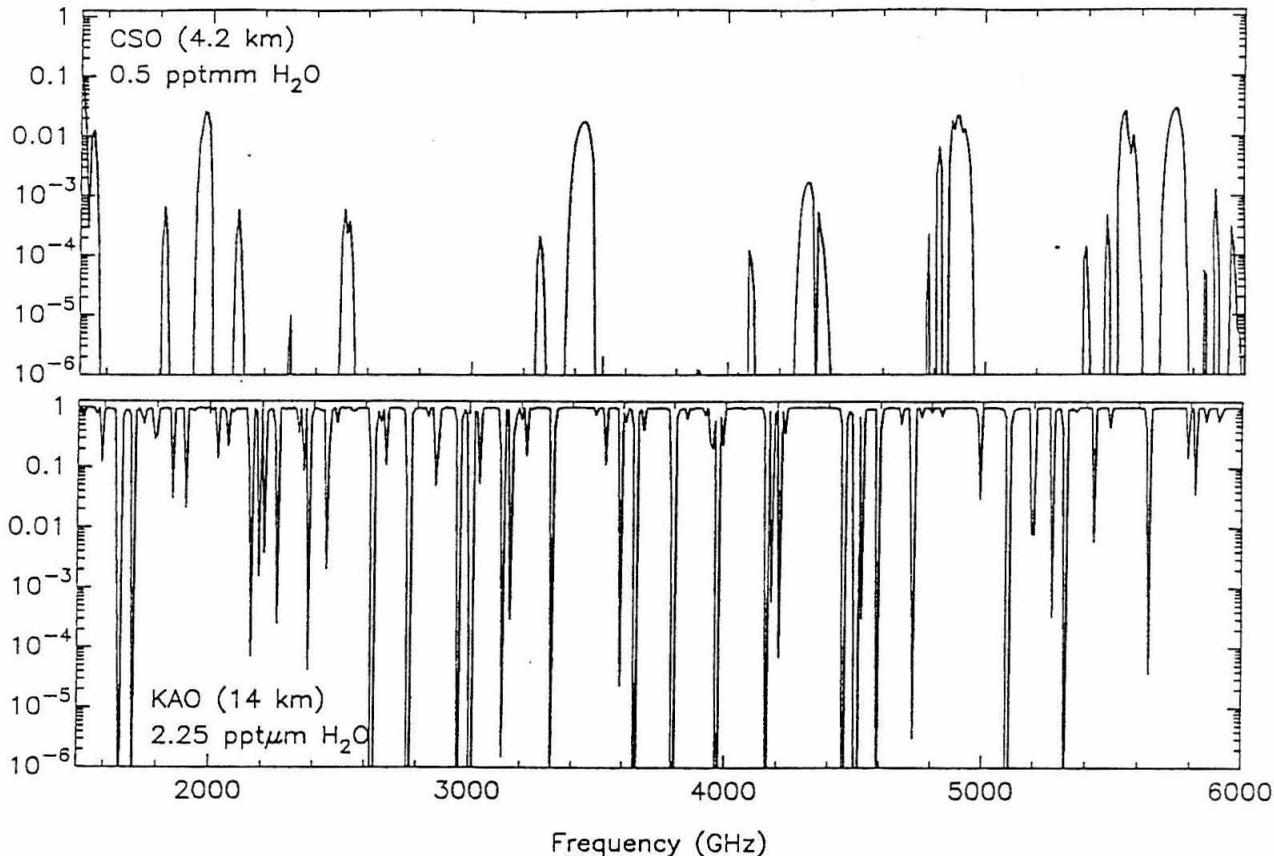


Figure 1.1 – Atmospheric transmission in the far-infrared from elevations of (top) 4.2 km, applicable to the summit of Mauna Kea, and (bottom) 14 km, (46,000 ft.) applicable to the flight ceiling of the KAO. Assumed zenith angle is 0° , and assumed water vapor column densities are (top) $1.7 \times 10^{21} \text{cm}^{-2}$ or 0.5 mm of precipitable H_2O , and (bottom) $7.3 \times 10^{18} \text{cm}^{-2}$ or 2.25 μm of precipitable H_2O .

shorter wavelengths and higher altitudes this is not true.) A simple Lorentzian lineshape is known to be somewhat in error in the distant wings of atmospheric lines; however, the correct lineshape is not known. Indeed, even with the approximation of a single effective effective temperature and pressures for each species, the determination of the true collision-broadened lineshape applicable to atmospheric lines is a complicated molecular dynamics problem, and still the subject of ongoing research. Burch (1968) showed that a Lorentzian, a Van Vleck – Weisskopf lineshape, and a Zhevakin – Naumov lineshape all significantly

disagree with experiment over some range of frequency and/or pressure. The collisional lineshape depends on the collision partner (Bignell, Saiedy, and Sheppard, 1963, Winters, Silver, and Benedict, 1964). A Lorentzian is a good first approximation, however, and can be shown on general grounds to be an upper limit on the absorption in the distant line wings.

It may be seen from figure 1.1 that, even for exceptionally good atmospheric conditions, useful amounts of atmospheric transmission cannot be obtained anywhere in the far-infrared. FIR astronomers must therefore resort to either spacecraft, balloon payloads, or airplanes as platforms for telescopes and instruments. Of these, airplanes are the most suitable for new and experimental instrument technologies. The primary facility for airborne astronomy is the Kuiper Airborne Observatory, a C-141 aircraft modified by NASA to house a 91 cm diameter, Cassegrain telescope. It is based at Moffett Field, California. The usual observing altitude is 12.5 km (41,000 feet). A fundamental limitation of such an observing platform is that, even at this altitude, residual atmospheric absorption is by no means negligible. (See the lower curve of figure 1.1.) Moreover, even when the transmission at the precise observing frequency is adequate, it is possible for the wings of strong nearby absorption lines to produce such irregular baselines that accurate spectroscopy is impossible.

Aside from this fundamental limitation, there are some practical drawbacks to observing on the KAO. Due to limited flight range and various other constraints on flight planning, total integration times on a given astronomical object cannot in practice exceed ~ 3 hours. Also, compared to a terrestrial observatory, the KAO is a very demanding environment in terms of vibration, acoustic noise, noise and grounding of electrical power supplies, and radio-frequency interference. Furthermore, there are strictly enforced limits on the weight, size, tipping moments, and methods of (physical) construction of instruments to be mounted on the telescope. Nonetheless, in the decade since its commission, a wide variety of far-infrared instruments have been constructed or adapted for operation on the KAO. The potential astronomical rewards are so great that the effort required to deal with the constraints and drawbacks of airborne observing are more than justified.

1.2 Astronomical motivation

Many spectral lines of great astrophysical importance lie in the FIR wavelength range from 50 μm to 200 μm . Foremost among these are the fine structure lines of several abundant species of atoms and atomic ions. In particular, our instrument was designed with the $^2P_{3/2} \rightarrow ^2P_{1/2}$ transition of singly-ionized carbon [CII] at 157 μm and the $^3P_2 \rightarrow ^3P_1$ and $^3P_1 \rightarrow ^3P_0$ transitions of neutral oxygen at 63 μm and 145 μm , respectively, in mind. Table 1.1, taken from a recent review, (Watson 1985), lists some of the other analogous lines of less abundant atomic species that also exist in this wavelength range. In addition to the atomic fine structure lines, there are also a great many low-lying rotational transitions of simple molecules that lie between 50 μm and 200 μm . We particularly mention the ground-state rotational transitions of the hydrides of oxygen (OH, $^2\Pi_{3/2} J = 5/2 \rightarrow 3/2$, 119 μm), carbon (CH, $^2\Pi_{3/2}(J = 3/2) \rightarrow ^2\Pi_{1/2}(J = 1/2)$, 149 μm), and deuterium (HD, $J = 1 \rightarrow 0$, 112 μm), and the ladder of rotational transitions of CO. It is beyond the scope of this thesis to comprehensively review the astrophysical significance and the interpretation of the intensities of these FIR transitions. Rather, the reader is referred to several recent reviews (Watson 1985, Watson 1984, Watson 1982, Phillips 1986, Genzel and Stacey 1985). The nature of the astrophysical sources does impose constraints on spectrometer design, however. Therefore, some astrophysical background is necessary in order to understand the motivation for our instrument's design.

The main astrophysical sources of fine-structure emission from low-excitation species, including [CII] and [OI], were long thought to be large regions of cool, atomic, relatively diffuse interstellar gas. (Note that, because the ionization potential of carbon is less than that of hydrogen, $11.3 < 13.6$ eV, it can be ionized by UV radiation soft enough to easily penetrate atomic regions ($912 \text{ \AA} < \lambda < 1102 \text{ \AA}$). Thus, the dominant ionization state of carbon in such regions is singly-ionized.) In some cases, these regions are distributed as sheaths or envelopes existing at the boundaries of denser, colder, molecular clouds: "photodissociation regions" (Tielens and Hollenbach 1985a, 1985b) fall into this category. In other cases, these cool, atomic regions coincide with the diffuse HI clouds that have been extensively studied in the 21 cm hyperfine line of atomic hydrogen. (Kulkarni and Heiles, 1986). The relative contributions of the various types of object to the total [CII] and [OI] emission of a galaxy are not known. The approximate proportionality found between

Table 1.1 – FIR astronomical lines (instrumental targets)

Species	Transition	Wavelength (μm)	n_{crit} (cm^{-3})	Astronomical detection
[CII]	$^2P_{3/2} \rightarrow ^2P_{1/2}$	157.741	3.0×10^3	Russel <i>et al.</i> 1980
[OI]	$^3P_0 \rightarrow ^3P_1$	145.527	1.5×10^5	Stacey <i>et al.</i> (1983)
[OI]	$^3P_1 \rightarrow ^3P_2$	63.1837	9.8×10^5	Melnick, Gull; and Harwit (1979)
[NII]	$^3P_1 \rightarrow ^3P_0$	203.9	4.8×10^1	not yet detected
[NII]	$^3P_2 \rightarrow ^3P_1$	121.889	3.1×10^2	not yet detected
[SiI]	$^3P_1 \rightarrow ^3P_0$	129.682	–	not yet detected
[SiI]	$^3P_2 \rightarrow ^3P_1$	68.474	–	not yet detected
[OIII]	$^3P_1 \rightarrow ^3P_0$	88.356	5.1×10^2	Ward <i>et al.</i> (1975)
[OIII]	$^3P_2 \rightarrow ^3P_1$	51.816	3.6×10^3	Melnick <i>et al.</i> (1978)
CH	$^2\Pi_{3/2}(J = 3/2) \rightarrow ^2\Pi_{1/2}(J = 1/2)$	149.09, 149.39 [†]	–	Stacey, Lugten, and Genzel (1986)
HeH ⁺	$J = 1 \rightarrow 0$	149.1373	–	not yet detected
OH	$^2\Pi_{3/2} J = 3/2 \rightarrow 1/2$	119.441, 119.234 [†]	–	Watson (1982)
HD	$J = 1 \rightarrow 0$	112.075	–	not yet detected

[†] Lambda doublet components (hyperfine structure ignored this table.)

[CII] and CO, $J = 1 \rightarrow 0$ integrated line intensities (Watson *et al.* 1985) suggests that photodissociation regions associated with molecular clouds are dominant, but the recent discovery of strong [CII] emission from dense, cold, molecular regions leaves the issue unsettled. Diffuse clouds have not yet been observed in the FIR fine structure lines.

The importance of these fine structure lines arises from the fact that, over a wide range of physical conditions, they provide the dominant cooling mechanism for cool, atomic interstellar gas (Dalgarno and McCray, 1972). A steep increase in the cooling function (cooling rate in erg/s-cm^{-3} as a function of temperature) at a particular temperature tends to cause a disproportionately large fraction of the interstellar medium to equilibrate at that temperature, and leads us to speak of a “phase” of the interstellar medium. For

example, the steep increase in the cooling function near 10,000 K due to population of excited electronic states of heavy elements by electron collisions (Spitzer, 1978, p. 133 ff.) leads to the fact that the range of electron temperatures in HII regions is so narrow, viz. some 6,000 – 12,000 (Turner and Mathews, 1984, Spitzer 1978, p. 80). Similarly, the increases in the cooling function due to excitation of the excited fine structure level of CII at $T \sim \Delta E/k = 92 K$ and of OI, $63 \mu\text{m}$ at $T \sim \Delta E/k = 230 K$ by atomic H collisions lead to a clustering in the temperatures of neutral diffuse regions in the neighborhood of 100 K. The low-excitation fine structure lines therefore are the natural spectroscopic “probes” of the diffuse atomic phase of the interstellar medium in much the same sense that the visible forbidden lines of heavy elements (e.g. [OIII] 5007 Å, 4958 Å, and 4363 Å, [SII] 6716 Å and 6730 Å) are natural probes HII regions, or the low-lying rotational transitions of CO are the natural probes of the molecular phase.

The [CII] and [OI] fine structure lines were first detected several years ago (Russell *et al.*, 1980 for [CII], Melnick, Gull, and Harwit, 1979 for [OI] $63 \mu\text{m}$). Since the spectral resolution available from then-current spectrometers was inadequate to resolve the lines from cool, diffuse regions, no kinematic studies have been made using them. Their main astrophysical application has been the use of integrated line intensities to derive densities and temperatures for the emitting gas. In atomic regions, these transitions are primarily excited by collisions with neutral hydrogen atoms. At high enough densities, the collisional de-excitation rate of the upper level becomes comparable to the spontaneous radiative de-excitation rate (Einstein A-coefficient), and the level populations are said to be “thermalized”. As shown in table 1.1, this occurs at a much lower density for [CII] than [OI]. Thus the ratio of fluxes $I(157\mu\text{m})/I(63\mu\text{m})$ can be, and has often been, used as an accurate measure of density over a fairly large range, approximately $10^3 - 10^6 \text{ cm}^{-3}$. Similarly, it turns out that the intensity ratio of the two [OI] lines, $I(63\mu\text{m})/I(145\mu\text{m})$, is relatively insensitive to density, and forms a good measure of temperature over the range 100 – 500 K (see Watson 1985). The optical depth manifests itself in the absolute intensity of the lines. This interpretation of line ratios assumes all optical depths are $\ll 1$, an assumption which has recently been found to be invalid in some sources. However, the optically thin case is still useful as an illustration of the information available in the fine-structure line intensities.

What are the implications of these astrophysical considerations on spectrometer design ? Firstly, because the temperature is so low, the thermal linewidth will be very small; $T = 100$ K corresponds to only $(kT/m)^{1/2} \sim .25$ km/s for carbon. This sets the minimum frequency scale which would ever be of astrophysical interest. Because the thermal linewidth is so low, the lines are expected to be primarily Doppler-broadened by the large-scale velocity structure of the source. The characteristic scale of this velocity structure is not known. Indeed, it is one of the obvious subjects of study for a very high resolution spectrometer. On the basis of comparison with molecular clouds, however, which are also primarily broadened by large-scale velocity gradients, one expects linewidths of $\sim 1 - 5$ km/s. In the $200 - 50 \mu\text{m}$ range of our instrument, 1 km/s corresponds to $5 - 20$ MHz. For determination of accurate baselines, a frequency coverage of at least ~ 3 times the linewidth is generally required. This leads to a requirement on the spectral coverage of our instrument of from ~ 15 MHz for observations of the longest wavelength lines from the most quiescent sources, to ~ 300 MHz for observations of short wavelength lines from somewhat more active sources.

A second implication for instrument design is that, since the densities of the emitting regions of interest are usually high enough to thermalize [CII], ($n > n_{crit}$, see table 1.1), the brightness temperature of the $157 \mu\text{m}$ line (in the Rayleigh-Jeans approximation) is simply $T_B = T(1 - e^{-\tau})$, where τ is the optical depth. Thus, the brightest sources are expected to have brightness temperatures nearly equal to the kinetic temperature, namely ~ 100 K. For subthermally excited lines, or for sources of lower density or optical depth, or for sources significantly cooler than $h\nu/k$, brightness temperatures will be lower.

The signal actually available at the input port of a spectrometer is described by the antenna temperature, T_A , which is equal to the brightness temperature only for sources with large enough angular sizes to fill the telescope beam. (In fact, the complete relationship between T_B and T_A , including the proper corrections for forward and backward spillover, source coupling efficiency, etc. is somewhat more complicated – see Kutner and Ulich, 1981 – but it does not affect the argument here.) The beamsize of the KAO is diffraction-limited in the far-infrared, and is thus proportional to λ/D , where $D = 91$ cm is the primary mirror diameter, with a constant of proportionality of order unity. The value of the constant depends on the

precise definition of beamsize and the particular aperture illumination chosen. The optical configuration chosen for our receiver has, in a Gaussian beam formalism, (Goldsmith, 1982), an edge taper of 13 db, a fairly standard value. This leads to a FWHM beam diameter of $\theta_{FWHM} = 42.0'' \left(\frac{\lambda}{157\mu\text{m}} \right)$. Even at fairly substantial galactic distances (say, ~ 1 kpc) most of the diffuse atomic envelopes of molecular clouds (Russel *et al.*, 1980) and the diffuse HI clouds (Kulkarni and Heiles, 1986) have substantially larger angular extents than this.

In short, the primary target of our instrument, namely low-excitation, atomic fine structure emission from cool, diffuse, HI clouds, photodissociation regions, and molecular cloud haloes, involves measurement of selected ~ 100 MHz wide stretches of spectrum at various frequencies throughout the $50 \mu\text{m} - 200 \mu\text{m}$ range. The minimum channel width of interest is given by the thermal linewidth of the source, and is numerically $\sim 1.25 - 5$ MHz. We, of course, desire the maximum sensitivity possible, but at the very least, the minimum detectable brightness temperature must be ≤ 10 K to achieve reasonable signal-to-noise ratio on even the brightest sources. Available integration times are 3 hours or less.

For some of our instrument's secondary targets, these numbers could differ. The main astrophysical source of atomic fine structure emission from the high-excitation species in table 1.1 (e.g. [OIII], [NIII]) are galactic HII regions. These typically consist of an early-type star whose copious output of Lyman continuum photons ($\lambda \leq 912 \text{ \AA}$) ionizes the interstellar medium out to some maximum, "Strömgren", radius, where the photoionization just barely balances the recombination. They have been extensively studied in the radio continuum and in optical recombination lines for decades. Compared to the optical lines, the far-infrared line intensities have the advantages of less sensitivity to electron temperature and lack of reddening correction. The [NIII] $57 \mu\text{m}$ and [OIII] $88 \mu\text{m}$ and $52 \mu\text{m}$ lines have been used to study nitrogen and oxygen abundances in many individual HII regions, allowing the dependence of N/O on galactic radius to be determined (Lester *et al.*, 1986). The line ratio for ions with two fine-structure transitions, e.g. [OIII], is a good measure of electron density. As far as instrumental requirements go, the main difference between these lines and the low-excitation fine structure lines is the increased thermal linewidth. At the typical temperatures of 10^4 K, the thermal linewidth is an order of magnitude greater than

for cool HI regions, and the minimum channel width of interest is unlikely to be less than 15 MHz. Thus, for these regions, the incentive to build an ultra-high resolution (heterodyne) spectrometer is somewhat less than for cooler sources.

The FIR molecular lines listed in table 1.1 have a wide variety of applications in the study of the kinematics and chemistry of molecular clouds. In some cases (e.g. NH_3 , OH), there exist low-energy transitions of the same molecules due to inversion or lambda-doubling, which alone have generated whole subfields of radio astronomy. The FIR pure rotational transitions can provide new information for these studies. The high-J transitions of CO have been used (Watson 1982) to probe regions of shock-excited gas produced by high-velocity outflows embedded deep within molecular clouds. The rotational transitions of the light hydrides can be used to supplement the kinematic studies of clouds made in mm-wave transitions of heavier molecules. They can also be used to derive molecular abundances, which may then be compared with the predictions of extensive chemical models (e.g. Prasad and Huntress, 1980, 1982) that have been developed on the basis of observations of heavier, less abundant species whose transitions lie in the mm-wave region. In general, the instrumental requirements for observing the molecular lines are similar to those which apply to the low-excitation fine structure lines.

The case of HD is somewhat special. Its abundance is uniquely important, because the molecular component of the cosmic deuterium abundance is the component which is by far the most poorly known. The cosmic deuterium abundance is, in turn, one of the few available tests of primordial nucleosynthesis models. Previous efforts to determine the HD abundance in molecular clouds via UV absorption line observations (York and Rogerson 1975) have been confined to the handful of clouds with strong background stars and suitable amounts of extinction. Studies based on abundance ratios (determined from mm-wave line intensities) of heavy molecules with their deuterated counterparts are highly dependent on the modeling of the cloud chemistry to separate out the effect of chemical fractionation. Measurements of the intensity of the $J = 1 \rightarrow 0$ rotational line would be a much more reliable and direct technique. Despite extensive efforts with relatively high resolution (~ 350 MHz) incoherent spectrometers, the line has not yet been detected. The problem is not so much one of instrumental signal-to-noise as it is a matter of highly irregular baselines

and blending due to the presence of two moderately strong atmospheric lines of HDO and H₂¹⁷O, only 660 MHz below, and 670 MHz above, the HD line at 2.6768 THz. Aside from space-based spectrometers, the best hope for detecting this important line is increased spectral resolution.

1.3 General instrumental strategies

The question of whether, at any given frequency and resolution, coherent (i.e. heterodyne) or incoherent spectroscopy is more sensitive is in general somewhat complicated, and depends on the assumptions made regarding detector and pre-amplifier noise, number of channels simultaneously observed, background temperature, and background emissivity (see Phillips and Watson 1985 for a calculation in the case of LDR). However, for resolutions sufficient to study the narrowest velocity structure of interest in the lines from cool clouds, namely $\nu/\Delta\nu \sim 10^6$, heterodyne detection is always superior. This is true despite the facts that a single-detector heterodyne receiver is sensitive to only one quadrature of the incoming electric field (namely that in phase with the local oscillator), and to only one polarization (namely that parallel to the LO). The reason for this superiority of coherent detection is that, above a certain resolution, there is in practice a fairly direct tradeoff between resolution and throughput for the optical filters (e.g. Fabry-Perot interferometers) used in incoherent spectrometers.

The principle of heterodyne detection is to add, or “mix”, the weak signal beam to be detected with the much stronger beam from a local oscillator (LO) at a nearby frequency. The detector element, or “mixer”, produces an output signal – in our case a photocurrent – which, at the signal and LO frequencies, is non-linearly related to the incident field. The non-linearity produces signals at the second harmonics of the signal and LO frequencies, at the sum and difference frequencies, and, depending on the non-linearity, possibly at higher order rational combinations of the signal and LO frequencies. The detector’s response is not sufficiently fast to follow the signals at the sum and second harmonic frequencies, either because of intrinsic physical mechanisms or because of the existence of small parasitic impedances at high frequencies. The standard terminology from radio engineering is to speak of the input signal wave as the “RF” (radio frequency) signal, and the difference frequency wave as the “IF” (intermediate frequency) signal. The IF is low enough that standard electronic amplifiers, filters, oscillators, etc. operate, and thus,

for mm-wave and higher frequency receivers, $\nu_{IF} \ll \nu_{RF}$.

Before describing the various broad design strategies that are available for construction of high frequency heterodyne receivers, it is useful to review the formalism for specifying instrument performance in heterodyne spectroscopy. The basic result is the Dicke equation (Dicke 1946), which states that the signal-to-noise ratio obtained in the measurement of a signal of radiation temperature T_S is

$$S/N = C \frac{T_S}{T_N} \sqrt{\Delta\nu T}, \quad (1.1)$$

where T_N is the instrumental noise temperature, $\Delta\nu$ the pre-detection, i.e. channel, bandwidth, and T the total integration time. C is a numerical factor of order unity which depends on the details of the source and instrumental noise spectra at the IF frequency (Tiuri 1964). The factor of $\sqrt{\Delta\nu T}$ accounts for the fact that measurements of the IF signal made at times separated by $\Delta t > (2\Delta\nu)^{-1}$ may be considered, in the sense of the Nyquist sampling theorem, statistically independent. Thus, $2\Delta\nu T$ is the number of statistically independent measurements of the IF signal made in the course of the integration time T . Note that the equivalent noise temperatures discussed here, and throughout this thesis, are Rayleigh-Jeans equivalent temperatures. That is, they are simply a shorthand notation for power per unit bandwidth, in units of temperature. An alternative convention used by some workers in the field is to quote the true physical temperature of a blackbody that would radiate the same power per unit bandwidth. The two conventions are equivalent if $h\nu \ll kT$, i.e. in the Rayleigh-Jeans limit. For near quantum limited detectors at sufficiently high frequencies, such as ours, this is a non-negligible correction (see §2.3).

Equation 1.1 is the basis for comparison of the performance of receivers with narrowband versus wideband IF's. To illustrate, consider two instruments of equal noise temperature, which are used to observe the same spectral line with the same resolution $\Delta\nu$. One receiver is assumed to be intrinsically limited in its IF bandwidth B , so that it can only obtain sensitivity T_N on one such channel at a time, i.e. $B = \Delta\nu$. The other is assumed to have a broader IF bandwidth, $B' = N\Delta\nu$. Using an appropriate "backend", or IF spectrometer, such as an array of N filters and detectors, the latter is thus able to observe N channels simultaneously. By what factor is this latter spectrometer superior to the former? The answer depends on B_s , the bandwidth of the source spectrum that is of interest (i.e. the source linewidth plus

required baseline). If $B_s \geq B'$, then for each stretch of spectrum of width B' , the broadband instrument can spend N times as much integration time on each channel as the narrow system. Thus, the overall signal-to-noise of the spectrum measured by the broadband instrument is $N^{1/2}$ times higher. If $B_s < B'$, however, some fraction of the broadband system's bandwidth is wasted by integrating at frequencies that do not contain interesting emission. The broadband system's advantage in signal-to-noise ratio is then only $(B_s/B)^{1/2} < N^{1/2}$. The improved signal-to-noise ratio for broadband spectrometers is sometimes referred to as the "multiplex advantage". Thus, if the comparison of two real receivers were really as described above, then, of course, no one would ever build the narrowband spectrometer. In practice, however, narrowband spectrometers often have much lower noise temperatures than wideband ones, in which case the multiplex advantage of the latter can be offset, or even more than offset, by the ratio of noise temperatures. The appropriate figure of merit in comparing the two is $T_N^{-1} B^{1/2}$ for observations of sources broader than either system's bandpass, and simply T_N^{-1} for systems narrower than either system's bandpass (i.e. for single-channel observations). For sources of intermediate width, the ratio of the wideband system's signal-to-noise ratio to the narrowband's is $\frac{T_N}{T'_N} \left(\frac{B_s}{B}\right)^{1/2}$, where (un-)primed quantities refer to the wide(narrow)band system.

Considering the fact that the bandwidths of interesting sources vary a considerable amount, the above distinction between wide and narrowband systems is somewhat vague. There are three somewhat more specific distinctions which may be used to classify high frequency receiver designs. The local oscillator may be fixed in frequency (or only step-tunable) and the IF bandpass wide enough to accommodate a full source spectrum, or a scanning local oscillator may be used with a narrowband IF. Practically speaking, the gist of the distinction is in the nature of the LO. In addition, the detector element may be either a bulk mixer or an antenna coupled device. Finally, the mixers may be classified (for lack of better words) as either photon, i.e. "power", detectors or "diode-like" detectors. The core of the last distinction is really the presence of a low-frequency cutoff for power detectors, and its absence for diode-like detectors. At any given frequency, there are various advantages and drawbacks to all of these possibilities.

The easiest distinction to understand is that between bulk and antenna-coupled mixers. For visible

and infrared wavelength heterodyne receivers, bulk mixers are the rule. Only one heterodyne receiver has been built for astronomical spectroscopy in this wavelength range, a 10 μm spectrometer based on HgCdTe photodiode mixers (Betz 1972). The many receivers which have been built for heterodyne light-wave communication in the visible and near-IR (see Kazovsky 1986, and references therein) likewise employ photodiode or (intrinsic) photoconductor mixers with physical dimensions $\gg \lambda$. At millimeter and submillimeter frequencies (say $\nu < 800$ GHz), antenna-coupled mixers are nearly universal. The mixers themselves are most commonly either SIS (superconductor-insulator-superconductor) tunnel junctions (Wengler 1987, Phillips and Rutledge 1986, Phillips and Woody 1982), or small-area Schottky diodes (Held and Kerr 1978a, 1978b). Their typical linear dimensions are much less than a wavelength, and the incident electric fields are coupled onto the mixer either by suspending it across a waveguide or by placing it at the feed of some kind of antenna. The main problem involved in extending antenna-coupling structures to higher frequencies is simply the difficulty of fabricating waveguides and antennas with the necessary accuracy ($\ll \lambda$) in physical dimensions. Furthermore, the performance of both SIS junctions and Schottky diodes is degraded at very high frequencies due to the basic device physics. Nonetheless, the primary efforts to apply heterodyne techniques to the submillimeter and FIR have concentrated on extending the low-frequency (i.e. mm-wave) approach upward in frequency, and Schottky diodes have become the standard devices for pioneering ever higher frequencies. Indeed, concurrently with our project, there have been three other efforts to construct far-infrared heterodyne receivers and apply them for the first time to astronomical spectroscopy, (Betz and Zmuidzinas 1984, Röser *et al.* 1986, Chin 1987), and all three employ antenna-coupled Schottky diodes as mixers (indeed, all three use the identical type of antenna, a longwire in corner-cube). Very recently, one of these instruments has succeeded in obtaining astronomically useful spectra from the KAO (Betz and Zmuidzinas 1987).

The performance of Schottky diodes is much degraded above a few hundred GHz, primarily because of parasitic capacitance which, even for the smallest area devices which can be fabricated, greatly dominates the barrier conductance at high frequencies. The best laboratory performance which has been reported for a Schottky diode receiver at the frequencies of interest to us is 17,000 K (single-sideband) at 119 μm . The aim of our project was to investigate an entirely different strategy, one which was suggested

by the excellent performance achieved by single crystal, extrinsic Ge:Ga photoconductors in incoherent, or “direct detection” applications in the far-infrared. These devices have largely superseded bolometers for astronomical instruments, both for broadband photometers, as in the IRAS satellite, and for moderate resolution spectrometers, e.g. the UC Berkeley Tandem Fabry-Perot Interferometer (Storey, Watson, and Townes, 1980). Unlike Schottky diodes, germanium photoconductors must be cooled to liquid helium temperature in operation. Typically, quantum efficiencies of 10 % – 20 % are inferred from measurements of the signal-to-noise ratio in background-limited direct detection experiments (Watson, 1982). If their heterodyne performance is quantum-limited, such a figure for the quantum efficiency implies a vast improvement in noise temperature over a Schottky diode receiver. As discussed in more detail in chapter 2, the response of Ge:Ga photoconductors is ordinarily limited to wavelengths shortward of 120 μm . It has been found that application of a large uniaxial stress along the [100] crystal axis alters the band structure in such a way as to extend the photoconductive cutoff to $\lambda < 200 \mu\text{m}$ (Kazanskii, Richards, and Haller, 1977). This forms the long wavelength limit to our instrument.

The distinction between “diode-like” and “power” mixers is partly a matter of different languages spoken by researchers working on mm-wave and optical heterodyne receivers. Analysis of mm-wave receivers generally begins with the current-voltage (I-V) characteristic of the device, (approximately exponential in the case of Schottky diodes), from which conductance or admittance matrices, and thence conversion gains and noise temperatures are calculated. With optical and infrared heterodyne reception, one generally speaks of the direct-detection responsivity, that is, the response of detector current to incident squared-voltage (i.e. power), as fundamental. This distinction is partly artificial, since, after all, the lowest-order non-linearity of an exponential diode I-V curve is quadratic, and therefore, for sufficiently small signals, diodes can be, and often are, used as direct (power, i.e. square-law) detectors. The true distinction between the two types of device lies in timescales. Aside from parasitics, the I-V curve of a diode is the same at DC as at the operating frequency, and direct detection is due to rectification of the RF waveform according to the DC I-V curve. For “power” type detectors, the DC I-V curve may be non-linear, but there is no relation between that non-linearity and the responsivity at the RF frequency. Rather, macroscopic currents do not flow at the RF, and the mechanism for direct detection is the inducement of transitions between the internal

states of the detector. There is a low-frequency cutoff to direct detection of the RF, corresponding to the difference in energy levels of the detector's internal states. The power-detectors, i.e. photodiodes and photoconductors, are thus often spoken of in a somewhat vague and mystical way as being "intrinsically quantum mechanical" devices, while the diodes are spoken of as "intrinsically classical". In fact, the true correspondence between quantum mechanical and classical detection is more subtle. Diodes can also be "quantum mechanical", depending on whether the voltage scale of the I-V curve's non-linearity is greater or less than $h\nu/e$. For insufficiently sharp I-V curves, or in the limit of $\nu \rightarrow 0$, significant sampling of the non-linearity only occurs for signals with many photons' worth of energy, i.e. for classical waves. For practical purposes, the lack of a low-frequency cutoff is the basic distinction between diode-like and power detectors.

Since diode-like detectors respond to the instantaneous RF voltage, the maximum possible IF bandwidth is roughly equal to the RF bandwidth – i.e., in practice the IF bandwidth is never limited by the detector but is rather determined by the bandpass of the IF amplifiers and filters. This allows great flexibility in the design of the LO and IF systems. For power-detector mixers on the other hand, there is always some physical process in the mixer itself which limits the IF bandwidth. For the photodiodes and photoconductors used in visible and near-IR heterodyne receivers, this mixer bandwidth is a few GHz. Since $\nu_{IF} = |\nu_{RF} - \nu_{LO}|$, a limited IF bandwidth implies that the LO must be tunable to within ν_{IF} of the desired signal frequency. In the far-infrared this is a major constraint on the local oscillator.

There are no very suitable candidates for LO's in the far-infrared. Indeed, the lack of useful and convenient sources of electrical power at these frequencies is unquestionably the most important technological obstacle to full exploitation of the far-infrared, not only for astronomy, but for other applications as well. Multiplied solid-state oscillators and multiplied klystrons roll off above a few hundred GHz; at 1000 GHz, $< 1 \mu\text{W}$ is typical (Erickson, 1987). Carcinotrons and backward-wave oscillators extend to slightly higher frequencies, but are physically cumbersome, heavy, very power-hungry, notoriously unreliable, and extremely expensive. By far the most useful device for far-infrared power generation is the FIR molecular gas laser (Chang and Bridges 1970). This consists of a $10 \mu\text{m}$ CO_2 laser, which is used to pump a

vibrational transition of some simple molecule (methanol, ammonia, and difluoromethane are three of the most common species), which fills a far-infrared optical resonator. The CO₂ pump produces an inversion in the populations of the rotational sublevels of the vibrationally excited manifold. This yields gain for the laser transition, which lies between the optically pumped sublevel and the next lower sublevel in the vibrationally excited manifold. The physical principles of FIR laser operation are discussed at greater length in chapter 3. With rather heroic expenditures of effort and expense, FIR lasers have yielded power outputs on the order of 1 W (continuous-wave), but for reasonably sized devices, 1 – 50 mW is more typical. Their great drawback is that they operate only at fixed frequencies corresponding to the molecular energy spectra. Several thousand lines are known at present, and more are being discovered all the time.. However, the typical separation of known laser lines in the 100 μm region is 10 – 20 GHz, and if one restricts oneself to the stronger lines, the typical separation is correspondingly larger.

Table 1.2 lists the main astronomical lines at which our receiver is targeted, and for each astronomical line, the nearest one or two FIR laser lines, (from the compilation of Knight 1986), their power levels and frequency offsets. As may be seen, the LO situation is much better for some of our target lines than for others. For example, the 119 μm ground-state rotational line of OH is only ~ 8 GHz from a methanol transition which produces one of the strongest laser lines in the entire far-infrared. The HD 112 μm line, on the other hand, only has a rather weak laser line nearby, and an offset of some 30 GHz is required before a strong line is found. For the [OIII] 52 μm line, the nearest laser transition is 104 GHz away. The consequence of this technological limitation of FIR lasers is to make the dichotomy between broadband and narrowband detectors much wider. A wide bandwidth detector such as a Schottky diode or SIS junction can directly use the FIR laser line as its LO, with the frequency offset between it and the astronomical line made up by centering the IF bandpass at the offset frequency listed in table 1.2. For a detector whose intrinsic bandwidth is significantly less than this frequency offset, however, a tunable local oscillator must be used. Thus, the bandwidth of Ge:Ga photoconductors is a critical number for the design of an instrument such as ours. At the outset of this project, it was thought possible that, with the proper impurity concentrations, Ge:Ga photoconductors might achieve bandwidths on the order of GHz, comparable to near-IR and visible photodiodes and photoconductors. Unfortunately, this turned out not to be true. We have made extensive

Table 1.2 – FIR laser lines adjacent to astronomical targets

Astronomical target		Laser line		Laser power (mW)		Offset (GHz)
species	ν_{ast} (GHz)	species	ν_l (GHz)	Knight list [†]	Our laser [‡]	$\nu_{ast} - \nu_l$
[CII]157 μ m	1900.54	CH ₂ F ₂	1891.27	20	1.8	9.27
[OI]145 μ m	2060.07	CH ₃ OD	2058.14	.1	–	1.93
[OI]63 μ m	4744.77	¹³ CH ₃ OH	4751.34	–	.35	-6.57
[NII]122 μ m	2463	CH ₂ F ₂	2447.97	10	.3	15
[OIII]88 μ m	3393.0	¹³ CH ₃ OH	3411	–	–	-18
[OIII]52 μ m	5785.8	NH ₃	5681.8	–	–	104
CH 149 μ m	2010.8	CH ₃ NH ₂	2027.75	10	–	16.9
	2006.8					-20.9
HeH ⁺ 149 μ m	2010.2	CH ₃ NH ₂	2027.75	10	–	-17.6
OH 119 μ m	2510.0	CH ₃ OH	2522.8	20	1.6	-12.8
	2514.4					-8.4
HD 112 μ m	2674.93	CH ₂ DOH	2664.06	1	–	10.9

[†] Knight (1986)

[‡] see table 3.3

studies of photoconductor bandwidths as a function of donor and acceptor concentrations, and of bias (see §2.3) on a wide variety of detectors. The highest (3 db) bandwidth measured was 65 MHz, and this was only obtained with a large sacrifice in responsivity. Detectors with “useful” sensitivities generally have bandwidths ≤ 10 MHz. Thus, the use of a scanning LO in our instrument is unavoidable.

Unfortunately, the lack of continuously tunable sources of FIR power is extremely acute. At present, the best performance has been achieved by the “brute-force” technique of generating continuously tunable sidebands by modulating the beam from a FIR laser. The achieved power output from such a sideband generator is orders of magnitude lower than the full laser power, but, on the other hand, the higher responsivity of Ge photoconductors implies an LO power requirement that is also orders of magnitude

lower than for Schottky diodes. Thus, the crux of this project has been the experimental determination and theoretical understanding of two quantities, the LO power requirement for quantum-limited heterodyne performance from germanium photoconductors, and the power capability of a reasonably sized (i.e. small enough to fit on the KAO) sideband generator based on a FIR molecular gas laser.

1.4 Overall design of our instrument

Without further ado, we now describe the overall design of our receiver, and its implementation for both the laboratory and KAO versions of the instrument. Figure 1.1 shows a functional block diagram of the receiver. A walk through the instrument from the point of view of an LO photon begins with the CO₂ laser. This is a commercial, RF-excited waveguide model. It runs sealed off, and is relatively light and compact, a necessity for an airborne system. It produces approximately 12 W (CW) on its strongest lines. Two additional components have been added to stabilize its output, an optical isolator to prevent “pulling” by 10 μm radiation reflected back off the far-infrared cavity, and an optoacoustic cell filled with the same molecular gas used in the laser cavity. The optoacoustic cell is used to actively tune the CO₂ cavity length, and thus the precise 10 μm frequency, into resonance with the molecule’s pump transition. The pump beam, circularly polarized after passing through the isolator, is focussed and directed into the FIR cell. This is a 1 meter long, cylindrical, dielectric waveguide with flat metal mirrors at both ends. Hole coupling is used at both input and output. The cavity length, which is mechanically referred to a set of four invar rods, can be manually tuned, as can the orientations of the input and output mirrors, without breaking the vacuum of the FIR cell. The cell is filled with typically 100 – 300 mtorr of the lasing gas. A complete description of the principles of operation, construction details, and measured performance of our far-infrared laser is given in chapter 3. The maximum power output of the FIR laser is approximately 2 mW.

Tunable sidebands are generated on the FIR laser carrier by using a small-area Schottky diode in a corner-cube antenna (Krautle, Sauter, and Schultz, 1977) as a reflective modulator. The incident far-infrared beam induces a traveling wave on a short (a few λ) length of wire, or “whisker”, electrochemically sharpened at its tip. The tip contacts the Schottky diode. The traveling wave is partially reflected by the

Signal

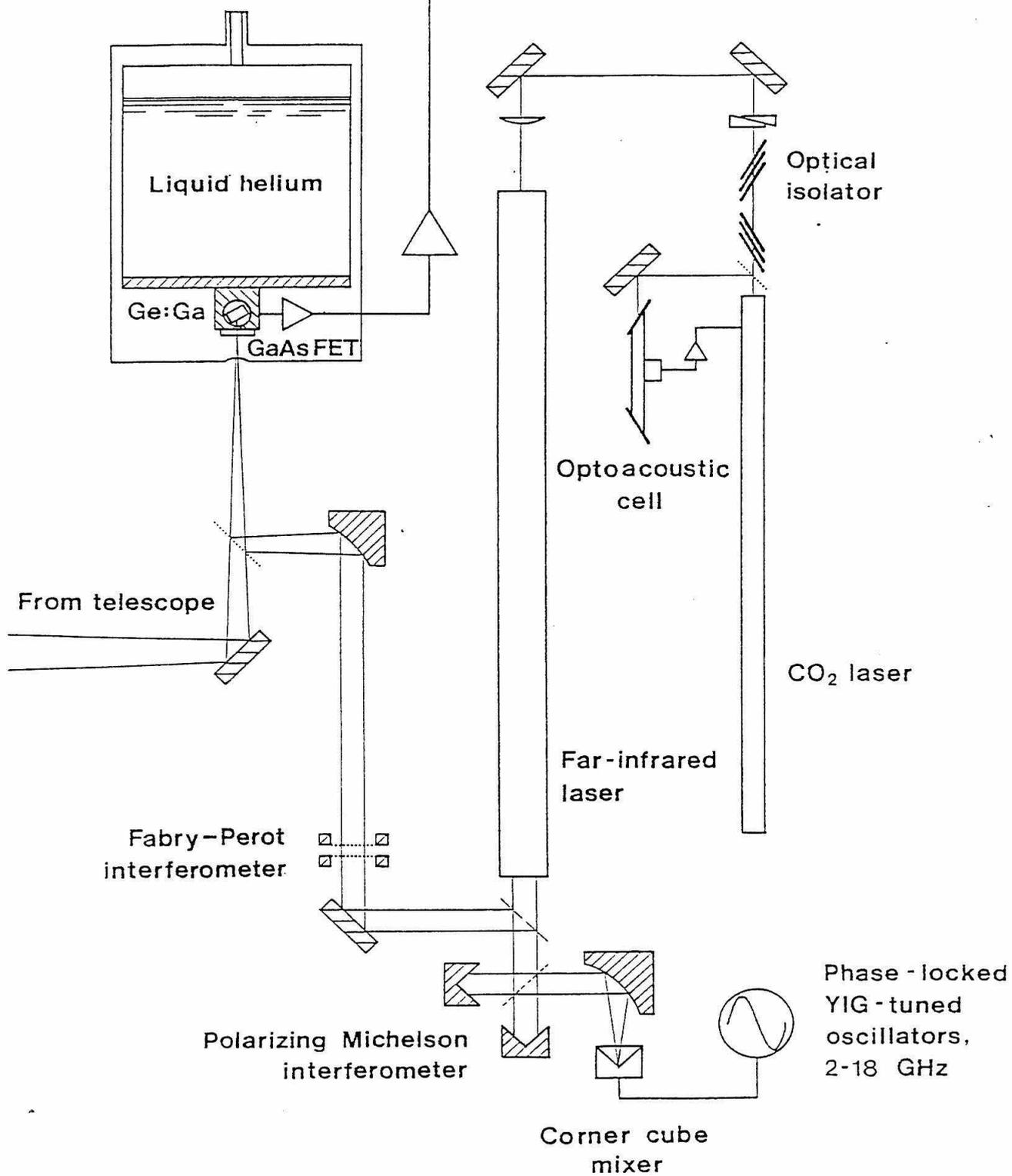


Figure 1.2 - Functional block diagram of our receiver

impedance discontinuity between the whisker and the diode, and the reflected wave is then re-radiated. The reflection coefficient off the diode depends on the diode impedance, and therefore the applied diode bias. The diode bias is modulated at the microwave frequency offset required to bring the FIR laser frequency and the astronomical frequency into coincidence. This frequency is generated by a set of commercial YIG-tuned (Yttrium Iron Garnet) microwave oscillators interfaced to the observing computer so as to allow computer controlled sweeps of the LO. Thus, the re-radiated wave consists, in frequency-space, of a strong carrier at the laser frequency (since the modulation depth from the diode is in practice very low) and weak sidebands, one of which is at the desired LO frequency. The carrier and sidebands are separated by a combination of a polarizing Michelson interferometer (PMI) and a tunable Fabry-Perot interferometer. The principle of the PMI is to differentially rotate (actually retard) the polarizations of the carrier and sidebands. The response of a corner-cube antenna is highly polarized; it is sensitive to radiation in which the E-field is in the plane containing the whisker and the dihedral "corner". If the (nominally) linear polarization of the laser output is parallel (normal) to the corner-cube polarization, then the path length difference of the PMI is chosen to be an even (odd) number of half-wavelengths at the carrier frequency. This maximizes the coupling of the laser onto the diode. The path length difference is also chosen to be an odd (even) number of half-wavelengths at the sideband frequency, however. Therefore, the re-radiated carrier from the corner-cube is coupled back into the laser cavity with the same polarization as the original laser beam, but the re-radiated sideband emerges from the PMI with polarization normal to the laser. The sideband is then split off with a wire-grid polarizer placed between the laser and the PMI.

The PMI by itself does not produce adequate rejection of the carrier, however, nor does it distinguish between the desired and undesired sidebands. These functions are performed by a piezoelectrically (or manually) tunable Fabry-Perot interferometer. It is a copy of the scanning Fabry-Perot incorporated in the UC Berkeley Tandem Fabry-Perot Interferometer (Storey, Watson, and Townes 1980), and was built in house. The transmitted local oscillator beam is then diplexed with the signal from the telescope using a simple mylar beamsplitter, and then directed into the cryostat. The cryostat contains various low-pass filters to eliminate the the background of visible, near-IR, and mid-IR radiation incident of the detector, and a single-crystal Ge:Ga photoconductor mounted inside an optical integrating cavity. In many cases, a

single-stage, low-noise GaAsFET transistor IF pre-amplifier was also mounted in the cryostat. The cryostat is operated at liquid-helium temperature (4.2 K) ordinarily, and at pumped liquid-helium temperature (~ 1.5 K) for the stressed detectors (i.e. for operation to $200 \mu\text{m}$.) In a few cases, a room-temperature pre-amplifier was used. The IF signal is then passed through additional (room-temperature) amplification, filtered, and processed by one of a variety of available backend devices. In many cases, a simple bandpass filter and RF power meter served the needs of the experiment perfectly well. The filters, detectors, integrating cavities, and pre-amplifiers are described in detail in §2.2.

In most of our laboratory experiments, not all of the components indicated in figure 1.2 were necessary simultaneously. For example, to measure the modulation bandwidths of photoconductors, the Fabry-Perot was removed, the YIG oscillators replaced with a low frequency VCO and a commercial RF spectrum analyzer used as the backend. The actual experimental configuration used in each experiment is described in chapters 2 – 4 along with the purpose and results of each individual experiment. Note that for some of the components indicated in figure 1.2, several versions were built and used for different experiments. For example, three different cryostats have been constructed for this project, one, (D-49 in our nomenclature), for heterodyne and high-frequency modulation lab experiments, one (D-69) for low-background, low frequency, direct detection applications in the lab, and one with a large helium capacity for use in the airborne receiver. For our sideband generator, we have used Schottky diodes both from Bell Labs – batch N280(91) fabricated by M. Schneider's group (Schneider, 1982) – and from R. Mattauich's group at the University of Virginia (batches 1E4 and 1E12). The latter had smaller areas and yielded better performance, and were therefore used in most of our lab experiments. With one minor exception, all of our photoconductors were fabricated by E. Haller's group at Lawrence Berkeley Laboratory. Detectors from several different batches have been tested (see table 2.2). Two versions of corner-cube antenna have been built, one quite standard design with fixed dimensions, and one in which the distance from whisker to dihedral angle was tunable (see §4.1). As for the laser, all our earlier experiments were conducted with the CO₂ laser, FIR cell, and isolator separately mounted on an optical table with the optical configuration easily adjustable.

In the summer of 1986, our laboratory experiments were suspended, and all efforts were concentrated

on adapting the instrument for installation on the KAO. The optical table was replaced with a rigid, welded aluminum, space frame in which the CO₂ and FIR lasers were permanently mounted. The optical design required to obtain good overlap between the telescope and LO beams, and an appropriate illumination pattern of the telescope primary by the beam from the cryostat, was calculated and the required focussing mirrors fabricated. All of the optics for the airborne receiver were implemented with off-axis parabolic and elliptical mirrors made by the technique described by Erickson (1979). Our electronics was mounted in KAO compatible chassis, and in many cases rebuilt from scratch, and the instrument was interfaced to a microcomputer for data recording and (simple) analysis. In retrospect, and considering the great complexity of the instrument and the lack of adequate testing for some of its components, this "packaging" of the spectrometer for the KAO was premature. However, in September 1986, an attempt was made to mount the instrument on the KAO telescope. The aim was not to attempt spectroscopy, but rather to use the direct (attenuated) beam from the laser to make heterodyne observations of the continuum from an astronomical source. One of the recently fabricated (and untested) vacuum windows failed shortly after takeoff, and the flight was aborted. A second attempt to perform astronomical observations from the KAO was made the following February (1987). The physical packaging was somewhat simplified, and various other improvements were made. However, the sideband generator was not performing adequately, and again, the aim was only to make continuum observations at the laser wavelength. On this flight, the receiver functioned properly when mounted on the telescope with the airplane on the ground. However, once airborne, problems related to ground loops and interference on the power supply lines developed in both the detector bias circuitry and in a rack of electronics used for chopper motor control and Schottky diode bias. Also, one of the experimenters fell ill in flight, and therefore no attempt was made to diagnose or repair the electronic problems, and the flight was aborted $\sim 1/2$ hour after reaching altitude. The electronics problems have not been reproduced on the ground.

In summary, there remain both practical and fundamental obstacles to having a working astronomical spectrometer. The practical problems, like the faulty vacuum window on our first flight and the malfunctioning electronics on our second flight, though highly visible, (and of course highly embarrassing), are in some sense less important because they can almost certainly be easily fixed. The fundamental problems, on

the other hand, do not have any easy solution. The primary fundamental problem is inadequate LO power produced by the sideband generator. In our lab experiments, we found that $\sim 2 \mu\text{W}$ of LO power was required for optimal performance from the Ge:Ga mixers. The maximum power output we have actually measured from the sideband generator is 9.5 nW. There is thus a very large improvement in performance required. The shortage of LO power is exacerbated by other fundamental problems. The laser can produce spontaneous pulsations and excess broadband noise within its homogeneous gain linewidth (~ 10 MHz) due to intrinsic non-linear dynamical effects (see §2.3). These can be tuned away, but only by sacrificing power. Similarly, the detectors can exhibit non-linear dynamical effects that cause excess noise at high biases (near breakdown). Again, these can be tuned away by reducing the bias voltage, but this entails a sacrifice of responsivity, and therefore increases the LO power requirement. Thus, the LO power shortage is an acute problem. The low sideband power is caused by a combination of low FIR laser power, low sideband conversion efficiency, and low throughput of the Fabry-Perot, of which three, the sideband conversion efficiency is the largest loss, -39 db.

After the February flight and before beginning the preparation of this thesis, I conducted a few additional experiments and some numerical analyses in an effort to account quantitatively for the low sideband conversion efficiency. These were successful. The measured antenna efficiency and Schottky diode parasitics lead to a predicted conversion efficiency which agrees with the measured value to within the experimental errors (see chapter 4). Using this understanding of the loss mechanisms, we can come to some conclusions regarding the prospects for future improvement. The main point is that no single improvement will yield anything near the required $2 \mu\text{W}$ of sideband power. Smaller Schottky diodes with higher impurity concentrations can in theory yield as much as 20 db improvement in conversion efficiency, but this is obtained only for diode diameters of $.4 \mu\text{m}$, a highly ambitious (though not inconceivable) advance over the current state-of-the-art. An improvement of perhaps 10 db is more realistic in the near term. An improvement of a factor of 6 - 8 in the Fabry - Perot transmission can probably be obtained fairly easily by replacement of the metal mesh mirrors used in the sideband experiments with ones incorporating slightly coarser (lower reflectivity) mesh. The "brute force" technique of increasing the laser power by simply obtaining a higher power CO_2 laser could also yield a large improvement in sideband power. How

large an improvement is not certain, but a factor equal to the increase in CO₂ laser powers (which could be as much as a factor of 5) is a reasonable lower limit. Improved FIR laser output couplers might yield higher FIR powers with the same CO₂ pump power; however, some experimentation has already been done along this line for our laser, without encouraging results. Finally, improved far-infrared filtering of the background incident on the detector could yield a reduction in the LO power required for optimum heterodyne noise temperature. Thus, with some combination of these improvements, it would be possible to fully realize the potential of Ge:Ga photoconductors for low-noise heterodyne spectroscopy. Except for the Fabry-Perot mirrors, however, none of these improvements is particularly easy, quick, and cheap.

1.5 – Engineering formalism for description of Ge:Ga heterodyne performance

The standard formalism for characterizing the performance of photoconductors, (see the review by Bratt, 1977), was developed primarily for direct detection applications, but applies equally well in our case. It does not explicitly refer to any of the microscopic physical processes in the semiconductor, but rather subsumes them into empirical quantities which can be directly determined by experiment. The response of the photoconductor to a beam of incident FIR radiation of power P , modulated at angular frequency ω , is given by the photocurrent, i , or current responsivity, S ,

$$\begin{aligned} i &= \left(\frac{P}{h\nu} \right) \frac{e\eta G}{(1 + \omega^2 \tau_r^2)} \\ S \equiv i_{DC}/P &= \left(\frac{e}{h\nu} \right) \eta G \end{aligned} \quad (1.2)$$

Here, and in the rest of this thesis, e is the fundamental electric charge, h is Planck's constant, and $\nu = \omega/2\pi$ is frequency. Thus, $P/h\nu = \dot{N}$ is the rate of incidence of FIR photons. The quantum efficiency, η , is the (wavelength dependent) fraction of incident photons which are converted into mobile charge carriers, i.e. holes, since all our detectors are p-type. It is less than unity due both to losses in optically coupling the incident radiation into the detector, and to ill-understood internal losses in the germanium crystal. G is the photoconductive gain, given by

$$G = \tau_r/T_{tr} = \tau_r \left(\frac{\mu E}{l} \right), \quad (1.3)$$

where τ_r is the recombination-limited lifetime of the photogenerated holes, and T_{tr} is the mean transit time for carriers to traverse the detector from contact to contact. Also, μ is the drift mobility, E the bias field,

and l the interelectrode distance. The photoconductive gain can be either greater or less than unity. An intuitive way of thinking about it is as follows. Once a carrier is photoexcited, it drifts toward the negative detector contact. If it does not suffer recombination before getting there, then, when it arrives, it is detected by the external circuit (pre-amplifier) and immediately replaced with a carrier injected at the opposite contact. The injected carrier can in turn drift through the crystal, be “counted” at the negative contact, be replaced, and so on, until carrier recombination occurs within the crystal. A single photoexcited carrier travels, on average, G times through the external circuit, producing a correspondingly larger (or smaller, if $G < 1$) output signal. This is a highly picturesque explanation, but it can be justified by a proper rate equation analysis.

From a physical point of view, the basic material constants are η , μ , and τ_r , while the basic engineering parameters are the S , G , and the 3-db bandwidth $B = 1/2\pi\tau_r$. From equations 1.2 and 1.3, one would expect that, in the absence of dark current due to e.g. hopping or thermal generation of carriers, the DC I-V curve should be Ohmic, with a conductance proportional to incident FIR flux. This is indeed the case at low bias. However, the recombination time increases at high bias due to carrier heating, which leads to an I-V curve resembling that shown in figure 1.3, which rises between quadratically and cubically with voltage. Often, as in figure 1.3, the Ohmic region is negligibly small. Above a certain “breakdown” field, the rate of impact ionization of neutral acceptors by field-accelerated carriers equals the recombination rate, and the current increases by many orders of magnitude. As discussed in a long series of recent papers, (Westervelt and Teitsworth, 1985, Teitsworth and Westervelt, 1984, 1986, Gwinn and Westervelt, 1986), the strongly non-linear coupling between the electric field and the carrier concentration near breakdown can lead to chaotic behavior and/or spontaneous pulsations at high biases.

Aside from such anomalous noise sources, there are two fundamental sources of noise in our detectors, thermal (Johnson) noise, and generation-recombination (g.r.) noise. Due to the low temperature and high impedances at which the photoconductors are typically operated, thermal noise can generally be neglected in comparison with g.r. noise. The latter is the analogue of shot noise in e.g. a photodiode or vacuum

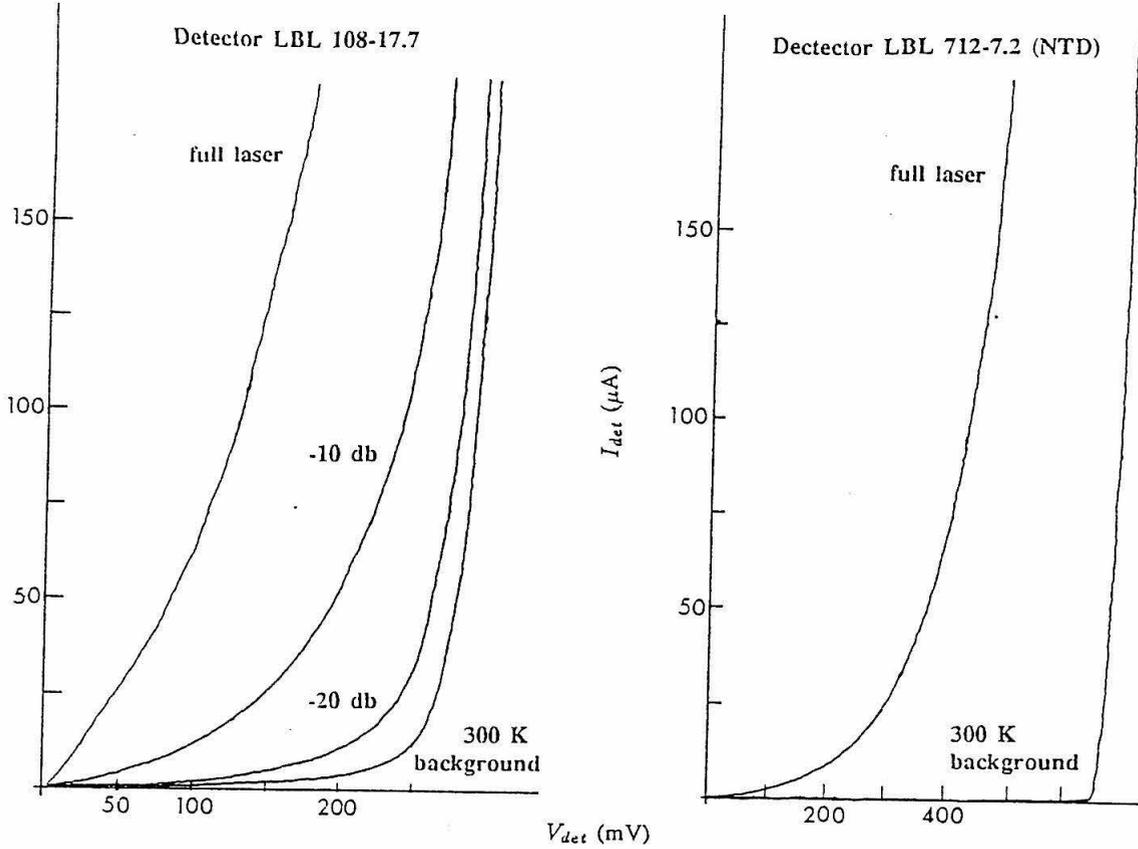


Figure 1.3 – Typical I-V curves for two of our detectors, with varying amounts of far-infrared radiation incident.

tube. Its spectral density is given by

$$\langle i^2 \rangle_{gr} = 4eIG(1 + \omega_{IF}^2 \tau_r^2)^{-1} \text{ A}^2/\text{Hz} \quad (1.4)$$

$$\propto G^2$$

where I is the direct current through the detector, which consists of components due to photocurrent from various sources, thermally generated current, hopping current, etc. The factor of 4 replaces the factor of 2 in the usual shot-noise equation $\langle i^2 \rangle = 2eI$ because generation and recombination are independent random processes which each generate noise current.

Using equations 1.2 – 1.4, it is not difficult to calculate the fundamental limit on heterodyne sensitivity for a photoconductive mixer, and the conditions under which the fundamental limit is achieved. Given a monochromatic incident signal field E_S at angular frequency ω_S , and a coincident LO field E_{LO} at ω_{LO} ,

the low-frequency component of the incident FIR power may be expressed as

$$P = \frac{4\pi}{c} \overline{|E_S \cos \omega_S t + E_{LO} \cos \omega_{LO} t|^2} \quad (1.5a)$$

$$= \frac{4\pi}{c} \{ E_S E_{LO} \cos(\omega_S - \omega_{LO})t + E_{LO}^2/2 + E_S^2/2 \} \quad (1.5b)$$

$$= 2\sqrt{P_{LO}P_S} \cos \omega_{IF}t + P_{LO} + P_S \quad (1.5c)$$

where terms oscillating at frequencies $(\omega_S + \omega_{LO})$, $2\omega_S$, and $2\omega_{LO}$ have been dropped. The bar denotes a time average over many RF periods. The assumption that the signal is monochromatic implies that no other heterodyne signal appears at ω_{IF} due to downconversion of power in the other sideband. Thus, our derivation applies to single-sideband conversion. This heterodyne signal power is converted into photocurrent via the responsivity. Thus, the first term in (1.5c) yields the desired IF signal, whose RMS value is

$$\langle i^2 \rangle_{sig} = 4S^2 P_{LO} P_S \overline{\cos^2 \omega_{IF}t} = 2S^2 P_{LO} P_S \quad (\text{A}^2). \quad (1.6)$$

The second and third terms of (1.5c) contribute DC photocurrent. We assume $P_{LO} \gg P_S$ so that the third term can be neglected. Although the LO photocurrent appears at DC, fluctuations due to the g.r. noise it induces appear throughout the IF passband, as given by (1.4), and contribute noise to measurements of the desired signal. If this is the dominant noise source, then the overall signal to noise ratio is given by

$$\begin{aligned} S/N &= \frac{\langle i^2 \rangle_{sig}}{4eI_{LO}G\Delta\nu} = \frac{\langle i^2 \rangle_{sig}}{4S^2(h\nu/\eta)P_{LO}\Delta\nu} \\ &= \frac{P_S}{2(h\nu/\eta)\Delta\nu}, \end{aligned} \quad (1.7)$$

where $\Delta\nu$ is the IF channel width used in the measurement. It is customary to measure signal strengths in radio astronomy in terms of equivalent Rayleigh - Jeans radiation temperatures, i.e. power per unit bandwidth (expressed in units of temperature). Thus, $P_S = 2kT_S\Delta\nu$, (where the factor of two accounts for the fact that blackbodies emit equally into both polarizations), and the signal-to-noise ratio of (1.7) may be expressed as the ratio of signal temperature to an equivalent instrumental noise temperature, T_N , (see equation 1.1), where

$$T_N(SSB) = \frac{h\nu}{\eta k}. \quad (1.8)$$

This noise performance, limited by the g.r. noise due to the LO-induced photocurrent, is referred to as “ideal”, or “quantum-limited”. For a quantum efficiency of unity, it corresponds to the fundamental limit on coherent detection (or amplification) imposed by the uncertainty principle (see Caves 1982). Equation (1.8) is applicable to signals which are present in one sideband only. Noise temperature is usually measured, however, using signals from blackbody radiators at various temperatures, which fill both sidebands. In this case, ideal performance corresponds to a noise temperature in Kelvins which is a factor of 2 lower. It is important to recognize that the noise temperature which is relevant for spectroscopic applications can be different for different IF frequencies and source spectra. For narrow bandwidth detectors, such as ours, there is virtually always “interesting” signal (either spectral line emission or required baseline) in both sidebands simultaneously. For high IF receivers, e.g. those based on Schottky diode mixers, the spectral line (and nearby baseline) of interest usually lie entirely in one sideband. Thus, the relevant comparison between the two types of spectrometer is between $T_N(DSB)$ of the low IF frequency system and $T_N(SSB)$ of the high IF system.

The various significant contributions to system noise temperature when the LO - induced g.r. noise is not completely dominant, and the conditions under which ideal performance is achieved may be easily understood from figure 1.4. The mean squared current spectral density in the IF (in A^2/Hz) is plotted against the FIR signal temperature. In the ideal case, the slope is, according to equation (1.6), proportional to the LO power and the square of the responsivity. The noise temperature defined by equation (1.8) is simply the source temperature that would be required to increase the output power by an amount equal to its zero-signal value; graphically it is simply (the absolute value of) the x-intercept of the output curve. The standard hot/cold load technique of measuring noise temperature is based on this picture. The output power is measured twice, with the detector illuminated by blackbody radiation at two different, accurately known temperatures. The noise temperature may then be computed from the ratio of output powers without having to know the total system gain, which is generally not as easy to determine accurately. A dimensionless conversion gain may be defined as the ratio of the signal temperature in the IF to the signal temperature in the incident FIR beam. It is obtained by converting the squared current spectral

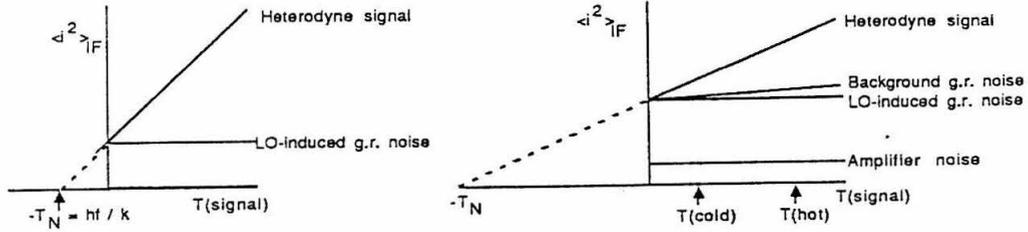


Figure 1.4 – Contributions to the IF output power spectral density as a function of signal temperature.

density to IF power with the detector impedance; i.e.

$$G = T_{IF}/T_{RF} = R_d \langle i^2 \rangle / kT_S = 2S^2 R_d P_{LO}. \quad (1.9)$$

The most important contributions to system noise besides LO-induced g.r. noise, i.e. the factors that often prevent one from achieving ideal performance in the real world, are amplifier noise and g.r. noise due to the broadband FIR background incident on the detector. Amplifier noise is independent of both LO power and signal temperature. The g.r. noise due to the background is independent of LO power, but proportional to the temperature of the background. To some degree, therefore, the background noise mimics true heterodyne signal, and if it is significant, it must be subtracted out in a hot load/cold load measurement of heterodyne noise temperature. The condition for background g.r. noise to be insignificant is simply $P_{LO} \gg P_{bgnd}$. The condition for amplifier noise to be insignificant (compared to LO-induced g.r. noise) is

$$\frac{\langle V_N^2 \rangle}{R_d^2} + \langle i_N^2 \rangle \approx \frac{\langle V_N^2 \rangle}{R_d^2} \ll 4eI_{LO}G \quad (1.10)$$

where $\langle V_N^2 \rangle$ and $\langle i_N^2 \rangle$ are the mean squared voltage and current noise spectral density of the amplifier. The first equality assumes, as is the case in practice with our amplifiers and detectors, that the amplifier noise has a characteristic impedance much greater than the detector impedance. Note that there is an implicit

dependence in (1.10) on LO power through the detector impedance. In particular, at high LO powers, the detector impedance is driven so low that the condition (1.10) no longer holds, and the system noise temperature is degraded due to the increased importance of amplifier noise.

1.6 – Highlights of our results

This project has yielded a number of significant new results on various aspects of far-infrared technology. Although, as discussed earlier, the implications for the prospects for an astronomically useful receiver are somewhat discouraging in the near term, this does not detract from the accuracy or importance of our positive results. The latter are discussed in detail in chapters 2 – 4. Here, we offer only a brief summary, bearing in mind that such a cursory listing glosses over many important details and complicating factors. Beginning with our experiments on far-infrared laser sideband generation, we have established the following results :

- A new technique for the measurement of V_{FB} , the flat-band, or “built-in”, voltage of small-area Schottky diodes has been developed and applied to our diodes. The ordinary method is to measure the diode capacitance as a function of DC voltage. The standard application of Poisson’s equation to the epilayer (Torrey and Whitmer, 1948, p.72) yields the depletion region width, and thence the barrier capacitance as a function of diode voltage, which turns out to vary as $C_b \propto (V_{FB} - V)^{-1/2}$. Thus a plot of the relation between $1/C_b^2$ and V yields a straight line whose x-intercept is V_{FB} . Our technique is closely related to the standard technique, but does not require measurement of the very small capacitances (of order 1 fF) which are typical of high frequency diode, and which are difficult to measure accurately. It turns out to be quite simple to show that, under certain conditions, the video (i.e. direct-detection) voltage responsivity of an antenna-coupled Schottky diode varies with capacitance as

$$S_V \propto (1 + \omega_{FIR}^2 R_{eq}^2 C_b^2)^{-1}$$

(see equation 4.11). Under realistic conditions, R_{eq} , (defined in equation 4.9) is approximately equal to the antenna impedance R_a . The inverse square dependence of responsivity on capacitance combines with the inverse square-root dependence of capacitance on voltage from Poisson’s equation to yield a video responsivity that falls linearly with voltage. The voltage at which the linearly extrapolated responsivity

equals zero is V_{FB} . The conditions that must be met for this to apply are $\omega R_{eq} C_b \gg 1$ and $R_a \ll R_b \ll R_L$, where R_b is the barrier resistance of the (ideal) diode and R_L the load impedance of the video circuit (bias circuit, oscilloscope, etc.). The technique has been applied to our 1E4, Univ. of Virginia Schottky diode, and the experimental data and a linear fit are shown in figure 4.9. Using a somewhat indirect method involving the predicted saturation current, the derived value of $V_{FB} = 1.03 \pm .02V$ was checked for consistency with the other adopted diode parameters. It was found that the agreement was excellent, considerably better than that achieved in the attempts by other workers, who lacked this determination of V_{FB} , to determine consistent sets of diode parameters (see Crowe and Mattauch, 1987).

– We have made the first direct (i.e. not on a scale model) measurement of the main beam efficiency of a corner-cube antenna, and compared it with the “conventional” theory. The measurements were made at $\lambda = 214\mu\text{m}$ (1.4 THz), using our earlier “fixed dihedral” corner-cube design. During this particular measurement, the whisker length was 3.25λ and the distance from the dihedral $.63 \lambda$. The theoretically predicted value of η_{beam} is 47 %. The measured value is 44 %. The complete theoretical and experimental beam patterns are shown in figure 4.6. Examining the complete patterns, it is clear that the extreme closeness of the predicted and measured efficiencies is partly fortuitous, but the level of agreement between overall patterns is still reasonably good. The efficiencies quoted are in agreement with the rough estimate of $\sim 50 \%$ for $L = 4 \lambda$, $s = 1.2 \lambda$, by Harris (1986).

– We have developed an accurate quantitative model of the sideband conversion efficiency of a Schottky diode in a corner-cube antenna, and compared its prediction with our experimental results. We have also investigated the dependence of the predicted conversion efficiency on the diode parameters, so as to determine guidelines for future diode optimization. The results of the modeling have been reported by Grossman (1987), and are included as section §4.2 of this thesis with only minor modifications. The model treats the combination of corner-cube and Schottky diode as a reflective modulator. The high frequency parasitics are treated according to the theory of Champlin and Eisenstein (1978), and the effect of the vanishing of the depletion region at voltages above V_{FB} is incorporated according to the formulation of Crowe and Mattauch (1987). It is found that, in contrast to the low-frequency behavior, phase modulation

dominates amplitude modulation by a large factor for all realistic diode parameters. It is also found that performance is degraded at frequencies well below the plasma resonance in the undepleted epilayer due to a second resonance caused by the interaction of the effective inductance due to carrier inertia with the barrier capacitance. The predicted conversion efficiency is quite insensitive to the values of the antenna impedance, diode substrate size, temperature, and Schottky barrier height. For the measured parameters of our 1E12 diode at a wavelength of $119 \mu\text{m}$, (2.52 THz), where most of our sideband experiments were done, the model predicts a conversion efficiency due to the diode alone of -28.0 db. To obtain the total conversion efficiency, this figure must be multiplied by the square of the antenna coupling efficiency (\approx main beam efficiency). The main beam efficiency of the corner-cube was not measured at 2.52 THz, but from comparison of beam scans at 1.4 and 2.52 THz and the measured beam efficiency at 1.4 THz, we crudely estimate a main beam efficiency of 30% at 2.52 THz. This leads to a predicted total sideband conversion efficiency of -38 db. The experimentally measured value is -39 ± 2 db (see table 4.4). The model predicts an improvement in diode conversion efficiency of nearly 20 db over the 1E12 diode when the doping level is raised from $2 \times 10^{17} \text{cm}^{-3}$ to $\sim 5 \times 10^{17} \text{cm}^{-3}$ and the diode diameter is reduced from $1.4 \mu\text{m}$ to $0.4 \mu\text{m}$. The complete dependence of conversion efficiency on diode radius and epilayer doping is shown as a series of contour plots, one for each frequency, in figure 4.17.

Turning now to our results on germanium photoconductors as heterodyne mixers,

– We have, in collaboration with I. S. Park of Lawrence Berkeley Laboratory, made an extensive study of the bandwidth, photoconductive gain, and direct detection responsivity of heavily compensated Ge:Ga photoconductors. A series of 8 such detectors were fabricated by the technique of neutron transmutation doping (Haller 1984) and tested. The results are reported by Park *et al.* (1987), but are also described in somewhat greater detail as part of section §2.3 of this thesis. Recombination bandwidths were measured by two methods. Firstly, the FIR laser beam was directly modulated at MHz frequencies using the Schottky diode/corner-cube combination, and the rolloff of the photocurrent signal was measured as the modulation frequency was swept. A Lorentzian fit to the data then yielded the bandwidth. Using this technique, the variation of bandwidth with bias voltage (due to carrier heating) was measured for one detector. The result,

shown in figure 2.10, is in agreement with the theoretical expectation of a constant bandwidth at low bias and a bandwidth $\propto E^{-3/2}$ in the hot carrier regime.

Bandwidths were also determined by measuring the rolloff of the detectors' g.r. noise spectra. These measurements revealed the existence of sporadic excess noise at low frequencies due either to the laser (in which case the excess noise scaled with the square of the laser power) or to the detector when the detector bias was near breakdown. Figure 2.8 shows examples of measured noise spectra with and without the presence of excess laser noise. Figure 2.13 shows an example of the extremely strange detector noise spectra that are sometimes possible near breakdown. These effects could generally be tuned away or subtracted out, after which the detector bandwidths measured by the two techniques agreed very well, as shown in figure 2.11, which shows the dependence of bandwidth on compensating impurity concentration. The highest measured value was 65 MHz. As a side benefit of our detector characterization, the dependence of bandwidth on N_D leads to an inferred recombination cross-section of $\sigma_r = 3.2 \times 10^{-13} \text{ cm}^2$ at 4.2 K, in close agreement with the theory of Brown and Rodriguez (1967), but a factor of ~ 30 lower than the 4.2 K value predicted by Abakumov, Perel', and Yassievitch (1977). Unless the carrier temperature is a great deal (factor of 4) higher than the lattice temperature, this appears to cast strong doubts on the latter theory. We have also used three different methods to determine photoconductive gains for this set of detectors, the most direct being simply measurement of the amplitude of the detector g.r. noise (see table 2.4). There are definitely some discrepancies in the derived values, but overall, the inferred photoconductive gains seem to lie between .02 and .08, with the most heavily compensated detectors generally having the lowest gains, as expected.

– Using several other detectors, also fabricated at LBL, but not in the same NTD series, we have studied the heterodyne noise performance of germanium photoconductors and its relation to other measured detector parameters. Nearly all our heterodyne experiments were performed at $119 \mu\text{m}$, the wavelength of the strong methanol laser line. In order to achieve reasonable LO power levels, heterodyne measurements were made using the attenuated laser beam rather than the sideband as the local oscillator. Comparing the performance of two detectors whose responsivities and photoconductive gains, (the latter again determined

from the amplitude of the g.r. noise), differed by a factor of 5, the scaling of the LO power requirement with G^2 was roughly confirmed (see table 2.9). Our best achieved noise temperature was obtained with a detector (LBL 108-17.7) with $B = 3$ MHz, $G = .17$, and $S = 3.9$ A/W at $119 \mu\text{m}$. As shown in the tuning curves of figure 2.18, the lowest (double-sideband) noise temperature obtained, uncorrected for the background g.r. noise or the Planck correction, was 590 K, at an LO power of $1.6 \mu\text{W}$, and a detector voltage (current) of 160 mV ($10 \mu\text{A}$). After applying the two corrections, this value is raised slightly to 655 K, corresponding to a heterodyne quantum efficiency (according to equation 1.8) of 9.2 %. The latter value compares quite well with the value $\eta = 10$ % obtained from the measured responsivity and photoconductive gain. In figure 1.5, this noise performance is compared to the best reported noise temperatures for various competing technologies throughout the submillimeter and far-infrared. As may be seen, the advantage in noise temperature of the germanium photoconductor over the Schottky diode is a factor of 26. Even assuming the case most favorable to the Schottky receiver, namely a source bandwidth greater than the 500 MHz IF bandwidth typical of Schottky systems, the photoconductor-based system would still have a factor of 2.0 advantage in total signal-to-noise ratio.

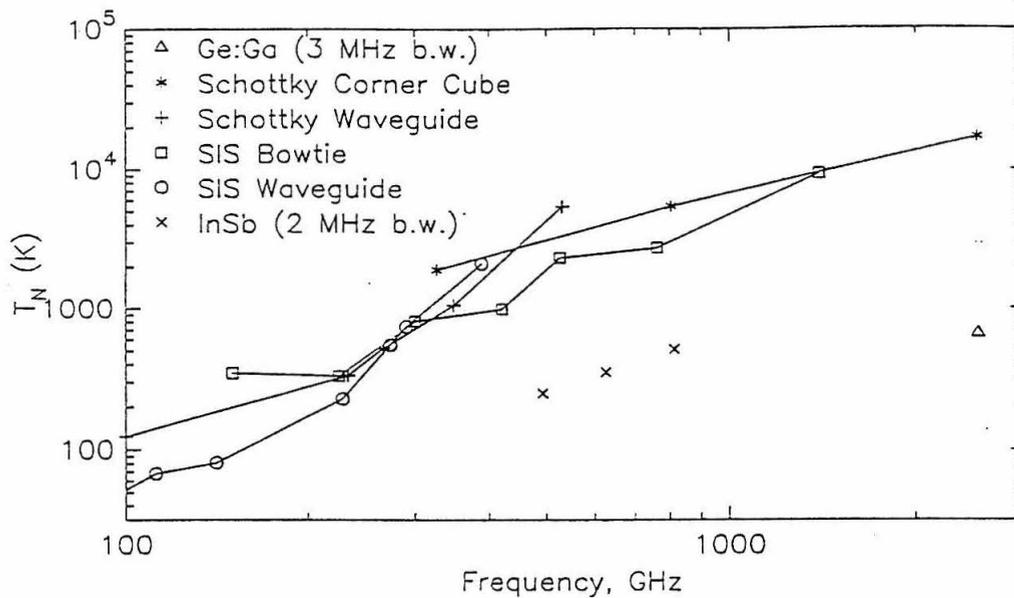


Figure 1.5 – Best reported heterodyne noise performance, as of September 1987, of various submillimeter and far-infrared receiver technologies. Double-sideband noise temperatures are plotted for the narrow-band systems, the germanium photoconductor and the InSb hot-electron bolometer, single-sideband for the wideband Schottky and SIS systems. References are : for InSb bolometer, Brown, private communication, for Schottky corner-cube, Röser et al. 1986, for SIS bowtie, Wengler 1987, for SIS waveguide, and Schottky waveguide, see references in Wengler 1987, p.71.

Chapter 2 – Ge:Ga Photoconductors as Heterodyne Mixers

It is the use of a germanium photoconductor as the front-end mixer which most fundamentally distinguishes our receiver from competing ones, and which has determined most of the other unique features of the system. In particular, the relatively low bandwidth of the photoconductor, determined, by the carrier recombination lifetime, has dictated the use of a scanning local oscillator, which is necessarily of much lower power, and is far more complex, than a fixed LO. As discussed in the introduction, however, the high responsivity of a photoconductor can in many cases compensate, or even more than compensate, for these disadvantages.

In this chapter, the detailed characteristics of the photoconductors used in our system are discussed. First, we review the current theoretical understanding of the spectral response, mobility, breakdown field, and recombination time of extrinsic photoconductors. In some cases, these properties can be understood in terms of an extremely simple and intuitive model of the impurity centers as solid-state analogues of the hydrogen atom. In the course of the theoretical review, we also touch briefly on some of the previous experimental work on photoconductor characterization. We conclude the theoretical review with a very brief discussion of the reduction in impurity ionization energy due to uniaxial stress, an effect our receiver relies on for its response at wavelengths longer than $120\ \mu\text{m}$. In the following section, we describe the experimental details of our two systems, both the laboratory system we used for measuring detector bandwidths and heterodyne noise temperatures, and the airborne system we developed for astronomical observations at $\lambda < 200\ \mu\text{m}$. First, the audio frequency circuit used for DC bias and direct detection is described. Then the intermediate frequency (IF) circuitry, which operated in the range of .1 – 100 MHz, is described. It consisted of a cryogenically cooled GaAsFET amplifier for the higher speed detectors, and a room-temperature Si JFET amplifier for the slower detectors. Then the “RF circuit” is described, i.e. the integrating cavity in which the detector was mounted and the far-infrared filters used to exclude unwanted high frequency radiation.

In the last section we describe our experimental results on all the detectors for which we have data. We first discuss our experimental determinations of detector bandwidths. The most reliable and comprehensive

data come from a series of highly compensated detectors (LBL boule 729) specifically grown for high speed applications by I. S. Park. As a side benefit of this bandwidth characterization, we derive a value for the hole recombination time at 4.2 K from the dependence of bandwidth on minority impurity concentration. This value, 3.2×10^{-13} cm², is compared to two of the theories discussed in §2.1. Next, we discuss our data on the detectors' photoconductive gain, responsivity, and spectral response. These include not only the 729 series detectors, but also the low compensation unstressed detectors used for noise temperature measurements and the stressed detector used in the 150 μ m receiver. Finally, we discuss our data on heterodyne noise temperature. The heterodyne measurements were made using the attenuated laser carrier, rather than the tunable sideband, as the local oscillator. This was the only way to obtain power levels adequate to perform sensible measurements, but it means that our noise temperatures were obtained at uninteresting astronomical wavelengths. The best heterodyne noise temperature we achieved is 655 K, double-sideband. The measurement demonstrates that Ge:Ga photoconductors are by far the most sensitive heterodyne mixers for this wavelength region. This noise performance is 9.2 % of the quantum limit.

2.1 Theory

An intuitive understanding of far-infrared photoconduction may be obtained from figure 2.1, which displays the simplest features of the band structure of germanium in the [100] direction. At liquid-helium temperatures, thermal generation across the intrinsic gap is negligible. In far-infrared applications, photogeneration must be prevented by external filtering of the visible and near-infrared component of the incident radiation ($h\nu > 1.1$ eV). The levels important for far-infrared detection are localized levels of group III or group V impurities. As all our detectors are p-type, we specialize throughout this chapter to the case of group III majority impurities. The impurity nucleus is fixed at a lattice site that would ordinarily be occupied by a germanium atom. In the absence of radiation, three of the four bonds with nearest neighbor germanium atoms are completed, and one of the four bonding orbitals with a nearest neighbor remains unfilled. A sufficiently energetic photon can cause a valence electron from a neighboring Ge-Ge bond to migrate to the Ge-impurity unfilled bond, leaving a localized negative charge density in the neighborhood of the impurity nucleus (an A^-) and a mobile vacancy (hole) in the periodic array of germanium nuclei

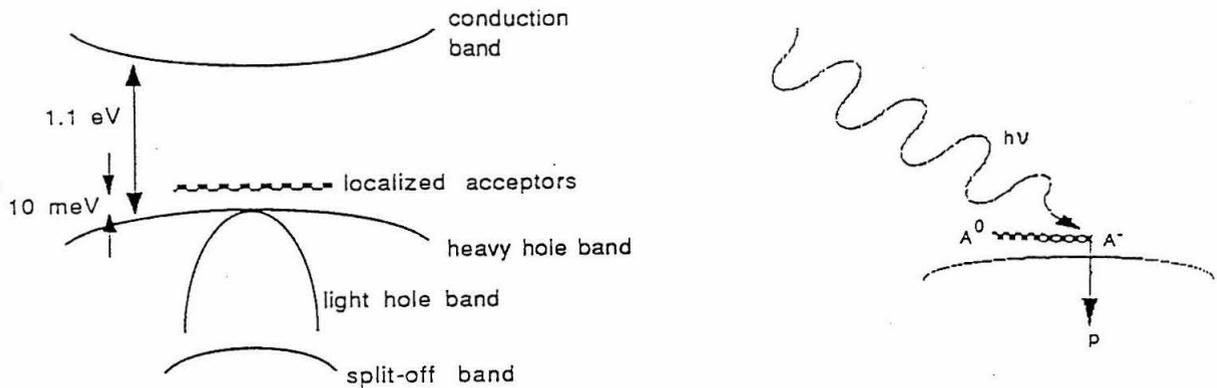


Figure 2.1 – The (simplified) band structure of germanium, and a schematic illustration of the process of extrinsic photoconduction.

and valence electrons. This process may be visualized, as in figure 2.1 as the photo-ionization of a bound state of the impurity nucleus plus four valence electrons (A^-) and a hole.

A surprisingly accurate analytic treatment can be made based on the analogy between the impurity center - a bound state of an A^- and a hole - and the hydrogen atom. The theoretical justification, experimental verification, and limits of applicability of this approach are comprehensively reviewed by Ramdas and Rodriguez (1981). The basic idea is very simple : the impurity center provides a nearly central, $1/r$ potential for the hole, so the spectrum of eigenstates is expected to be the same as that of a hydrogen atom. It is rescaled, however, to account for the facts that a) the impurity is embedded in a dielectric medium, and b) the hole has an inertial mass (acceleration in response to an applied force) different than that of a free electron due to the periodic background potential of the germanium lattice. On length scales greater than an interatomic spacing, electric fields are reduced by a factor of the dielectric constant ϵ , ($= 16$ for germanium,) due to the polarizability of the lattice. The mass of the bound charge is reduced by a factor of $\langle m^* \rangle$, an average of the effective mass tensor (normalized to the free electron mass) over direction. The appropriate directional average for computing the energy spectrum is the conductivity effective mass (Sze 1981) given by $\langle m^* \rangle = 3(1/m_1^* + 1/m_2^* + 1/m_3^*)^{-1}$, (where m_i^* is the component of the effective mass tensor along its i th principal axis), and numerically equal to .34 times the free electron

mass. Thus, the hydrogen atom analogy yields :

$$V(r) = -\frac{e^2}{r} \rightarrow -\frac{e^2}{\epsilon r} \quad (2.1a)$$

$$m \rightarrow \langle m^* \rangle \quad (2.1b)$$

$$13.6 \text{ eV} = \frac{me^4}{2\hbar^2} \rightarrow \frac{\langle m^* \rangle e^4}{2\hbar^2 \epsilon^2} \approx 10 \text{ meV} \quad (2.1c)$$

$$912 \text{ \AA} \rightarrow 124 \text{ } \mu\text{m} \quad (2.1d)$$

This simple scaling argument accounts for the well known long-wavelength cutoff of conventional (i.e. unstressed) photoconductors at $\sim 120 \text{ } \mu\text{m}$.

The hydrogen atom analogy assumes an impurity potential $U(r) = -\frac{e^2}{\epsilon r}$ everywhere. This is a good approximation because the spatial extent of even the ground state impurity wavefunction, or “scaled Bohr radius”,

$$0.53 \text{ \AA} = \frac{\hbar^2}{me^2} \rightarrow \frac{\hbar^2 \epsilon}{\langle m^* \rangle e^2} = 44 \text{ \AA}, \quad (2.1e)$$

covers a great many (of order 10^5) germanium atoms, and their screening effect may therefore be approximated as that of the bulk material. Within a few interatomic spacings of the impurity atom, however, the dielectric screening will be reduced, and the impurity’s potential well will be deeper than $-e^2/\epsilon r$. States with a large amplitude near the origin (i.e. s states, and in particular the ground state,) are therefore expected to be slightly more tightly bound.

In addition to the ground state, the hydrogen atom analogy predicts the details of the spectrum of excited states. A large number of these excited states have been identified with observed peaks in photoconductivity spectra (Ramdas and Rodriguez 1981, McMurray *et al.* 1986). As one proceeds to shallower excited states, their spatial extents become larger. Eventually, their wavefunctions begin to overlap, and the isolated hydrogen atom is no longer a valid analogy. Wavefunction overlap also imposes an upper limit on the impurity density for which the hydrogen atom analogy is valid. Above this critical concentration, the impurity states merge into a band and become de-localized. This critical concentration, numerically, in the range of 10^{16} cm^{-3} , is considerably higher than the doping of any of our detectors, but is highly relevant to the operation of a newly developed device, the blocked impurity band (BIB) detector.

In addition to the spectrum of energy eigenstates, other properties of the photoconductors that are relevant in our application are the mobility, both its absolute magnitude and its electric field dependence, the breakdown field, or impact ionization cross-section, and the carrier lifetime, or recombination cross-section. We discuss each of these in turn. Most of the basic theoretical and experimental work on these basic physical properties of semiconductors was done in the 50's and early 60's. Recently, there has been some renewal of interest in the basic physics (that is, apart from interest in device optimization), due to the usefulness of extrinsic photoconductors as a laboratory testbed for non-linear dynamics. In their analysis of the non-linear dynamics of the devices, Westervelt and Teitsworth (1985, henceforth WT,) in order to justify the equations of motion they adopt, review many aspects of the basic theory. In some cases we follow their treatment, and in all cases we note the formula they adopted for each of these properties. In some cases, however, examination of the original papers reveals that the formulae adopted by Westervelt and Teitsworth are not justified, or do not apply to our detectors. This probably does not affect the qualitative behavior of the solutions of their equations of motion, however, or their basic conclusions.

Mobility

The mobility of relatively impure germanium, such as ours, is dominated at liquid-helium temperature by ionized impurity scattering. The ionized impurities exist because there is always some finite concentration of compensating impurities, whether deliberately introduced or not. Thus, in p-type germanium, there are positively charged donors, D^+ , (mainly phosphorus in our detectors), and an equal number of ionized acceptors, A^- . In addition, there are, of course, the extra neutral acceptors, A^0 , that are "left over". When the detector is far from saturation, a very small fraction of the latter are ionized due to photoexcitation. In the "scaled Bohr atom" picture, ionized impurity scattering is the analogue of Rutherford scattering of an electron off a charged ion. It is an elastic process; in it, the carrier's momentum is relaxed, but its energy is conserved.

Another important process is scattering off acoustic phonons. At higher temperatures, roughly $T > 20$ K in germanium, and for very low compensation material ($N_D < 10^{12}$ cm $^{-3}$), acoustic phonon scattering dominates ionized impurity scattering and is the mechanism that determines the total mobility.

It is important even for our detectors, at liquid helium temperatures, because it is an inelastic process, and is the dominant source of *energy* – relaxation for the carriers. It is therefore critical in determining the recombination cross-section, (since carrier recombination is an inelastic process), and in determining the effective temperature of the holes.

The theoretical description of ionized impurity scattering was first worked out by Conwell and Weisskopf (1950). They derived the following formula for the conductivity :

$$\sigma_I = n \frac{2\epsilon^2 m^{*2}}{3n_I e^2 kT} \left(\frac{m}{2\pi kT} \right)^{3/2} \left[\ln \left(1 + \frac{36\epsilon^2 d^2 (kT)^2}{e^4} \right) \right]^{-1} \int_0^\infty v^7 e^{-\frac{mv^2}{kT}} dv \quad (2.2)$$

where $d = \frac{1}{2}n_I^{-1/3}$ is half the mean distance between the ionized impurity centers, whose concentration is n_I , and n is the concentration of free carriers. Using $\sigma_I = ne\mu_I$, and evaluating the integral yields the mobility due to ionized impurity scattering :

$$\mu_I = 8\sqrt{2} \left(\frac{\epsilon^2 m^*}{n_I e^3} \right) \left(\frac{kT}{\pi m^*} \right)^{3/2} \left[\ln \left(1 + \left(\frac{3\epsilon kT}{n_I^{1/3} e^2} \right)^2 \right) \right]^{-1}, \quad (2.3)$$

as quoted by Debye and Conwell (1954). The formula quoted by WT differs from this in that the squared term in the argument to the logarithm is only taken to the first power, and in that a factor of $2\sqrt{2}$ is missing from the prefactor. These appear to be simple errors in transcription of the Conwell and Weisskopf result. They amount to a fairly serious numerical discrepancy, of about a factor of 5, for a representative one of our detectors. The original Conwell and Weisskopf result is smaller, and much closer to our measured Hall mobilities.

The Conwell and Weisskopf result is purely classical, and, (oft-cited though it is), it incorporates some rather crude assumptions. It results simply from application of the Rutherford formula to a collection of scattering centers that are assumed to be independent and non-overlapping. The integration over impact parameter is arbitrarily cut off at half the mean separation between the scattering centers. A more sophisticated, quantum mechanical treatment was developed by Herring, (private communication quoted by Debye and Conwell, 1954) and independently by Brooks (1951). It assumes that the potential for each scattering center is again Coulombic, but is screened by the mobile carriers, of density n . The potential from a random spatial distribution of such scattering centers is Fourier analyzed in order to derive the scattering

matrix element. The result is :

$$\mu_I = 8\sqrt{2} \left(\frac{\epsilon^2 m^*}{n_I e^3} \right) \left(\frac{kT}{\pi m^*} \right)^{3/2} \left[\ln(1+b) - \frac{b}{1+b} \right]^{-1}, \quad (2.4)$$

where

$$b \equiv \frac{6}{\pi} \left(\frac{em^*}{n} \right) \left(\frac{kT}{\hbar e} \right)^2.$$

This is identical to the Conwell-Weisskopf formula except for the term in brackets. The variation of mobility with compensating impurity concentration, as predicted by these models, is shown in figure 2.2. The Conwell-Weisskopf formula is the dashed line, while the solid lines show the Brooks-Herring result for various carrier concentrations. The Hall mobility measurements on our 729 series detectors are also shown. When $n \approx n_I$, i.e. when most of the free carriers are due to photoionization of neutrals rather than to compensation, then the two formulae yield similar values, although the Brooks - Herring result predicts a somewhat faster rolloff of mobility with ionized impurity concentration. At very low carrier concentrations, however, the reduced screening implies a substantially lower mobility (by about an order of magnitude) in the Brooks - Herring treatment. The Brooks - Herring result implies that the mobility weakly depends, through the carrier concentration, on FIR illumination. The dependence may be described in practical units by noting that the lower limit on carrier concentration is given by thermal excitation :

$$n \geq N_A e^{-\frac{E_f}{kT}}. \quad (2.5)$$

The carrier concentration due to FIR illumination may be estimated by

$$n = \eta \dot{N} \tau_r, \quad (2.6)$$

where η is the quantum efficiency, \dot{N} is the rate of incidence of FIR photons, and τ_r the recombination time. Typical values for our detectors might be $\eta = .1$, $\tau_r = 5$ nsec, and $N_A = 2 \times 10^{14} \text{cm}^{-3}$. For these parameters, the lowest curve shown in figure 2.2, (for $n = 200 \text{cm}^{-3}$), corresponds to .8 nW at $100 \mu\text{m}$. Since the highest power levels relevant to our application are some 6 - 7 orders of magnitude larger than this, even the weak dependence of the Brooks - Herring formula predicts a somewhat significant effect.

The theory of acoustic phonon scattering developed by Bardeen and Shockley (1950) predicts a

Mobility vs. Compensation

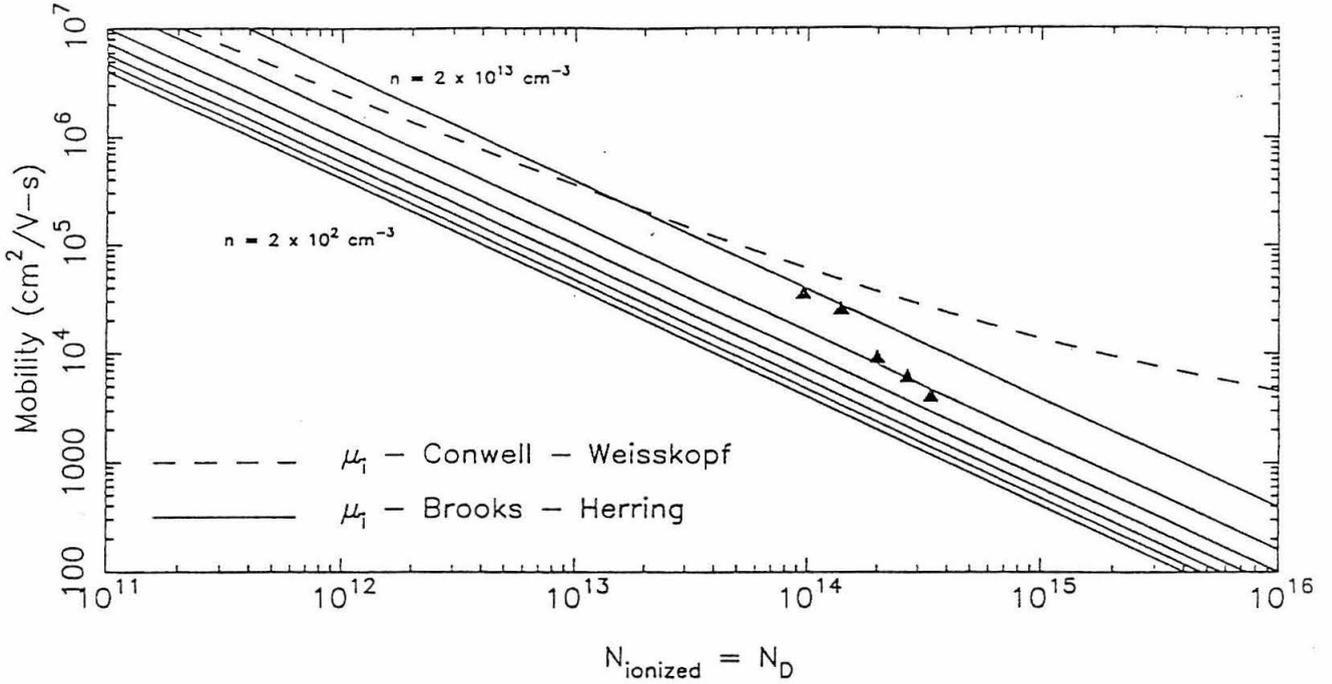


Figure 2.2 – Compensation dependence of the mobility due to ionized impurity scattering, showing the theoretical results and the Hall mobilities measured on our 729-series detectors (I. S. Park, private communication.) The lowest concentration for which the Brooks-Herring result is shown, $n = 2 \times 10^2 \text{ cm}^{-3}$, corresponds to the density of thermally generated carriers at 4.2 K (equation 2.5). Successive curves are separated by a factor of 100 in carrier concentration.

mobility that decreases with the 3/2-power of temperature :

$$\mu_{ph} = \frac{\sqrt{8\pi}}{3} \frac{\hbar^4 c_l}{\varphi_n^2 m^{*5/2} (kT)^{3/2}} \quad (2.7)$$

where c_l is the average longitudinal elastic constant and φ is the shift of the conduction band edge per unit fractional dilation. The numerical result used by WT is a fit to the experimental data of Norton and Levenstein (1972) on Cu-doped germanium :

$$\mu_{ph} = 4 \times 10^7 T_K^{-3/2} \text{ cm}^2/\text{V} - \text{s}. \quad (2.8)$$

The $T^{-3/2}$ dependence of the acoustic phonon mobility and the $T^{3/2}$ dependence of ionized impurity scattering combine to produce a temperature dependence for the overall mobility which peaks at the tem-

perature where μ_I and μ_{ph} are approximately equal. Measured Hall mobilities as a function of temperature for our detectors obey this dependence quite closely, peaking at about 20 K. At 4.2 K, typical measured mobilities are $\sim 10^5$ cm²/V-s for the more lightly compensated detectors, and $\sim 10^4$ cm²/V-s or less for the more heavily compensated ones. The measured values of the acoustic phonon component, μ_{ph} , for our detectors, obtained by extrapolation from the high temperature regime, are somewhat lower than the $4 \times 10^7 T_K^{-3/2}$ mentioned above, and have a weak dependence on doping (as was found by Norton and Levenstein, 1972). For our detectors, μ_{ph} varies from $\sim .8$ to $2 \times 10^7 T_K^{3/2}$ for N_A in the range 1.2 to 10×10^{14} . Note that Hall mobilities differ from drift mobilities by a constant factor which depends on the carrier heating (Sze 1981, p. 34), but which is generally close to unity. Due to the difficulty of accurately calculating this correction, however, (equal to $\langle \tau^2 \rangle / \langle \tau \rangle^2$, where τ is the carrier scattering time and brackets denote an average over the carriers' energy distribution function), we have simply applied the measured Hall mobilities directly in the analyses of this chapter.

So far, we have only discussed the low-field mobility. Under the conditions in which we generally operate the photoconductors, however, carrier transport is a much more complicated problem, because the carriers are "hot". That is, the width of their velocity distribution, (i.e. their temperature), is greater than that of a thermal distribution at the lattice temperature T_l . The field dependence of mobility in n-type germanium was extensively studied at 4.2 K by Koenig, Brown, and Schillinger (1962, henceforth KBS). By studying the current-voltage characteristics of their samples with very high speed pulses, they were able to separate out the variation of mobility and carrier density at "high" (i.e. close to breakdown) fields. They found that the mobility was constant at low fields and smoothly approached $\mu(E) \propto E^{-1/2}$ at high fields, as predicted for acoustic phonon scattering, (Schockley 1951). This behavior may be understood in terms of drifted Maxwellian velocity distributions. At zero applied field, the drift velocity $v_d = 0$, and the "typical" carrier velocity, i.e. the velocity that characterizes the carrier kinetic energy, is $v_T = \left(\frac{2kT_l}{m^*}\right)^{1/2}$. At small fields, the center of the distribution function is given by $v_d = \mu_I E$, and the kinetic energy (i.e. broadening) added to the distribution is characterized by $\frac{m^*}{2}(\mu_{ph} E)^2$. Even though $\mu_{ph} > \mu_I$, the added kinetic energy is still only a small fraction of the thermal kinetic energy. At high fields, $\mu_{ph} E > \left(\frac{2kT_l}{m^*}\right)^{1/2}$, the shape of the distribution is significantly perturbed. The relation between velocity and field is found by

equating the rate at which energy is imparted to the carriers by the field with the rate at which energy is lost via inelastic collisions. I.e.

$$eE v_d = \left(\frac{1}{2} m^* v_T^2 \right) \left(\frac{v_T}{l_i} \right). \quad (2.9)$$

Here, l_i is the inelastic mean free path. The drift velocity is given by the acceleration due to the field over an elastic scattering time :

$$v_d = \frac{eE}{m^*} \left(\frac{l_e}{v_T} \right), \quad (2.10)$$

so that, combining 2.9 and 2.10,

$$\rightarrow v_T \propto E^{1/2} \quad v_d \propto E^{1/2}.$$

Another way of expressing this behavior is to say that, in the hot-carrier regime, the effective carrier temperature, T_h , is proportional to field.

Breakdown Field

“Breakdown” in germanium photoconductors refers to an increase in carrier concentration of many orders of magnitude which occurs over an extremely narrow range of bias field, above some critical “breakdown field”, E_b . It results from impact ionization of neutral acceptors by field-accelerated carriers. The process is essential in Westervelt and Teitsworth’s modeling of the chaotic dynamics of photoconductors, because it provides a highly non-linear coupling between the equations of motion for carrier concentration and electric field.

No adequate first-principles calculation of the breakdown field exists in the temperature and compensation regime that applies to our detectors. The theoretical treatments which have been developed define breakdown to occur when the rate of impact ionization, which increases strongly with field, equals the rate of recombination, which decreases relatively slowly with field. The rate of impact ionization is the difficult part of the calculation, since it involves the convolution of the (velocity-dependent) impact ionization cross-section with the carrier velocity distribution. Thus, the result can depend critically on the high-energy tail of the distribution, (since $E_I \gg kT_h$). Zylberstejn (1962) calculated E_b under the assumption that acoustic phonon scattering was dominant. This assumption makes the theory tractable

because it decouples the problems of determining the carrier distribution function, (for which Zylberstejn used the results of Stratton, 1957,) and impact ionization. That is to say, carriers in the high-energy tail of the distribution function, that have $E > E_I$, lose most of their energy to acoustic phonons, and the energy exchange that occurs in impact ionization may be treated as a small perturbation which does not affect the distribution itself. Zylberstejn derived values of breakdown field in the range of $E_b =$ a few V/cm for mobilities $\mu_{ph} \sim 10^6 \text{ cm}^2/\text{V} - \text{s}$. Zylberstejn's theory compared very well with the data of KBS. Their detectors covered a range of compensation fairly comparable to our detectors, but, in addition to being n-type rather than p-type, their overall (i.e. majority) doping levels were much lower than ours. Specifically, their detectors had majority concentrations of 1×10^{12} to 3×10^{13} , about two orders of magnitude lower than the range covered by our detectors. (See table 2.2.) Thus, Zylberstejn's acoustic phonon assumption was much better satisfied for their detectors. Cohen and Landsberg (1967) also calculated the effect of compensation on breakdown field, in the limit that $E_I \ll kT$, and with a somewhat more general technique than Zylberstejn's. Their analysis is quite complicated, but the result they derive is that E_b varies from about 20 to 200 V/cm, as the compensation ratio, $\frac{N_D}{N_A}$ is increased from $\sim 1.$ to $\sim .9$. They do not present any results for the temperature dependence of E_b .

WT adopt an empirical approach based on a drifted Maxwellian velocity distribution. The total recombination rate is proportional to N_{A^-}/N_A , and the impact ionization rate to N_{A^0}/N_A . This determines the dependence on compensation. The drifted Maxwellian yields an impact ionization coefficient of

$$\kappa = \kappa_0 \frac{x^{1/2}}{1+x} \left[\exp \left(\frac{E_I}{kT} \left(\frac{1-x}{1+\alpha x} \right) \right) + 1 \right]^{-1} \quad (2.11)$$

where

$$x \equiv \frac{m^* v_d^2}{E_I},$$

is the kinetic energy due to the carriers' drift velocity, and

$$\alpha \equiv \frac{2\eta}{3} \left(\frac{E_I}{kT} \right).$$

The empirical constant η measures the ratio of spreading to drift in the velocity distribution. Near breakdown, and acoustic phonon scattering dominates both spreading and drift, and $\eta \sim 1$. The recombination

rate adopted by WT depends on field (see below) according to $r = r_0 \left(1 + \frac{\eta+1}{3} \left(\frac{E_I}{kT}\right) x\right)^{-3/2}$, which yields the following condition for breakdown :

$$\left(\frac{x^{1/2}}{1+x}\right) \left(1 + \frac{\eta+1}{3} \left(\frac{E_I}{kT}\right) x\right)^{3/2} \left[1 + \exp\left(\frac{E_I}{kT} \left(\frac{1-x}{1+\alpha x}\right)\right)\right]^{-1} = \left(\frac{r_0}{\kappa_0}\right) \left(\frac{N_D}{N_A - N_D}\right) \quad (2.12)$$

The ratio of cross-sections, $\frac{r_0}{\kappa_0}$, is essentially a free parameter in this treatment, but is expected to be near unity. WT adopt a value of 1/2. The key feature of equation 2.12 is that the left hand side is an *extremely* steep function of x in the range of interest. It varies from 10^{-2} to greater than unity as x is raised from 1/4 to 1/2. Therefore, the results of this treatment are not badly approximated by the prescription that breakdown occurs at a constant value of x , (which we shall denote x_{crit} , and take to be roughly unity,) independent of compensation. For large changes in compensation there will be a logarithmic change in the value of x_{crit} , but for most of our detectors this is negligible. In short, WT's formulation predicts

$$\mu E_{br} = x_{crit} \left(\frac{2E_I}{m^*}\right)^{1/2} \approx 1 \times 10^7 \text{ cm/s}. \quad (2.13)$$

Thus, virtually all the dependence of breakdown field on compensation arises through the mobility, in WT's formulation.

Recombination Time,

The theoretical prediction of recombination cross-section in germanium is a complicated story. The complexity is due to the fact that recombination is an inelastic process, and therefore, the binding energy must be carried away by acoustic phonons (optical phonons are much too high in energy, $E_{opt} = 37 \text{ mV}$.) Since the impurity ionization energy is much greater than the energy of a typical phonon, $E_I \gg kT$, direct recombination into the ground state is a multi-phonon process, with correspondingly minute cross-section. Therefore, the dominant process is capture into highly excited states followed by a cascade of single phonon scatterings and emissions, as the carrier gradually diffuses into the ground state. Roughly speaking, there are two streams of theoretical work on this process of carrier recombination. The first was begun by the "giant trap" theory of Lax (1960), and has since been modified and revised by many authors, e.g. Ascarelli and Rodriguez (1961), who developed a quantum mechanical version, and Brown and Rodriguez (1967), who also incorporated the phonon polarization. The second, which takes a fundamentally different approach

to the problem, has been pursued by several workers in the Soviet Union, and has been comprehensively reviewed by Abakumov, Perel', and Yassievitch (1977, henceforth APY).

The Lax theory and its extensions are ugly theories. They all involve a summation over the impurity's excited states of $\sum_{n=2}^{\infty} \sigma_n \beta_n$, where β_n is the "sticking probability", the probability that an impurity in state n will eventually decay into the ground state rather than be ionized. The result of Lax's original treatment was

$$\sigma_r = \left\{ \frac{\pi}{9} \left[\ln \left(\frac{\gamma}{1.78 \delta} \right) + \frac{\delta}{\gamma} \right] \right\} \frac{1}{l_i} \left(\frac{e^2}{\epsilon k T} \right)^3 \left(\frac{m^* s^2}{k T} \right)^2 \quad (2.14)$$

where

$$\gamma \equiv \frac{k T}{m^* s^2}.$$

Here, s is the speed of sound, l_i is the mean free path for acoustic phonon scattering, and δ is a dimensionless upper cutoff to the integral for the binding energy, which must be solved for numerically in terms of γ . For practical cases, $4 < \delta < 10$. The term in brackets is actually an approximation valid only when $\gamma/\delta \gg 1$. However, Lax gives numerical solutions to his equations, which show that, in fact, at $\gamma/\delta = 2$, (corresponding approximately to $T = 4.2$ K,) equation 2.14 is still accurate to better than 10 %. Numerically, Lax's theory, (using his quoted values for parameters such as s , m^* , etc.), predicts $\sigma_r = 1.3 \times 10^{-12}$ cm² at 4.2 K. The results of Brown and Rodriguez (1967) cannot be expressed in an analytic form, but their numerical solution predicts a somewhat lower value, $\sigma_r = 3.5 \times 10^{-13}$ cm².

Abakumov, Perel', and Yassievitch (1977) state flatly that the Lax theory and its extensions are wrong. The basic difference between APY's theory and Lax's, stated in terms of Lax's language, has to do with the sticking probability at large n . Lax's theory supposes that the first capture event takes place to a level which has binding energy of order kT , i.e. $n \sim 5$. Lax assumed that recombination through higher excited states could be neglected since, for them, the sticking probability approached zero. It is true that the sticking probability must approach zero for large n , but in order to justify truncating the summation, it is necessary to assume (as Lax tacitly did) that the sticking probability approaches zero faster than the cross-section approaches infinity. APY contend that this stronger assumption is not valid, and that the dominant process is recombination through very highly excited states, with binding energy $\ll kT$. APY's treatment is based

on the Pitaevskii method for treating recombination in gases. Their final result, is extremely simple and intuitive. Indeed, it is *precisely* the scaled hydrogen atom analogue of the cross-section first calculated by J. J. Thompson (1924) for recombination of electrons onto hydrogen ions. The result can be written down by inspection :

$$\sigma_r = \pi r_T^2 \left(\frac{r_T}{l_0} \right), \quad (2.15)$$

where

$$r_T = \frac{e^2}{\epsilon k T}$$

is the radius from a center at which a carrier's binding energy would be kT . Thus, it is the radius at which a captured carrier has, on average, an even chance of staying bound. The factor $\frac{r_T}{l_0}$ is the probability that the carrier will collide with an acoustic phonon while it is within a range r_T of the center. For a capture to occur, such a collision is necessary in order to carry off the excess energy. Thompson interpreted l_0 to be the mean free path, i.e. he assumed that the energy and momentum relaxation times were the same and equal to the scattering time. The Russian theory is the same except that it recognizes that what is important is the energy relaxation time. It identifies l_0 as the mean distance traversed by a carrier in one energy relaxation time., i.e. $l_0 = v_T \tau_i$. l_0 is independent of temperature, and is related to the inelastic mean free path by a factor $l_0/l_i = \frac{kT}{2m^*s^2}$, since, in each phonon scattering the carrier loses energy $\sim m^*s^2$, so that it takes $\frac{kT}{2m^*s^2}$ scatterings for the carrier to lose $\sim 1/2$ its energy, $(kT/2)$. In short, APY's theory leads to

$$\sigma_r = \frac{4\pi}{3} \frac{1}{l_0} \left(\frac{e^2}{\epsilon k T} \right)^3 = \frac{1.1 \times 10^{-9}}{T_K^3} \text{ cm}^2 \quad (2.16)$$

where the numerical value assumes an energy relaxation length of $l_0 = 4.3 \times 10^{-3}$ cm (APY 1977). The factor of $4/3$ arises from a geometric average over paths within r_T of the scattering center.

Apart from a numerical factor of order unity, the Lax and APY theories differ by a factor of $\left(\frac{m^*s^2}{kT} \right)^2$. The question of the appropriate speed of sound to use is somewhat problematical, but assuming it lies somewhere between the values given by Lax for the longitudinal and transverse speeds, then $.3K < \frac{m^*s^2}{k} < .8K$. APY adopt a value of .73 K. Thus, at 4.2 K, there is at least a factor of 30 difference

between the recombination cross-sections predicted by the two theories. We therefore expect it to be relatively easy to discriminate between them experimentally.

Finally, we note what the theories have to say about the field dependence and temperature dependence of the recombination time. Both theories predict a very steep increase in the recombination cross-section, and therefore a steep decrease in the bandwidth, as the detectors are cooled. APY, and the intuitive, scaled hydrogen atom view, predict a T^{-3} dependence of the cross-section, and therefore a $T^{-5/2}$ dependence of the bandwidth. Lax's theory predicts T^{-4} for the cross-section at high temperatures ($\gamma/\delta \gg 1$) and T^{-3} at low temperatures, where the bracketed term in equation 2.15 is no longer a valid approximation. This corresponds to a bandwidth varying as $T^{-7/2}$ at high temperatures and $T^{-5/2}$ at low temperatures. Brown and Rodriguez's modification of the Lax theory predicts a somewhat gentler temperature dependence : $\sigma_r \sim T^{-3}$ at high T and $\sigma_r \sim T^{-2}$ at low temperature. An obvious implication of any of these theories is substantially reduced bandwidth for stressed detectors, since they must be cooled to $T < 2$ K.

As with impact ionization, the field dependence of the recombination cross-section is governed by the total carrier velocity v_T , which rises only as the carriers are heated, at high bias. Thus, the recombination rate will be approximately constant at low bias, while at higher biases, v_T will go up and the cross-section go down. Lax's theory predicts a dependence of cross-section on total carrier kinetic energy, U, of $\sigma_r \propto U^{-2}$, i.e. $\sigma_r \propto v_T^{-4}$. Therefore, the recombination bandwidth is expected to vary as $\sigma_r v_T \propto v_T^{-3} \propto T_h^{-3/2}$. Since the carrier temperature is independent of field at low bias and proportional to field in the hot carrier-regime, this amounts to a bandwidth that is independent of field at low bias, and that varies as $E^{-3/2}$ at high bias.

Effect of uniaxial compression

In the absence of stress, the valence band maximum of germanium is fourfold degenerate, and the "split-off" band, separated by the spin-orbit interaction, is twofold degenerate. The bands may be thought of (in the tight binding approximation) as superpositions of $^2p_{3/2}$ and $^2p_{1/2}$ orbitals of the individual atoms. The "split-off" band may be ignored in FIR applications since its separation from the valence band maximum (290 meV) is so much larger than the extrinsic gap. The valence band maximum is comprised of two degenerate bands with different curvatures (i.e. different effective masses), the "light" and "heavy"

hole bands, each of which is in turn twofold degenerate. This latter degeneracy is due to the inversion symmetry of the diamond-type crystal structure.

When a uniaxial stress is applied, the degeneracy between the light and heavy hole bands is lifted. Since the inversion symmetry of the unit cell is preserved, each band remains twofold degenerate. Pikus and Bir (1960) calculated the effect of stress on the valence band structure by adding a “deformation potential”, $\sum_{ij} V_{ij} \epsilon_{ij}$ to the usual periodic crystal potential and applying degenerate perturbation theory to the Bloch solutions. Here, ϵ_{ij} is the tensor describing the fractional deformation of the unit cell. Their result for the hole energy surfaces was

$$E_{l,h} = Ak^2 + a\Delta \pm \sqrt{E_k + E_{ek} + E_e} \quad (2.17)$$

where the plus sign describes the light holes and the minus sign the heavy. E_k is the usual expression (e.g. Kittel p. 224) describing the shape of the energy surfaces at zero deformation :

$$E_k^2 = B^2k^4 + C^2(k_x^2k_y^2 + k_x^2k_z^2 + k_y^2k_z^2). \quad (2.18)$$

E_{ek} describes the change in the shape of the energy surfaces with deformation, and consists of a sum of terms quadratic in k and linear in ϵ . The term of interest to us is E_e , which describes the shift in the maximum ($k = 0$) point of the energy surface with stress :

$$E_e = \frac{b^2}{2} [(\epsilon_{xx} - \epsilon_{yy})^2 + (\epsilon_{xx} - \epsilon_{zz})^2 + (\epsilon_{yy} - \epsilon_{zz})^2] + d^2 [\epsilon_{xy}^2 + \epsilon_{xz}^2 + \epsilon_{yz}^2]. \quad (2.19)$$

The material constants b and d describe the energy shifts due to compression and strain respectively. Jones and Fisher (1970) determined their values by fitting the piezospectroscopic shifts of several (excited state) lines of p-type impurities in germanium. They obtained $b = -2.0 \pm .2$ eV and $d = -3.6 \pm .7$ eV for Ge:Ga, in agreement with the earlier, less direct measurements of Hall (1962). The direction of the energy shift is that the light hole energy increases and the heavy hole energy decreases.

The practical result of the splitting of the valence band maximum is to reduce the acceptor binding energy E_I . At the maximum stress that may be safely applied to the crystal, approximately 6000 - 8000 kg/cm², E_I is reduced from 10 meV to approximately 6 meV (Kazanskii, Richards, and Haller, 1977).

Thus, the long-wavelength cutoff for extrinsic photoconduction may be extended to about $200 \mu\text{m}$. The high responsivity and excellent NEP of the devices is not degraded by the application of stress. A practical consideration is that the reduction in binding energy is sufficient to make thermal excitation of carriers significant at 4.2 K. Numerically, it may be seen from equation 2.5 and 2.6 that the thermal excitation at 4.2 K is comparable to photoexcitation at incident power levels on the order of a microwatt. Therefore, for stressed operation, it is necessary to cool the detectors to pumped liquid-helium temperature ($\sim 1.5 \text{ K}$).

2.2 Experimental Details

The way in which our detectors are interfaced to the rest of the system is illustrated in figure 2.3. The diplexed local oscillator (LO) and signal beams pass through a vacuum window, typically a thin sheet of mylar, and are focused onto the entrance aperture of an optical “integrating cavity” in which the detector is mounted. Low-pass filters at cryogenic temperature block out the near-infrared and visible component of the incident radiation, thereby eliminating intrinsic photoconduction. They also reduce the far-infrared background due to room-temperature blackbody radiation. One of the detector contacts is electrically grounded. The other is soldered to the IF line, which is brought out of the integrating cavity via an electrical feedthrough. The high frequency component of the detector current, (i.e. the IF signal), is fed, through a blocking capacitor, into the first stage IF pre-amplifier. The IF signal is thence brought out of the cryostat and further amplified and processed. The DC and audio frequency components of the detector current are coupled to the audio circuit through a pair of fairly large (compared to the detector impedance) resistors. This somewhat complicated electronic arrangement is dictated by the relatively high impedance of the detectors under normal operating conditions. Because of this, parasitic capacitance to ground is a significant problem.

In this section, we describe the detailed construction and performance of the audio circuits, IF circuits, detector cavities, and FIR filters used in our experiments. In some cases, the versions actually used turned out not to be optimal because the LO power level turned out to be so much less than originally expected. For our laboratory measurements of bandwidth, photoconductive gain, and heterodyne noise temperature, unstressed detectors, a cryogenic GaAsFET pre-amplifier, and an early, non-optimized version of the audio circuit were used. For the airborne receiver, a stressed detector and detector cavity were used. The stressed detector had much lower bandwidth than the unstressed detectors; therefore, a room-temperature Si JFET pre-amplifier was used.

Audio circuit

The low-frequency circuit shown in figure 2.4 was used to provide DC bias to the detector and to monitor, at audio frequencies, the detector photocurrent. In some cases IC 2 was replaced by a battery and

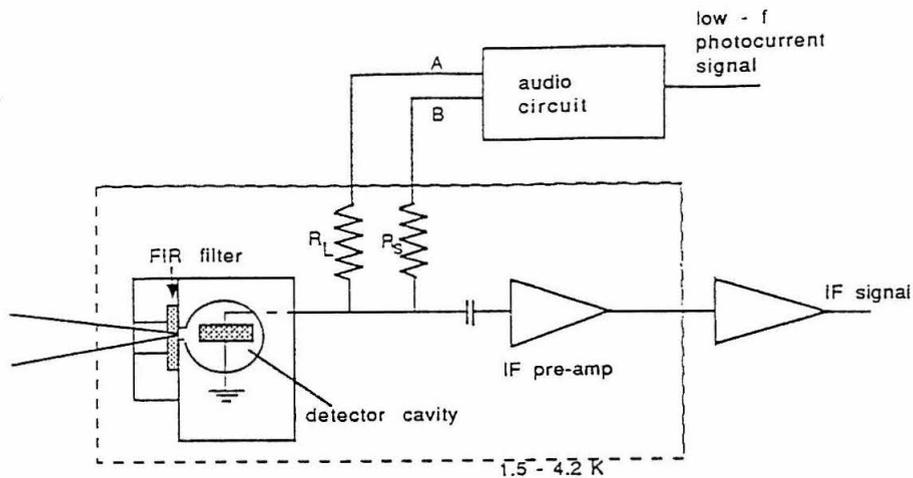


Figure 2.3 – Incorporation of Ge:Ga photoconductors into our receiver system. For the airborne version, the first-stage IF pre-amplifier is at room-temperature, and the detector cavity is designed to apply a large uniaxial stress on the detector.

an adjustable resistive voltage divider whose output was connected to R_L . This same circuit is often used for the biasing of superconducting tunnel junction mixers (Woody, private communication.) The effect of this circuit is to hold the (low-frequency) detector voltage fixed, regardless of detector impedance. The detector voltage is sensed through R_S and fed back to the inverting input of IC 1, which provides whatever current and voltage are necessary to hold the detector voltage equal to the control voltage appearing at the non-inverting input. The two low-noise, JFET-input, buffer amplifiers (IC 2 and IC 3), and the differential amplifier produce an output voltage V_{RL} equal to the programmed gain, G , times the voltage appearing across the load resistor. The latter is just the detector current times R_L , since the input impedance of the op-amps is much greater than the detector impedance. The key characteristic of this circuit is that it presents a low impedance (i.e. voltage bias) to the detector at DC and audio frequencies, and a high impedance, approximately $R_L \parallel R_S \approx R_L$, at the IF frequency. A voltage bias is not absolutely essential to the performance of the instrument, but it is desirable for the interpretation of photoconductor measurements. This is because the mobility is a function of electric field, rather than current density, so that if the photoconductor were current-biased, the mobility, and therefore the photoconductive gain, would depend

on incident FIR power level in a complicated way. Generation-recombination (g.r.) noise level would not be linearly proportional to laser power, and interpretation of much of our data would be considerably more difficult. A high impedance at the IF frequency is required, firstly to prevent the addition of significant current noise to the IF signal, and secondly, to prevent the audio circuit from shorting out the relatively high impedance IF signal. At high frequencies, the feedback of the audio circuit is rolled off by C_1 , as well as by the internal compensation of IC 1 (whose gain typically falls below unity at 5 MHz or below,) so that the detector simply sees the resistor R_L to ground.

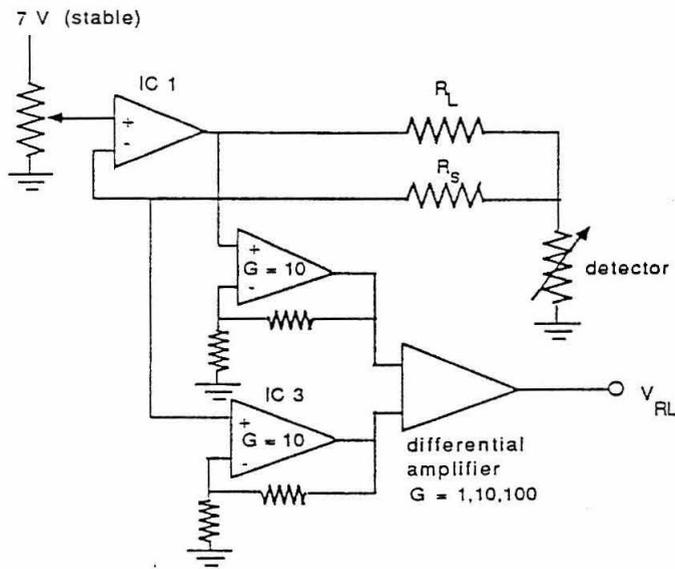


Figure 2.4 – Circuit used to provide DC detector bias and to monitor the detector current at audio frequencies.

The resistors R_L and R_S must themselves have low parasitics, and must be located as close as possible to the mixer in order to avoid stray capacitance to ground due to the bias lines. We use ordinary microwave metal film resistors in chip form (TRX Corp., Attleboro Falls, MA) which are located inside the preamp case, just before the preamp's input blocking capacitor. Their resistance changes significantly when cooled to 4.2 K, but can be measured *in situ* by removing the room-temperature bias circuit and measuring the resistances from the two dewar feedthroughs (points A and B) to ground, and between A and B, taking

care to measure all resistances at the same current.

The direct detection noise-equivalent power (NEP) is an important parameter of the system because, in practice, the power emitted by the sideband generator is so low that its detection and optimization requires fairly high sensitivity. To give a numerical idea of the requirement, we note that the highest sideband power, measured at the detector, that we have yet achieved is 9.5 nW. Since this figure doesn't allow for the loss due to a beamsplitter, and since the process of optical alignment generally begins with a considerably lower signal, it is clear that the direct detection NEP of the system ought to be no greater than, say 10^{-11} W Hz^{-1/2}. This is, of course, several orders of magnitude worse than the performance of a system optimized for direct detection, based either on cryogenic bolometers or photoconductors, but it is well beyond the capability of any room-temperature detector.

The direct detection NEP is simply the RMS noise voltage of the audio circuit divided by the voltage responsivity. The voltage responsivity is the inherent current responsivity of the detector times the load resistance R_L . The RMS noise voltage of the audio circuit, referred to the input of the buffer amplifiers, is given by

$$\langle V_{RL}^2 \rangle = (4eI_dG)R_L^2 + \langle V_a^2 \rangle \quad (2.20)$$

The first term in this expression is the g.r. noise due to the detector current. In a system optimized for direct detection, the detector current is determined by the level of room-temperature background incident on the detector, and this term dominates. The amplifier noise, $\langle V_a^2 \rangle$, is the sum of the Johnson noise due to R_L and R_s , the input voltage noise of the buffer amplifiers, and the voltage noise appearing at the non-inverting input of IC 1. (When a battery bias is used, the latter is replaced by the room-temperature Johnson noise of the voltage divider.) In the bias circuit built for the airborne receiver, $\langle V_a^2 \rangle$ has been measured as the detector resistance, R_L , and R_s were varied, and the expected dependences verified. Under typical operating conditions, the noise is dominated by the input voltage noise of the buffer amplifiers (LF-356's) and is approximately 25 nV Hz^{-1/2}. This implies that

$$\text{NEP} = \frac{\langle V_a^2 \rangle}{R_L S} = \frac{1.3 \times 10^{-12}}{\left(\frac{R_L}{3.1k\Omega}\right)\left(\frac{S}{3.7A/W}\right)} \text{ W Hz}^{-1/2}, \quad (2.21)$$

where S is the current responsivity of the detector. We see that for the parameters of the airborne receiver, (shown in parentheses), the direct detection NEP satisfies our requirement, but only by a factor of a few. The NEP given by equation 2.21 is equal to the background-limited value for a background power of $\sim 5 \mu\text{W}$, (assuming $G = .23$, our current best estimate for the detector in the airborne system.) Further improvement in NEP is possible only by reducing the background power below this value with better FIR filtering and increasing R_L .

IF Circuitry

Both the signal and the detector's g-r noise appear as current generators in parallel with the detector impedance, the IF amplifier's input impedance, and the parasitic capacitance to ground. (The audio circuit may be ignored so long as R_L and R_s are much greater than the impedance of this parallel combination.) Under normal operating conditions, our mixers present an impedance much greater than 50Ω to the external circuit, and therefore require a high-input impedance IF pre-amplifier. A pre-amplifier located outside the dewar can pose a serious problem, because the cable leading to it from the detector will contribute a substantial parasitic capacitance. Assuming the length of the cable, L , is much less than a quarter wavelength at the IF frequency, it appears as a lumped capacitance of approximately 1 pF/cm , for typical 50Ω characteristic impedance cable. This produces an RC rolloff in the detector response of $\nu_{3 \text{ db}} = 160\text{MHz}/L_{\text{cm}}$ which, for practical cable lengths limits the IF to below 10 MHz. To achieve higher IF bandwidths, one is therefore driven to cryogenically cooled amplifiers, simply by the requirement of physically locating them as close to the detector as possible. On the other hand, both GaAs field-effect transistors (GaAsFETs) and some silicon metal-oxide-semiconductor FETs (MOSFETs) – the only readily available devices which operate at liquid helium temperatures – have considerable low-frequency, or “1/f”, components to their noise spectrum below about 10 MHz. As mentioned earlier, these considerations have led to a two-track strategy in our IF electronics design, with a cooled GaAsFET amplifier used for the high IF frequency applications, such as measuring the bandwidths of the compensated detectors, and a room-temperature amplifier used when a few MHz of bandwidth or less was required, such as in the airborne system with a stressed detector.

The stressed detector has a bandwidth of approximately 400 kHz, so a room-temperature amplifier is most suitable for it. The amplifier we have used is described at length elsewhere (Brown 1984), so we omit any detailed description of its design here. Its most relevant performance characteristics are listed in table 2.1. It is based on a parallel array of 8 stages, each of which has a silicon J-FET at the input which is operated in source-follower mode, so as to bootstrap out most of the gate-source capacitance, followed by a low-noise bipolar transistor operated in common-emitter mode. The only aspect of the amplifier's performance which is slightly tricky is that the amplifier will oscillate whenever the source capacitance is more than about half the amplifier's input capacitance, i.e. ~ 15 pf. Therefore, it is still essential to minimize the length of the IF cable leading out the cryostat.

Table 2.1 – Preamplifier Performance

	300 K Si J-FET	4.2 K GaAsFET
Bandpass (ignoring RC)	70 kHz - 15 MHz	500 kHz - 500 MHz
Power gain	27 db + $\log_{10} \left(\frac{Z_{IF}}{50\Omega} \right)$	6 db + $\log_{10} \left(\frac{Z_{IF}}{50\Omega} \right)$
Voltage noise, $\langle V_a^2 \rangle^{1/2}$.42 nV-Hz ^{-1/2}	.29 nV-Hz ^{-1/2} †
$(\langle V_a^2 \rangle / \langle i_a^2 \rangle)^{1/2}$	150 k Ω	$\gg R_1$ ‡
Input capacitance, C_a	28 pF (+ C_{cable})	~ 3 pF

† ignoring 1/f noise (see text)

‡ R_1 , as indicated in figure 2.5, is the resistor which sets the amplifier input impedance. Its value is varied for optimizing different experiments. The highest value we have used is 2.7 k Ω .

Our unstressed detectors have considerably higher bandwidths than our stressed detector. For our laboratory tests on them, therefore, a cryogenically cooled, single-stage GaAsFET amplifier was built. Such amplifiers have been, at least until the recent development of HEMT's, the premier devices for ultra-low noise applications in radio astronomy, and an extensive literature exists on them. (The definitive review of the device characteristics of GaAs MESFETs is Pucel, Haus, and Statz, 1975. The design and performance of specifically cryogenic amplifiers based upon them is reviewed by Weinreb, 1980.) Nearly

all of this work has been done at microwave frequencies, however, and relatively little data exists on either amplifier design or device characteristics at frequencies below 100 MHz, where we wished to use them. (Exceptions are Su, Rohdin, and Stolte, 1983, Petersen, Gupta, and Decker, 1983, and Richards, *et al.* 1986.) Our application is highly unusual due to the combination of low frequency, large fractional bandwidth, cryogenic operation, and high source impedance.

The key issue in achieving good noise performance from GaAsFET amplifiers at microwave frequencies is the impedance matching of the source and the FET. In practice, we have little control over the source matching in our application, simply because of our large fractional bandwidth and the low frequency. These eliminate any possibility of using stripline transformers, stub tuners, etc. which are the common tools at microwave frequencies. A more fundamental problem with GaAsFETs, however, is their “ $1/f$ ” noise. Measured low frequency noise spectra have an overall f^{-1} spectrum of noise power versus frequency, but, in addition, there is often a “bump” in the spectrum at a frequency of a few hundred kilohertz, so that at frequencies near the knee, (i.e. a few MHz) the slope is somewhat steeper than f^{-1} (Liu and Das, 1983). There is little published data on the low-frequency noise of these devices, particularly at cryogenic temperatures, and even less understanding of its sources. It is generally attributed to traps in the channel (Pucel, Haus, and Statz 1975, Hughes 1986.) Our strategy was simply to build the amplifier and test it for suitability in our system. We found that at 4.2 K the $1/f$ -knee (i.e. the frequency at which the total noise was double its value at high frequencies within our passband) was located at about 15-20 MHz. The noise is actually steeper than $1/f$ at the lower end of our passband. At room temperature, the $1/f$ noise was roughly unchanged in amplitude and spectral shape, but because the noise level at high frequency was higher at room temperature, by a factor of about 5, the “knee” frequency was reduced to about 10 MHz.

A schematic of the amplifier actually built is shown at the left of figure 2.5. It is the simplest circuit conceivable, a single-stage FET in common source configuration. R_D is chosen to yield a 50 Ω output impedance, for matching to the output cable and second stage amplifier, a commercial, low-noise, 50 Ω , bipolar amplifier (Miteq Corp., Hauppauge, N.Y.). The circuit was fabricated from stripline on dielectric substrate (RG Duroid, Rogers Corp. Tucson, AZ). Capacitors and resistors were low-parasitic,

microwave types in chip form (American Technical Ceramics, TRX Corp., respectively.) The FET used was a Mitsubishi MGF 1412, a standard device for ultra-low noise microwave applications (Weinreb, 1980). Other FET's tested gave poorer performance. The primary practical problem in debugging the amplifier was elimination of parasitic high-frequency oscillations. Because the FET has gain up to ~ 20 GHz, care must be taken to eliminate positive feedback to the input which can occur through small parasitic elements of the circuit components, (e.g. resistors that look like capacitors at some frequency,) through the grounds, capacitively coupled through the amplifier case, etc. It was for this reason that microwave components were used throughout, and that most of the amplifier case was lined with microwave absorber. It was found empirically, however, that the most effective means of suppressing such oscillations was inclusion of the small inductor at the FET gate, L_c . Generally, three or four turns of thin wire on a radius of 1-2 mm, (an estimated inductance of 5-10 nH) was sufficient. An additional practical problem had to do with power dissipation. The FET, being merely soldered to the traces of the circuitboard, was not very well coupled thermally to the helium bath. As a result, heat could be conducted down the IF lead and raise the temperature of the detector. To eliminate this, rather draconian thermal grounding, with external copper braids, of the connector which carried the IF line, was required.

The noise spectrum of the amplifier at room temperature and at 4.2 K, measured with ambient and liquid nitrogen-cooled 50Ω loads at the input, is shown at the right of figure 2.5. The base noise temperature is about 30 K. In addition, there is the $1/f$ noise component, with a knee at ~ 20 MHz. The voltage gain with 50Ω source and load impedances is given simply by $g_m R_2$. Thus, the power gain, at 4.2 K under these conditions is about 6 db (it varies somewhat with bias.) This is not the actual power gain during operation, however. Most of the amplifier's power gain lies in the impedance transformation it performs, bringing the high impedance signal, $R_{IF} = R_{det} \parallel R_a$, down to 50Ω . The total power gain is

$$G = G_{50} \left(\frac{R_{IF}}{50\Omega} \right) \approx 6 \text{ db} + \log_{10} \left(\frac{R_{IF}}{50\Omega} \right), \quad (2.22)$$

which might be some 20 db for typical detector impedances. This amplifier, whose performance is summarized in table 2.1 and figure 2.5, is the one that was used for all our measurements of mixer bandwidth and heterodyne noise temperature.

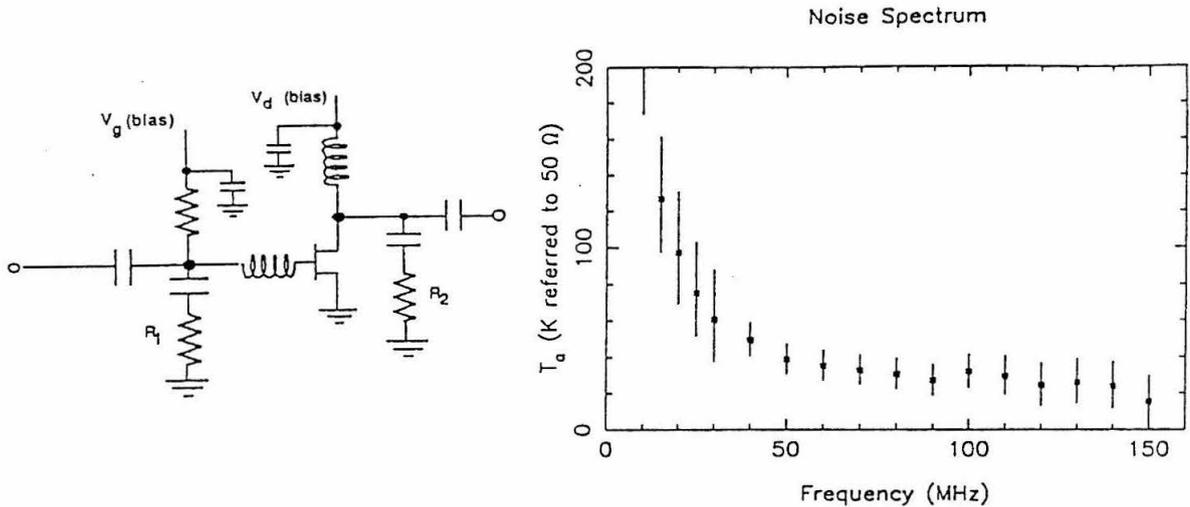


Figure 2.5 – (left) Schematic of the cryogenic GaAsFET pre-amplifier used in our laboratory experiments. (right) Measured noise of the amplifier at 4.2 K, with 50 Ω source and load impedances.

There are various strategies that might improve the IF performance for future versions of the receiver. Three possibilities for obtaining improved performance in high frequency (≥ 10 MHz) IF applications are replacement of the cooled MGF 1412 with a better GaAsFET or a high electron mobility transistor (HEMT), replacement of the cooled amplifier with a cooled transmission line transformer, and modification of the present low-frequency room-temperature amplifier so as to reduce its input capacitance and “bootstrap” out the cable capacitance. Although HEMTs have achieved microwave noise temperatures typically a factor of 3-4 lower than conventional GaAsFETs, the low frequency noise characteristics of the two devices are essentially identical (Das *et al.* 1985). An inherent problem with transmission line transformers is that the correct impedance transformation would be achieved for only one particular value of detector impedance, which would severely constrain the tuning of detector bias and LO power. Elimination of the cable capacitance using feedback from a room-temperature amplifier (“bootstrapping”) is possible, but would require very careful control of the phase versus frequency of the feedback signal, since, at the upper end of our desired passband, the cable’s electrical length would be approaching a quarter wavelength.

The most promising possibility, however, is the development of conventional GaAsFET transistors with lower $1/f$ noise. Recently, considerable success has been achieved in explaining the low frequency noise spectrum of GaAsFETs as a superposition of a few trapping levels (the most important lies at .75 eV) in a FET channel which has an inherent temperature gradient (Hughes, 1986). This understanding has been exploited to grow more trap-free GaAsFET channels, and dramatic reductions in $1/f$ noise have been reported. Hughes (1986) has found a reduction in the knee frequency of his devices from 40 MHz to approximately 1 MHz, and a reduction in the noise spectral density at 1 MHz of a factor of 20. Unfortunately, these devices were not available to us in time to be incorporated into our IF amplifiers.

The requirement on IF noise performance is given by comparison of the amplifier input noise with g.r. noise due to the detector current. Ideally, that current is dominated by the photocurrent due to the incident local oscillator power rather than by current due to room-temperature background radiation or thermal excitation, but as far as the IF amplifier is concerned, the source of the detector current is irrelevant. Explicitly, the requirement is :

$$\frac{\langle V_a^2 \rangle}{R_{IF}^2} \leq \langle i_{gr}^2 \rangle = 4eIG \quad (2.23)$$

$$R_{IF} = R_a \parallel R_d$$

Numerically, it turns out that this IF pre-amplifier requirement is fairly similar for the airborne receiver and for the receiver used in our laboratory measurements. There are wide variations depending on detector bias and incident FIR power level (see §2.3, on our detector results,) but roughly speaking, a typical detector current for the airborne configuration might be $1 \mu\text{A}$, and a typical photoconductive gain, .25. The detectors used in the laboratory system generally have photoconductive gains a factor of 5 – 10 lower, but are illuminated with higher powers. Typical photocurrents in our laboratory measurements were in the range of $10 \mu\text{A}$. In both cases, detector impedances were typically several $\text{k}\Omega$, but in the laboratory experiments, we often used an amplifier input impedance, R_a , that was somewhat lower. That is, noise performance was deliberately sacrificed for the sake of a flatter IF passband (higher R_{IFC} frequency). The usual IF impedance was several hundred Ω for bandwidth measurements, and $2.7 \text{ k}\Omega$ for the heterodyne noise measurements, in which optimum noise performance was desired. $1 \text{ k}\Omega$. These values lead to requirements on the noise spectral density of the pre-amplifier of $\langle V_a^2 \rangle^{1/2} < 4.0 \text{ nVHz}^{-1/2}$ for the airborne

system and $\langle V_a^2 \rangle^{1/2} < 2.0 \text{ nVHz}^{-1/2}$ for the lab system during noise measurements. Comparing with table 2.1, we see that, currently, the IF pre-amplifiers are not significantly limiting the system performance. For the lab measurements on detectors with the lowest photoconductive gain, and at lower incident power levels, this conclusion is somewhat marginal. Likewise, if the room-temperature background incident on the airborne detector were reduced by a large factor, improvement in the IF amplifier would be desirable. Of course, if a microwatt of LO power were available, as was originally intended, neither background reduction nor improvement in the IF amplifiers would be necessary.

Cavity Design

The design of the detector cavity and mixer block is important to the system performance for two reasons. Firstly, the receiver optics, i.e. the coupling of the detector to the telescope and the local oscillator beams, depends critically on the cavity design, and secondly, the responsive quantum efficiency depends on the cavity design. Here we discuss only the latter. The responsive quantum efficiency, η_r , is defined as the ratio of the rate of photogeneration of mobile carriers to the rate of incidence of FIR photons onto the cavity aperture. It consists of two factors, an “optical quantum efficiency”, η_o , defined as the fraction of incident photons which are absorbed in the photoconductor, and another efficiency, which we shall denote η_c , which is the fraction of absorbed photons which generate mobile carriers. The latter is not unity due to the existence of other, ill-understood, loss mechanisms in the photoconductor, e.g. excitation to bound excited states, phonon production, etc. It is not affected by the cavity design, however, and is simply a property of the material.

The optical quantum efficiency, η_o , depends on the cavity design in a complicated way because the mean absorption length of an FIR photon – typically 5 mm for an acceptor concentration of $n_a = 10^{14} \text{ cm}^{-3}$ – is comparable to or larger than typical detector dimensions. (This is in contrast to intrinsic photoconductors, whose typical absorption lengths are on the order of microns.) For this reason, some sort of optical cavity is required in order to obtain a longer effective path length through the photoconductor by means of multiple reflections. Two configurations have been widely used. One is the “endfire” detector, in which the incident photons travel along the long axis of the crystal, are totally internally reflected off the beveled rear face,

and make further reflections off the long walls – in this case, the detector itself acts as the optical cavity by means of total internal reflection. The other is the “integrating cavity,” in which a rectangular detector is mounted in a cavity, all of whose interior surfaces are good reflectors, and which has a small entrance aperture. Very little work has been done on systematically comparing these two designs or on finding the optimal detector and cavity dimensions for either. As a result of this lack of data, most workers in the infrared detector community hold very firm convictions as to the best method of cavity design.

The cavities we have used are of the “integrating cavity” variety. The usual way of explaining how they work is based on ray-tracing in the geometric optics limit. By employing a cylindrical cavity and orienting the detector faces to be skew to the plane of the aperture, the angles and positions of the photons may be considered randomized. In that case, so long as the aperture’s area is small compared to the detector dimensions, and so long as the cavity walls are perfect reflectors, “the only place the photons can be absorbed is the detector,” and the optical quantum efficiency is $\eta_o = 1$, *independent of the cavity’s shape and size.*

In fact, it is an open question whether or not this geometric optics viewpoint is legitimate. It implies that the optical efficiency increases monotonically as the entrance aperture’s radius decreases. This is the case because the only important quantity (in this viewpoint) is the mean number of passes a ray makes through the cavity before escaping back out the entrance aperture. One therefore expects that smaller apertures are always better. However, once the aperture size reaches the diffraction limit for the incident optical beam, $A \sim \lambda^2/\Omega$, the geometric optics approximation is invalid. For smaller apertures, the diffraction spot of the incident beam spills over the edge of the aperture and is partially vignetted. Furthermore, for such small aperture radii, the finite thickness of the aperture becomes significant. In our airborne cavity, the thickness of aluminum between the inner cavity and the outer face of the mixer block, though made as small as conventional machining techniques would allow, .075 mm, was still $3\lambda/4$ at 100 μm . Thus, at small radii, the entrance aperture is not a thin screen, but rather a short length of waveguide.

Thus, a sensible prescription for achieving maximum optical efficiency, and the one that we adopted, is to choose the radius of the aperture to just barely satisfy the condition of negligible vignetting for the

input beam. For the airborne system at $158 \mu\text{m}$, the input beam, treated as a Gaussian, has a waist size of $w_0 = .22 \text{ mm}$ ($f/3.7$ at the 3 db points). This has led to an aperture size of $r = .44 \text{ mm}$. However, there is an inherent inconsistency in this common-sense prescription. Because diffraction is guaranteed to be a significant effect when the prescription is used, the geometric optics approximation cannot be taken as valid. Therefore, there is no basis for believing the optical efficiency to be independent of cavity size and shape. Ideally, one would like to have the diffraction spot much smaller than the aperture area, so that geometric optics is valid and there is no dependence on cavity shape. In turn, one would also like the aperture area to be much smaller than the detector area, so that (in the geometric optics view) each ray makes many passes through the detector before escaping back out the cavity, and therefore the efficiency is high. These two conditions cannot simultaneously be satisfied without using excessively large detectors, which are undesirable due to their low photoconductive gain.

As far as theory goes, the only way to improve on this prescription would be to solve for the resonant modes of the cavity and for the modes of the input aperture, viewed as a waveguide, to compute their coupling by means of overlap integrals of the electric field in the aperture, to decompose a Gaussian input beam into a superposition of these modes, and then to compute the energy density of each mode in the volume occupied by the detector. Obviously, this would be a very ambitious undertaking. It would be easier to explore the dependence of efficiency on cavity size and shape experimentally. Small feedhorns could be placed outside the entrance aperture to vary the coupling between the free space mode and the cavity modes. Or, even better, a cavity in which one wall's position was tunable *in situ*, i.e. a cavity with a backshort, could also be used to explore, and perhaps exploit, the breakdown of the geometric optics prescription.

The mixer blocks we have actually used in the system are illustrated in figure 2.6. The unstressed design, shown on the left, is a copy of the integrating cavity arrangement used in many low-background direct detection applications, such as the Berkeley tandem Fabry-Perot spectrometer (Watson 1982). The detector is soldered by one of its metallized contacts onto a brass post using low temperature indium solder. The mixer block is intended for "side-looking" dewars, such as our low-background direct detection dewar

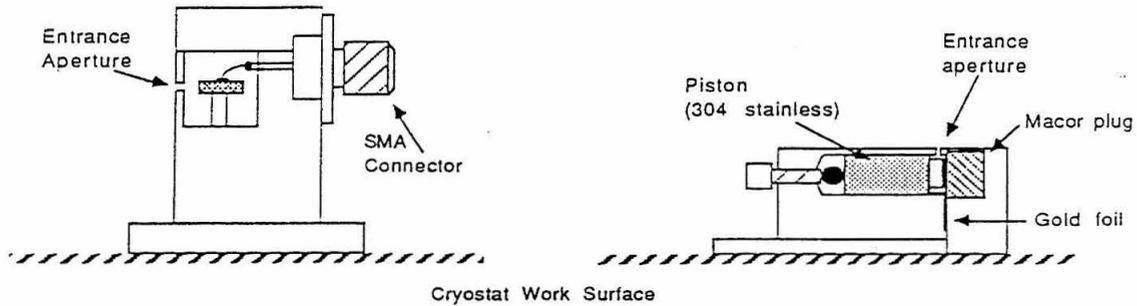


Figure 2.6 – Mixer blocks used (left) in our lab experiments with unstressed detectors, and (right) in the airborne receiver with a stressed detector.

(D-69) and our laboratory heterodyne dewar, (D-49). It is machined from oxygen-free copper and gold-plated in order to improve the detector's thermal grounding. Electrical contact to the live electrode is made via a short length of copper wire soldered on one end to the metallized contact, and on the other end to an SMA feedthrough (EMC Technology Corp., Cherry Hill, NJ). In D-49, a short SMA connector then transmits the IF signal (and DC bias) to the amplifier case.

The total parasitic capacitance of this arrangement is estimated to be about 3 pF, based on the rolloff of the measured amplifier noise spectrum with the SMA connector (length ~ 2 cm) in place, but no mixer block attached. This estimate is quite crude, since the rolloff is complicated by the presence of the amplifier's $1/f$ noise, which is hard to subtract out reliably; however, it agrees with our expectation based on the physical dimensions involved. The mixer block and detector probably add a small additional parasitic capacitance, but it is expected to be negligible by comparison. It is probably dominated by the capacitance between the center conductor of the SMA feedthrough and the hole in the mixer cavity which it slips into, since the dimensions were chosen to provide as close a fit as possible consistent with not shorting the detector to ground. This was done to prevent stray light leakage onto the detector in low-background applications. The

diameter of the entrance aperture machined into the cavity was 2 mm, which is almost certainly oversized for many applications, but there is the capability of mounting reflecting screens with various sized entrance apertures across the front of the mixer block. In most of the heterodyne experiments, however, the screens were not used.

The mixer block illustrated at the right of figure 2.6 was designed for the airborne dewar, which is down-looking, and can be used with either stressed or unstressed detectors. The grounded electrode of the detector rests upon a stainless steel piston which is driven from its far side by a screw and ball bearing. In order to avoid unnecessary stray capacitance, the surface against which the opposite electrode is pressed is not made of a thin layer of insulator backed by metal, as in stressed mixer blocks for direct detection, but is rather made of a large piece of Macor (a machinable ceramic manufactured by Corning Glass Works.) Stress is applied to the detector primarily through the differential thermal contraction of the germanium crystal and the surrounding mixer block, which is made of 6061 aluminum. As has been the experience of other workers, however, (Watson 1982, e.g.) the differential thermal contraction is not quite enough by itself to move the photoconductivity edge to 200 μm . Some amount of pre-stressing at room temperature is necessary. We have found empirically that applying enough stress to reduce the room temperature resistance of the sample by 5 % is sufficient. Figure 2.7 shows photoconductivity spectra obtained with the airborne system, with only the stress due to differential thermal contraction resulting in a photoconduction edge at $\sim 135 \mu\text{m}$, and with the 5 % pre-stress applied, resulting in an edge at 200 μm . The ripples are standing waves due to imperfect anti-reflection coating of the filters. The cutoff at 100 μm is due to the mesh-on-dielectric lowpass filter used.

In this mixer block, small pieces of .001" thick gold foil are placed between the detector contacts and the surfaces against which they are pressed. They serve several purposes. Firstly, because the gold flows slightly under pressure, they distribute the stress more evenly across the crystal. Secondly, the upper piece of gold forms the electrical contact to the IF and DC circuitry. It is soldered to a short length of wire, which leads out through a hole in the cavity "lid". Finally, it was found that adequate heatsinking of the detector was impossible without the gold foil between the grounded contact and the stainless steel piston.

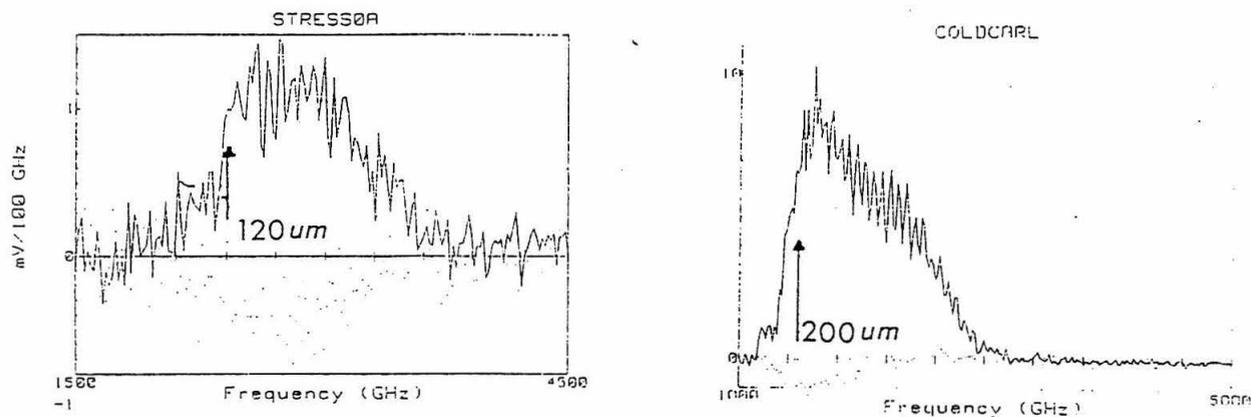


Figure 2.7 – Photoconductivity spectra obtained with a Fourier transform spectrometer on our stressed detector (LBL 82-4.6) with only the stress due to differential contraction applied (left), and with pre-stress applied at room-temperature (right). The former spectrum was taken at 4.2 K, where the high thermally induced g.r. noise reduced the signal to noise ratio.

This is not surprising, since both Macor and stainless steel are extremely poor thermal conductors at 4.2 K, and since the thermal path through the stainless steel piston must pass through the small area junctions between the spherical ball bearing and the flat piston and screwhead. Note also that the background power and the desired LO power are both of the order of microwatts, some two to three orders of magnitude larger than typical incident powers in low-background direct detection applications. Therefore, the gold foil placed between the detector's grounded contact and the piston was shaped and inserted so as to climb up the side of the cavity wall and then be firmly pressed between the mixer block and its lid. The thermal conductance of this heat path was crudely measured in the following way : A known amount of thermal energy was deposited in the detector by biasing it above breakdown for a known amount of time. The time constant for the detector to cool back down to the bath temperature was then measured by monitoring its resistivity as a function of time, and comparing with the resistivity versus temperature curve. The latter was

independently measured by cooling the bath from 4.2 to 1.5 K extremely slowly (over the course of, say, two hours,) and monitoring the detector resistivity and helium vapor pressure. The thermal time constant measured in this way turned out to be ~ 200 sec at 4.2 K. The heat capacity (i.e. specific heat C_V times the volume) of the detector plus stainless steel piston could then be estimated using the standard values of Debye temperature and electronic specific heat constant (*Materials at Low Temperatures* ed. Reed and Clark, chap. 4). The thermal conductance was then estimated from :

$$S = \frac{C_V V}{\tau} \approx 8 \mu\text{W/K}, \quad (2.24)$$

at 4.2 K. The same procedure at 1.5 K yielded $\sim 4 \mu\text{W/K}$, though with considerably larger uncertainty.

For the incident power levels of $\sim 1 \mu\text{W}$ which we expect (the background load in the final configuration for the airborne system is estimated to be 370 nW) this thermal conductance is adequate. Also, the heatsinking can be tested by checking the ratio of DC photocurrent with the detector facing room-temperature and liquid-nitrogen temperature loads (outside the dewar). A standard rule-of-thumb is that this ratio should be greater than two if the filters and detector are sufficiently cold, and the stressed detector configuration does indeed satisfy this criterion.

Great pains were taken to minimize the parasitic capacitance to ground of the IF signal line. With a cooled preamp, the total capacitance, including .25 pF from C_{gs} of the MGF 1412 transistor, would come to some 1.3 pF, corresponding to $\nu_{3db} = 120$ MHz for a 1 k Ω detector. Unfortunately, this effort turned out to be unnecessary since our stressed detector had such a narrow bandwidth that RC rolloff was not a significant problem.

FIR Filtering

Filtering the incident FIR radiation is important both for the NEP in direct detection and for the heterodyne performance. Because the filters are at cryogenic temperature, their emissivity is not important. Imperfect filter transmission simply attenuates the signal and local oscillator powers, but does not itself add noise. The attenuation of the local oscillator increases the LO power requirement, and the attenuation of the signal acts to lower the effective quantum efficiency for heterodyne detection. If the LO power transmitted

through the filter is much greater than the incident room-temperature background, then only the reduced effective quantum efficiency is important. However, in the LO-starved regime in which we have found ourselves, the heterodyne performance depends on the background power in a more complicated way. This is because the incident background level determines the detector resistance. Therefore, the importance of amplifier noise depends on incident background level. Also, the useful detector bandwidth will in some cases be RC limited, and therefore will depend on incident background level.

We have used two different types of lowpass (i.e. long-wavelength transmitting) filter. The first are interference filters made from capacitive grids evaporated on dielectric substrates. The second are restrahlen filters made from single-crystals of various halide salts. The capacitive grids were fabricated by J. Keene and are described at length by Whitcomb and Keene (1980). A capacitive grid is simply an array of metallic squares with dimensions and spacing comparable to the cutoff wavelength; it is the Babinet complement of the "inductive grid" formed by a mesh of metal wires. The grid periodicity, and to a lesser extent the "filling factor" of the squares, determines the cutoff frequency. Typically, several grids, each evaporated on a separate piece of polyethylene, are sandwiched together in order to suppress secondary transmission peaks in the stopband. The stressed detector scan shown in figure 2.7 was taken with a capacitive grid in place whose nominal cutoff wavelength was $100 \mu\text{m}$. As may be seen, the cutoff is quite sharp, and residual transmission in the region $10 \mu\text{m} < \lambda < 65 \mu\text{m}$, in which the grid is the only filtering element, is too small to be measured with our setup.

In some cases, we have used restrahlen filters in transmission as short wavelength blocking filters. As an historical aside, we mention that the restrahlen filter is one of the unsung heroes of modern physics. Originally developed at the Univ. of Berlin in the 1890's, (Rubens and Nichols 1897), it was the technical breakthrough that first allowed accurate measurement of the blackbody spectrum at wavelengths longward of the Wien's law regime. They are thus responsible for Planck's famous hypothesis about the analytical form of the spectrum, and therefore, indirectly, for the development of quantum theory. (See Pais, 1982, p.364 ff) The filters have been described from a modern point of view by Armstrong and Low (1973, 1974). They are based on the photoexcitation of transverse optical phonons for suppressing transmission

in the stopband. Because the crystals have relatively high indices of refraction, they must be anti-reflection coated to maintain high transmission in the passband. We have made all our anti-reflection coatings from polyethylene, on account of its low loss in the far-infrared, its ready availability in a large variety of thicknesses and its relatively good match to the refractive indices of most crystals. Our technique for applying it to the crystals is described by Watson (1982) and consists of melting the polyethylene sheet onto the crystal while the two are held under vacuum (to prevent bubbles from forming) and are sandwiched between layers of mylar. In addition to the capacitive grids and restrahten filters, we frequently also employ a liquid nitrogen temperature blocking filter of crystal quartz, coated with black polyethylene. The combination has high transmission in the FIR since the polyethylene serves as a good anti-reflection coating for the quartz. The carbon black in the polyethylene blocks all transmission of near-infrared and visible radiation from the detector.

The detector's geometric field of view is limited to 0.28 steradians by the aperture in the helium temperature radiation shield. The size of the aperture was chosen to satisfy the usual rule-of-thumb (Goldsmith 1982) for avoiding "significant" vignetting of a Gaussian beam, i.e. aperture radius greater than twice the $1/e$ radius of the beam, for the beam incident from our final focusing mirror, at $\lambda = 157 \mu\text{m}$. (The criterion is slightly better satisfied at $118 \mu\text{m}$.) The detector aperture's diameter is 0.89 mm. The power emitted by a room temperature blackbody with this throughput is $P_{bg} = 370 \mu\text{W}$ for a filter cutoff of $100 \mu\text{m}$, as seems approximately to apply to the grid filter used in taking the FTS spectrum shown in figure 2.7. With a cutoff of $65 \mu\text{m}$, which applied in some of our D-49 experiments in which only quartz and black polyethylene were used, the corresponding background is $P_{bg} = 1.3 \mu\text{W}$.

2.3 Experimental Results

We have tested a number of detectors for their performance in heterodyne applications, namely for their recombination bandwidths, photoconductive gains, and responsivity. Our data are most extensive and reliable on the matter of bandwidths. At the outset of the project, it was not known what bandwidth, even in order of magnitude, could be expected from the mixers. It was even hoped that multi-gigahertz bandwidths might be achievable with proper levels of compensating impurities, thereby avoiding the need

Table 2.2 – A Rogues Gallery of Detectors

Detector	N_A cm ⁻³	N_D cm ⁻³	Fabrication (NTD or conv.)	Dimensions L x W x H (mm)	Orientation (E-field direction)
LBL 496-5.5	3×10^{14}	9×10^{13}	NTD	3 x 1 x .5	[311]
LBL 729-6.0(20)	3×10^{14}	1.4×10^{14}	NTD	3 x 1 x .5	[311]
LBL 729-13.0(20)	3×10^{14}	1.9×10^{14}	NTD	3 x 1 x .5	[311]
LBL 729-9.4(21)	6×10^{14}	2.6×10^{14}	NTD	3 x 1 x .5	[311]
LBL 729-14.6(21)	6×10^{14}	3.3×10^{14}	NTD	3 x 1 x .5	[311]
LBL 729-9.4(22)	1×10^{15}	3.8×10^{14}	NTD	3 x 1 x .5	[311]
LBL 729-14.6(22)	1×10^{15}	4.5×10^{14}	NTD	3 x 1 x .5	[311]
LBL 729-17.0(22)	1×10^{15}	5.1×10^{14}	NTD	3 x 1 x .5	[311]
LBL 712-7.2	1.2×10^{14}	4×10^{13}	NTD	3 x 1 x 1	-
LBL 108-17.7	2×10^{14}	$\sim 10^{13}$	conv.	3 x 1 x 1	[311]
LBL 82-4.6	2×10^{14}	$\sim 10^{12}$	conv.	2 x 1 x 1	[100]
Eagle-Pitchard	-	-	conv.	3 x 3 x 3	-

for sideband generation to provide the local oscillator. A complete list of all detectors on which we have any data whatsoever, together with their impurity concentrations, physical dimensions, and crystallographic orientations, is given in table 2.2. All but one of the detectors were fabricated at Lawrence Berkeley Laboratory by Prof. E. Haller's group. The naming convention for the LBL detectors is that the first number designates the boule from which the wafer was cut, while the second number designates the distance, in centimeters, of the wafer from the head of the boule. The boules are Czochralski grown from material that has previously been zone-refined, and there is a large impurity concentration gradient from the head to the tail of the boule. In all cases, the detectors listed in table 2.2 represent different *materials*; for each material, we have several individual samples, and in many cases more than one sample of a particular material was tested.

Detectors 712-7.2, 496-5.5, and the entire 729- series were doped by the method of neutron transmutation doping, (NTD), which is described in detail by Haller (1984). Detector 496-5.5 and the 729

series were fabricated under more uniform conditions and tested more systematically than any of our other detectors. They were grown at LBL by I.S. Park as part of a project to systematically study the influence of majority doping level and compensation level on those detector parameters most important for heterodyne applications, particularly recombination bandwidth. The results of our measurements on this series of detectors are discussed first in this section. They are also reported in Park, Haller, Grossman, and Watson (1987). Following this, we discuss our results on detector 82-4.6. This is our only detector whose crystallographic orientation makes it suitable for stressed operation out to 200 μm . Therefore, it is the detector used in the airborne system. Then, we briefly discuss our results on a detector fabricated by the Eagle-Pitchard Corp. and acquired through JPL (courtesy of H. Pickett and J. Farhoomand.) It was tested for bandwidth and responsivity only, and, once determined to be inferior to LBL 108-17.7 for our application, no further tests were made on it. Finally, we discuss our results on detectors 108-17.7 and 712-7.2. These measurements include not only determinations of bandwidth, photoconductive gain, etc., but also our measurements of heterodyne noise temperature. Detector 712-7.2 was doped by NTD, while 108-17.7, originally intended for direct detection applications, was doped by the conventional method. The optimum noise temperatures of the two detectors are comparable, but detector 712-7.2 has much higher bandwidth and therefore requires higher LO power. The heterodyne noise temperature of 655 K at 119 μm was measured on detector 108-17.7, with $\sim 1.6 \mu\text{W}$ of LO power.

Detector 496-5.5 and the 729 series detectors

The 729 series detectors were fabricated from a boule of single-crystal germanium that was phosphorus-doped in the range 5×10^{13} to $5 \times 10^{14} \text{cm}^{-3}$. The concentration of compensating p-type impurities was estimated to be $1 \times 10^{12} \text{cm}^{-3}$ from variable-temperature Hall effect measurements. Detector 496-5.5 was fabricated from a boule of single-crystal germanium with very low donor concentration, $\leq 10^{13} \text{cm}^{-3}$. Neutron transmutation doping (NTD) was used to add further dopants to several 0.5 mm thick wafers of these crystals. NTD of germanium produces Ga acceptors and As and Se donors by neutron capture of Ge isotopes followed by nuclear electron capture ($^{71}\text{Ge} \rightarrow ^{71}\text{Ga}$) or beta decay ($^{77}\text{Ge} \rightarrow ^{77}\text{Se}$), respectively (Haller *et al.* 1984). The ratio of donors to acceptors created by NTD is given by isotopic abundance

and neutron capture cross-sections. It leads to a compensation ratio of approximately $(As + 2Se) / Ga = 0.4$ (Haller *et al.* 1984). All concentrations of our samples are given in table 2.2. Three neutron fluences were used, indicated by the number in parentheses after the detector designation, leading to three series of samples, each having a specific concentration of Ga and varying concentrations of compensating donors. After NTD, the Ge wafers were annealed at 400 C for 6 hours in an Ar atmosphere to remove the radiation damage caused mainly by fast neutrons, and to activate impurities (Palaio 1983). Ohmic contacts were produced by implantation with boron ($1 \times 10^{14} \text{cm}^{-2}$ at 25 keV and $2 \times 10^{14} \text{cm}^{-2}$ at 50 keV). These implant doses lead to degenerately (i.e. metallicly) doped contact areas. 200 Å of Pd and 8000 Å of Au were then sputtered on the implanted surface. The final size of the detectors was $3.0 \times 1.0 \times 0.5 \text{ mm}^3$ with opposite electrodes on the $1 \times 3 \text{ mm}^2$ surfaces. In all cases the resulting material is p-type, so our measurements refer to the photoactivation of the gallium sites and transport of holes.

A complete characterization of the photoconductors would consist of a determination of the quantum efficiency as a function of wavelength, and of the recombination time and mobility, which are both functions of bias field. These three material parameters are related to three directly observable detector properties, namely the wavelength-dependent responsivity, the bandwidth, and the photoconductive gain. For each detector in this series that was characterized, we have made measurements of mobility, as determined by variable-temperature Hall effect, of 93 μm and 118 μm direct detection responsivity, of bandwidth, as determined both from the frequency rolloff of g.r. noise and from the frequency rolloff of response to directly modulated FIR radiation, and of photoconductive gain, as determined by the amplitude of g.r. noise. Obviously, there is some redundancy in these measurements. This has the desirable feature of allowing us to perform some checks on the measurements and the method of analysis, and has given us some confidence that the measured detector properties are indeed related to material parameters in the canonical way.

The variable temperature Hall effect measurements were performed in order to determine the free-carrier mobility at 4.2 K. Van der Pauw geometry was used with a sample size of $7 \times 7 \times 1 \text{ mm}^3$. Ohmic contacts were prepared by boron implantation on four corners of the sample.

The 93 μm current responsivity measurements were performed at LBL with a setup very similar to our low-background dewar, D-69. The detector, cooled to 4.2 K, was situated in a conventional integrating cavity and a series of cold filters, including a fixed-spacing, narrow bandwidth Fabry-Perot filter, was used to reduce the incident background power to a very low level. This was estimated to be about 10^{-13} W, which corresponds to 5×10^7 photons/sec. The detector was voltage biased and the optical input signal was chopped between 300 K and 77 K blackbodies at 20 Hz. The resulting photocurrent was amplified in a transimpedance amplifier and synchronously detected.

All the laser measurements were also performed at 4.2 K. In some cases, the CO_2 laser frequency was dithered and the optoacoustic lock loop used for long-term laser stability. In other cases, the open loop stability was adequate and the CO_2 laser lock loop was de-activated. In all cases, the FIR power output was frequently monitored with the pyroelectric detector, and, if necessary, spurious detector data rejected and the laser retuned. Absolute FIR power levels were determined from the pyroelectric detector's responsivity (see §3.2). The laser stability and the accuracy of the power calibration are significant sources of uncertainty in our final results.

All our laser measurements were performed using the strong methanol laser line at 118.84 μm . At the time of the measurements, the laser's maximum power output on this line was about 0.5 mW. The 119 μm responsivity was determined from current-voltage curves measured with the laser focused onto the detector cavity's entrance aperture and with the laser blocked (i.e. 300 K blackbody radiation incident on the detector.) Linearity was checked by inserting attenuators into the laser beam and comparing the drop in detector photocurrent with that in the pyroelectric signal. Neither detector 496-5.5 nor any of the 729 series detectors showed any non-linearity down to the limits of our measurement (a few percent,) even at the highest laser power. Note that detector 583-4.6, on the other hand, an ultra-high responsivity detector optimized for direct detection (Haegel 1983), was tested at the time of our sideband measurements (see §4.4, particularly figure 4-26,) and found to saturate at an incident power of $\sim .5 \mu\text{W}$ (3 db compression point).

Two techniques were used to measure detector bandwidths. The more direct method consisted of

modulating the FIR laser beam and measuring the amplitude of detector photocurrent at the modulation frequency. Assuming the modulated FIR laser power remains constant, the photocurrent amplitude ought to roll off as $(1 + (\omega\tau_r)^2)^{-1}$ as the modulation frequency is swept. We used the laser sideband generator as our far-infrared modulator. The ultra-low capacitance Schottky diode (obtained from R. Mattauch at the Univ. of Virginia, batch no. 1E12), and corner-cube antenna (Krautle, Sauter, and Shultz 1977), were the same ones used in generating the 6-8 GHz sidebands described in chapter 4. As described in more detail there, the FIR laser beam induces a traveling wave on the longwire antenna, which is then partially reflected off the Schottky diode and re-radiated. The reflection coefficient is a (generally non-linear) function of diode voltage. Thus, applying an RF signal across the diode at ν_{mod} will produce a reflected FIR beam that is modulated (generally non-sinusoidally) at ν_{mod} .

For these measurements, much lower RF frequencies were required than in the sideband generation. Therefore, the Fabry-Perot was removed from the usual sideband generation setup, the polarizing Michelson interferometer was tuned to zero pathlength difference, and the 2-18 GHz YIG oscillators were replaced with a 5-100 MHz VCO. The same voltage ramp that was used to sweep the VCO was also used to drive the internal VCO of the spectrum analyzer with which the amplified detector photocurrent was measured. The raw data then consisted of X-Y plots of the spectrum analyzer output versus the VCO driving voltage. Because the two VCO's were not precisely matched in linearity, the peak spectrum analyzer response would gradually drift away from the actual modulation frequency as the two were scanned. Even though the maximum resolution bandwidth of the spectrum analyzer (3 MHz) was always used, it was found that, in practice, the longest single sweep that could be made without yielding a spurious (and highly non-Lorentzian) rolloff was ~ 45 MHz. This made accurate bandwidth determinations of the fastest (22-series) detectors impossible by this method. In addition, because the depth of modulation of the FIR beam was very small, (due to poor conversion efficiency of the Schottky diode and non-ideal throughput of the optics,) the broadband amplifier noise was not completely negligible, especially for the lower responsivity detectors. Therefore, an X-Y plot of the spectrum analyzer output with the laser blocked was taken immediately before each data scan, and subtracted during data analysis. Our raw X-Y plots of the spectrum analyzer output were digitized and fitted to Lorentzians to determine a modulation bandwidth.

Non-flatness in the amplifier chain following the detector also introduced systematic error into our measurements. The gain vs. frequency was measured during initial characterization of the apparatus, however, and found to be flat to better than 0.8 db over the 5-100 MHz range. An additional source of uncertainty was non-flatness in the RF drive power applied to the Schottky diode, due to losses in the bias tee used for some of the early measurements. The insertion loss of the bias tee versus frequency was measured, and for those data scans taken before the bias tee was replaced, the data scans were divided by the insertion loss. This represented an overcorrection, however, since it was found in separate measurements that, at the RF drive power used (usually $\sim +5$ dbm), the strength of the modulated FIR power grew slower than linearly with RF drive power. In other words, the Schottky diode was to some degree saturated by the RF drive. The full correction, however, only produced a 10-15 % increase in the bandwidth derived from Lorentzian fits to the data, so the additional uncertainty introduced by the diode saturation was not large compared to other sources of uncertainty.

Our final series of experiments consisted of measurements of the g.r. noise spectra of the detectors. This was done by removing the corner-cube modulator and polarizing Michelson interferometer, and focusing the FIR laser output directly onto the detector cavity entrance aperture. Spectra of the amplified photocurrent signal with the laser on, laser attenuated by various amounts, and laser blocked were recorded and digitized. The g.r. noise may be described by the relation (Bratt 1977) :

$$\langle i_{gr}^2 \rangle = \frac{4ei_d G}{1 + (\omega \tau_r)^2} \quad (\text{A}^2/\text{Hz}) \quad (2.25)$$

where i_d is the DC detector current, (dominated by the photocurrent from the laser), and G the photoconductive gain. The conversion from noise spectral density at the amplifier output (what the spectrum analyzer measures,) to RMS photocurrent is given by

$$P_{out} = A \left(\frac{R_d R_a}{R_d + R_a} \right) \langle i^2 \rangle \quad (\text{W}/\text{Hz}) \quad (2.26)$$

where A is the power gain of the amplifiers, R_a the input impedance of the first stage pre-amplifier, and R_d the differential impedance of the detector, given by the slope of the I-V curve with laser on. For these experiments, the cooled GaAsFET amplifier was used, since the detector bandwidths were well over

10 MHz. The amplifier's input impedance, set by a single resistor, was generally set to $200\ \Omega$ for these measurements, in order to ensure that RC rolloff would not contaminate the carrier-lifetime rolloff, even though this meant a significant reduction in signal-to-noise ratio. Thus, the rolloff frequency of the g.r. noise spectrum directly yielded the detector bandwidth, and the absolute level of the noise spectral density yielded the photoconductive gain.

There were two significant source of uncertainty in this measurement. The simplest is merely the calibration of amplifier and spectrum analyzer gains, and the uncertainty in the detector impedance derived from the I-V curves. We estimate that these uncertainties total < 2 db. They are only relevant to the determination of the photoconductive gain, of course, and not to the determination of the bandwidth.

The other source of uncertainty was the sporadic existence of low-frequency (< 10 MHz) noise on the laser. As described in chapter 3, if it was discovered in real time, it was always found to be possible to retune the laser so as to eliminate the noise, albeit with some sacrifice in laser power. In about one third of our measured spectra, however, it was not realized until the analysis stage that a single Lorentzian would not adequately fit the spectrum. In these cases an additional low-frequency component (arbitrarily taken to be another Lorentzian) was added to the model. The model was further constrained, however, by the fact that laser noise varies quadratically with laser power while detector g.r. noise only varies linearly. In all such cases there were spectra taken at three or more power levels so that it was possible confirm that the low-frequency excess was indeed due to laser noise. The remaining noise did fit a single Lorentzian spectrum, and scaled linearly with laser power, so we are confident it was indeed detector g.r. noise. Furthermore, the bandwidths derived are quite consistent with what would be expected by comparison with other measurements of the same detector at different biases, that did not show excess noise. Nevertheless, the subtraction of excess noise does introduce significant uncertainty into the determination of the recombination bandwidth, for those spectra for which it was necessary. As an example of our raw data, and the quality of the fits, we show in figure 2.8 the measured noise spectra for two cases, one in which a single Lorentzian provided a good fit, and one in which an additional low-frequency component was present.

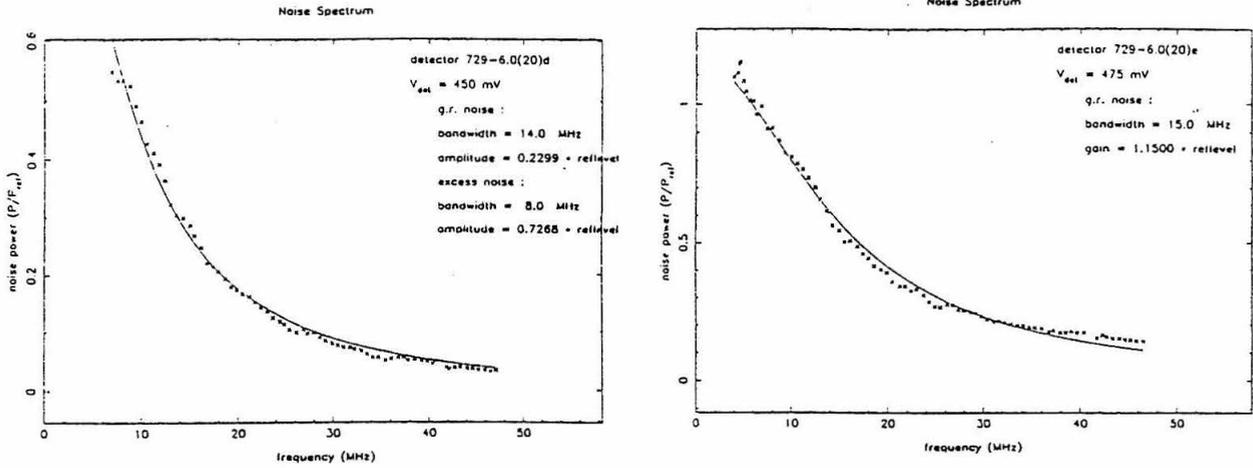


Figure 2.8 – Measured noise spectra of a Ge:Ga detector with 119 μm laser illumination. Laser power is about .3 mW. Spectrum at the right is well fit by a single Lorentzian. An additional low-frequency component is present in the spectrum at the left.

We begin the discussion of our results with the breakdown field. Table 2.3 lists the data, while figure 2.9 displays it in the same form that Zylberstejn used to compare his theory with KBS's data, namely a graph of E_{br} versus $(N_{maj} - N_{min})/N_{min}$. The effective mobility listed in column 3 of the table is simply that which would allow the kinetic energy in the *drift* velocity to account for the impact ionization, i.e.

$$\frac{m^*}{2} (\mu_{eff} E_{br})^2 \equiv E_I. \quad (2.27)$$

The first conclusion to be drawn from the data is that the empirical approach adopted by WT fails miserably. As discussed in §2.1, WT's treatment would predict that μ_{eff} be proportional to the total drift mobility, with a constant of proportionality of $x_{crit} \approx 1$. As may be seen from table 2.3, the dependence of μ_{eff} , (i.e. E_{br}), on compensation is much weaker than the dependence of μ_{Hall} . Furthermore, the magnitude of μ_{eff} is some two orders of magnitude larger than μ_{Hall} . The failure of WT's approach is not really very surprising, however. Predicting the breakdown field from the drift mobility can be viewed as an extrapolation of the carrier distribution function from the velocity where it peaks out to its wings, where the carriers have sufficient kinetic energy for impact ionization. For the low compensation, high

Table 2.3 – Breakdown Fields of NTD GE:Ga Detectors

Detector	Measured E_{br}	μ_{eff} (from eqn 2.27)	Hall mobility
	V/cm	$\text{cm}^2/\text{V} - \text{s}$	$\text{cm}^2/\text{V} - \text{s}$
496-5.5	7.9	1.3×10^6	3.5×10^4
729-6.0(20)	11.4	8.9×10^5	2.5×10^4
729-13.0(20)	17.8	5.7×10^5	9×10^3
729-9.4(21)	14.6	7.0×10^5	6×10^3
729-14.6(21)	18.0	5.6×10^5	4×10^3
729-9.4(22)	20.2	5.0×10^5	-
729-14.6(22)	23.1	4.4×10^5	-
729-17.0(22)	27.2	3.7×10^5	-
712-7.2	6.2	1.6×10^6	$\sim 3 \times 10^4$
712-7.2	μ_{ph} (Hall measurement extrapolated to 4.2 K)		$\sim 2 \times 10^6$

mobility detectors WT had in mind, this is not a terribly large extrapolation. In our detectors, the ionized impurity scattering is much stronger, the velocity at which the carrier distribution peaks is much lower, and therefore the extrapolation is a much larger one. It is therefore natural to expect much greater sensitivity to the assumed shape of the distribution function.

Comparing our data with Zylberstejn's graph (figure 2.9) we find remarkably similar behavior. For low amounts of compensation ($N_{min}/N_{maj} < .1$) the breakdown field is nearly independent of compensation, with a value of 2 - 5 V/cm. At about $(N_{maj} - N_{min})/N_{min} = 5$, there is a "knee" in the curve of breakdown field, with E_{br} increasing rapidly at lower values. At any given compensation, however, there can still be considerable (a factor of 2) scatter in E_{br} from one detector to the next. In addition to the similarity in the shapes of the dependence in Zylberstejn's graph and our data, the agreement in the numerical value of E_{br} is also good. Our highest breakdown detectors (the 22 series) have a factor of two higher E_{br} than KBS's, but all our other detectors lie within the range spanned by KBS's. The empirical

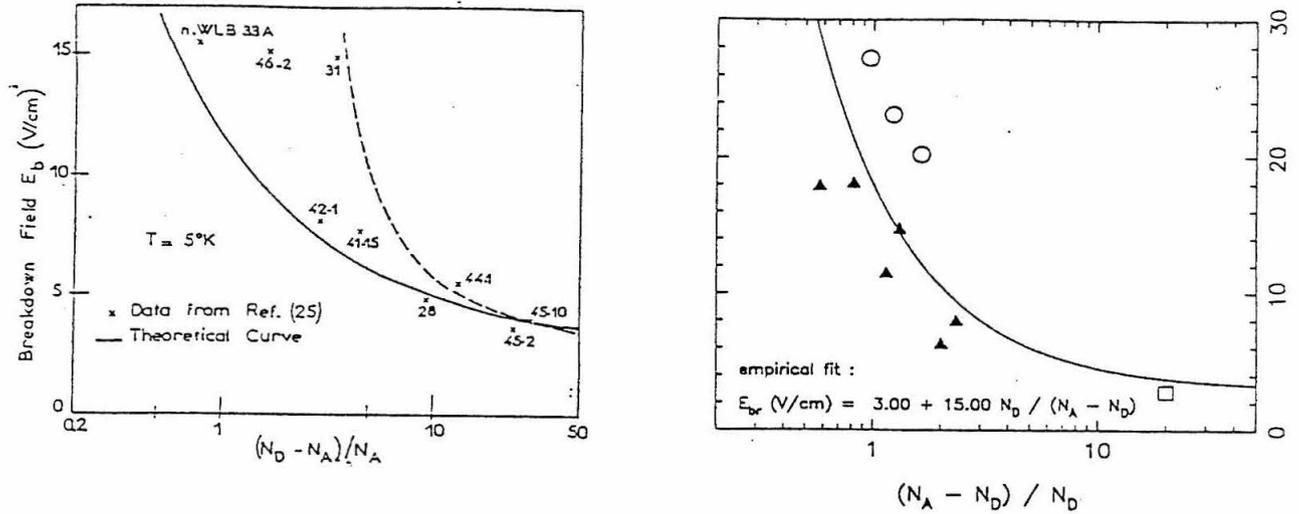


Figure 2.9 – Breakdown field versus compensation. (left) Data from the (n-type) detectors of KBS compared to the theories of Zyberstejn (1962) (solid line) and Koenig (dotted line). (right) Data from our detectors. Hollow circles are the highest neutron fluence (series 22) detectors. Square is detector 108-17.7 (not NTD). Note the factor of two vertical scale change between the two graphs.

fit we have used in our calculations on detector optimization (chapter 1) is also indicated.

As mentioned in §2.1, Zylberstejn’s theory assumes that acoustic phonon scattering dominates the mobility, which is not the case for our detectors. However, the agreement between that theory’s predicted breakdown fields and our measured values suggests that it is indeed the acoustic phonon mobility, not the total mobility, which determines E_{br} . In fact, the full apparatus of Zylberstejn’s theory is not necessary to arrive at this conclusion. Examining table 2.3, we see that for detector 712-7.2, the effective mobility deduced from the breakdown field agrees (to within 25 %) with the acoustic phonon mobility obtained by extrapolating $\mu_{H_{all}}$ down to 4.2 K. Furthermore, from KBS’s mobility measurements, we expect a weak dependence of the acoustic phonon mobility on total doping level, with higher doping corresponding to lower μ_{ph} . We find that the effective mobility deduced from E_{br} falls by about a factor of 4 as the majority impurity concentration is raised from 1.2×10^{14} (712-7.2) to 1×10^{15} (729-17.0(22)). Apparently, the effective mobility for determining the breakdown field tracks the acoustic phonon mobility quite closely.

Two other experimental observations regarding E_{br} should be mentioned. We have noticed a small (typically 5 - 10%) but significant reduction in the breakdown field of all our detectors when they are illuminated with the full laser power ($\sim .5$ mW). Also, we have tested one NTD detector (729-6.0(20)) at a temperature of 1.5 K as well as 4.2 K. We found no significant ($> 10\%$) change in the breakdown field when the temperature was reduced. This is inconsistent with the simple identification of μ_{eff} with μ_{ph} , since the latter varies as the $-3/2$ power of lattice temperature. However, it is consistent with Zylberstejn's full theory, which predicts an extremely weak dependence of breakdown field on temperature. In conclusion, we find that the breakdown fields of our detectors are determined by acoustic phonon scattering, the dominant inelastic process in the carrier transport. Attempts, such as WT's, to predict E_{br} from the total drift mobility fail for heavily compensated detectors. The breakdown field at 4.2 K may be predicted fairly accurately from equation 2.12 by simply identifying μ_{eff} with the acoustic phonon mobility. More accurate prediction of E_{br} as a function of compensation, or of temperature can be made by applying the full theory of Zylberstejn (1962).

We now turn to our results on recombination bandwidth. Figure 2.10 shows our modulation measurements of bandwidth versus bias field for a single detector. The indicated error bars are somewhat conservative estimates of the range over which a subjectively "good" fit of the data to a single Lorentzian could be obtained. As explained earlier, the uncertainties are almost entirely systematic, so a more quantitative error estimate cannot really be made. The bandwidth definitely falls off with increasing bias. The form of the dependence agrees very well with the theoretically expected one (see §2.1), namely a bandwidth that is independent of bias at low fields, and that falls off as $E^{-3/2}$ at high fields, where the carriers are hot. It is also consistent with the $E^{-1.8}$ dependence found by KBS with the technique of pulsing the detector from breakdown into the sub-breakdown regime, and measuring the decay of the conductivity. Note that because of the presence of chaotic dynamics near breakdown, (WT), the results of the pulse technique by themselves could be open to doubt. Although our measurements are in excellent agreement with the theoretical expectation, they are not really accurate enough, nor do they extend to low enough field, to really test the theoretical dependence in great detail. Indeed, considering all the uncertainties, a bandwidth varying inversely with bias field is also a quite acceptable empirical approximation to our data over the

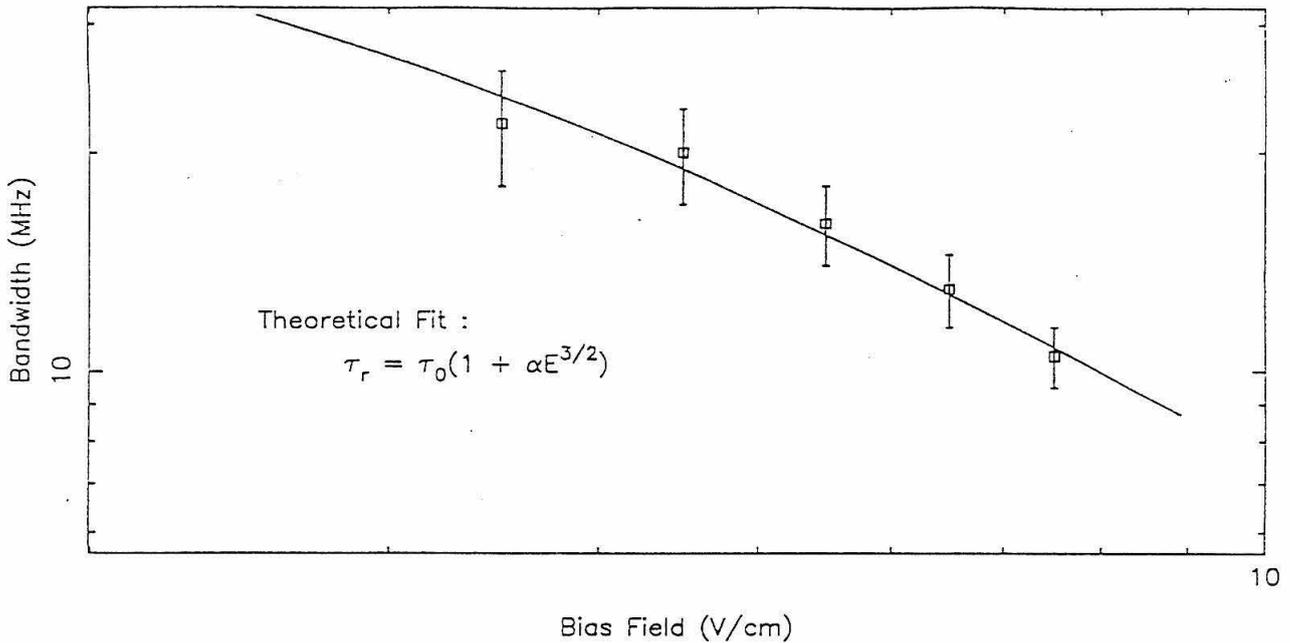


Figure 2.10 – Field dependence of the recombination bandwidth, measured on detector 496-5.5. The rolloff at high field is due to carrier heating.

range of practical interest, $0.5 \leq E/E_{br} \leq 1$.

One of the main reasons for our wanting to determine the dependence of bandwidth on bias is to be able to normalize $B(N_D)$, the bandwidth versus compensating impurity concentration, to a single value of E/E_{br} . Physically, the bandwidth in the low-bias limit, where the carriers are thermalized, would probably be the most fundamental quantity to examine. However, the low-bias limit is not the regime in which the photoconductors are used in practical applications, nor is it a region in which we can, with our techniques, measure the bandwidth with any accuracy. (In both cases, the responsivity is too low.) Therefore, we have normalized all our bandwidth measurements to a bias $E = 0.8E_{br}$ using the empirical approximation of $B \propto E^{-1}$ described above. The bandwidths were actually measured at biases that varied from about .7 to .95 times the breakdown field, so this normalization never amounted to more than about a 15 % correction. The results of our measurements, using both the modulation and the g.r. noise techniques, are displayed in figure 2.11.

These bandwidth measurements are a central result of this work. It is clear that the two techniques

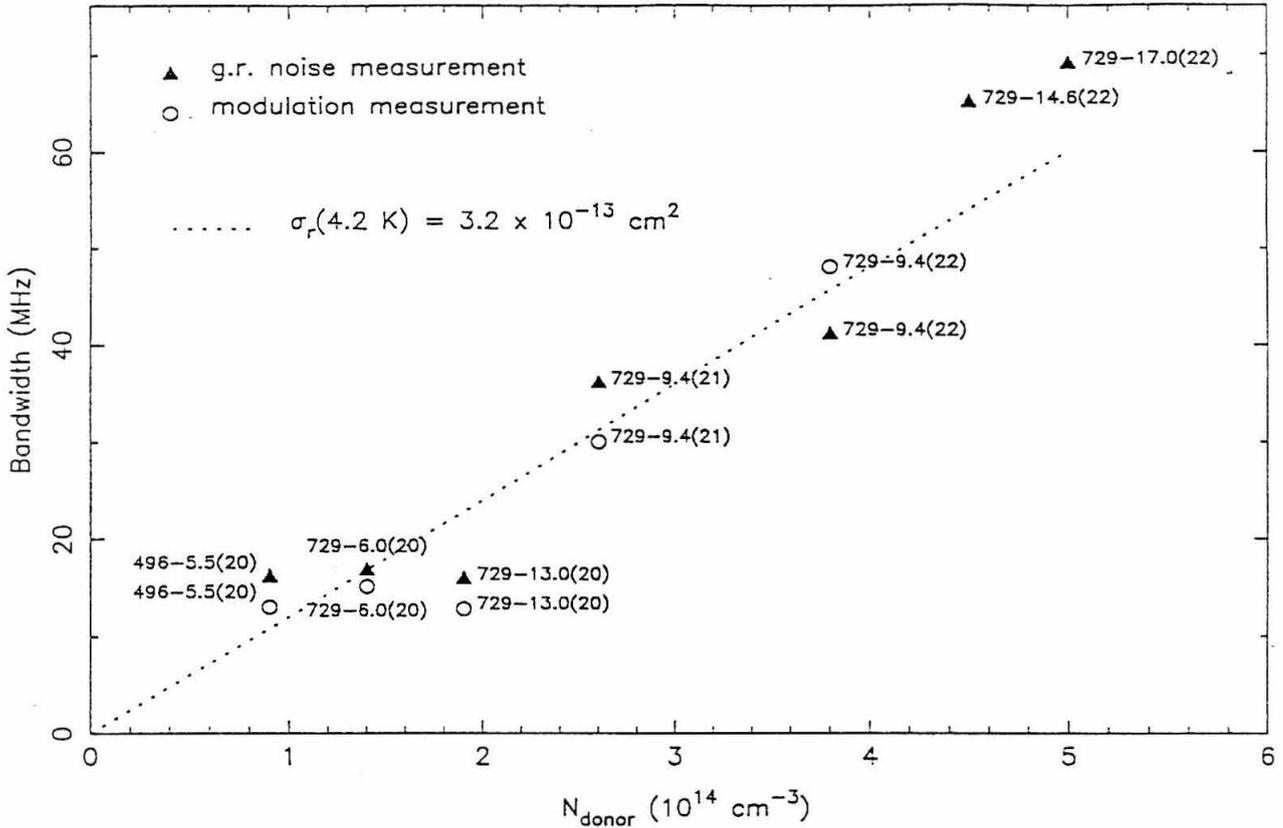


Figure 2.11 – Measured recombination bandwidth as a function of minority impurity concentration, for the 729-series NTD detectors.

employed to measure the bandwidth agree fairly well. The (22) series detectors have the highest bandwidths of any we have measured – some 60 MHz. Measurements on detectors 108-17.7 and 583-4.6 which were made during this same run of experiments yielded only upper limits to their bandwidths. As expected, there appears to be an approximately linear relation between bandwidth and donor concentration. The slope of the relation is a measure of the recombination cross-section via

$$\sigma_r = \frac{B(N_D)}{N_D v_T} = \frac{B(N_D)}{N_D} \left(\frac{3kT_h}{m^*} \right)^{-1/2} \quad (2.27)$$

Taking $T_h = 4.2 \text{ K}$ we obtain $\sigma_r = 3.2 \times 10^{-13} \text{ cm}^2$. This will be a slight overestimate, however, since at 0.8 times the breakdown field, the carriers are undoubtedly somewhat hotter than the lattice. Based on the measured bias dependence of bandwidth (figure 2.10) we do not expect the bandwidth, and therefore σ_r , in the thermalized regime to be more than perhaps a factor of two greater than our present determination.

The values of σ_r predicted by the various theories discussed in §2.1 are : $1.3 \times 10^{-12} \text{ cm}^2$ for Lax's classical cascade theory, $3.5 \times 10^{-13} \text{ cm}^2$ for Brown and Rodriguez's quantum mechanical extension of the cascade theory, including phonon polarization, and $1.4 \times 10^{-11} \text{ cm}^2$ for APY's theory, which is (in its final result) a scaled version of the Thompson recombination cross-section. Clearly, our data favor the Brown and Rodriguez theory. However, to be fair, we note that since we do not know for certain the true carrier temperature in our experiments, the extremely steep dependence of σ_r on temperature may be used to make any of the theories fit the measured cross-section. For APY's theory, a carrier temperature of $T_h = 15K$ would have to be hypothesized (taking $\sigma_r \sim T_h^{-3}$), which seems implausible. For Lax's theory, $T_h = 5.8K$ (taking $\sigma_r \sim T_h^{-4}$) would suffice.

The interpretation of our data on breakdown fields and bandwidths has been fairly straightforward. We now turn to the more ambiguous part of our data on the 729 series detectors, that relating to responsivities and photoconductive gains. In figure 2.12 we plot, for each of the 729 series detectors, the responsivity measured at $93 \mu\text{m}$ with the low background setup at LBL, and the $119 \mu\text{m}$ responsivity measured with the FIR laser. The responsivities shown in the plot were measured at biases of .7 to .9 times the breakdown field. The exact bias was slightly different for each detector, contributing to the scatter in the plotted data. Nonetheless, two facts are immediately apparent. Firstly, the responsivity values, especially for the most heavily doped samples, are quite low compared to material used for low background direct detectors. This is expected, and is due to both shorter free carrier lifetime and lower carrier mobility. That is, the ionized acceptors (created by the compensating donors) constitute both the recombination centers for free carriers and the dominant scattering mechanism at 4.2 K. Thus, at high levels of compensation, the photoconductive gain is reduced both because the recombination time is shortened and because the transit time is lengthened.

The other fact apparent from figure 2.12 is that the $119 \mu\text{m}$ responsivity shows a much weaker falloff at high doping levels than does the $93 \mu\text{m}$ responsivity. A possible explanation for this lies in the fact that the $119 \mu\text{m}$ laser wavelength happens to lie on the steep "edge" of the photoconductivity spectrum, just longward of the peak. The responsivity, (through the quantum efficiency) is an extremely strong function

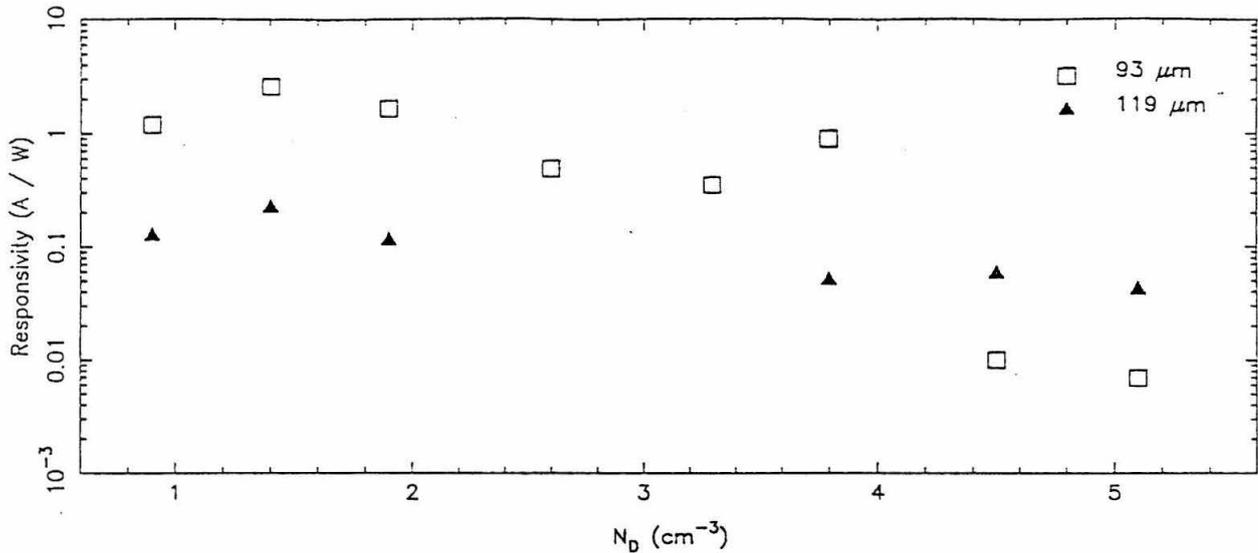


Figure 2.12 – Direct detection responsivities for the 729-series detectors measured at 93 μm and 119 μm .

of wavelength in this region, and two orders of magnitude difference between 93 μm and 118 μm is easily possible if the edge is sharp. It is known that the width of spectroscopic features in extrinsic germanium increases with concentration (Ramdas and Rodriguez, 1981). As the impurity concentration increases, the impurity wavefunctions begin to overlap, causing the highly excited state transitions to broaden and merge into the ionization edge. The width of the ionization edge itself also broadens. This effect would explain our results. For the highly doped detectors, the ionization edge would be relatively broad and the drop in responsivity due to the lower photoconductive gain could be largely compensated for by the increased quantum efficiency. This does not explain, of course, how the 119 μm responsivity could actually be higher than the 93 μm responsivity for an individual detector; it only explains why the rolloff of responsivity with doping should be slower at 119 μm than at 93 μm . It is clear that there is also some other effect which has produced anomalous responsivity measurements for the 22 series detectors (the three most heavily doped samples in the plot), probably in the 93 μm rather than the 119 μm measurements.

Evidence that some sort of wavefunction overlap effect was operating in the detectors was also provided by the 93 μm measurements of NEP and dark current. The series (22) detectors had relatively high dark currents, typically ~ 10 nA at usual bias levels. The measured dark currents were in agreement with the

Shlovskii and Efros (1984) formulae for hopping conductivity. With the low background filter in place (incident $\dot{N} \approx 10^7$ photons/s at $93 \mu\text{m}$), the shot noise from this dark current dominated the signal, and responsivity and NEP values could not be accurately measured. Therefore, for these three detectors, the $93 \mu\text{m}$ responsivities shown in figure 2.12 were obtained by removing the Fabry-Perot bandpass filter and using only a low-pass blocking filter. The series (21) and (20) detectors had dark currents that were lower than the (22) series, but much larger than would be expected given their lower doping levels. The reason for this is not known. Their measured NEP's were also dark-current limited rather than background-limited. Therefore, the direct detection NEP measurements cannot be used to derive detector quantum efficiencies.

Given these measurements of bandwidth, there are three methods by which we may estimate the photoconductive gain. Method 1 is to take the measured recombination time and Hall-effect mobility, and compute $G = \frac{\tau\mu E}{l}$. Note that no normalization to $\frac{E}{E_{br}} = 0.8$ is required in this case; we simply use the measured bandwidth and the bias field at which the measurement was made. A limitation to this method, however, is that at high acceptor concentrations, such as those for the (22) series detectors, the variable temperature Hall-effect measurements are not directly interpretable in terms of mobility, because of the presence of significant hopping conductivity. (This is why no μ_{Hall} is listed for these detectors in table 2.3.) The method should work properly for the (20) and (21) series though. The second method derives the photoconductive gain from the amplitude of the measured g.r. noise, as described by equation 2.25. The third method is to *assume* a responsive quantum efficiency, η_r , and then convert the measured responsivities to photoconductive gains. There is, of course, no particular value of η_r which is known a priori to be correct, so this last method is not quite on the same footing as the other two. At least at $93 \mu\text{m}$, however, it is perhaps reasonable to assume that η_r will be nearly the same for all the detectors, since they are all mounted in identical integrating cavities and there is no complication due to the photoconductive edge. Therefore, for method 3, what we have done is chosen the quantum efficiency at $93 \mu\text{m}$ and $119 \mu\text{m}$ so that the derived photoconductive gain has the value of .04 for detector 496-5.5, in agreement with methods 1 and 2.

The results of these three methods for the 729 series detectors are given in table 2.4. The bias field

Table 2.4 – Photoconductive Gains of NTD Detectors

Detector	E (V/cm)	τ_r (nsec)	$G_1 = \frac{\tau_r \mu E}{l}$ (method 1)	G_2 (g.r. noise) (method 2)	$G_3(119\mu m)$ (method 3, $\eta = .03$)	$G_3(93\mu m)$ (method 3, $\eta = .4$)
496-5.5	5.5	10.0	.039	.042	.042	.040
729-6.0(20)	9.8	12.2	.060	.064	.077	.087
729-13.0(20)	15	9.4	.025	.056	.040	.055
729-9.4(21)	12	4.55	.007	.051	.012	.016
729-14.6(21)	11	4.5	.004	-	-	.012
729-9.4(22)	18	4.0	-	.044	.017	.030
729-14.6(22)	20	2.65	-	.029	.020	3.3×10^{-4}
729-17.0(22)	25	2.65	-	.028	.015	2.3×10^{-4}

and recombination times listed are those which corresponded to the g.r. noise measurements from which G_2 was derived. G_1 is thus the photoconductive gain predicted (on the basis of μ_{Hall}) at the particular bias at which G_2 was measured. Unfortunately, the 93 μm responsivities were generally measured at slightly different bias levels. This is expected to introduce some scatter, at, say the 30 % level, into the comparison of G_3 and G_2 . We also note that for some detectors, calibration of the absolute power level of the IF was not done, and it was necessary to assume that the calibration had remained the same since its last measurement. These systematic uncertainties make it difficult to come to firm conclusions on the basis of the data in table 2.6. What we can say is the following : Firstly, there is some tendency for the most heavily doped detectors to have lower measured photoconductive gains, as expected. However, the range in G_2 is only slightly more than a factor of two over the seven detectors for which we have measurements. Secondly, for the most lightly doped detectors (496-5.5, and the two 20-series detectors,) there is fairly good agreement between the values of G obtained by the three different methods. This may be regarded as evidence (albeit not very compelling evidence, in light of the uncertainties,) that the adopted values of quantum efficiency, $\eta(119 \mu m) = .03$ and $\eta(93 \mu m) = .4$, are correct. Thirdly, the measured values of photoconductive gain are in the range of $\sim .03 - .06$, about an order of magnitude lower than typical

values for detectors optimized for direct detection.

There are two gross discrepancies in the values of G calculated by the different methods. For detector 729-9.4(21), both method 1, and method 3, using either the $93\ \mu\text{m}$ or $119\ \mu\text{m}$ responsivity, yield photoconductive gains much lower than that derived from the measured g.r. noise. This suggests that this detector may produce large amounts of excess noise. The spectrum of this detector's noise was quite accurately Lorentzian, but its I-V curve showed hysteresis and highly unusual structure within about 20 % of the breakdown voltage. For detectors 729-14.6(22) and 729-17.0(22), the photoconductive gains derived from the $93\ \mu\text{m}$ responsivity were much lower than G_2 . This is due to the extremely low $93\ \mu\text{m}$ responsivity measured for these detectors (see figure 2.12), which, as mentioned earlier, we do not entirely understand.

The gain-bandwidth product is a commonly quoted figure-of-merit even though it is not expected to be truly constant as a function of doping. As expected, we find it is only a weak function of impurity concentration in the range covered by our detectors, varying from .67 MHz for 496-5.5 to 1.75 MHz for 729-9.4(21) and 729-9.4(22).

Before moving on from the analysis of the 729 series detectors, we must discuss the excess detector noise we have seen in some cases. The most blatant example we have found of excess noise in the IF due to the detector is shown in figure 2.13. It consists of a set of g.r. noise spectra of detector 496-5.5 at a selection of biases. The breakdown voltage for this detector was 310 mV. As usual, this varied slightly with illumination. From g.r. noise alone we would simply expect a series of Lorentzian spectra whose rolloff frequencies decreased with increasing bias as found in the modulation measurements of figure 2.10. What is actually found is something rather bizarre. The spectra are indeed properly Lorentzian up to a bias of ~ 270 mV, with approximately the same bandwidths as found in the modulation measurements. Above this bias, a noise "bump" appears in the spectrum, whose frequency and amplitude vary somewhat with bias. The noise bump is completely reproducible, and is not due to instability in the IF circuitry. This was verified by changing the laser power and bias voltage simultaneously, in such a way that the detector presented identical impedances to the IF. Completely different noise spectra were measured in the two cases, demonstrating that the noise was not associated with the IF.

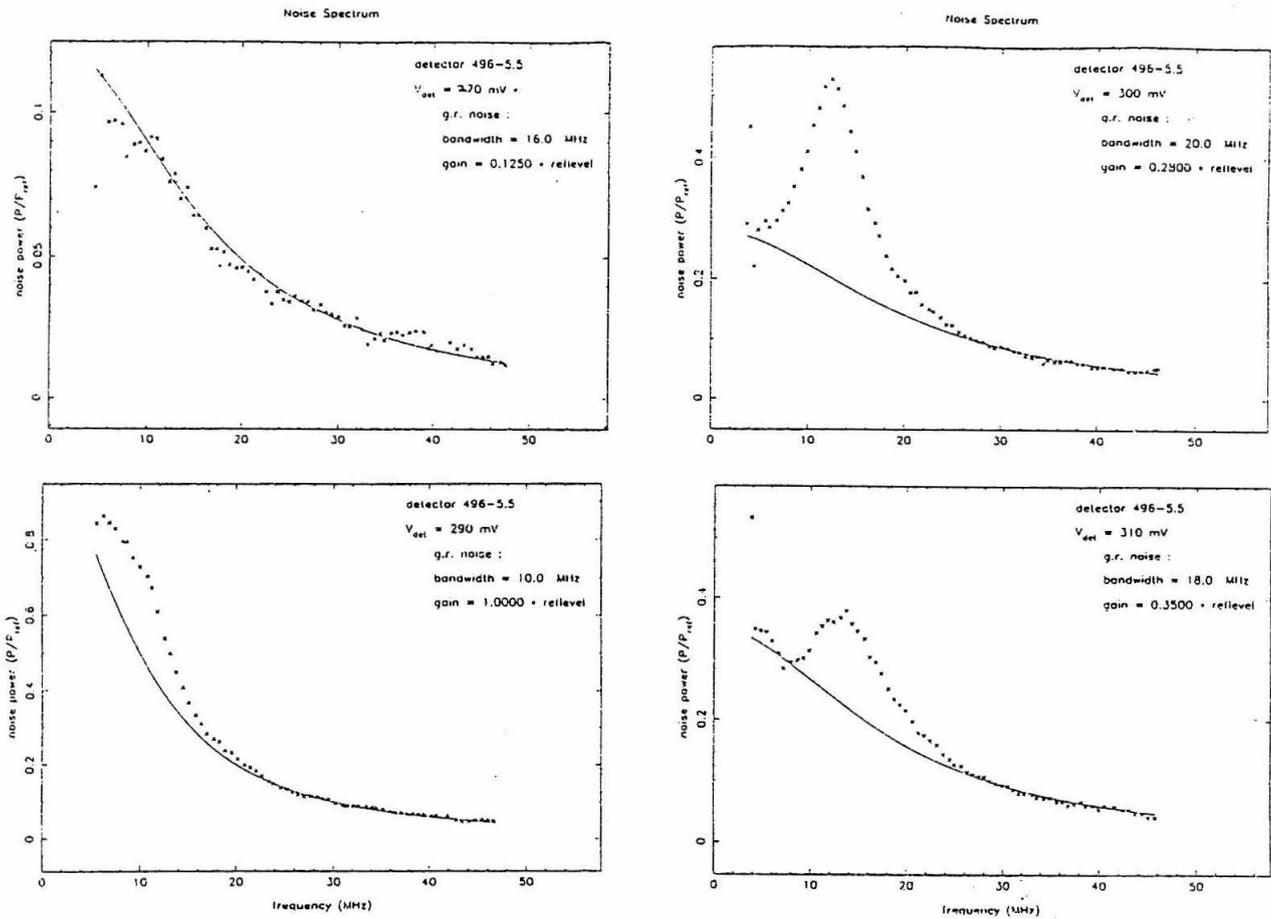


Figure 2.13 – Non-Lorentzian excess noise displayed by detector 496-5.5 at biases near breakdown. It may be attributable to non-linear dynamical effects similar to those studied by WT.

We do not know for certain what the explanation is for “bumps” in the noise spectra like those of figure 2.13. The fact that it seems to be limited to values quite near breakdown suggests that it is related to the non-linear dynamics, period doubling, deterministic chaos, etc. studied at length by Westervelt and Teitworth(1984, 1985, 1986a, 1986b) in Ge:Ga photoconductors at audio frequencies. The primary differences between our detectors and the ones they studied, which are standard LBL detectors optimized for low-background direct detection, are the recombination times and dielectric relaxation times (i.e. impedances). Their detectors have much lower compensation levels than our NTD detectors, and therefore have much lower recombination bandwidths and much slower dielectric relaxation times because of the higher impedances. These are the two parameters that in their model determine the frequency scale

of the dynamical behavior. It thus seems likely that all the phenomena they find at audio frequencies will be shifted up into our IF band, just as the desired photoconductive signal is. If this hypothesis is correct, then there is no way simply adjust the doping in order to avoid “chaotic” noise. So long as the doping is adjusted to match the recombination bandwidth to the desired IF bandwidth, it will also shift the frequency scale of the chaotic dynamics into the IF. However, the chaotic dynamics is not really a fatal problem in practice. It simply puts an upper limit on the bias that may be applied to the detector. So far, we have not found any evidence that the region of chaotic dynamics ever extends more than $\sim 40\%$ lower than breakdown, (usually it is considerably less). Thus, the excess noise can be tuned away, at the price of a somewhat lower responsivity and therefore higher LO power requirement. It also seems likely that some of the anomalously large amplitude, but still Lorentzian shaped, IF noise produced by detector 729-9.4(21) may be attributable to the same effect, and that the detector simply should have been biased slightly lower during the measurement in order to get an estimate of the true g.r. noise.

To summarize our test on the 729 series of detectors, we have obtained data on Hall-effect mobility, data on responsivity from low-background direct detection, data on recombination bandwidths from the response to directly modulated FIR radiation, and data on bandwidths and photoconductive gains from g.r. noise spectra. There are definitely some aspects of the high-frequency performance of some of the detectors which do not fit into the standard picture of photoconductor performance outlined in chapter 1. In particular, the (20) and (21) series of detectors have excessively high dark currents, which prevent them from having background-limited NEP's at low backgrounds. Also, some of the detectors displayed excess noise in the frequency range of our IF when biased near breakdown. In one case, the excess noise had a Lorentzian spectrum with rolloff given by the carrier recombination time, i.e. it appeared just like g.r. noise except that its amplitude was several times higher than it ought to have been, given the photoconductive gain estimated by other means. In other cases, the spectrum was highly non-Lorentzian. It is conjectured that this excess noise at high biases is related to the chaotic non-linear dynamics studied by Westervelt and Teitsworth.

Despite these non-ideal aspects of the detectors, we have found that the overall picture described in

chapter 1 is confirmed. Increasing the concentration of compensating impurities increases the bandwidth and reduces the photoconductive gain in approximately linear fashion, as shown in figure 2.11. Bandwidths as high as 60 MHz have been obtained at donor concentrations of $5 \times 10^{14} \text{ cm}^{-3}$. The measured dependence of recombination bandwidth on bias is consistent with the theoretical expectation of a bandwidth that is independent of bias at low fields, and which falls off as $E^{-3/2}$ at high bias, when the carriers are heated. The experimental errors, however, are large enough to accommodate the $E^{-1.8}$ dependence reported by KBS, or even a simple E^{-1} dependence. From the slope of the bandwidth - donor concentration relation, we deduce a recombination cross-section of $3.2 \times 10^{-13} \text{ cm}^2$ at $E = .8E_{br}$, and $T = 4.2 \text{ K}$, in excellent agreement with the extended "cascade theory" of recombination developed by Brown and Rodriguez (1967), but about a factor of thirty smaller than the value predicted by APY. For heterodyne applications, this series of detectors has provided valuable information on the tradeoff between mixer bandwidth and local oscillator requirement, but itself lies too far toward the high-bandwidth, low gain end of the spectrum to be directly useful in our receiver, with the very low level of local oscillator power currently available.

Detector 82-4.6

We have only one detector, LBL 82-4.6, that is properly oriented for stress, and which is therefore suitable for the airborne system. Because there has been no question of choosing between various possible detectors, our aim has been simply to measure its properties under operating conditions, rather than to try to attempt a careful study of their dependence on e.g stress and temperature. This detector has a lower concentration of compensating impurities than any of our NTD material : $N_A = 2 \times 10^{14}$, $N_D = 1 \times 10^{13} \text{ cm}^{-3}$. As a result, its bandwidth is lower, its breakdown field lower, and its photoconductive gain and responsivity higher than those of our NTD detectors. The breakdown field is further lowered due to the reduction in impurity ionization energy caused by the stress.

The breakdown field for this detector, under stress and at 1.5 K, is $\sim .3 \text{ V/cm}$, approximately an order of magnitude lower than the lowest of the NTD detectors. We have made no careful measurements of the detector's properties with laser illumination. However, from the known throughput of the system, and the measured spectral response (e.g as shown in figure 2.7), we can compute the power incident on the detector

the g.r. noise is $G = .23$, about a factor of five higher than the NTD detectors. This is completely in line with what would be expected on the basis of scaling from the most lightly doped of the NTD detectors, 496-5.5. Using $G = \tau_r \frac{\mu E}{L}$, and recalling that this detector has twice the interelectrode spacing of the NTD detectors, we find that a mobility of $\mu = 2.3 \times 10^5 \text{ cm}^2/\text{V} - \text{s}$ is implied. This is a factor of ten higher than 496-5.5, consistent with the reduction in donor concentration. From the photoconductive gain of .23 and the measured responsivity, we estimate a responsive quantum efficiency of $\eta_r = 15 \%$ for this detector/cavity combination. As discussed earlier, it is clear from these results that the system noise in the present configuration is limited by the g.r. noise from the room-temperature background.

Eagle-Pitchard detectors

In the early period of the project we performed some limited tests, all at $119 \mu\text{m}$, on samples of Eagle-Pitchard Corp. material obtained through JPL. The geometry of these detectors was not optimal in that the interelectrode distance was 3 mm, a factor of six larger than the LBL NTD detectors. Furthermore, they were tested in a rather different cavity geometry, in which the ratio of active detector area to aperture area was relatively low, namely 2.5. The setup used for measuring the g.r. noise spectrum was slightly different than that used in most of our other tests. In place of the spectrum analyzer at the output of the IF amplifier chain, there was a double balanced RF mixer (Hewlett-Packard 10534A), driven by a second LO of frequency DC - 10 MHz. The downconverted signal was amplified in a bandpass amplifier, of center frequency 100 kHz and $Q = 10$, and the output rectified and measured in an RF power meter (Hewlett-Packard 436A). A measured spectrum, which displays a 3 db rolloff frequency of 3.8 MHz, is shown in figure 2.15 . The amplitude of the g.r. noise, $\langle i^2 \rangle = 2.4 \times 10^{-24} \text{ A}^2/\text{Hz}$ at 5 MHz, leads to photoconductive gain of $G = .11$ at a bias of $E = 2.1 \text{ V/cm}$. The incident laser power was not carefully monitored during these measurements. Assuming an incident power of $\sim .5 \text{ mW}$, however, (which should be correct to within a factor of 2, based on typical powers obtained around the time of these tests,) yields a responsivity of .2 A/W and a quantum efficiency of 1.9 % . Of course, since we have no independent way of estimating the photoconductive gain for this detector, it is possible that the presence of excess noise may be causing us to overestimate G and therefore underestimate η .

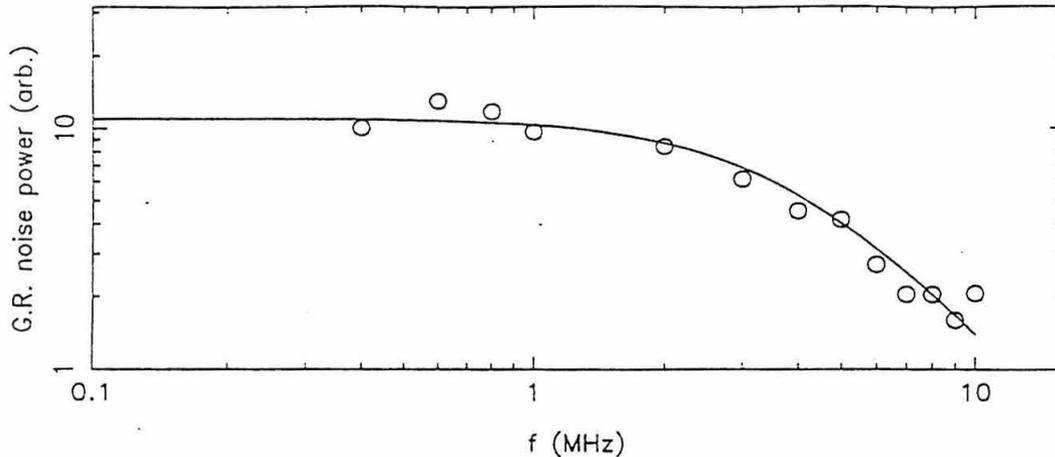


Figure 2.15 – G.R. noise spectrum of Eagle-Pitchard detector, displaying 3.8 MHz detector bandwidth.

LBL 712-7.6 and LBL 108-17.7 – Heterodyne noise temperature

For detectors 712-7.6(NTD) and 108-17.7, we have data on heterodyne noise temperature as well as the usual data on bandwidth, g.r. noise, and responsivity. The data may be divided into three sets based on the range of heterodyne noise temperatures that were achieved. The first set of experiments were all performed on detector 108-17.7 and yielded double-sideband (DSB) noise temperatures of 20,000 - 30,000 K. It was later found that there were two major problems that degraded the noise performance during these measurements : very low transmission of the restrahlen blocking filter that was used, and a degraded noise temperature of the GaAsFET preamp, due to poor electrical connections that developed after repeated thermal cycling. The first set of data, before either of these problems were fixed, is not totally superseded by the later data, however, because it was the only instance in which heterodyne noise temperatures were measured at wavelengths besides 119 μm . The second set of data was taken after the salt filter was replaced with a capacitive grid lowpass filter, and noise temperatures were in the range of 5000 K, both for detector 108-17.7 and 712-7.6. G.R. noise spectra taken in this condition were used to determine the bandwidth and photoconductive gain of 712-7.6. The corresponding parameters of LBL 108-17.7 were determined both from this series of measurements and from later modulation measurements. Detector 712-7.2 has considerably higher bandwidth and lower photoconductive gain than 108-17.7, and comparison

of the heterodyne measurements of the two detectors provides accurate and direct confirmation that the heterodyne performance depends on bandwidth, photoconductive gain, and LO power in the canonical way. Finally, the excess noise from the pre-amplifier was eliminated, and measured receiver temperatures on detector 108-17.7 fell to 500 - 1000 K, at 119 μm . The best noise performance we have measured is 590 K from raw hot and cold IF noise powers, which translates into $T_N(DSB) = 655\text{K}$, after correction for both the background g.r. noise and the Planck spectrum of the hot and cold loads, both of which make the noise temperature appear artificially low. We present tuning curves for this measurement, (fig. 2.18), showing the dependence of $T_N(DSB)$ on LO power for fixed bias current and fixed bias voltage.

The method we have used to measure heterodyne noise temperatures is shown in figure 2.16. The laser output is focused by a polyethylene lens and diplexed with the signal beam by a 25 μm thick, flat, mylar beamsplitter oriented at 45° to the incident laser. The (nominal) polarization of the laser was linear and normal to the plane of incidence of the beamsplitter. The transmission of the beamsplitter was separately measured (at each of our operating wavelengths) with the pyroelectric detector. The signal consisted of alternately room-temperature (295 K) and liquid-nitrogen cooled sheets of AN 72 eccosorb (Emerson and Cumings Corp., Hawthorne, CA). It is assumed that the effective blackbody radiation temperature of the eccosorb at liquid-nitrogen temperature is 95 K, as found by Erickson (1985) at 300 GHz. Except for a few special experiments, the heterodyne measurements were performed with the IF bandpass determined by a homemade 10-pole LC filter. Its measured characteristics are: center frequency = 5.0 MHz, 3 db frequencies = 4.2, 5.8 MHz, effective bandwidth (including insertion loss) = 1.5 MHz. Noise temperatures were then derived by the usual Y-factor technique, i.e.

$$T_N(DSB) = \frac{YT_h - T_c}{1 - Y} \quad (2.28)$$

where

$$Y = \frac{P_{IF}(T_c)}{P_{IF}(T_h)}$$

G.R. noise spectra were measured either by connecting the IF to a commercial spectrum analyzer whose internal LO was controlled by an external sweep, as in the measurements of the 729 series detectors, or by sending the IF signal into an RF mixer driven with a second LO, as in the Eagle-Pitchard detector

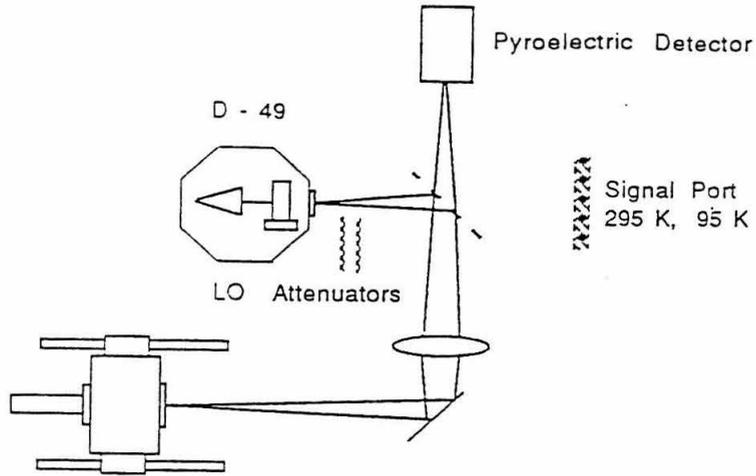


Figure 2.16 – Experimental setup for measuring heterodyne noise temperatures of *Ge:Ga* photoconductors.

measurements.

Table 2.8 lists the measured noise temperatures obtained in the first set of experiments at 118 μm , 96 μm , and 70.5 μm , and the tuning conditions under which they were obtained. Because the noise temperature is so high, (i.e. the heterodyne signal-to-noise so low,) the errors in the noise temperature are very large. It is apparent that, within these errors, the noise temperature is the same at all three wavelengths.

Table 2.8 – Initial measurements of $T_N(\lambda)$ (DSB)

Wavelength	70.5 μm	96.5 μm	118.8 μm
Detector bias	260 mV	260 mV	240 mV
LO power	27 μW	35 μW	130 μW
P_{gr}/P_N	.41	.45	.31
Beamsplitter transmission	.64 [†]	.80	.65
$T_N(\text{DSB})$	$20,800 \pm \begin{smallmatrix} 10,600 \\ 5,200 \end{smallmatrix}$	$30,300 \pm \begin{smallmatrix} 10,000 \\ 6,000 \end{smallmatrix}$	$24,200 \pm \begin{smallmatrix} 5,000 \\ 3,500 \end{smallmatrix}$

[†] discrepant with theoretically expected transmission of .78

The second series of heterodyne measurements provides clear and direct evidence that the guidelines

regarding detector optimization that have been discussed in chapter 1 are indeed correct. Detector LBL 108-17.7, our “benchmark”, lightly compensated detector, and LBL 712-7.2, our most lightly doped NTD detector, were tested for T_N under identical experimental conditions. In order to interpret the noise temperature measurements it is essential to determine the basic detector parameters. Figure 2.17 shows our determinations of the two detectors’ bandwidths. The data on the right, for detector 108-17.7, were obtained in later experiments, in which the sideband generator was used as a FIR modulator in the same way as was done for the 729-series detectors. The detector bias was 140 mV, the operating point which was found to yield optimum noise temperature in the heterodyne measurements. The data are well fit by a Lorentzian with $\nu_{3db} = 3$ MHz, about an order of magnitude less than the typical 729-series detectors.

No modulation measurements were performed on detector 712-7.2, but a crude estimate of its bandwidth was obtained from a series of g.r. noise spectra measured with varying laser power levels, one of which is shown at the left of figure 2.18. Unfortunately, these measurements, like many of the 729-series measurements, suffered from the presence of substantial excess noise at low frequencies. In addition, comparison of the spectra at different laser power levels shows clear evidence of saturation. That is, the IF noise power did not scale linearly with incident laser power level above a certain level, about 20 on the scale of figure 2.18. The reason for this is currently not understood. The DC detector current did scale linearly at all incident laser levels. Power levels at all the IF amplifiers were far below their respective saturation levels. The data shown in figure 2.18, however were taken at a low enough laser level that saturation is only significant for the excess low-frequency noise, and can be neglected in fitting the true g.r. noise spectra. Unfortunately, the bandwidths of 712-7.2 and of the low-frequency excess noise are comparable, so that the uncontaminated part of the g.r. noise spectrum lies entirely in the regime of 6 db/octave rolloff. Thus, considering these data alone, the 12 MHz fit to the g.r. noise spectrum is really only an upper limit rather than an accurate determination of the detector bandwidth.

Two additional facts, however, lead us to believe that 12 MHz is not far from the true detector bandwidth. By scaling the bandwidth from our next most lightly doped NTD detector, 496-5.5, ($\nu_{3db} = 16$ MHz at $E = .8E_{br}$), and taking into account the fact that 375 mV is only about .6 times the breakdown

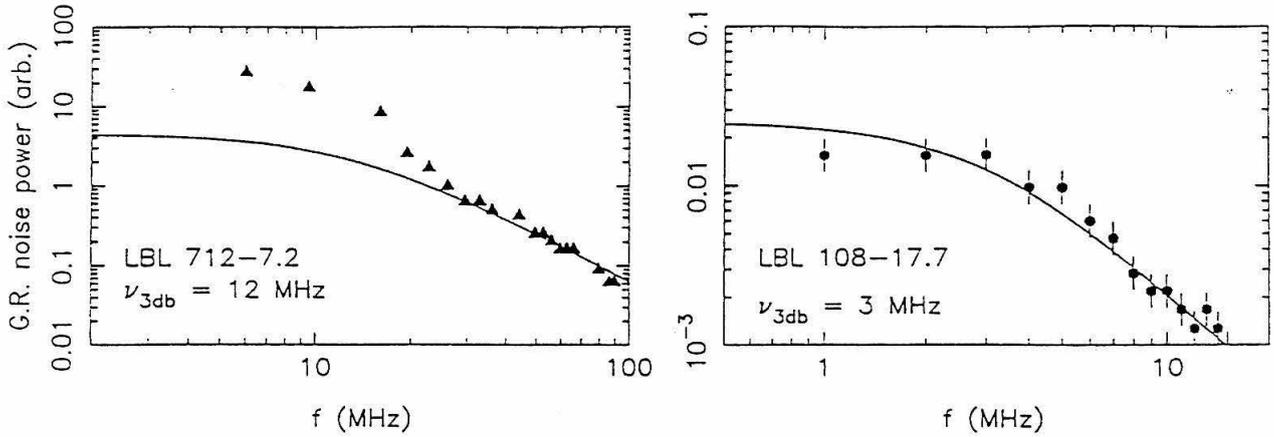


Figure 2.17 – Bandwidth determinations of detectors 712-7.2 and 108-17.7

voltage of 712-7.2, we would expect a bandwidth of 9.8 MHz. (The assumption behind this scaling is that the recombination cross-section is the same for both detectors.) Furthermore, a few measurements of heterodyne noise temperature were made for this detector at IF frequencies higher than 5 MHz by mixing the IF with a second LO. With a second LO frequency of 20 MHz, i.e. with the two sidebands located at (first) IF frequencies of 15 and 25 MHz, the mixer temperature was ~ 5000 K. With a second LO frequency of 15 MHz, it was only 2600 K. These values are not much higher than the T_{mixer} found at $\nu_{IF} = 5$ MHz, i.e. with no second LO, namely 2300 K (see table 2.9), implying a fairly high (> 10 MHz) detector bandwidth. Note that these figures, and those given in table 2.9, are *mixer* noise temperatures. I.e., they were obtained by applying equation 2.28 with a Y factor given by

$$Y_{mixer} = \frac{P_{IF}(T_h) - P_{IF}(LO\ blocked)}{P_{IF}(T_c) - P_{IF}(LO\ blocked)} \quad (2.29)$$

The noise temperature defined in this way neglects the contribution to receiver noise due to the IF amplifier. It also neglects that due to g.r. noise from the detector's dark current, or from its response to room-temperature background radiation, so it is slightly different from the commonly understood T_{mixer} which excludes only the IF amplifier contribution. However, these additional noise sources were very small in this particular set of measurements and can be neglected.

The responsivities and photoconductive gains of the two detectors were determined from measurements

Table 2.9 – Measured parameters of Ge:Ga mixers used in heterodyne tests

	LBL 108-17.7	LBL 712-7.2 (NTD)
E_{br}	2.8 V/cm	6.15 V/cm
Bias, V_d	140 mV	375 mV
Responsivity, $S(119\mu m)$	3.85 A/W †	.80 A/W
Bandwidth, B	3.0 MHz	12 MHz
Photoconductive gain, G	.17	.037
Responsive quantum efficiency, η_r	10 %	10 %
Receiver noise temperature, T_N	4800 K	4700 K
Mixer noise temperature, T_{mixer} ‡	1000 K	2300 K
LO power, P_{LO}	13 μ W	210 μ W*
Heterodyne noise temperature, η_{het}	9.2 %	4.0 %

† low power limit; 3 db saturation power is $\sim 130 \mu$ W.

‡ obtained from IF power levels with $P_{IF}(LO \text{ blocked})$ subtracted.

* highest LO power obtainable.

of g.r. noise made at the same time as the heterodyne experiments. The IF amplifier impedance during all the heterodyne measurements was set to 2.7 k Ω . A summary of the derived detector parameters is given in table 2.9. The responsivity of detector 108-17.7 was found to be somewhat saturated at our highest laser powers. We estimate the power at which the responsivity is reduced to half its small-signal value, (the “3 db compression point”), to be $\sim 130 \mu$ W. As expected, the higher bandwidth detector, 712-7.2, also has lower photoconductive gain and lower responsivity. The measured responsivities and (independently) measured photoconductive gains of the two detectors are in the same ratio, namely about a factor of five higher for detector 108-17.7. The implied responsive quantum efficiencies, η_r , are 10 % for both detectors. Since the two detectors have identical dimensions and were tested in the same integrating cavity and optical configuration, it is highly gratifying that the derived quantum efficiencies are in fact identical, and equal to a reasonable value.

At the end of table 2.6 are listed the best measured receiver and mixer temperatures for the two detectors, and the LO powers at which they were obtained. Because of the excess pre-amplifier noise during these measurements, the receiver temperatures are much higher than the mixer temperatures. The most striking thing about the comparison of the two detectors is the LO power that was required to achieve the best noise temperature. As expected, this was much higher for the lower gain detector, 712-7.2. Indeed, the LO power of $210 \mu\text{W}$ listed in table 2.9 is actually the highest LO power that was obtainable during the experiment, so it is really only a lower limit to the optimum P_{LO} . Since the noise temperature obtained at this power was not vastly different from that found for 108-17.7, however, it seems likely that $210 \mu\text{W}$ was not far from the optimum value. As discussed in chapter 1, the LO requirement is expected to vary as G^2 , so scaling from detector 108-17.7, we would expect a value of optimum P_{LO} of $275 \mu\text{W}$, in reasonable agreement with observation.

The reason that receiver temperature increases at higher than optimal LO powers is the reduction in detector impedance. Beyond a certain point, R_{IF}^2 falls faster than $\langle i_{signal}^2 \rangle \propto P_{LO}$ rises, so that both the LO-induced g.r. noise voltage and the signal voltage become less significant relative to the amplifier noise voltage. This is not indicated in the table, but it was quite obvious in the experiments on detector 108-17.7. When more than $\sim 25 \mu\text{W}$ of LO power are incident on the detector, removing an attenuator decreases the total IF power. Furthermore, from measurements of the I-V characteristic during the heterodyne experiments we can relate this behavior to the detector impedance directly. Under the conditions given in table 2.9, ($V_{det} = 140 \text{ mV}$, $P_{LO} = 13 \mu\text{W}$), the detector impedance was $1.6 \text{ k}\Omega$, about .6 times the amplifier input impedance. Driving the impedance lower than this with higher LO power degraded the receiver noise temperature. For detector 712-7.2, the full laser power of $210 \mu\text{W}$ was only sufficient to drive the impedance down to $3.2 \text{ k}\Omega$, about twice the impedance at which 108-17.7 had its optimum performance. This suggests that the optimum P_{LO} for 712-7.2 may have been as much as a factor of two (but not more) greater than the experimental value of $210 \mu\text{W}$.

Before moving on from these measurements, we note that the optimal *mixer* temperatures for the two detectors do not agree, as we would expect they should. The reason for this discrepancy, about a

factor of two, is unclear. There is also a somewhat smaller discrepancy between the optimal mixer noise temperature for 108-17.7 in these experiments (1000 K) and in the final series of experiments. The tuning of V_{det} and P_{LO} was done quite crudely in the experiments summarized in table 2.9, which may explain part of the latter discrepancy. The mixer temperatures given in the table include the 35 % signal loss due to the beamsplitter. When this is removed, the values of T_{mixer} may be converted to heterodyne quantum efficiencies. As shown in the table, this leads to $\eta_{het} = 9.2$ % for detector 108-17.7, in good agreement with the responsive quantum efficiency found from the comparison of g.r. noise and responsivity. For LBL 712-7.2, the factor of two higher mixer temperature carries through to a corresponding discrepancy between the responsive and heterodyne quantum efficiencies.

Our final set of heterodyne experiments were made after the excess GaAsFET pre-amplifier noise had been eliminated by re-flowing several solder joints. Its noise performance was then nominal, as summarized in figure 2.5 and table 2.1. In this condition, the pre-amplifier noise comprised only ~ 25 % of the total IF power with LO applied. That is, the receiver was well into the ideal, LO-induced g.r noise-dominated regime, and the receiver noise temperature and mixer noise temperature (as defined by equation 2.29), were more nearly equal. All the measurements made in this condition were done on detector LBL 108-17.7 and at $\lambda = 119 \mu\text{m}$.

Figure 2.18 shows a pair of tuning curves for the receiver noise temperature obtained during these experiments. For the solid curve, the LO power was varied (by inserting a series of FIR attenuators) while holding the detector voltage fixed. For the dashed curve, the detector voltage was increased at lower LO powers so as to hold the detector current fixed. This has the effect of reducing the variation in detector impedance as the LO power is varied. The behavior of the two tuning curves can be understood quite simply. At high LO powers, the detector impedance dominates (i.e. is smaller than) the pre-amplifier input impedance, and both LO-induced g.r. noise *voltage* and heterodyne signal *voltage* are reduced in comparison to amplifier noise. Thus, the receiver noise climbs rapidly, although simultaneously reducing the detector bias (and therefore the responsivity), slows down the increase in noise temperature. Likewise, reducing the LO power below the point ($\sim 2 \mu\text{W}$ for this experimental configuration), where the detector

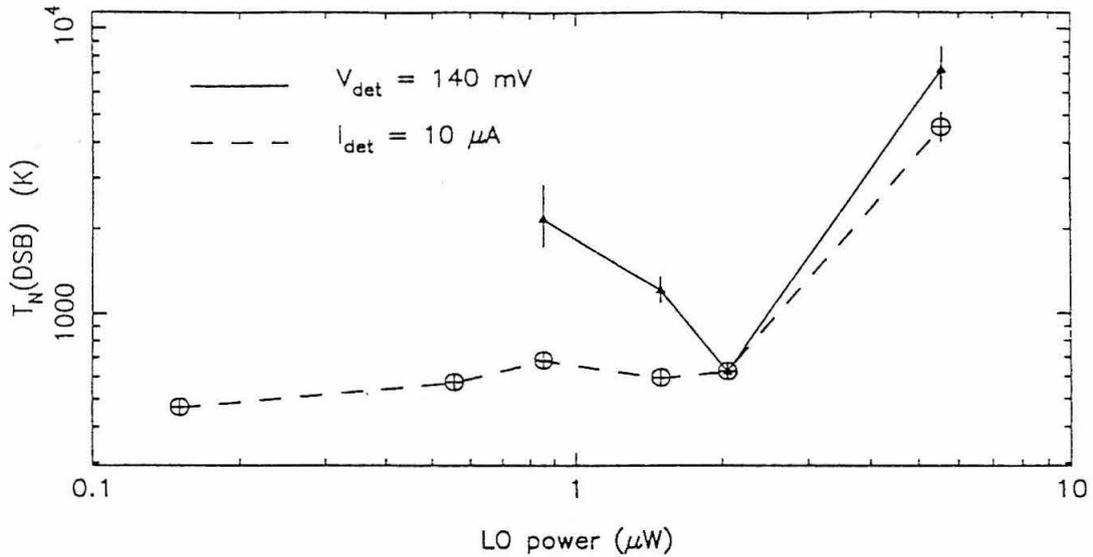


Figure 2.18 – Heterodyne tuning curves for detector 108-17.7, showing our lowest achieved noise temperature, $T_N = 590 \text{ K}$, at $P_{LO} = 1.6 \mu\text{W}$, uncorrected for the background g.r. noise or the Planck spectrum of the hot and cold signal loads.

and amplifier impedances are matched also reduces the signal relative to the amplifier noise, and causes the receiver noise to rise sharply. This is simply loss of conversion gain due to low LO power. Increasing the detector's responsivity by means of increased bias can almost completely compensate for this, however. The constant current curve rises very little below the optimum LO power. There is a limit to the extent that increased responsivity can compensate for low LO power, however. This limit is due to the g.r. noise induced by the room temperature background radiation (see chap. 1). Since this g.r. noise is reduced when a cold load is introduced into the signal beam, it to some degree mimics true heterodyne signal. By mindless application of equation 2.28, it is possible to infer quite respectable heterodyne noise temperatures with zero LO power applied. This effect occurs fairly often with other types of high sensitivity heterodyne receivers, e.g. superconducting tunnel junctions (Wengler, private communication), as well.

In our experiments, the effect may be separated out by measuring the IF power not only with hot and cold loads at the signal port and the LO applied, but also with hot *and* cold loads at the signal port and the LO blocked. This was not done for all the measurements shown in figure 2.18, unfortunately, but it was done for a series of measurements at $V_{\text{det}} = 140 \text{ mV}$ and varying LO powers taken during the same

experiment. At $V_{det} = 140$ mV and $P_{LO} = 2 \mu\text{W}$, i.e. at the minimum of the solid curve in the figure, the effect was fairly small. The receiver noise temperature increased from its raw value of 625 K (shown in the figure) to only 670 K when the effect was subtracted out. Because the IF power with LO blocked and a hot signal load was only $\sim 15\%$ higher for the next lower point on the constant current curve, ($V_{det} = 160$ mV, $P_{LO} = 1.6 \mu\text{W}$), than at 140 mV, it is unlikely that the correction is much larger there. At the lowest two points of the dashed curve, however, the hot, LO-blocked IF power is a factor of 2 – 3 higher than at 140 mV, and it is quite possible that nearly all the observed “signal” is in fact due to background g.r. noise. For these points, the correction could be huge, and we therefore do not consider the measured receiver noise temperatures to be reliable.

Thus, our best, reliably measured receiver noise temperature is from the $10 \mu\text{A}$ ($V_{det} = 160$ mV, $P_{LO} = 1.6 \mu\text{W}$) point. The raw, uncorrected, noise temperature at this point is 590 K. Applying the same correction for the background g.r. noise as was found for the neighboring point on the curve yields 635 K. A final effect that must also be corrected for is the “Planck correction”, namely, the fact that the radiation density in W/Hz is not precisely proportional to temperature, since $119 \mu\text{m}$ is not really that far into the Rayleigh-Jeans regime. It may be taken into account by replacing $T_h = 295$ K and $T_c = 95$ K in equation 2.28 with

$$\begin{aligned} T'_h &= \frac{h\nu}{k} \left(e^{\frac{h\nu}{kT_h}} - 1 \right)^{-1} = 239 \text{ K} \\ T'_c &= \frac{h\nu}{k} \left(e^{\frac{h\nu}{kT_c}} - 1 \right)^{-1} = 47 \text{ K}. \end{aligned} \quad (2.30)$$

This yields our final best estimate for the true receiver noise temperature :

$$T_N(DSB) = 655 \text{ K} \quad (2.31)$$

at $P_{LO} = 1.6 \mu\text{W}$. Taking the beamsplitter transmission into account yields for the heterodyne quantum efficiency

$$\eta_{het} = \frac{h\nu}{2kT_N(DSB)} = 14\%, \quad (2.32)$$

in rough agreement with the quantum efficiency derived from the photoconductive gain and responsivity.

Our final figure for receiver noise temperature is a factor of ~ 25 better than the best reported noise temperature for a Schottky diode receiver at the same wavelength, viz. 17,000 K (Röser *et al.* 1986).

As discussed in chapter 1, however, a simple comparison of noise temperatures puts narrow-bandwidth receivers such as ours in the most advantageous light possible. (Some might say an unfairly advantageous light.) It compares the signal-to-noise ratios that would be realized in observing a source that is unresolved in both receivers' instantaneous bandwidths. Taking the opposite extreme case, of a source which fills the bandwidths of both receivers, the multiplex advantage regains the Schottky receiver a factor of the square root of the ratio of bandwidths (500 MHz vs. 3 MHz). Even in this case, the Ge:Ga photoconductor is superior, by about a factor of 2 (see figure 1.8). In the astronomically more realistic intermediate cases, the advantage of the Ge:Ga photoconductor would lie between these two factors. The great drawback of the photoconductor is the requirement of a continuously tunable local oscillator.

The power requirement on the local oscillator is two to three orders of magnitude less severe for the Ge:Ga photoconductor than for the Schottky diode receiver; only some $1.6 \mu\text{W}$ were required to yield our optimum noise temperature. At present, however, this is not enough to make up for the requirement of continuous tunability. Undoubtedly, this situation will improve in the next few years, at which time it may be expected that Ge:Ga photoconductors will come to realize their full potential as ultra-low noise mixers. As for reducing the LO power requirement in present-day systems, the heterodyne measurements we have reported here, as well as our experiments on the g.r. noise spectra of other detectors (82-4.6, the 729 series), have shown that, depending on the detector's photoconductive gain, the local oscillator power requirement is determined by either the FIR background power or the IF amplifier noise. At present, the requirement imposed by the background is the more stringent one, both in our laboratory system with detector 108-17.7, and in the airborne system with detector 82-4.6. With improved FIR filtering of the background, it may be possible to reduce the LO power requirement to the level where it could be fulfilled by a present-day sideband generator. Some degradation in noise temperature would no doubt be introduced, since real narrow-band filters (resolving power > 100 would be desired) do not have unity transmission, but this would be a small consideration compared to the reduction in background g.r. noise.

Chapter 3 – Far-infrared Laser

The heart of our receiver's local oscillator is the far-infrared laser. The optically pumped molecular gas laser was discovered by Chang and Bridges in 1970. In the decade or so following its discovery, a great deal of work was done on theoretically understanding and empirically optimizing its steady-state performance. Complete far-infrared laser systems are now available commercially from a number of sources. The commercial models, however, are generally quite expensive, and, in any case, are not sufficiently compact for mounting on the KAO. It was therefore decided that our laser would be built in-house. It was designed and constructed by D. M. Watson. For us, as for most FIR laser users nowadays, the laser was intended to be a scientific tool rather than an object of study in its own right. Indeed, except for the important aspect of spontaneous pulsations, instabilities, and noise, the devices have really passed beyond the stage of being a subject of basic physics research in their own right. Furthermore, even from an engineering point of view, our data on the performance of our laser is not extensive or careful enough to really add much to the state-of-the-art of FIR laser technology.

Thus, this chapter is a relatively short one. The physics of FIR laser oscillation is reviewed at only the very basic level necessary to understand the motivation of our laser's design. The theory review naturally divides itself into discussions of the molecular gain medium and of the optical cavity. Next, we discuss the details of our laser system's construction, first the 10 μm CO₂ laser which forms the pump, and which was bought commercially, then of the FIR cavity, and finally, of the two components used for amplitude stabilization of the FIR output, namely the optical isolator and the optoacoustic lock loop. The performance of the latter two components is also described. In the third section, we discuss the measured performance of the system, in particular the power and far-field mode patterns of the output at various FIR wavelengths. In a final section, we discuss in some detail one of the non-linear dynamical effects to which FIR lasers are subject, the Lorenz instability. The experimental identification of Lorenz-type spontaneous pulsations and deterministic chaos in far-infrared lasers has been made only quite recently, (Högenboom *et al.* 1985, Weiss and Brock, 1986,) considerably too late, in fact, to be taken into account in the design of our system. These effects are of great interest in their own right, but are still not well understood. Their

existence has important implications for laser applications, such as ours, in which the noise spectrum of the laser within the rather low bandwidths of the carrier is important. They also have important implications for the development of more compact, higher pressure FIR lasers. We have not studied these phenomena systematically, but, as we shall show, we have good reason to believe that our laser system exhibits the same effects identified by others as being of Lorenz type.

3.1 Basic Theory – The gain medium

The gain cycle of an optically pumped far-infrared laser is illustrated in figure 3.1. The strong electromagnetic field of a $10\ \mu\text{m}$ CO_2 laser, of frequency Ω_{CO_2} , is tuned to resonance with a transition between a rotational state in the lasing molecule's ground ($v=0$) vibrational manifold and a rotational state in its first excited vibrational manifold ($v=1$). We shall denote the former state 0 and the latter 1, in a standard 3-level atom treatment. The lasing molecule must have a permanent electric dipole moment, so that the pump and laser transitions are electric-dipole allowed. The lasing transition is a pure rotational transition from state 1 to a lower rotational state, which we denote 2, in the same ($v=1$) vibrational manifold. Within each vibrational manifold, a Boltzmann distribution of the rotational level populations is maintained by binary molecular collisions, which occur with rate γ . (This does not apply, of course, to states 0, 1, and 2, whose populations are strongly affected by interaction with the electromagnetic field.) The cycle is completed by vibrational de-excitation at rate Γ via collisions, either binary molecular collisions or collisions with the walls of the laser cavity. Usually, wall collisions dominate. The vibrational de-excitation is the rate-limiting step in the cycle.

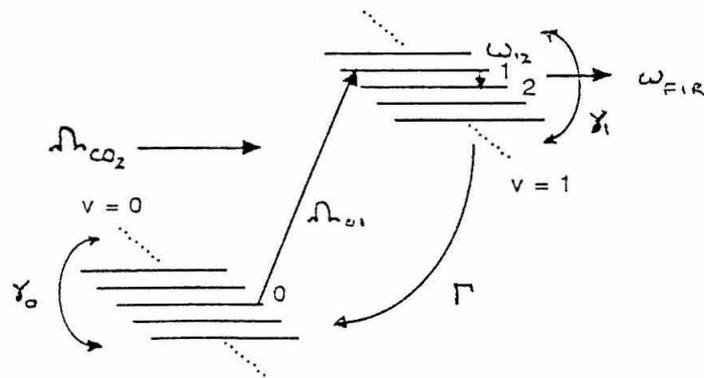


Figure 3.1 – FIR laser cycle

It is extremely difficult to compute from first principles the gain coefficient α (in cm^{-1}) of the FIR transition. For one thing, the quantum number assignments for states 0, 1, and 2 are unknown for most laser lines, including all the difluoromethane (CH_2F_2) lines we have used. Even when the state assignments are known, however, as they are for the 119 μm and 70.5 μm methanol lines, accurate calculation of α is not possible. The most thorough attempt was made by Heppner *et al.* (1980), whose predicted gain coefficients came out a factor of ~ 5 higher than the observed values. There are two basic difficulties. The first is accurate evaluation of the partition function, which is needed in order to determine the level populations. The second is the fact there is no single mechanism that strongly dominates the broadening of either the pump or lasing transitions.

Table 3.1 lists the gain coefficients and pump absorption coefficients experimentally determined by Heppner *et al.* (1980) for the two methanol lines we have used the most – those at 119 μm and 70.5 μm . The low values of pump absorption at typical operating pressures are characteristic of far-infrared laser lines, and are the reason why quite long and bulky laser cavities are required. In practice, it is very difficult to construct cavities that have high (i.e. greater than a few) finesse at 10 μm and are still acceptable in their far-infrared properties, so the only real hope for miniaturizing the systems lies in raising the operating pressure.

Table 3.1 – Parameters of the 119 μm and 70.5 μm methanol laser lines

Parameter	119 μm	70.5 μm	Comments
CO ₂ pump	9P(36)	9P(34)	
Frequency offset	25 MHz	-25 MHz	CO ₂ pump - FIR absorption
$ \mu_{12} $	0.66 D (= 0.66×10^{-18} esu-cm)	0.36 D	Laser transition dipole moment
$ \mu_{01} $	0.12 D	0.105 D	Pump transition dipole moment
β_{pump}	$1.5 \pm .2 (\text{m} - \text{torr})^{-1}$	$2.3 \pm .3 (\text{m} - \text{torr})^{-1}$	pump absorption coefficient
α_{max} (FIR)	0.29 m^{-1}	0.17 m^{-1}	max. FIR gain coefficient
p_{max}	150 mtorr	200 mtorr	pressure for maximum gain

The upper limit on operating pressure, according to the conventional theory, (as developed by Tucker, 1974), for the case of a homogeneously broadened FIR transition, is set by the following criterion for positive gain :

$$\frac{h\nu}{kT} \left(\frac{\gamma}{\Gamma} \right) f_u < 1. \quad (3.1)$$

Here, f_u is the Boltzmann factor for the fraction of the $v=1$ molecules which, in thermal equilibrium, would be in state 1. The factor in parentheses is the ratio of rotational relaxation rate (dominated by molecular collisions,) to vibrational relaxation rate (dominated by wall collisions). Since the rotational relaxation rate is proportional to pressure, and the vibrational relaxation rate is inversely proportional, through the diffusion constant, to pressure, equation 3.1 sets an upper limit on the operating pressure. Physically, equation 3.1 expresses the competition between collisional redistribution of the $v=1$ manifold, which acts to thermalize the laser level populations, and vibrational de-excitation of the $v=1$ manifold, which fuels the pump transition and therefore acts to invert the laser levels.

The low operating pressure of FIR lasers has two other highly undesirable consequences, besides limiting the gain coefficient and therefore requiring meter-long cavities. Both have to do with the low homogeneous linewidth, typically 3 - 10 MHz, resulting from low pressure. The first problem is the low tuning bandwidth, i.e. the fact that the FIR cavity's resonant frequency, ν_c , which determines the laser's precise output frequency, can only be tuned over the range for which the gain is near its peak value, and the gain width is determined by the width $\Delta\omega_{12}$ of the molecular transition. In most other gas lasers, helium-neon lasers for instance, pressure broadening dominates $\Delta\omega_{12}$ by a large factor. The pressure-broadened linewidth is so large, in fact, that it exceeds the spacing of the cavity's longitudinal modes, $\frac{c}{2L}$, and the laser will lase regardless of the cavity tuning. FIR lasers don't have this convenient feature, and, even in cases where pressure broadening still dominates other sources of broadening in $\Delta\omega_{12}$, they can only be tuned within a few MHz of the molecule's natural transition frequency.

What is even worse, however, is that at these low pressures, the pressure-broadening may be only comparable to, or even less than, the other broadening mechanisms of ω_{12} . It is important to distinguish two different types of broadening mechanisms for laser transitions. "Homogeneous" broadening refers to

processes in which the probability distribution of a molecule's emitting a laser photon at a given frequency offset ($\omega - \omega_{12}$) from the center frequency is the same for every molecule. Examples are pressure-broadening and "natural" or lifetime broadening. "Inhomogeneous" broadening mechanisms involve different subsets of the molecules in the upper state each having a slightly different frequency of peak emission. Examples are Doppler broadening and AC Stark effect broadening. The problem associated with low pressure-broadened linewidth is the existence of various types of instabilities. It is far beyond the scope of this chapter to describe these instabilities in any detail. In any case, the current understanding of them is very incomplete, and, in some cases, still quite controversial. At the end of this chapter we shall discuss one instability, the Lorenz instability, but only to the extent necessary to relate our experience to that of other workers' observations. In general, the regimes in which instabilities can occur are described by ratios of the linewidths due to various broadening mechanisms, and/or of the passive cavity linewidth. The reason we have chosen to discuss the Lorenz instability in more detail is that it is the only instability which would not be automatically quenched if, through some future technical advance, it became possible to operate FIR lasers at much higher pressures than presently (pressure-broadened linewidth \gg all other gain linewidths).

Since the width of the pump absorption is always dominated by Doppler broadening, (collisions are much less effective at inducing vibrational transitions than rotational ones,) the circulating pump radiation only interacts with molecules within two narrow velocity intervals. Thus, the gain spectrum of the laser transition resembles figure 3.2. A small fraction of the molecules in the upper laser level have undergone collisions since being excited to $v = 1$, and therefore have velocities near zero. Most of the active molecules, however have velocities for which the Doppler shifts of either the forward or backward-going pump beams bring the pump frequency into exact resonance with the vibrational transition frequency. When the cavity frequency is tuned to lie between the two peaks, instabilities can occur due to beating or switching between the FIR beams emitted by the two sets of emitting molecules (Abraham *et al.* 1985, Lefebvre, Dangoisse, and Glorieux, 1984). Clearly, such an instability will be eliminated if the pump is tuned to lie precisely on resonance. Alternatively, the instability will disappear if the linewidth of the FIR transition, due e.g. to pressure-broadening, is much greater than the frequency separation of the two groups of molecules.

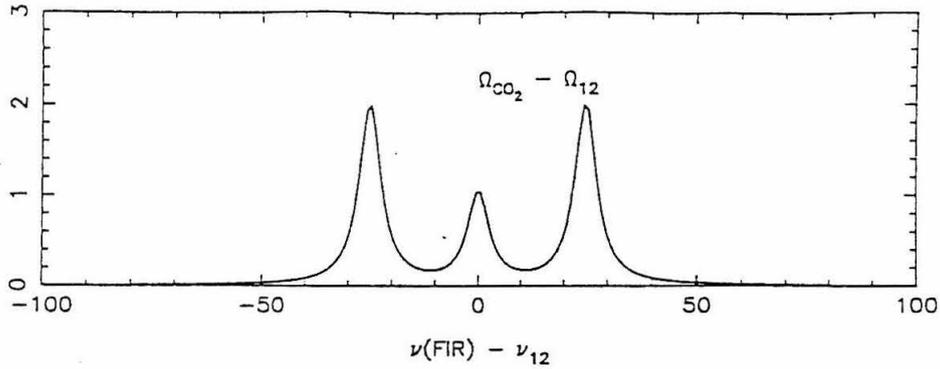


Figure 3.2 – Gain spectrum of the laser transition. A frequency offset of 25 MHz, and a purely pressure-broadened linewidth of 6 MHz FWHM, corresponding to the 119 μm methanol line have been assumed.

A similar double-peaked gain spectrum, and similar instabilities, are possible if the AC Stark effect, or “Autler-Townes” effect, is important. This effect is a general consequence of the quantum-mechanical interaction of a two-level system with a strong electric field. Assuming the interaction Hamiltonian may be written as $H' = -\mu E$, where μ is the 2-level system’s electric dipole moment, and that the pump is on resonance, it is then a textbook exercise (e.g. Yariv 1975) to show that the probability of finding the system in the upper (or the lower) of the two states (i.e. the diagonal components of the density matrix,) oscillates harmonically with time at the Rabi frequency, $\omega_R = \frac{\mu E}{\hbar}$. This splits the upper and lower pump levels (levels 0 and 1) by the Rabi frequency. The splitting of level 1 is then manifested by a double-peaked gain spectrum for the laser transition, and it may be shown (Heppner *et al.*, 1980 and references therein) that the separation of the gain peaks is given by $\delta\omega = 2\omega_R(\omega_{12}(\omega_{10} - \omega_{12}))^{1/2}$, which is of order ω_R . Heppner *et al.* (1980) directly measured many such double-peaked gain profiles and used them to accurately determine the dipole moments of the pump transitions. It is clear that this effect is most important under conditions of high pump intensity, i.e. either high CO₂ laser power or a small CO₂ beamwidth in the active medium. As with Doppler broadening, the effect becomes irrelevant, in terms of generating instabilities, if the homogeneous (i.e. pressure-broadened) linewidth is much larger than the splitting.

In table 3.2 we list the important molecular parameters and formulae that enter into the FIR broadening of the 119 μm line, along with typical values under our operating conditions. No single mechanism

Table 3.2 – Laser line-broadening mechanisms

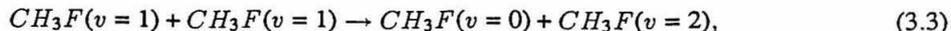
Mechanism	FWHM linewidth
Pressure-broadened linewidth	$\Delta\nu_p = \gamma_1/\pi = 5.9 \text{ MHz} \times (p/200 \text{ mtorr})$
Rabi frequency of pump	$\nu_{Rabi} = \frac{ \mu_{01} E_{pump}}{h} = 1.7 \text{ MHz } I_{W/cm^2}^{1/2}$
Doppler-broadened linewidth	$\Delta\nu_D = \frac{c}{\nu} \left(\frac{2 \ln 2 kT}{m} \right)^{1/2} = 2.8 \text{ MHz}$
Cavity linewidth	$\dagger \Delta\nu_c = c\Im(\gamma) = \frac{c\ell}{2L} = 4.5 \text{ MHz} \left(\frac{\ell/3\%}{L/100\text{cm}} \right)$

\dagger Here, $\Im(\gamma)$ is the the amplitude loss coefficient (in cm^{-1}), and ℓ the one-way cavity power loss.

dominates the broadening by a very large factor. AC Stark broadening (splitting), Doppler broadening (splitting), and pressure broadening all produce linewidths that are comparable with one another and comparable with the passive cavity linewidth. It is this fact more than any other that makes accurate theoretical analysis of the FIR laser performance intractable.

Having exhibited the substantial drawbacks of low operating pressure in FIR lasers – low gain coefficient, low tuning bandwidth, and instability – we note that, very recently, it has been reported that much higher pressure operation may be possible in some cases. Everitt, Skatrud, and DeLucia (1986) have reported lasing at the $\sim .1 \text{ mW}$ level on a line of $^{13}\text{CH}_3\text{F}$, at pressures as high as $\sim 3 \text{ torr}$. The specific line is not mentioned, but it is presumably the $1221 \mu\text{m}$ line pumped by the 9P(32) CO_2 pump. The conditions of the experiment were such that the cutoff pressure predicted by conventional theory was 400 mtorr. The higher gain coefficient allowed them to operate with a cavity length of only 15 cm. In some cases, they also observed lasing in a cavity only 5 cm long. They directly measured a tuning bandwidth of 24 MHz at 400 mtorr, and deduced a potential tunability of 100 MHz at the full 3 torr operating pressure. Furthermore, they found the maximum pressure in their experiment was limited only by the available pump power, and they predict a maximum operating pressure of $\sim 25 \text{ torr}$ and tuning bandwidth of $\sim 1 \text{ GHz}$ for a 100 W pump. The theoretical explanation they give for the breaking of the “vibrational bottleneck” in their experiment is the following. In the high pump intensity limit, which they achieved in practice by focussing the CO_2 beam rather than by using a particularly high power laser, the population in the

vibrationally excited state becomes sufficiently high that collisional processes such as



which have small energy defects, can be highly probable. Inclusion of the higher vibrational levels ($v \geq 2$) thus can substantially increase the vibrational de-excitation rate to the ground state, since each molecule in the $v=n$ state which diffuses to the wall and is de-excited, has actually de-excited a total of n molecules – itself plus the $n-1$ $v=1$ molecules that were necessary to excite it. Whether such processes are generally significant in other laser molecules remains to be seen, but on the face of it, there seems to be no reason why they shouldn't be. If we may offer here a (wholly unsupported) general opinion, it appears to us that, if optically pumped molecular FIR lasers have any long-term future in astronomical applications, particularly in space-borne astronomical applications, then that future lies with these ultra-high pressure lasers rather than with the conventional types.

Electromagnetic modes

There have conventionally been two types of resonant cavity used for FIR lasers, those based on confocal, semi-confocal, or near-confocal cavities, in which a free-space mode is set up between reflecting mirrors, and those based on waveguide modes. It is generally agreed (see e.g. the review of Hodges, 1977) that the former are inferior for most applications. The reasons are poor mode control and inconveniently large transverse dimensions. The large transverse dimensions are dictated by the necessity of getting low beam divergence over the ~ 1 meter absorption length required by the molecule's pump absorption coefficient. At far-infrared wavelengths, this leads to transverse dimensions of 10 - 20 cm. Since the angular alignment of the mirrors must be good to within the single-mode diffraction limit of the transverse cavity dimension, the free-space-propagation type of resonant cavity is much more sensitive to misalignment than the waveguide type.

The more commonly used resonant cavity employs an electromagnetic mode which, over some portion of its extent, is guided. Most commonly, the waveguide is a hollow cylinder of dielectric or of metal. The modes of hollow dielectric waveguides in the "overmoded" limit of radius $a \gg \lambda$, were first derived

by Marcatili and Schmeltzer (1964). The lowest order modes are illustrated in figure 3.3. In our system, (and in most other FIR laser applications as well,) we attempt to tune the resonator so that laser oscillation occurs only in the lowest order “hybrid” mode, EH_{11} . This mode is desirable because it is linearly polarized and has its maximum amplitude at the center. It may be shown (Degnan 1974) that the overlap integral of the EH_{11} mode with a Gaussian beam is maximized for a Gaussian beam waist size of $w_0 = .6435a$. The complex propagation constants for the EH_{nm} modes, γ_{nm} , are given by

$$\Re(\gamma_{nm}) = k \left[1 - \frac{1}{2} \left(\frac{u_{nm}}{ka} \right)^2 \left(1 + \frac{2}{ka} \Im(\kappa) \right) \right] \quad (3.4a)$$

$$\Im(\gamma_{nm}) = \frac{1}{a^3} \left(\frac{u_{nm}}{k} \right)^2 \Re(\kappa) \quad (3.4b)$$

where

$$\kappa = \frac{\rho^2 + 1}{2\sqrt{\rho^2 - 1}}$$

and ρ is the ratio of the dielectric waveguide’s complex index of refraction to the internal medium’s. The free space propagation constant is $k = \frac{2\pi}{\lambda}$, and u_{nm} is the m -th root of the Bessel function J_n . Equation 3.4b expresses the well known scaling behavior (Hodges, Foote, and Reel, 1977) of FIR waveguides, that the cavity loss is proportional to the square of the wavelength and the inverse cube of the waveguide radius. Thus, larger diameter waveguides are required at longer wavelengths.

Feedback in waveguide resonator systems is accomplished by mirrors at the two ends of the waveguide, which double as input and output couplers. Most of the ingenuity in FIR laser engineering has gone into dreaming up newer and cleverer types of input/output couplers, and several different types have been used. The goals in designing them are to achieve a), perfect coupling of the available $10 \mu\text{m}$ pump power into the cavity, b), a high cavity Q at $10 \mu\text{m}$, so that all the available pump power is absorbed in the gas, and c), a far-infrared cavity which couples out the optimal (in terms of overall output power) fraction of the circulating FIR power and which has negligible other losses. For the $119 \mu\text{m}$ line, the optimal output coupling fraction is 20 % (Julien and Lourtioz, 1980).

3.2 Construction details of our laser – CO_2 pump

Our CO_2 pump laser – a model RF 160GS – was designed and built commercially by Laakmann Electro-optics (San Juan Capistrano, CA). It is transversely excited by a 165 W radio-frequency wave at

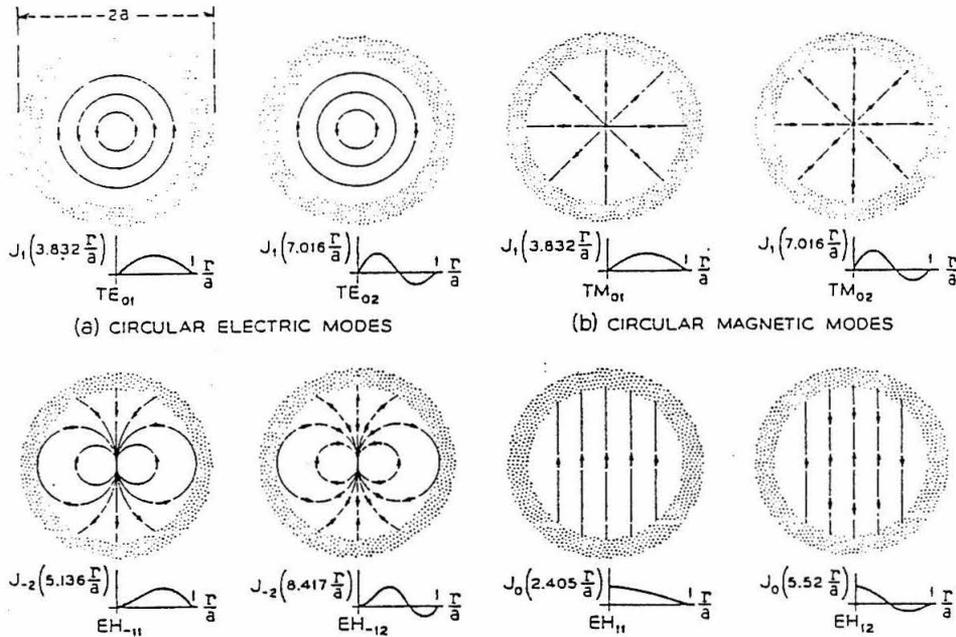


Figure 3.3 – Lowest modes of a highly overmoded dielectric waveguide, after Marcatili and Schmeltzer (1964). The electric field is plotted.

41 MHz (an industrial heating frequency band). The RF power is provided through ordinary high-power coaxial cables from an external rack-mounted power supply. In practice, 41 MHz power permeates our entire laboratory and every piece of equipment in it whenever the laser is on. Even after taking Draconian measures to isolate and filter our apparatus, we have not succeeded in completely eliminating a large 41 MHz signal from our IF passband. Since the front-end GaAsFET pre-amplifier can easily detect 10^{-15} W signals at this frequency, our failure to achieve total isolation is not surprising. We have been satisfied with achieving sufficient isolation to prevent saturation of any of our sensitive components, and have simply resigned ourselves to the existence of a “dead band” in the IF around 41 MHz. Since most of our mixers, including the ones we have used for the airborne system, roll off at considerably lower frequencies, this is not a serious limitation on overall performance. We note that other FIR laser systems used for astronomical heterodyne LO’s (Densing *et al.* 1985, Harris 1986) employ DC discharge-excited CO₂ lasers. Partly, this is for historical reasons; the early CO₂ lasers were all DC discharge-excited, and it is only in the last few years that the forefront of CO₂ laser engineering has shifted to RF excitation. In our application, DC discharges do avoid the problem of RF interference, but have the practical difficulties involved in dealing

with very high voltages (~ 10 kV).

Our CO₂ laser is fairly compact – 29 inches long – and is based on a ceramic (BeO) waveguide resonator. Most earlier CO₂ lasers were based on resonators employing free-space propagation. The tradeoffs are the same as for the FIR resonator : a free-space mode has larger transverse dimensions (a typical bore diameter might be 1 cm) and therefore higher sensitivity of the output mode to the mirror alignment. On the other hand, the loss is theoretically lower for free-space propagation. The bore of our CO₂ laser is rectangular, with an interelectrode spacing of 2 mm.

The laser had a forced-air cooling system when delivered. This was found to be inadequate for preventing sizable thermally-induced drifts in the output wavelength, and was replaced with a homemade liquid-cooled heatsink on which the laser was mounted. Ordinary automotive coolant is flowed through the heatsink and through a commercial cooling unit (Neslab Inc., Newington, NH). The laser cavity length is tuned by an ordinary voice coil on which the output coupler is mounted. The laser runs sealed off. Highest output power is obtained with a pressure of about 70 torr of laser mixture, although the power varies relatively slowly with pressure between 60 and 75 torr. A laser mixture that included a small fraction of xenon (CO₂ : N₂ : Xe : He = .15 : .15 : .10 : .60, obtained from Linde/VBA Inc.) yielded nearly a factor of two higher power than a similar mixture that lacked xenon. The operating pressure of 70 torr is considerably higher than that typically used in DC discharge-excited lasers. It leads to a relatively high available tuning bandwidth, ± 200 MHz around line center, according to the manufacturer's claims. This increases the probability of a chance coincidence with a molecule's vibrational transition, and makes a greater number of FIR lines available. The laser's output aperture is 1 mm in diameter, and we have found the mode to nearly always be a clean EH₁₁ pattern.

Tuning between different CO₂ transitions is accomplished via a micrometer-driven grating at the opposite end of the laser cavity from the output coupler. Verification of the CO₂ laser wavelength was accomplished with a standard 10 μ m spectrum analyzer (Optical Engineering Corp., Santa Rosa, CA). It was determined that the reproducibility of the grating orientation was sufficient that, once calibrated, the reading on the grating micrometer alone could be used to switch between lines, and the spectrum

analyzer could be dispensed with. CO₂ laser powers were measured on a standard thermocouple-based power meter (Coherent Technology, Auburn, CA). On its strongest lines, the CO₂ laser produces 11 - 12 W, corresponding to a peak efficiency of 7 % ($\frac{P_{laser}}{P_{RF}}$). On the 9P36 line used to pump the 119 μm methanol line, it typically produces 4.5 W.

Finally, we note that much higher performance CO₂ lasers than ours are commercially available. The present state-of-the-art in “low to medium” power CO₂ lasers is dominated by the RF-excited waveguide technologies developed at Hughes and at United Technologies. To give some concrete examples (without sounding too much like an advertisement for these companies’ products,) we note that United Technologies manufactures a laser, the “CADET” model, which delivers > 50 W on its strongest lines, 25 W on 9P36, and measures 48 cm x 14 cm x 10 cm, including its RF power supply. (See e.g. Newman and Hart, 1986.) Chiou’s study demonstrated an accurately linear relation between 10 μm pump power (above threshold) and FIR output power, although some researchers have reported (Farhoomand, private communication) a faster than linear dependence. It is clear, therefore, that the brute-force approach of simply using the most powerful available CO₂ laser that fits our compactness requirement would yield at least a factor of 5 improvement in local oscillator power.

Far-infrared cell

Our far-infrared cell is formed by a one meter long, hollow, fused quartz waveguide, with vacuum boxes at the ends in which the input and output couplers are mounted. The diameter of the waveguide is 17 mm, and was chosen on the basis of Chiou’s (1983) study of the power dependence of the 119 μm methanol line on waveguide diameter, pressure, and pump power. It turns out that the uniformity of the waveguide is quite critical when a circularly polarized pump beam is used. In this case, the two orthogonal linear polarization modes of the FIR output are degenerate. Theoretically, this should make it easy to control or “pin” the output polarization by deliberately introducing a very small amount of loss preferentially into one of the linear polarizations, e.g. by stretching a few parallel wires across the end of the waveguide. Without such a differential loss mechanism, one would expect a circularly polarized output. In practice, things did not work out so nicely. Our first waveguide was sufficiently non-axisymmetric (whether in physical radius,

surface roughness, or index of refraction we do not know,) that, when pumped with a circularly polarized $10\ \mu\text{m}$ beam, the FIR output was substantially elliptically polarized. Specifically, when a single polarizer was placed at the FIR output and rotated, variations of $\sim 20\%$ in transmitted power were seen, (see figure 3.4). Rotating the quartz tube rotated the position angle of the modulated component of transmitted power. Complete ellipsometry of the FIR output was not done; we know the relative amplitudes of the linear components of the FIR output but not their relative phases. Attempts to “pin” the output polarization with stretched wires across the end of this waveguide failed. The ratio of the linearly polarized components increased from 1.2 to only 1.5.

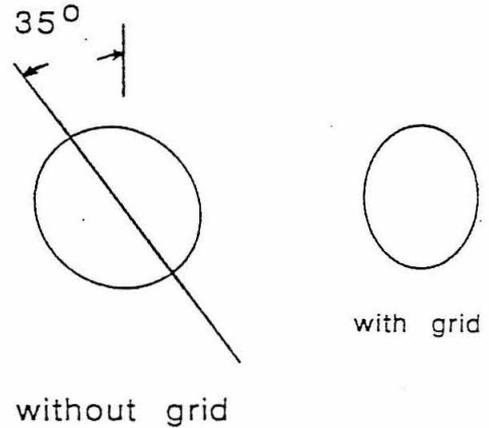
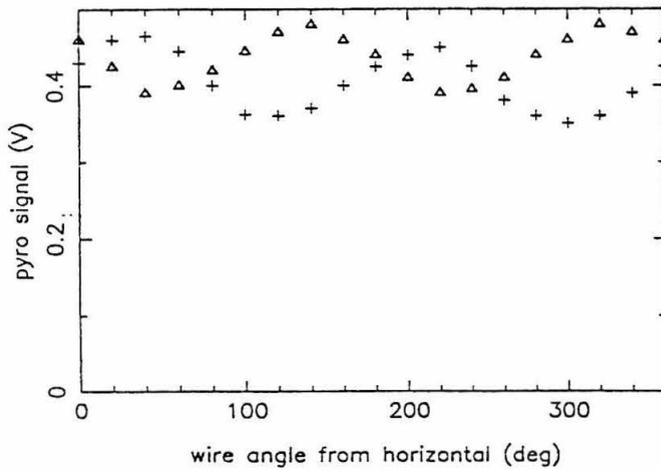


Figure 3.4 – (left) Variation of transmitted power through a linear polarizer at a given waveguide orientation (crosses), and with the waveguide rotated 45° (triangles); (right) Effect of (horizontally oriented) “pinning wires” on the FIR output polarization. The $10\ \mu\text{m}$ pump is circularly polarized in both cases.

Hole-coupling at the input and output of our FIR cavity was used. The mirrors are gold-coated optical flats (Spawr Corp., Covina, CA) with on-axis holes of radius 1 mm for both input and output. The diameter of the output coupling hole was chosen on the basis of the experiments of Julien and Lourtioz (1980). They constructed an output coupler from a Fabry-Perot etalon formed by two wafers of polished single-crystal silicon. The etalon spacing was adjustable *in situ*. By measuring the laser output as a function of spacing and then removing the coupler and separately measuring its transmission as a function of spacing, they were able to determine the optimal amount of output coupling. For the $119\ \mu\text{m}$ line this turned out to be

approximately 20 %. In fact, during the analysis period leading to the writing of this thesis, I found that the hole size selected was a factor of two smaller than the optimum. Some improvement in output power might be expected in a revised version.

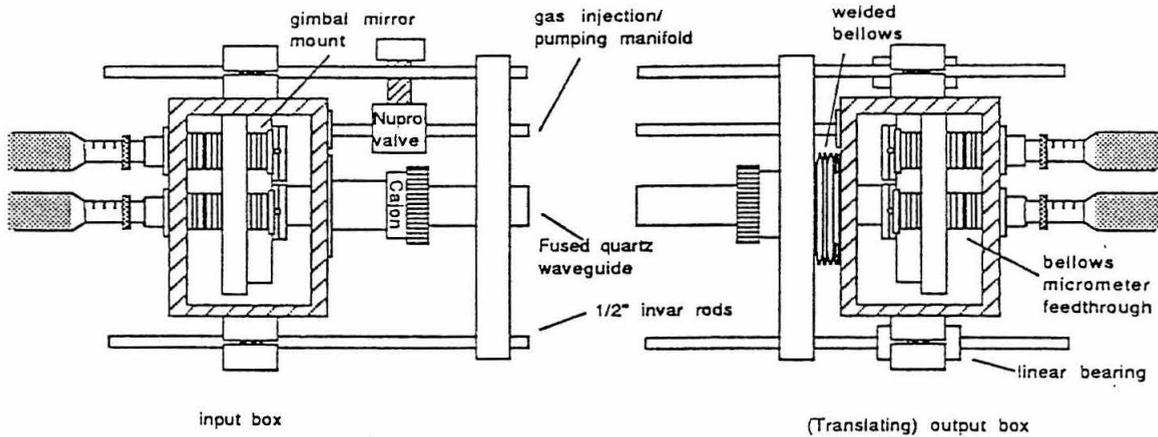


Figure 3.5 – Mechanical design of the far-infrared cavity

The overall mechanical design of the FIR cavity is illustrated in figure 3.5. The input and output coupling mirrors are each mounted in a two-axis gimbal stage (Burleigh Corp., Fisher, NY), which in turn is mounted inside a welded aluminum vacuum box. Angular adjustment in θ and ϕ of the two mirrors is done with ordinary micrometers which are mounted on the vacuum boxes with fittings which have been soldered onto a bellows arrangement, to allow translational motion without breaking vacuum. The input vacuum box is rigidly fixed to four 1/2" diameter invar rods, which form the basic cavity length reference. The invar rods are rigidly fixed at their other end to an aluminum plate on which a differential micrometer (not shown in the figure) is mounted to adjust the cavity length. The micrometer head then presses against the output vacuum box. Atmospheric pressure is sufficient to press the vacuum box back when the micrometer is backed off. A translatable seal between the waveguide and the output vacuum box is maintained by an ordinary vacuum fitting mounted on the fixed aluminum plate. Between this and the output box lies a stainless steel bellows. The weight of the vacuum box is supported through linear bearings by the four invar rods. The CO₂ beam is injected through a ZnSe vacuum window (II-VI Corp.,

Saxonburg, PA), anti-reflection coated for $10.6 \mu\text{m}$. The FIR output is extracted through a crystal quartz window. A polyethylene window cannot be used because the residual $10 \mu\text{m}$ radiation that passes through the output coupling hole would destroy it. Generally we have used z-cut quartz (i.e. extraordinary axis normal to the plane) for the output window. In some cases, however, an x-cut piece was used, whose thickness was such that it formed a quarter-wave plate at the laser frequency. This was used to produce a linearly polarized output when the pump (and therefore the intracavity FIR) mode was circularly polarized. This approach was adopted once “pinning” the output with differential loss was found not to work.

Laser stabilization

To improve the laser stability on timescales shorter than the optoacoustic lock bandwidth (i.e. $t < .01$ sec), our system incorporates an optical isolator between the CO_2 and FIR laser cavities. As described at length by Mansfield *et al.* (1981), even quite low levels of $10 \mu\text{m}$ power reflected back into the CO_2 laser cavity are sufficient to couple the two cavities, and to produce frequency “pulling”, and other non-linear effects, which make the CO_2 laser frequency and/or power highly unstable. Obviously, the effects are worst when the cavity is pumped on-axis. Other astronomical receivers that employ FIR lasers apparently do not have such severe coupled-cavity problems, but the reason is unclear. It may be due to slightly off-axis pumping (A. Betz, private communication.)

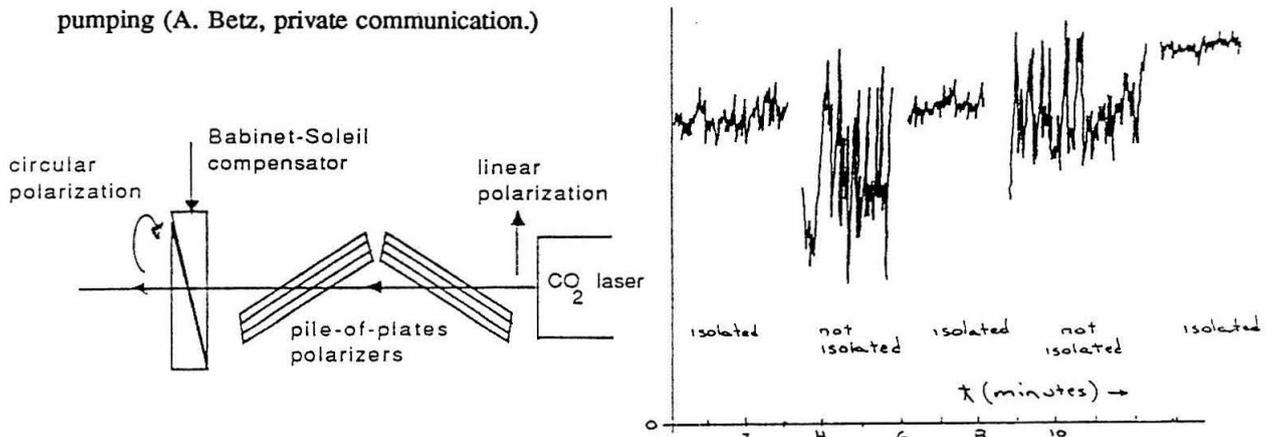


Figure 3.6 – (left) Schematic of the optical isolator, (right) FIR laser power over time, with and without optical isolation.

Our isolator consists of a pair of pile-of-plates polarizers oriented at Brewster's angle, followed by a Babinet-Soleil compensator (see figure 3.6). The arrangement of the pile-of-plates polarizers (each is

a set of three 1 mm thick ZnSe plates,) is chosen to produce no transverse offset of the transmitted beam. A Babinet-Soleil compensator (ours was manufactured by Special Optics Corp., Little Falls, NJ) is nothing but an adjustable phase retardation plate. Two pieces of birefringent material cut into mating prisms are mounted so that their (parallel) extraordinary axes may be rotated into any orientation relative to the polarization of the incident beam. In addition, a micrometer can drive one piece relative to the other so as to vary the pathlength through which the beam passes. Clearly, since all we wish to do is convert linear to circular polarization, this affords us more control of the polarization than we really need. However, a simple quarter-wave plate, which was used originally, was found not to have enough tunability (accomplished by rotation about the extraordinary axis), over the 9-11 μm band to allow us to cover all the FIR lines we wanted. An example of the effect of the isolator is shown at the right of figure 3.6. The CO_2 beam was chopped and the FIR signal synchronously detected with a 1 Hz bandwidth in this case, so the reduction in the high frequency noise is not indicated, but it was also quite marked.

The CO_2 laser came equipped by the manufacturer with a lock loop based on the total 10 μm power. A fraction of the laser output was split off by a BaF_2 beamsplitter into a pyroelectric detector. As the laser cavity length was modulated ("dithered"), the pyroelectric output was synchronously detected and used as a feedback signal. This lock arrangement had some drawbacks. Firstly, the front end amplifier for the pyroelectric signal did not have particularly low noise, which meant that fairly large dither amplitudes were required. More importantly, lock loops based on total power inherently have the disadvantages that a) because the 10 μm gain curve of the CO_2 laser is much broader than the far-infrared laser gas's absorption curve, very high precision is required of the lock loop in order to achieve significant stabilization of the FIR power, b) the correct frequency offset between the maximum absorption of the FIR gas and maximum 10 μm power must be found by hand. For these reasons, the Lakkmann lock loop was replaced by a homemade optoacoustic lock. It avoids problems a) and b) because the lock error signal is generated by the same molecular absorption that pumps the FIR laser.

Optoacoustic locking of FIR lasers has been discussed at length by Chiou (1983), Kavaya (1982), and Rosengren (1975). The basic principle of the optoacoustic detector (or "spectraphone") is illustrated in

figure 3.7. A sample of the CO₂ laser output is directed into a small sample cell containing the FIR laser gas. A small audio microphone (Mouser Electronics, Los Angeles, CA) mounted inside the sample cell detects the change in pressure caused by the molecular absorption and subsequent vibrational de-excitation, as the CO₂ laser frequency is tuned through the molecular absorption. As in the total power lock, the control signal for the CO₂ laser cavity length is the sum of a manually controlled DC signal, the dither signal, and the error signal. The error signal is simply the microphone output, synchronously detected in a commercial lock-in amplifier. It is clear that when the laser cavity is tuned below the frequency of peak absorption, the component of the microphone signal at the dither frequency will have 0° phase, whereas when the cavity is tuned above maximum absorption, it will have 180° phase. Precisely at the absorption peak, the microphone signal will be entirely at twice the dither frequency (except for noise), and the synchronously detected output will be zero. Thus, the lock loop naturally “finds” the correct offset between the CO₂ power peak and the FIR absorption peak, and a trial-and-error search for the correct offset is not required.

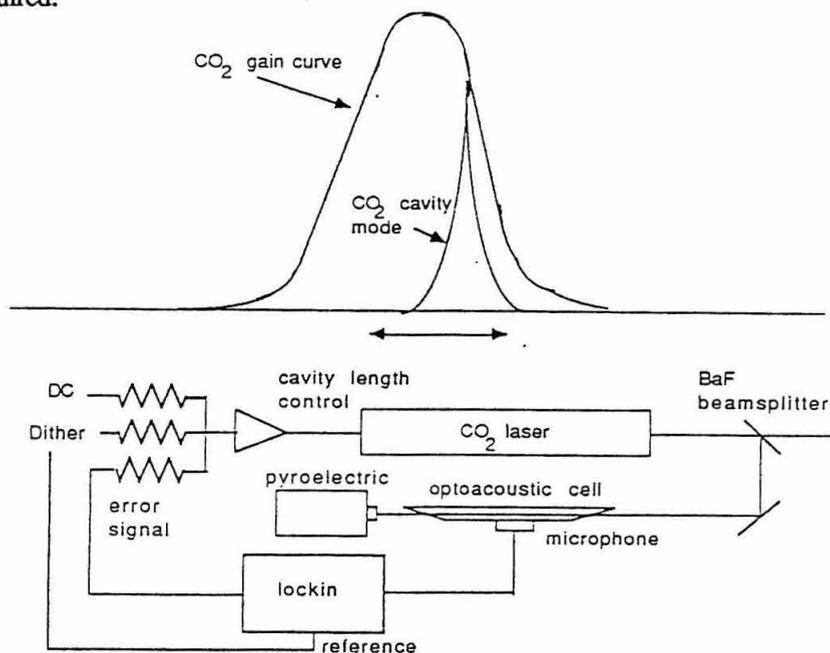


Figure 3.7 – Optoacoustic laser lock

When it is working properly, the performance of the optoacoustic lock is quite good. Peak-to-peak fluctuations in FIR power, over time periods of two hours, have been measured to be as low as 4 %. With

careful optimization, the residual dither in the FIR output can be made quite low – 5 % or less. More often during our lab experiments, however, a fairly large residual dither remained on the FIR signal when the lock loop was employed, even though it generally wasn't apparent on the (slow) pyroelectric detector. The size of dither required is determined by the noise level on the microphone signal and the relative widths of the FIR absorption in the sample cell and in the laser. The absorption in the sample cell is generally wider because it is run at higher pressure (typically around 750 mtorr) than the laser. It is still much narrower than the CO₂ gain curve, though. The size of the required residual dither during typical and optimized conditions is well-illustrated by figure 4.25, which shows oscilloscope traces of the output of our sideband generator, detected on a high sensitivity photoconductor. It was later found that much of the noise on the microphone signal was due to acoustic pickup of the mechanical motion of the dithered mirror. A stabler mount for the microphone improved this somewhat.

The optical layout of the CO₂ beam, and the physical arrangement of the CO₂ and FIR cavities in the airborne version of the instrument, are shown in figure 3.8. The CO₂ laser lies beneath the CO₂ amplifier tube. The folded optical configuration was designed to allow the insertion of an optical amplifier in order to boost the power of the CO₂ pump laser. The amplifier tube (kindly donated by A. Betz) consisted of a pyrex gas cell with cooling jacket, sealed at both ends with 10 μm ZnSe Brewster windows, which had formerly been used as an external mirror CO₂ laser cavity. In operation, it was filled with 10 - 20 torr of CO₂ laser mix and excited by a high-voltage DC discharge. Unfortunately, and for reasons which remain unclear, it never provided a power gain of more than 1.3 (1.2 db), and it was therefore removed from the setup. The loss budget of the CO₂ pump beam after a complete system alignment was as follows. On the 9P10 line which pumps the 159 CH₂F₂ transition, the power was 5.7 W after a single plane mirror at the CO₂ laser output. After the 5" ZnSe lens, it was 5.4 W. After the Babinet-Soleil compensator it was 5.0 W, and after the FIR laser cavity entrance aperture it was 3.8 W. Thus, only about two-thirds of the 10 μm power exiting the CO₂ laser aperture was available for pumping the FIR cavity.

3.3 FIR output : power and beam patterns

Our general purpose laboratory detector for the FIR laser consists of a commercial pyroelectric detector

Top View

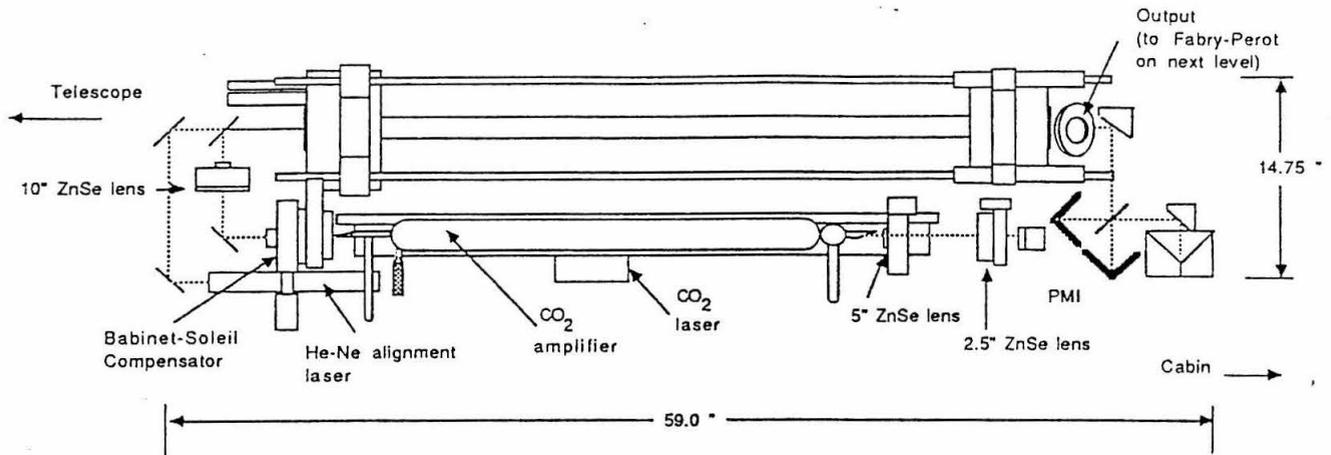


Figure 3.8 – Airborne laser layout

element (Moletron Corp., Santa Clara, CA) incorporated into a homemade feedback amplifier circuit. The active element is supplied with a coating (black paint of some kind) that is advertised to be highly absorbing throughout the visible, near-IR, and far-IR. The assembly is powered from 9V batteries and is physically mounted in a small ($\sim 3.5''$ by $1.5''$ diam.) all-metal package which provides extremely good shielding against radiative pickup. The latter is necessary since the pyroelectric signal has an impedance of $\sim 10^{10}\Omega$. The active element of the pyroelectric is only 1 mm square. Therefore, artificially low powers are measured when the FIR input beam is even only moderately slow, due to the diffraction spot spilling over the edge of the detector.

For nearly all our experiments, the pyroelectric is the fundamental FIR power reference. Unfortunately, there is an ambiguity of about a factor of four in its absolute calibration. The first absolute calibration of the pyroelectric was done at 6328 \AA (the He-Ne laser wavelength) by comparison with a commercial optical power meter (Newport Corp. model 815-SL) under the assumption that the responsivity was flat from the visible to the far-infrared. This yielded responsivities of 4000, 1300, and 140 V/W with load resistances of $R_L = 10^9$, 10^8 , and $10^7 \Omega$, respectively, at a chopping frequency of 13.5 Hz. The fact that responsivity does not scale with load resistance for the two highest values is due to the $R_L C$ rolloff of the pyroelectric.

With the highest value of R_L , the 3 db frequency of the rolloff was ~ 10 Hz, corresponding closely with the specified device capacitance of 15 pF, combined with ~ 5 pF of parasitic circuit capacitance. Using this calibration, the highest 119 μm laser power we have obtained is 1.6 mW.

Moreover, we have twice made direct comparisons between the pyroelectric signal produced by the laser on the 159 μm and 119 μm lines, *with identical laser alignment configurations*. In both cases, the 159 μm signal was about 30 % weaker than the 119 μm signal. After correcting the 159 μm signals for the overly slow input beams, however, the estimated 159 μm power was 10 % and 50 % greater than the 119 μm power in the two measurements. It is clear from these direct comparisons of the pyroelectric signals that the two laser lines have quite comparable strengths.

Much later, after most of our laboratory experiments were completed, the responsivity calibration was checked using the FIR laser output at 159 μm and a commercial, Peltier-element based FIR power meter (Scientech model 361). The 159 μm laser power was measured on the Scientech to be 1.8 mW. This agrees well with the previous calibration and comparison of the 119 μm and 159 μm lines, if we assume that at the time of the measurement, the laser's alignment was near optimal (which it should have been, since the measurement was performed immediately after a complete re-alignment.) On the other hand, the 159 μm laser power was also measured on the pyroelectric at the same time, and direct comparison with the Scientech reading indicated a responsivity of 1000 V/W for the pyroelectric. If this latter value is the correct one, then the peak 119 μm laser power measured in the early experiments must have been ~ 6.5 mW, and the laser must have been producing a factor of four less than its maximum power at the time of the Scientech measurement. Unfortunately, the precise optical configuration was not recorded, and the laser output beam not measured, during the Scientech measurement. The Scientech's active area is 1 inch in diameter. It is possible, therefore, that the 4000 V/W figure is correct and the low 159 μm pyroelectric signal was due to the laser power partially spilling over the edge of the active element.

Given this ambiguity in the pyroelectric's responsivity, it is impossible to give reliable figures for the laser output power. We have therefore made the conservative assumption that the early calibration is the correct one, and compiled all the measurements made in the course of our experiments of the total laser

power at various wavelengths. This is the best that can be done with the presently available data. The highest measured powers obtained on the various lines are listed in table 3.3. The pyroelectric signals were converted to powers in the table by assuming a responsivity of

$$S = 10^4 V/W \frac{\left(\frac{R_{FIR}}{10^9 \Omega}\right)}{1 + \left(\frac{f}{10 Hz}\right) \left(\frac{R_{FIR}}{10^9 \Omega}\right)}.$$

Under optimum conditions, the laser's output power is apparently ~ 2 mW on the strongest lines. Typical powers during most of the experiments were somewhat lower than those given in table 3.3. To give an idea of the typical powers, we present in figure 3.9 a histogram of all our power measurements on the $119 \mu\text{m}$ line. Each measurement going into the histogram is the maximum power level measured on a particular day. The possible reasons for less than optimal powers at any given time include low CO_2 pump power, poor alignment or transmission of the $10 \mu\text{m}$ optics, inadequate pumpout of the FIR cavity, imperfect alignment of the FIR cavity, as well as error in the measurement. Different peoples' understandings of the word "typical" vary, but to us it seems fair to say that, averaged over the course of our experiments, the "typical" $119 \mu\text{m}$ laser power was between .5 and 1 mW.

Far-field beam patterns of the laser were measured once over a complete two-dimensional grid, and several times over a pair of orthogonal one-dimensional scans. Reliable measurements were often difficult due to the presence of large standing waves between the laser output window and the pyroelectric detector. Coating all nearby exposed metal surfaces with eccosorb and tipping the pyroelectric by a large angle ($> 20^\circ$) from the incident beam made the beam patterns sufficiently reproducible for meaningful comparison with theory. Because the output coupling hole is much smaller than the waveguide diameter, ($\frac{r_0}{a} = .12$), it is a good approximation to treat the output hole as a uniformly illuminated aperture, for purposes of computing the diffraction-limited output. In this case, the angular pattern in the far-field is given by the usual Fraunhofer diffraction formula (Born and Wolf, p. 396) :

$$I(\theta) = I_0 \left(\frac{J_1(kr_0\theta)}{kr_0\theta} \right)^2, \quad (3.6)$$

which leads to

$$FWHM = .49 \frac{\lambda}{r_0}. \quad (3.7)$$

Table 3.3 – Maximum output power of the FIR laser

Wavelength	Molecule	CO ₂ Power	FIR Power
158.5 μm	CH ₂ F ₂	6 W (9P10)	1.8 mW [†]
118.8 μm	CH ₃ OH	5.1 W (9P36)	1.6 mW
96.5 μm	CH ₃ OH	6.4 W (9R10)	165 μW
77.9 μm	CH ₃ OH		200 μW
70.5 μm	CH ₃ OH	5.5 W (9P34)	200 μW
63.1 μm	¹³ CH ₃ OH	5.3 W (9P12)	350 μW
570.6 μm ‡	CH ₃ OH	8 W (9P16)	~ 100 μW
393.6 μm ‡	HCOOH	8.5 W (9R18)	40 μW
369.1 μm ‡	CH ₃ OH	8 W (9P16)	~ 20 μW
287.7 μm ‡	CH ₂ F ₂		40 μW
214.6 μm ‡	CH ₂ F ₂	5 W (9P34)	700 μW
164.6 μm ‡	CH ₃ OH	8 W (9P16)	~ 1 mW
122.5 μm ‡	CH ₂ F ₂	9 W (9R22)	300 μW
45.7 μm ‡	CH ₃ OH	8 W (9P16)	~ 1 mW

[†] Measured on Scientech (Peltier-element) power meter, all other lines were measured on pyroelectric.

[‡] Laser adapted for long wavelengths, with 25 mm waveguide diameter, and 5 mm output coupling hole.

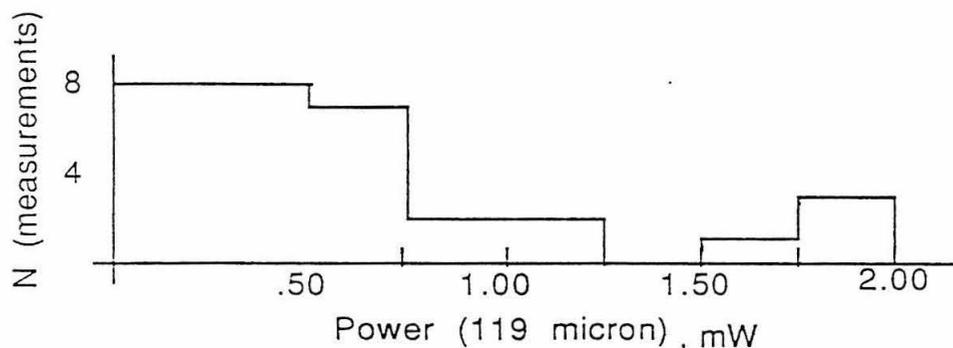


Figure 3.9 – Histogram of measured laser power levels on the 119 μm methanol line.

Each measurement going into the histogram is the maximum power obtained on a particular day.

The far-field of the laser output aperture is achieved at distances $z \gg \frac{r_0^2}{\lambda}$, i.e. $z \gg 8.5$ mm at $119 \mu\text{m}$. Thus, all physically accessible distances are well into the far-field.

In general, we find reasonable agreement between equation 3.7 and our measured patterns. Figure 3.10 shows the measured 2-D far-field pattern at $119 \mu\text{m}$. The beam is fairly circular and has a FWHM of $3.2^\circ \times 3.4^\circ$, in the vertical and horizontal directions, respectively. This compares very well with the value of 3.35° obtained from equation 3.7. With poor alignment of the FIR cavity mirrors, we have observed significant offsets ($1^\circ - 2^\circ$) between the peak of the FIR beam and the He-Ne alignment laser. The data of figure 3.10 were taken immediately after a complete laser re-alignment, and show no observable offset, however.

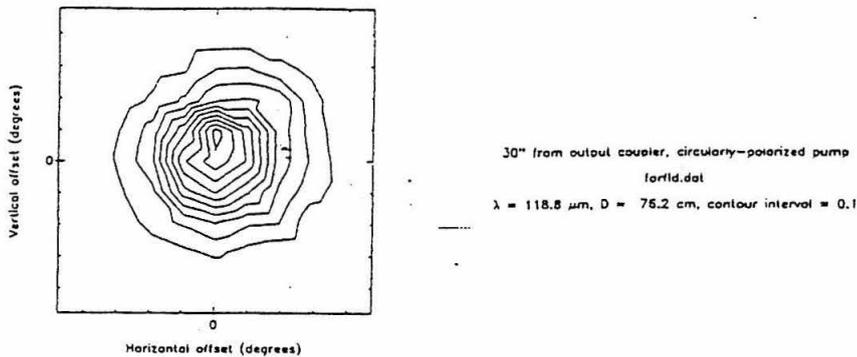


Figure 3.10 – Far-field intensity pattern of the laser at $119 \mu\text{m}$.

As a further comparison with theory, we show in figure 3.11 a set of 1-D scans of the far-field patterns at three different wavelengths, $70.5 \mu\text{m}$, $119 \mu\text{m}$, and $158.5 \mu\text{m}$. The measurements were all performed over the course of two consecutive days and therefore characterize a single, particular alignment configuration of the laser. Residual standing waves are apparent in the $119 \mu\text{m}$ and $158.5 \mu\text{m}$ measurements. In addition, it is clear that there is significant asymmetry in the underlying laser patterns as well. This could be due to imperfect mirror alignment or to the presence of a small component of a higher order waveguide mode. The FWHM widths of the $119 \mu\text{m}$ and $158.5 \mu\text{m}$ beams agree fairly well with equation 3.7, but the $70.5 \mu\text{m}$ beam is approximately 25 % wider than predicted. Also in figure 3.11, we show beam scans measured at two positions “inside” the polarizing Michelson interferometer, where the beam is nominally collimated. It is clear that, in fact, the beam is slightly diverging, but this is simply a matter of the paraboloid’s focal

length not being precisely correct. More disturbing are the facts that the beam is astigmatic (i.e. it has different amounts of divergence horizontally and vertically), that it is not very accurately symmetric, and that it is not single-peaked. Given the quality of this beam, it is not clear how meaningful a polarizing Michelson interferometer design based entirely on a Gaussian beam approximation would be.

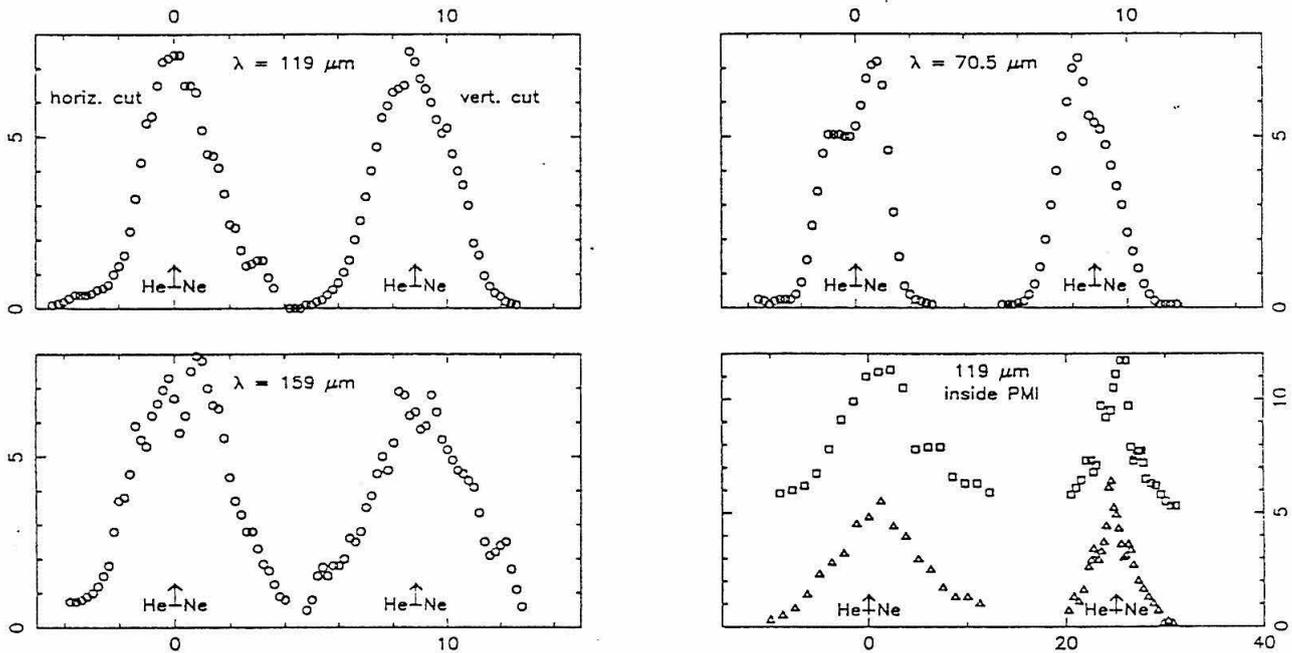


Figure 3.11 – One-dimensional scans of the laser intensity pattern at three wavelengths. FWHM beamwidths are $3.1^\circ \times 3.4^\circ$ (horiz. \times vertical) at $119 \mu\text{m}$, $2.7^\circ \times 2.6^\circ$ at $70.5 \mu\text{m}$, and $4.2^\circ \times 4.3^\circ$ at $159 \mu\text{m}$. (lower right) Beam scans taken “inside” the polarizing Michelson interferometer, where the beam is nominally collimated. The upper two scans are horiz. and vert. cuts measured 8 cm away from the collimating mirror, and indicate a FWHM beamsize of $8.9 \times 7.9 \text{ mm}$. The lower ones were measured 29 cm away and indicate a $9.4 \times 9.9 \text{ mm}$ beam.

3.4 The Lorenz instability : spontaneous pulsations and deterministic chaos in FIR lasers

The Lorenz equations were the first example discovered of a dynamical system whose time evolution is governed by a strange attractor. They were discovered by E. N. Lorenz in 1963, who developed them

as a zeroth-order model for convective turbulence in fluids. In recent years they have become a kind of archetype for a wide variety of non-linear dynamical systems. (For references to the early literature, see the book by Sparrow, 1982.) They are a set of three coupled, ordinary differential equations, which in dimensionless form are :

$$\frac{\partial x}{\partial \tau} = \sigma y - \sigma x \quad (3.12a)$$

$$\frac{\partial y}{\partial \tau} = rx - y - xz \quad (3.12b)$$

$$\frac{\partial z}{\partial \tau} = xy - bz \quad (3.12c)$$

where τ is a dimensionless time, and σ , r , and b are three real positive control parameters. In the model of a convective fluid, the spatial variations of velocity and temperature are decomposed into an infinite series of (spatial) Fourier modes. The dynamical variables x , y , and z of the Lorenz model represent a particular Fourier component of the velocity field, a horizontal Fourier component of the temperature field, and a vertical Fourier component of the temperature field, respectively. The control parameters σ , r , and b are the Prandtl number (ratio of kinematic viscosity to thermal conductivity), the Rayleigh number (actually the ratio of Rayleigh number to critical Rayleigh number), and a number related to the dimensionless size of the region under consideration. In some regions of control parameter space, the equations have a steady-state solution, in some regions the solutions are periodic, and in some they are chaotic, or “turbulent”. Despite a vast amount of study of the Lorenz equations, there remain regions of control parameter space in which the behavior of the solutions is not well understood.

There are many other physical systems besides a convective fluid which can be modelled by equations 3.12 . (See Sparrow, p. 4 for a long list, and p. 194 for an explicit derivation in the somewhat whimsical case of a water-wheel.) What we are concerned with is the analogy, first pointed out by Haken (1975), between the Maxwell–Bloch equations describing the dynamics of a gas laser and the Lorenz equations. We shall spare the reader the details of the calculation, but will describe the first principles from which the analysis begins, and will attempt to describe all the approximations involved and the physical meanings of the various parameters. Before launching into the equations, we wish to emphasize one thing. Even to a simple receiver-builder, who may not care – indeed who may be aggressively indifferent to the physics

of FIR lasers – and who only wishes to use them as a laboratory tool, the Lorenz instability is important. The reason is simple. The regime in which the Lorenz instability occurs is that of high pump intensity, high pressure, and fairly low cavity Q. This is precisely the regime toward which all the considerations discussed in §2.1 are driving FIR laser development. To achieve high far-infrared power, high tunability, the absence of other types of instability, and physical compactness, one is naturally driven to the high pump intensity, high pressure regime. Furthermore, choosing the output coupling fraction for maximum output power leads to fairly low-Q cavities.

Without further ado, we shall now derive equations 3.12 for the case of a homogeneously broadened FIR gas laser. The treatment closely follows Haken (1975), Risken and Nummedal (1968), and Riskin (1964). It is semi-classical; that is, the electric field is treated classically, but the atomic inversion and polarization are treated quantum mechanically. Equation 3.12(a) is the simplest of the three. It derives from Maxwell's equations in a dielectric medium. We assume the polarization of the atomic medium \vec{P} , is everywhere parallel to the laser field \vec{E} , so that a scalar treatment may be used. We also assume that the active medium is polarizable, but not magnetizable, and that *total* charge neutrality holds, i.e. $\vec{B} = \vec{H}$, and $\nabla \cdot \vec{E} = 0$. In that case, taking the curl of Faraday's law and substituting in the usual way yields a lossy, scalar wave equation for the electric field :

$$\frac{\partial^2 E}{\partial x^2} - \frac{\ddot{E}}{c^2} - \frac{4\pi\sigma_{\Omega}}{c^2} \dot{E} = \frac{4\pi}{c} \ddot{P}. \quad (3.14)$$

Here, the conductivity σ_{Ω} incorporates all the losses in the medium and the cavity. It is directly related to the passive cavity linewidth, viz. $\Delta\omega = 2\pi\sigma_{\Omega}$.

We next apply the “rotating wave” approximation, a standard technique in optical theory. The electric field (and similarly for the polarization field) is written

$$E = \tilde{E}(x, t)e^{i(kx - \omega t)}, \quad (3.15)$$

where k is the cavity wavenumber (which for our highly overmoded waveguide is extremely close to the free space value, $\frac{\omega_0}{c}$). \tilde{E} varies slowly in time relative to the optical frequency ω_0 and slowly in space relative to k. Therefore, in this approximation, second-order derivatives of \tilde{E} are ignored. Furthermore,

first-order derivatives in the damping term ($\frac{4\pi\sigma}{c^2}E$) and the coupling term ($\frac{4\pi}{c}\dot{P}$) of equation 3.14 are also ignored. This yields

$$\dot{\tilde{E}} = 2\pi i\omega_0\tilde{P} - 2\pi\sigma_\Omega\tilde{E} + \left[c\frac{\partial\tilde{E}}{\partial x} \right]. \quad (3.16)$$

It is not part of the rotating wave approximation, but the reduction of the laser equations to Lorenz form also involves the fundamental assumption of spatial homogeneity. Thus, the last term in 3.16 is ignored. The fact that the Lorenz instability occurs independently of any spatial inhomogeneity distinguishes it in a basic way from many other instabilities. The final step in the derivation of 2.13(a) is the normalization of \tilde{E} and \tilde{P} by their steady-state values. Setting $\dot{\tilde{E}} = 0$, we find

$$\tilde{E}_{CW} = \frac{i\omega_0}{\sigma_\Omega}\tilde{P}_{CW},$$

and therefore

$$\dot{E} = (2\pi\sigma_\Omega)P - (2\pi\sigma_\Omega)E, \quad (2.16)$$

where $E \equiv \frac{\tilde{E}}{\tilde{E}_{CW}}$ and $P \equiv \frac{\tilde{P}}{\tilde{P}_{CW}}$.

The derivation of the other two Lorenz equations is slightly more involved. An essential point about the Lorenz instability is that it involves the quantum-mechanical correlation of the upper and lower laser levels. Therefore, *a conventional rate-equation analysis of FIR laser operation cannot describe this instability*. One must begin either with the Schrodinger equation or with the equation of motion of the density matrix. We use the latter, and assume a) a two-level molecule (consisting of the two laser levels), and b) homogeneous broadening and an incoherent pump, both of which are described simply by empirical rate constants. The evolution of the density matrix is then given by

$$\frac{d\rho}{dt} = \frac{[H, \rho]}{i\hbar} - \begin{pmatrix} \gamma_{\parallel}\rho_{11} & \gamma_{\perp}\rho_{12} \\ \gamma_{\perp}\rho_{21} & \gamma_{\parallel}\rho_{22} \end{pmatrix} + \Lambda, \quad (3.17)$$

where

$$H = \begin{pmatrix} E_1 & -\mu E \\ -\mu E & E_2 \end{pmatrix} \quad \Lambda = \begin{pmatrix} \Lambda_1 & 0 \\ 0 & \Lambda_2 \end{pmatrix},$$

are the Hamiltonian and the term describing the pumping, respectively. The incoherence of the pump is expressed by the fact that Λ is diagonal. Note that the empirical relaxation rates for the diagonal and

off-diagonal components of the density matrix, γ_{\parallel} and γ_{\perp} , are not necessarily equal. Physically, both rates are dominated by 2-molecule collisions and are therefore proportional to pressure. We shall say more about them shortly, when we discuss numerical evaluation of the Lorenz control parameters. The energy difference of the two levels is $E_1 - E_2 = \hbar\omega_0$, and the model assumes a pure dipole interaction $\Delta H = -\vec{\mu} \cdot \vec{E}$, where $\vec{\mu}$ is the molecular dipole moment and \vec{E} the laser field. In this model, spontaneous emission is ignored (or rather, subsumed into the definitions of γ_{\parallel} and γ_{\perp} .)

The macroscopic inversion and polarization are defined by

$$\begin{aligned} S &= n(\rho_{11} - \rho_{22}) && \text{inversion} \\ P &= n\langle ex \rangle = n(\mu^* \rho_{12} + \mu \rho_{21}) && \text{polarization} \\ &= n\mu(\rho_{12} + \rho_{21}), && (3.18) \end{aligned}$$

where n is the concentration of active molecules in cm^{-3} . The equation of motion for the off-diagonal components of the density matrix leads directly to

$$\begin{aligned} \dot{P} &= -in\mu\omega_0(\rho_{12} - \rho_{21}) - \gamma_{\perp}P \\ n(\rho_{12} - \rho_{21}) &= \frac{i}{\mu\omega_0}(\dot{P} + \gamma_{\perp}P). \end{aligned} \quad (3.19)$$

Taking a second derivative for the off-diagonal components, and simply using the density matrix equation for the diagonal components, we then derive the basic equations of motion for the macroscopic variables :

$$\begin{aligned} \ddot{P} + 2\gamma_{\perp}\dot{P} + (\omega_0^2 + \gamma_{\perp}^2) &= -\frac{2\omega_0|\mu|^2}{3\hbar}ES \\ \dot{S} &= \frac{2}{\hbar\omega_0}E(\dot{P} + \gamma_{\perp}P) + \gamma_{\parallel}(d_0 - S). \end{aligned} \quad (3.20)$$

Here, μ^2 has been replaced in the usual way with $\frac{|\mu|^2}{3}$ to account for the implicit ensemble average over the molecular orientations (see e.g. Townes and Schawlow p. 23). Also, a macroscopic pump parameter has been defined by

$$d_0 \equiv \frac{n(\Lambda_1 - \Lambda_2)}{\gamma_{\parallel}}. \quad (3.21)$$

d_0 is the macroscopic inversion (the ‘‘saturated’’ inversion) that would occur in the absence of the laser field, $E = 0$.

The reduction of the basic matter equations (3.20) to the Lorenz equations now proceeds completely analogously to that of the wave equation for E. First, the rotating wave approximation is applied. This yields

$$\begin{aligned}\dot{\tilde{P}} + \gamma_{\perp} \tilde{P} &= -\frac{i|\mu|^2}{3\hbar} \tilde{E} \tilde{S} \\ \dot{\tilde{S}} &= \gamma_{\parallel} (d_0 - \tilde{S}) + \frac{2i}{\hbar} (\tilde{E} \tilde{P}^* - \tilde{E}^* \tilde{P}),\end{aligned}\quad (3.22)$$

where, as before, we have used tildes to denote the slowly varying quantities. The complex conjugates of E and P arise in the equation for the inversion because there is a time derivative involved in their product in (3.20). The term represents stimulated absorption, or, viewing the active medium macroscopically, it represents $\vec{E} \cdot d\vec{P}$ work which the field must do on the medium in order to invert it. The slowly varying dynamical variables, \tilde{E} , \tilde{P} , and \tilde{S} are complex and do not necessarily all have the same phase. However, the phase differences between them are not necessary for production of the Lorenz instability, so we specialize to the case $\arg(E) = \arg(P) = \arg(S) = 0$. In the linear stability analysis of Riskin and Nummedal (1968), the phase differences are included. They found this produced no qualitative difference; generally, the amplitudes become unstable before the phases do.

The final step in reduction of the laser equations to Lorenz form is the normalization of \tilde{E} , \tilde{P} , and \tilde{S} by their CW values. Furthermore, a normalized pump parameter is defined. I.e.

$$\begin{aligned}\tilde{S}_{CW} &= \frac{3\hbar\gamma_{\perp}\sigma\Omega}{\omega_0|\mu|^2} \\ \tilde{E}_{CW} &= \left[\left(\frac{\gamma_{\parallel}\hbar\omega_0}{4\sigma\Omega} \right) (d_0 - \tilde{S}_{CW}) \right]^{1/2} \\ \tilde{P}_{CW} &= -\frac{i\sigma\Omega}{\omega_0} \tilde{E}_{CW} \\ r &= \frac{d_0}{\tilde{S}_{CW}}.\end{aligned}\quad (3.23)$$

Including (3.16), the equation derived from the wave equation for E, we then obtain for the normalized dynamical variables, ($E \equiv \frac{\tilde{E}}{E_{CW}}$ etc.)

$$\begin{aligned}\dot{E} &= (2\pi\sigma\Omega)P - (2\pi\sigma\Omega)E \\ \dot{P} &= \gamma_{\perp}ES - \gamma_{\perp}P \\ \dot{S} &= \gamma_{\parallel}r - \gamma_{\parallel}S - \gamma_{\parallel}EP.\end{aligned}\quad (3.24)$$

These are seen to be identical to the Lorenz equations, 3.12, under the following set of identifications :

$$\begin{aligned}
 x &= \left(\frac{\gamma_{\parallel}(r-1)}{\gamma_{\perp}} \right)^{1/2} E \\
 y &= \left(\frac{\gamma_{\parallel}(r-1)}{\gamma_{\perp}} \right)^{1/2} P \\
 z &= r - S \\
 \tau &= \gamma_{\perp} t.
 \end{aligned} \tag{3.25}$$

The Lorenz control parameters are given by

$$\begin{aligned}
 b &= \frac{\gamma_{\parallel}}{\gamma_{\perp}} \\
 \sigma &= \frac{2\pi\sigma\Omega}{\gamma_{\perp}} \\
 r &= \frac{|\mu|^2}{3} \left(\frac{\omega_0}{\hbar\gamma_{\perp}\gamma_{\parallel}\sigma\Omega} \right) n (\Lambda_1 - \Lambda_2).
 \end{aligned} \tag{3.26}$$

For reasons which are mainly historical, the behavior of the Lorenz equations has primarily been studied as a function of r , for the values $b = 8/3$, and $\sigma = 10$. What, we may ask, is the parameter regime of interest for FIR lasers? In this case also, it is desired to study the solutions' behavior as a function of r , since this corresponds to varying the pump intensity, which is an easy thing to experimentally adjust. Note that we are particularly interested in the limit of large r , since, other things being equal, we would like to exploit the maximum pump power available. We shall not attempt to numerically compute a value of r for our laser, however, for the following reason. The pump rate $n(\Lambda_1 - \Lambda_2)$, is simply the number of vibrational excitations per second per unit volume. For our laser, there is some difficulty in defining the effective volume. This is because the CO_2 beam is tightly focused at the entrance to the cavity, but it approximately fills the waveguide diameter after ~ 1 meter of travel. Thus, r , which is simply the ratio of the actual pump rate to the threshold pump rate, is very large (probably 10 - 100) over a small, but significant, fraction of the active volume, and has a much more modest value ($r \leq 3$ perhaps) over the rest of the active volume. It is not clear how to deal with this spatial inhomogeneity.

As for the other two parameters, b is simply the ratio of relaxation rates for the diagonal and off-diagonal components of the density matrix. We have not found any authors who specifically address the

question of b 's numerical value in the context of FIR lasers. Pujol *et al.* (1987) claim that $\frac{\gamma_{\parallel}}{\gamma_{\perp}} = a$ few, for the 81.5 μm methanol line, but offer no justification or discussion. Intuitively, however, it appears to us that b should very nearly equal unity in FIR lasers, at least for the longer wavelength lines. Both types of relaxation are caused by molecular collisions. Suppose the initial molecular state is given, (in Dirac notation,) by $(a_1^2 + a_2^2)^{-1/2} (a_1|1\rangle + a_2e^{i\phi}|2\rangle)$. Then, γ_{\parallel} refers to collisions which change the relative amplitudes of the two components – i.e. $\Delta\left(\frac{a_1}{a_2}\right) \sim 1$, and γ_{\perp} refers to collisions which change the relative phase – i.e. $\Delta\phi \sim 1$ radian. In low-temperature systems, γ_{\parallel} can be much smaller than γ_{\perp} because changing the relative amplitudes requires a transfer of energy, while changing the relative phases does not. In our situation, however, it is a good approximation to say that *every collision is an energy-changing one*, because $\frac{kT}{\hbar\omega_0} > 1$. In this case, $\gamma_{\parallel} = \gamma_{\perp}$ = the pressure-broadened linewidth. Numerically, this is .025 MHz / mtorr, for the 119 μm line (see table 3.2).

The last parameter, σ , is simply the ratio of the cavity linewidth $\Delta\omega = 2\pi\sigma\Omega$, to the pressure-broadened linewidth. It is widely conjectured (see Sparrow p. 184), though apparently not analytically proven, that the Lorenz equations are globally stable for $\sigma < 1/3$, for *all* values of b and r . The condition for instability, $\sigma > 1/3$, often referred to as the “bad-cavity” condition, is the key reason why the instability was not observed in experimental laser systems until very recently. After the analogy between the laser equations and the Lorenz equations was pointed out, (Haken 1975,) it was generally felt that they could not apply to a real-world laser, because any cavity “bad” enough to satisfy $\sigma > 1/3$ would be so lossy that extremely high gain and pump power would be required to get the laser to lase at all – i.e., it was assumed that a “bad” cavity necessarily implied a high laser threshold. For most gas lasers this is true. Eventually it was realized, however, (Weiss and Klische 1984), that FIR lasers are rather exceptional in this regard. Their pressure-broadened linewidths are so low (compared to He-Ne lasers for example,) that a cavity can easily satisfy $\sigma > 1/3$ while still having relatively low loss in absolute terms. For example, for a homogeneous linewidth of 5 MHz, a wavelength of 119 μm , and a cavity length of $L = 1$ meter, the “bad cavity” condition implies a round-trip cavity loss of $1/3 \times 2 \left(\frac{2L\Delta\nu}{c}\right) = 2.2\%$, which is not a very “bad” cavity at all !

The best way to understand the present state of knowledge about the behavior of the Lorenz equations’

solutions is to examine a pair of diagrams of the $b - r$ plane in which the various regions of stability and instability are indicated. These diagrams, reproduced from the final chapter of Sparrow, are shown in figure 3.12. The behavior which is known analytically is shown on the left; behavior that is conjectured is shown on the right. These figures are a cut at $\sigma = 10$. As σ is reduced, the critical values of b which the H-curve (Hopf bifurcation) approaches at $r \rightarrow \infty$, and which the I-curve (first homoclinic explosion), A-curve, P-curve, and F-curve approach at $r \rightarrow \infty$, tend to zero. At $\sigma = 1$, the H-curve disappears, and at $\sigma = 1/3$, the other curves disappear. The usually quoted “bad-cavity” condition is that $\sigma > 1$. This is based on the linear stability analysis of Riskin and Nummedal (1968). That analysis amounted to a computation of the critical r value at which the H-curve was crossed. The result was

$$r > r_{crit} = 1 + \frac{(\sigma + 1)(\sigma + 1 + b)}{(\sigma - 1 - b)}. \quad (3.27)$$

The precise value is not important. The key point is that the laser must be quite far above threshold ($r > 8$) in order to cross the H-curve, and for the Riskin and Nummedal analysis to yield an instability. From the figures, however, we see that the Hopf bifurcation is not the whole story. In Sparrow’s (ibid. p. 189) words, “It should be noted, in particular, that turbulent behaviour can be observed in the Lorenz system at parameter values where the stationary points are still stable. [i.e. below the H-curve.] Most other finite dimensional models of turbulence require the stationary points to lose stability before turbulent behaviour can occur.” In other words, instabilities in the Lorenz model can set in at lower pump powers than predicted by equation 3.27. A conservative criterion for stability is therefore $\sigma < 1/3$, as we have stated, not $\sigma < 1$, as is usually quoted.

There have been several recent reports of the observation of Lorenz instabilities in FIR lasers (Högenboom *et al.* 1985, Weiss and Brock, 1986, Weiss 1985), on a number of different lines – e.g. CH_2F_2 , 117 μm , NH_3 , 81.5 μm , $^{15}\text{NH}_3$, 374 μm . These reports are quite preliminary, and because, in some cases, insufficient experimental details are provided to estimate the values of the Lorenz control parameters, and because it is unclear what the theoretical expectation would be even if the control parameters were known, it is hard to critically evaluate the agreement or lack thereof between theory and experiment. Weiss and Brock (1986) claim that the fact they see direct transition from CW operation to chaos with the cavity

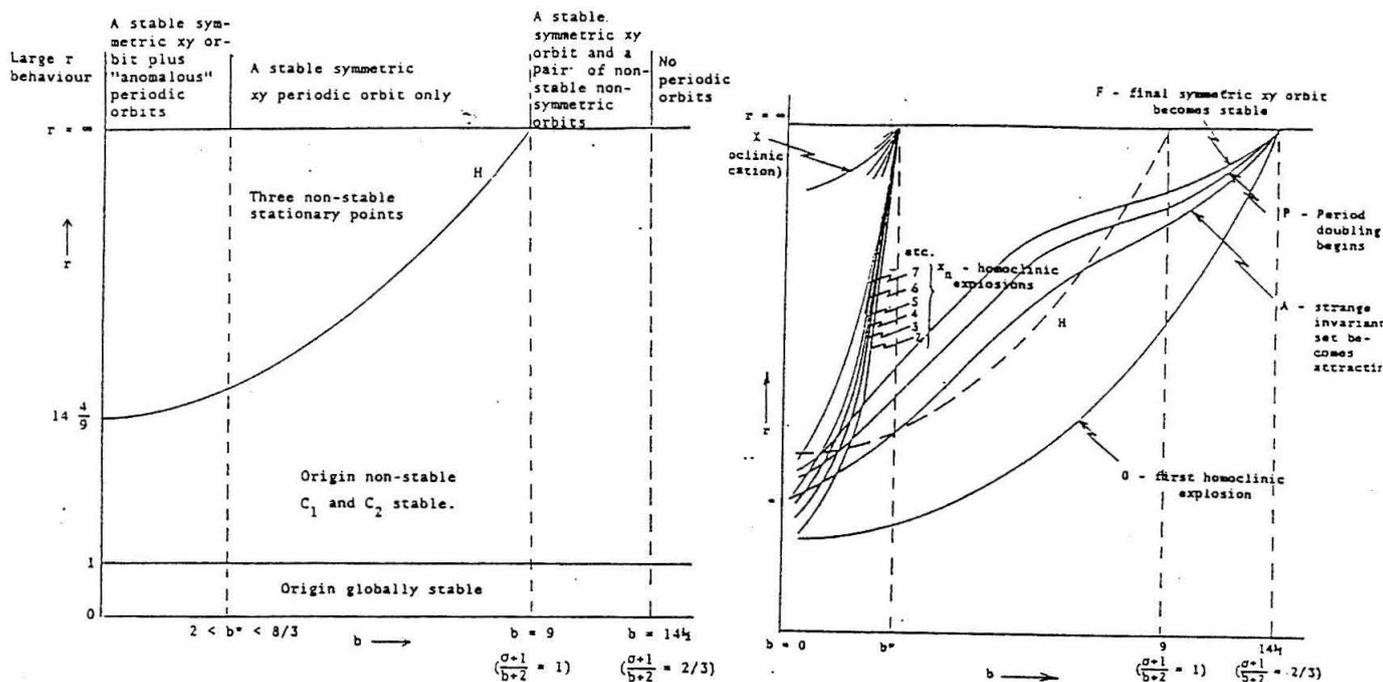


Figure 3.12 – Regions of stability and instability of the Lorenz equations in the b - r plane. For a FIR laser, b is close to unity, and r is the ratio of pump intensity to threshold pump intensity, and is therefore always > 1 .

tuned on the line center, and a period-doubling sequence to chaos with the cavity detuned, provides good evidence of the Lorenz nature of the instability. The same authors also claim that their measurements of the pump threshold for instability as a function of pressure, i.e. $r_{crit}(\sigma)$, also support the Lorenz model. In this case, however, they were only considering the Risken and Nummedal type of instability, i.e. the H-curve in figure 3.12, and they attributed all the other observed instabilities to the inhomogeneous component of the broadening. Apparently, the fact that the Lorenz equations can have chaotic solutions at lower r values than predicted by a linear stability analysis has not been appreciated by experimental workers.

In our system, we have seen a variety of pulsation and low-frequency noise effects almost from the first day we used the laser to illuminate a high-speed detector (i.e. a Ge photoconductor.) For a long while, these were all attributed to oscillations in the GaAsFET preamp which appeared at only some very specific value of photoconductor impedance, and therefore only at some very specific value of laser tuning which produced precisely the right amount of laser power. In some cases, GaAsFET oscillation undoubtedly was

the main problem. Eventually, however, we performed the obvious experiment, and observed the laser using a Schottky diode as a video detector at a time when the photoconductor indicated oscillations. This demonstrated that in many cases, the laser itself was naturally producing huge spontaneous pulsations in power. An example of the spectrum of the diode signal when the laser was pulsating on the 119 μm methanol line is shown in figure 3.13 along with some examples from the published spectra of Weiss and Brock (1986). We have not catalogued all the various sorts of behaviors we have seen, but qualitatively, our results seem to very much resemble those of Högenboom *et al.* and Weiss and Brock. By tuning the cavity length, or the orientation of the cavity mirrors (and therefore the Q of the cavity), we are able to tune the frequency of the laser pulsations. In the same ways, we can tune the harmonic content of the pulsations. In some cases, a period-doubling sequence can be seen in the spectra as the cavity or the pump frequency is tuned. In general, the spectrum is a series of harmonics whose envelope rolls off at a frequency of 5 - 10 MHz. We have seen fundamental frequencies ranging from as low as 20 kHz to as much as 4 MHz. The pulsations always seemed to be strongest at the tuning positions that yielded maximum average laser power. Tuning far enough away from peak power always made them disappear.

As was discussed in chapter 2, in the cases in which we have observed the low-frequency spectrum of the laser power carefully, on a Ge photoconductor, we have often observed low-frequency (i.e. $\nu < 5$ - 10 MHz) excess noise (see figure 2.35). Unfortunately, this noise can only be seen on the photoconductors, not the Schottky diodes, and the germanium photoconductors have their own chaotic, non-linear dynamics. As discussed in chapter 2, in order to separate the detector and laser noise, it is necessary to measure the power dependence of the noise spectra and decompose it into components that vary linearly and quadratically with laser power. This is a somewhat indirect technique, however. Because we did not make a systematic study of the laser noise, we cannot add much to the present state of knowledge about it, even though the latter is itself pretty meagre. Our one firm conclusion, and the most important matter from an operational point of view is that if laser pulsations or excess low-frequency laser noise can be discovered in real time, then it is always possible to tune them away, albeit with a sacrifice in total laser power.

Finally, we note that in terms of system performance, the low-frequency noise and instability is a

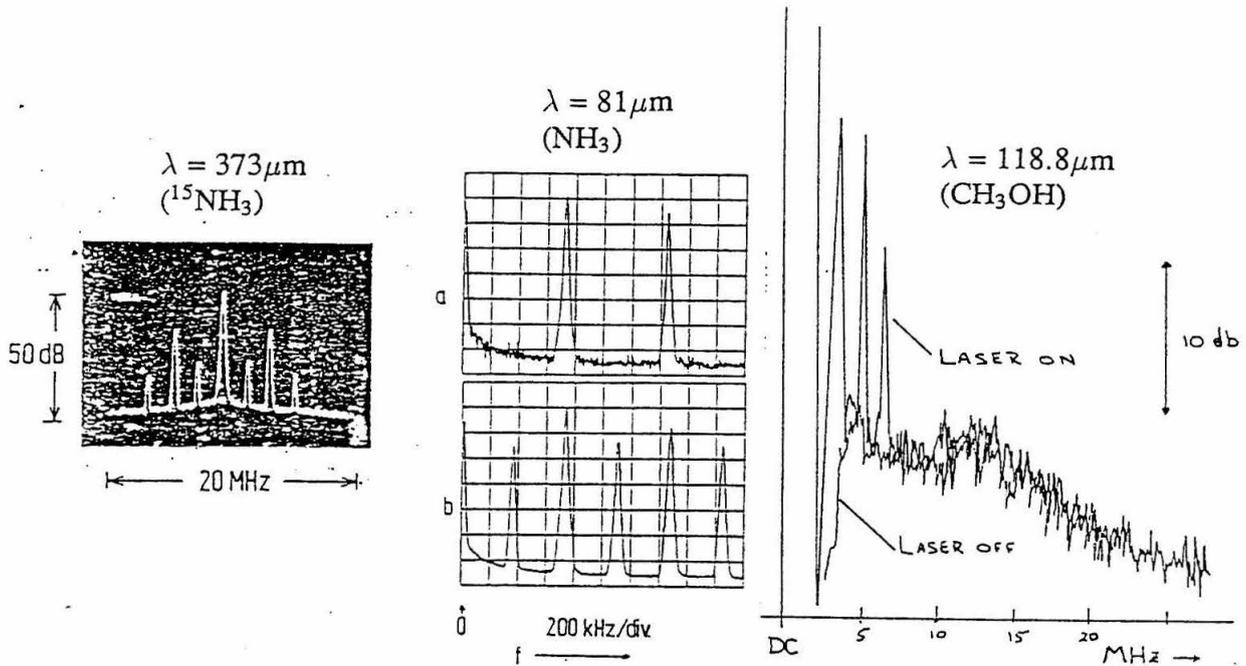


Figure 3.13 – Spontaneous pulsations on various FIR laser lines, $\lambda = 373 \mu\text{m}$ from Weiss (1985), $\lambda = 81 \mu\text{m}$ from Weiss and Brock (1986), $\lambda = 119 \mu\text{m}$, this work.

greater problem for our receiver than for others which employ high IF frequencies. For IF frequencies much greater than the pressure-broadened linewidth, these problems would certainly be less relevant. The problem of the laser instabilities was not realized early enough to be taken into account in our receiver design. We note that other scanning-LO, low IF frequency receivers (e.g. the InSb hot-electron bolometer, Phillips and Jefferts 1973,) have similar problems with low-frequency LO noise from multiplied klystrons. To summarize, the problem of laser instabilities and low-frequency noise is a fundamental problem with our receiver. Not enough is known about it to say for certain whether it is a fatal problem in our application. Since we have always found that tuning out the instabilities costs laser power, however, it appears, that its main effect is to exacerbate the receiver's other fundamental problem, low LO power, which we do know is fatal. The problem of LO production is the subject of the next chapter.

Chapter 4 – Sideband Generator

The approach we have adopted for generation of our receiver's local oscillator is one that has been used previously (Bićanić, Zuidberg, and Dymanus, 1978, Fetterman *et al.* 1978, Farhoomand *et al.* 1985) for other applications, such as molecular spectroscopy, requiring oscillation at arbitrary far-infrared frequencies and modest power levels. The idea may be implemented in various ways, but what is common to all of them is the following : A far-infrared laser, at angular frequency ω_{FIR} , is focussed onto a small-area Schottky diode at the feed of some sort of antenna. At the same time, a lower-frequency, higher power wave, (angular frequency ω_{RF} .) is also coupled onto the diode. The non-linearity of the diode's current-voltage (I-V) characteristic generates mixing products of the two waves $\omega_{mn} = m\omega_{FIR} + n\omega_{RF}$, where m, n are integers, which are coupled out of the diode via the same antenna structure by which the carrier was coupled in. Some sort of external filtering scheme is then required to separate the desired sideband from the (typically much stronger) carrier and all the other sidebands.

An additional feature common to all the various experimental implementations of this scheme is extremely low efficiency. For the single-sideband conversion efficiency, which we define as the power in one sideband to that in the incident carrier, Bićanić (1983) reports -58 db with an 890 GHz carrier, Farhoomand *et al.* report -45 db at 1890 GHz, while our own experiments at 2520 GHz (118.8 μm) have achieved approximately -40 db. Hitherto, very little effort has been devoted to understanding the reasons for such poor efficiency. For example, it is not known how much of the loss is attributable to low antenna efficiency and how much to losses in the diode. Nor is it known what the physical mechanisms are that contribute to the diode losses. Consequently, it is also not known what the ultimate limits to the diode losses are, nor how the diode parameters must be altered in order to achieve these limits.

In this chapter, we have a twofold purpose. We describe, component by component, the design and performance of the sideband generator we actually built, and we also analyze its operation theoretically. The theoretical analysis has been done in an effort to account for all our measured losses and to point the way to improving the performance of FIR laser sideband generators generally. Our implementation of the sideband generator consists of a corner-cube antenna, a high-frequency GaAs Schottky diode (obtained

from R. Mattauch's group at the Univ. of Virginia) at the corner-cube's feed, a polarizing Michelson interferometer, and a tunable Fabry-Perot interferometer used simply as a tunable bandpass filter. In the first section we describe the design and construction of our corner-cube antennas, and compare their theoretical and measured performances. In the next section, we make a detailed theoretical analysis of Schottky diodes as sideband generators, modeling the combination of diode and antenna as a reflective FIR modulator. The first part of this section describes the method we have used to determine the flat-band or "built-in" potential of the Schottky diodes, which is an important parameter entering the theoretical analysis. It is an extremely simple method, but to our knowledge has not previously been recognized. We therefore describe it in some detail, and also illustrate its application with experimental data from one of our diodes. In the next two subsections, we describe the theoretical derivation and numerical evaluation, respectively, of the diode's FIR reflectivity as a function of voltage, $\Gamma(V)$, as various parameters (frequency, diode radius, etc.) are varied. Then we discuss how the single-sideband conversion efficiency may be predicted from $\Gamma(V)$, and present a series of contour plots of the conversion efficiency in the diode radius - epilayer doping plane. In section §4.3 we discuss the polarizing Michelson and Fabry-Perot interferometers, and various details of the optics. Finally, in section §4.4, we discuss our measurements of the overall performance of the sideband generator. Our highest measured sideband power is 9.5 nW, measured in a single sideband at the output of the Fabry-Perot. It can probably be increased in a fairly straightforward way to about 70 nW by increasing the Fabry-Perot transmission with replacement of the mesh mirrors.

4.1 – Corner-cube Antennas

Two versions of corner-cube have been constructed. The simpler one, which has been used for most of sideband generation experiments, is illustrated in fig. 4.1 and is merely the conventional design, as described by e.g. Harris *et al.* (1986), scaled down to a design wavelength of $\lambda = 100\mu$. As shown in the figure, it incorporates three main sections: a base, or ground plane, and a pair of split dihedral reflectors. The distance from the 90° whisker bend to the ground plane is 4λ and the horizontal distance from bend to the dihedral is 1.2λ , as per the standard recipe of Krautle, Sauter, and Schultz (1977). The whiskers we use are frequently somewhat longer than 4λ , and the diode surface therefore lies slightly below the level of the ground plane, but no degradation in beam pattern or total coupling efficiency has been observed

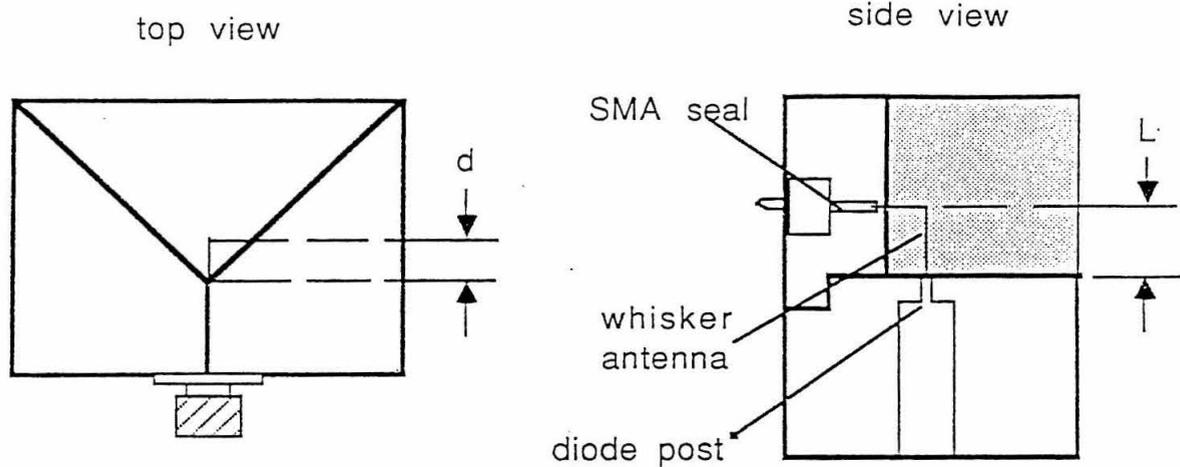


Figure 4.1 – Corner-cube antenna (first version)

because of this. The diameter of the hole into which the center pin of the SMA seal slips is chosen to yield a characteristic impedance of 50Ω , thereby minimizing the electrical length of the corner-cube's RF mismatch. The corner-cube was machined (Custom Microwave Inc., Longmont, Colorado) from brass, the reflecting surfaces were polished to an optical finish to facilitate optical alignment with visible (HeNe laser) light, and the entire assembly was gold plated.

The geometry of the contact between the whisker and Schottky diode is illustrated in figure 4.2. The epilayer is grown on the top surface of a degenerately doped GaAs substrate. An Ohmic contact is fabricated on the base of the substrate, with a surface layer of gold to facilitate soldering to the diode post. The epilayer is overcoated with a protective layer of SiO_2 approximately $.4 \mu\text{m}$ thick. The diode contacts, arrayed over the surface of the epilayer, thus lie in shallow wells in the SiO_2 . The Schottky barrier is formed by an electroplated layer of Pd, which is coated with an additional layer of gold. As the diode post is raised, the electrochemically sharpened whisker slips into one of the wells in the SiO_2 and contacts the diode pad within.

The whisker antennas are made of $.025 \text{ mm}$ diameter, Au(82%)-Ni wire (California Fine Wire Co.). They are fabricated by soldering a short length of wire onto the Kovar center-pin of an insulating SMA seal (EMC Technology, Cherry Hill, NJ), and then bending the wire over a stainless steel form. The form has

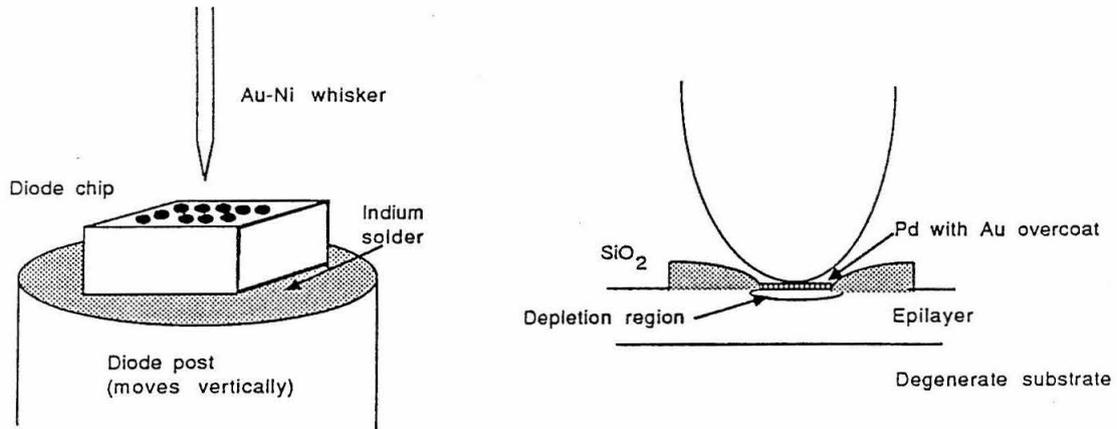


Figure 4.2 – Geometry of Schottky diode chip and whisker

a sharp edge machined on it at the precise distance from the front surface of the SMA seal that is required. The vertical section of the wire is then shortened to its proper length and pointed by electrochemical etching. A certain amount of trial and error was required to obtain whisker points sharp enough to contact our smallest diodes ($\sim 1.4\mu$ diameter). It was found, however, that good results ($> 50\%$ yield) could be obtained with an etch solution of 10% (by weight) sodium cyanide, 5% potassium ferricyanide, 85% distilled water and an etch voltage of ~ 14 V. After etching, the whisker is rinsed briefly in hydrochloric acid ($\sim 20\%$ solution), and then washed in methanol.

After the pointing, the whisker and seal are seated in one of the split dihedral pieces, the second dihedral piece lowered on top of the first, and the assembly bolted together. The dihedral assembly is then lowered onto the base and bolted fast, leaving the pointed whisker poised directly above, (or more usually, projecting slightly into,) the diode post hole. A curve tracer is then connected to the SMA output and the diode post slowly raised, using a specially-made fixture, until a contact is made. All the above steps, including soldering, bending, etching, assembly, and contacting, are performed under a stereo microscope. Once a contact has been made that has satisfactory slope parameter ($\frac{\eta kT}{e}$ per e-fold in diode current, or 60η mV/decade) and series resistance ($\leq 20\Omega$), the diode post is locked in place with a shaft clamp, and the entire assembly transferred to a small goniometer that allows fairly precise angular adjustment over a large range.

Two important changes have been made in our second corner-cube design (fabricated by the Zen

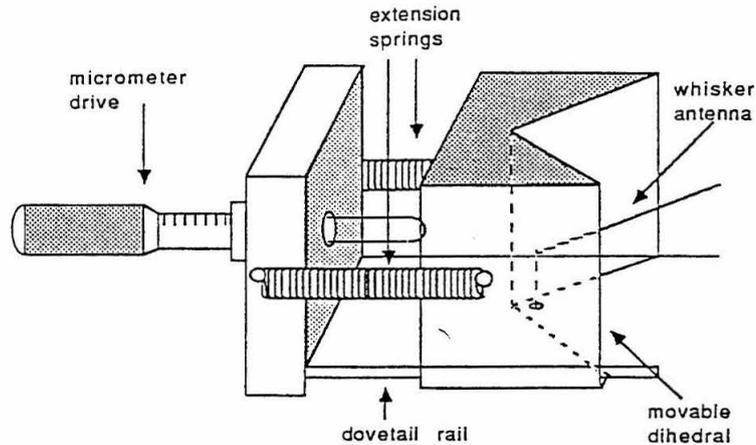


Figure 4.3 – Zen corner-cube

Machine Shop, Pasadena, CA). Firstly, it employs a moveable dihedral reflector, and secondly, the horizontal section of the whisker is oriented parallel to one of the dihedral faces (see fig. 4.3), rather than in the antenna's E-plane. The moveability of the dihedral provides an additional tuning element for optimization of the sideband generation efficiency. Partially, this is possible due to the dependence on dihedral position of the whisker's impedance, when viewed as a transmission line. A much larger effect, however, is the dependence of the antenna beam pattern on dihedral position due to variation of the relative phases of the whisker and its three images. A consequence of the moveability of the dihedral is that it is no longer possible to bring the horizontal section of the whisker in through it. This is desirable, in that the required clearance hole in the dihedral cannot be made small enough (with conventional machine tools) for one to be certain of its not perturbing the antenna beam pattern and efficiency. For example, in our first corner-cube, this hole is approximately 340μ in diameter, which is clearly significant at a wavelength of $100\mu\text{m}$. Furthermore, on the scale of $\frac{\lambda}{4}$, the two pieces of the dihedral reflector do not mate perfectly – the "crack" between them undoubtedly has some undesirable effect on the beam. (At longer wavelengths, e.g. the $370\mu\text{m}$ from which the first corner-cube's design was scaled, this would obviously be less of a problem.) In this respect, therefore, a whisker which comes in from the front of the corner-cube and allows an unsplit reflector has an advantage.

An important consideration in designing corner-cubes for our application is mechanical stability, i.e.

the susceptibility of the contact's electrical properties to mechanical shock and vibration. The ruggedness of a diode contact does vary somewhat from one whisker point to the next, but generally, any contact that is very good electrically will be adequately stable. Most likely, this simply reflects the fact that, when the whisker's tip is too large, bent, or irregularly shaped to slip cleanly into the diode well, so that it can only contact the electrode by a corner (this can be seen in many of the electron micrographs taken by Zimmermann *et al.* 1987), then the contact is degraded both in its electrical properties and in its mechanical stability. For both corner-cubes, we have found (not surprisingly) that the most delicate operation is the initial clamping of the diode post after the electrical contact is made. Once this has been successfully accomplished and the corner-cube mounted in its goniometer, the contact is usually insensitive to being picked up and set down on tabletops, light tapping on its mount with a blunt instrument, etc. Nonetheless, the time and tediousness involved in pointing and contacting whiskers makes a certain amount of paranoia worthwhile when handling the well-contacted ones.

Our second version of corner-cube incorporates a considerably longer horizontal section of whisker, and might therefore be expected to be less rugged. Indeed, based on our fairly limited experience with the new corner-cube, its contacts do appear to be somewhat more fragile than those made in the simpler corner-cube. In both cases, however, our actual experience on the KAO has been positive. In the flights we have made so far, the corner-cubes were either packed in their carrying case or held by hand during takeoff, and only bolted to the receiver once the plane was airborne. On the first flight, the new corner-cube's I-V curve remained unchanged throughout the flight. On the second flight, its I-V curve survived unchanged while it was mounted on the receiver, tuned, and removed from the receiver. However, the I-V curve degraded significantly in the course of extensive subsequent handling. The earlier version of corner-cube remained in its case during the flight and its I-V curve survived unchanged.

Although having the whisker come in from the front of the corner-cube has the advantage of a solid reflector, it is important in this case to avoid having the horizontal section of the whisker directly interfere with the antenna field. It is difficult to predict the size of this effect, but intuitively, we expect that the smallest perturbation will occur when the integral along the horizontal wire's length, of the parallel

component of the unperturbed electrical field, is minimized. Only the far field distribution of corner-cube antennas has ever been studied, but it is intuitively clear that the total field distribution in the E-plane resembles that shown in fig. 4.4. Although the near field is normal to the horizontal wire, and the far field is spatially well separated from it, it is possible that in the transition region, the parallel component of electric field is significant. Therefore, the second corner-cube was designed to have the horizontal section of the whisker come in at 45° to the E-plane, and parallel to one of the dihedral faces. (At the faces, of course, the parallel electric field must rigorously vanish whether in the near, far, or transition regions).

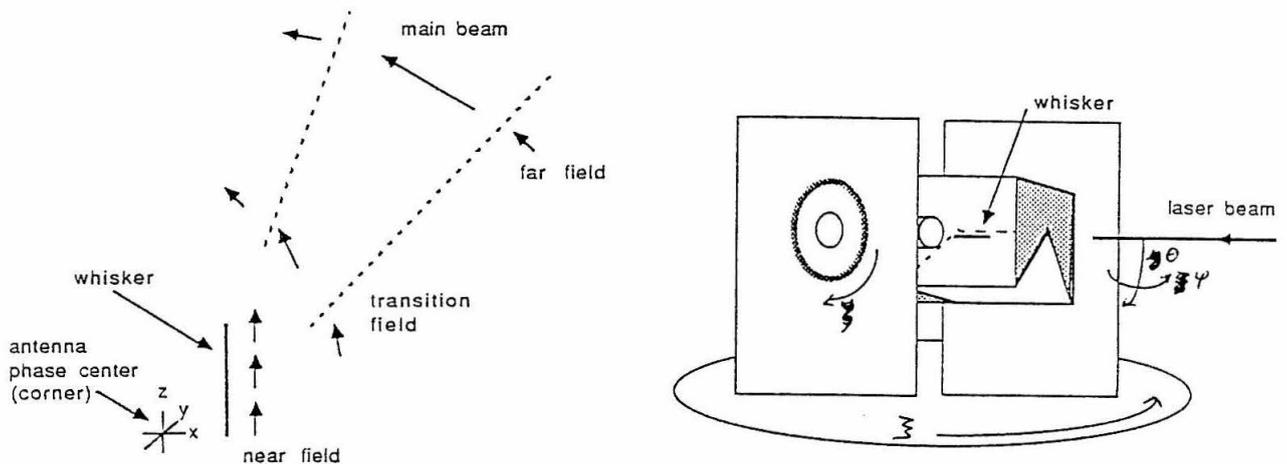


Figure 4.4 – (left) Schematically, the near-field, transition-field, and far-field polarization and distribution; (right) corner-cube coordinate systems

Because the confrontation of theory and experiment for the FIR beam patterns of corner-cubes is of general interest, and because it is important to the performance of our sideband generator, and (mainly) because we have some good experimental data on it, we shall now briefly review the theory of the beam patterns. The coordinate systems we shall use to describe the theoretical and experimental properties of corner-cube beams are summarized in fig. 4.4. The origin coincides with the corner, where the dihedral reflector intersects the ground plane. The dihedral angle is taken to be the z-axis, and the whisker antenna is parallel to it and separated from it by a distance d along the x-axis. The usual spherical coordinates in this frame are denoted (θ, ϕ) , and the antenna E-plane is given by $\phi = 0$. In the laboratory, the corner-cube

is mounted in a two-axis goniometer arrangement and the laser incident at a fixed angle, as shown at the right of figure 4.4. The reading on the y -goniometer axis is denoted ζ and that on the z -goniometer axis ξ . The relationship between the (θ, ϕ) in the whisker frame at which the laser is incident and (ζ, ξ) is :

$$\begin{aligned}\cos \theta &= \cos \zeta \cos \xi \\ \cos \phi &= \frac{\sin \zeta \cos \xi}{(1 - \cos^2 \zeta \cos^2 \xi)^{1/2}}\end{aligned}\quad (4.1)$$

The first step in understanding the properties of the corner-cube antenna is a consideration of the beam of a single longwire antenna without reflectors. Beginning with a traveling wave current distribution

$$\vec{I} = \begin{cases} I_0 e^{i(kz - \omega t)} \hat{z} & z < L, \quad x = 0, \quad y = 0 \\ 0 & \text{elsewhere,} \end{cases}\quad (4.2)$$

it is an elementary exercise (see e.g. Jasik, 1961) to derive the radiated power distribution in the far field ($|r| \gg L^2/\lambda$). It is

$$P = \frac{2I_0^2}{r^2 c^2} \frac{\sin^2 \theta}{(1 - \cos \theta)^2} [1 - \cos(kL(1 - \cos \theta))],\quad (4.3)$$

where $k = \frac{2\pi}{\lambda}$ is the free space propagation constant and L the whisker length. This pattern is azimuthally symmetric, and has a succession of conical lobes separated by nulls, one lobe for every half-wavelength of whisker length. The envelope of the antenna lobes is strongly peaked toward the forward lobes ($\theta \leq \pi/2$). The fact that the forward lobes are favored is due to the traveling wave, as opposed to standing wave, character of the current distribution (4.2). Although it is certainly not obvious that a traveling wave distribution should be applicable, it was found experimentally many years ago that the forward-backward asymmetry did indeed exist (Mattarese and Evenson 1970). Furthermore, it was found that the number of lobes and nulls in the pattern exactly corresponded to an effective whisker length equal to the geometric distance from the diode to the first bend in the whisker. The antenna behaved as though the portion of the whisker beyond the bend did not exist.

The extension of this treatment to a longwire antenna in a corner-cube is conceptually straightforward, but algebraically tedious. By adding together the electric fields due to traveling wave currents such as (4.2)

located at $x = \pm d$, $y = 0$, and the negative of (4.2) located at $x = 0$, $y = \pm d$, (to account for the 180° phase shift at a reflecting surface, and then taking the squared absolute magnitude, one obtains :

$$P(\theta, \phi) \propto \frac{\sin^2 \theta}{(1 - \cos \theta)^2} (1 - \cos(kL(1 - \cos \theta))) [\cos(kd \sin \theta \cos \phi) - \cos(kd \sin \theta \sin \phi)]^2 \quad (4.4)$$

This analysis was used by Krautle, Sauter, and Schultz (1977) in their original paper to conclude that $L = 4\lambda$; $d = 1.2\lambda$ was the optimal configuration. The complete beam pattern for this standard prescription is shown in fig. 4.5. As has been frequently remarked (e.g Goldsmith, 1982) the largest sidelobes do not lie in either of the principal planes. Nonetheless, the pattern is still, by any standard, quite good, with a main beam efficiency of 58%. We compute the efficiency in practice by approximating the main beam as an ellipse in the ζ - ξ plane, centered on the peak and extending to the 5% contour.

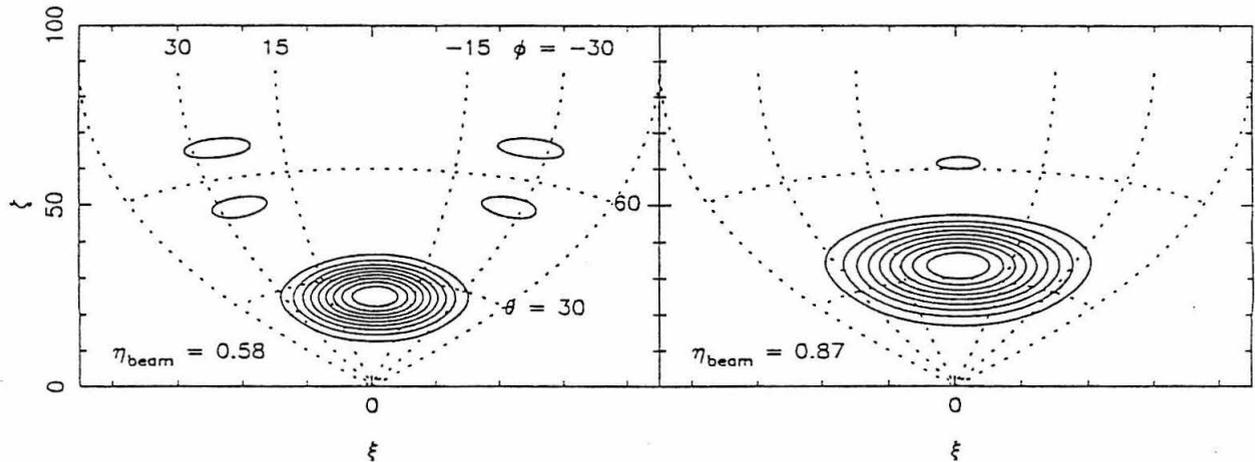


Figure 4.5 – Theoretical beam patterns : (left) the standard recipe, $L = 4\lambda$, $d = 1.2\lambda$, (right) standard corner-cube used at 50 % longer than the design wavelength, i.e. $L = 2.7\lambda$, $d = .8\lambda$. In both cases, contours are every 10 %. Dotted lines indicate the usual spherical coordinates in the whisker's frame of reference.

Now, Krautle, Sauter, and Schultz (1977) settled on $L = 4\lambda$ $d = 1.2\lambda$, simply by examining 1-D scans of the beam through the principal planes. Since their work, a very large number of researchers have adopted this recipe in the design of their corner-cubes. There has been extensive work on low-frequency

scale models, (Sauter, Schultz, and Wohleben, 1983), in order to better understand the relation between theoretical and experimental (1-D) beam profiles. There has also been work (Vowinkel 1986) on including various additional effects (and free parameters) into the models in an effort to improve the agreement. In particular, Vowinkel has added a standing-wave component to the current distribution, and a component of traveling-wave current in the horizontal section of the whisker. The latter innovation is somewhat curious in view of the Mattarese and Evenson (1970) result described above, and in view of the claim, based on an 8 GHz scale model, of Fetterman *et al.* (1978) of 13 db of FIR isolation due to the inductance in the 90° bend. According to Vowinkel, one of the results of including the horizontal wire current is the presence of a substantial ($\sim 10\%$) cross-polarized component to the radiated field away from the E-plane. At the peak of the main lobe, the cross-polarized component goes to zero, however. The final result of Vowinkel's extended theory is that for $L = 4\lambda$, $d = 1.1\lambda$, the main beam efficiency is found to lie between 50 % and 60 %, in agreement with the simple result predicted by equation 4.4.

In our own modeling of beam patterns, we have tried including two effects absent in the conventional analysis. The first is the presence of a ground plane at $z = 0$ which acts to reflect the small reverse sidelobes of the conventional pattern into the forward direction, where they coherently combine with the forward pattern. It is modeled in the obvious way, with four additional image currents at $-L < z < 0$. For whiskers an integral number of wavelengths long, the positions of the lobes and nulls do not change, due to the symmetry of the single longwire antenna pattern under reflection about the $z = 0$ plane. The relative amplitudes of the various lobes do change somewhat when the ground plane is included, but the effect on the main beam efficiency is small. The other effect we have included is the truncation of the image antennas to less than the full whisker length, L . This was intended to model the fact that in our actual corner-cubes, the whisker usually is slightly longer than 4λ , and therefore extends slightly below the level of the ground plane, whereas the image antennas can only extend down to the ground plane, and are always exactly 4λ long. Somewhat surprisingly, this also was found to have relatively little effect of the overall efficiency.

The implicit aim of all this work on corner-cube antennas is to improve their efficiency. It is remarkable

therefore, that from the point of view of main beam efficiency, equation 4.4 does not at all imply that $L = 4\lambda$, $d = 1.2\lambda$ is optimal. Indeed, it is quite surprising that this fact has not been noticed in the course of a decade of work on corner-cubes, but it is nonetheless true. Figure 4.5 illustrates it graphically. The theoretical efficiency is increased from 58 % to 87 % by scaling L and d down by one-third. The largest sidelobe then appears in the E-plane. By examining the pattern only in the principal planes, one would mistakenly conclude that the pattern was worse than the standard recipe. Of course, the main beam efficiency is only a part of the overall coupling efficiency. By scaling down L and d to improve η_{beam} , it is possible that the driving point impedance of the antenna may be changed in such a way as to decrease the total efficiency. The optimal driving point impedance of the antenna depends on the impedance of the device at the feed, however. We therefore regard optimizing the antenna impedance as a separate question from that of the main beam efficiency.

We have made a number of measurements of our corner-cubes' antenna patterns, using the video signal from the far-infrared laser. They have been made in a variety of optical configurations and at several FIR wavelengths. Unfortunately, they are all subject, to some degree, to various systematic errors. These can be caused by imperfections in : the shape (i.e. spatial purity) of the beam emitted from the far-infrared laser output coupler, the stability of the far-infrared laser power on timescales comparable to the intervals between measurements, the polarization purity of the laser (since the parallel- and cross-polarized components of the corner-cube response generally will have different angular distributions,) the angular size of the beam focussed onto the corner-cube, alignment of the polarizing Michelson interferometer, if used, vignetting and/or stray reflections of laser power by any optics intervening between the laser and corner-cube, and "pulling" of the far-infrared laser cavity as the corner-cube orientation is varied, due to overcoupling of the laser output. At one time or another, every one of these effects has been a significant problem. The shape and angular size of the laser beam enter due to the fact that (roughly) the measured responsivity as a function of (ζ, ξ) is the convolution of the true corner-cube pattern and the input beam distribution. The optics of the polarizing Michelson interferometer were designed so as to match the beam incident on the corner-cube to the corner-cube's intrinsic beam, since in this condition, the power coupling of the two is maximized. In this condition, however, the incident beam is fast enough to significantly broaden the

intrinsic corner-cube beam. Likewise, the problems of vignetting, stray reflections, optical alignment, and polarization purity are all most severe in the case of measurements made with the polarizing Michelson interferometer.

Our most careful and unambiguous measurements were made at a wavelength of 214 μm , on the first version of corner-cube (design wavelength 100 μm), and with the polarizing Michelson interferometer omitted. The optical setup is shown in fig. 4.6. The laser is not focussed at all. This produces a loss in the signal of nearly 20 db, compared to a matched beam, but nevertheless, the uncertainty due to ordinary Gaussian noise is still small compared to residual systematic errors. The laser was chopped at a low audio frequency and the corner-cube video signal synchronously detected. The pyroelectric signal was also synchronously detected, and the two demodulated outputs digitized and recorded on the computer. The optoacoustic lock was disabled for these measurements, and instead, the corner-cube's signal was simply normalized by the pyroelectric's. The efficacy of the this procedure was checked by holding the corner-cube orientation fixed and monitoring the two signals over a time interval of several minutes, while the CO₂ laser frequency drifted free. Drifts in FIR laser power of 30 %, typical of the free-running laser stability after a long warm-up, could be canceled out to a level of < 5 %, while the independently measured noise level was some 3 % of the peak signal.

In measuring the beam pattern, a polarizer was inserted in front of the pyroelectric, and oriented to transmit the corner-cube's nominal (E-plane) polarization. The purpose of this was to ensure that, if the laser polarization varied systematically as the corner-cube orientation was changed, due to laser cavity "pulling," the effect would also be canceled out in the normalization. This effect was indeed observed at the ~ 10 % level in preliminary measurements made without the polarizer. The laser polarization, with the corner-cube removed, was measured separately by rotating the polarizer in front of the pyroelectric. The cross-polarized component was ~ 5.5 % of the total power. The laser beam profile was also measured, by taking one-dimensional scans of the pyroelectric signal in the two directions normal to the beam, with the pyroelectric located well into the far field of the output coupler ($\frac{\lambda z}{\pi w_0^2} \approx 40$). The beam was found to be diverging at approximately 3° (FWHM). This is sufficiently slow that it may be treated as a plane wave.

No deconvolution, therefore, was applied to the measured pattern. The corner-cube beam was sampled at 5° intervals in the outer regions, and at 2° intervals in the neighborhood of the main beam. Between the measured points, the map was interpolated using the relaxation algorithm (i.e. every point updated to be the average of its nearest neighbors.)

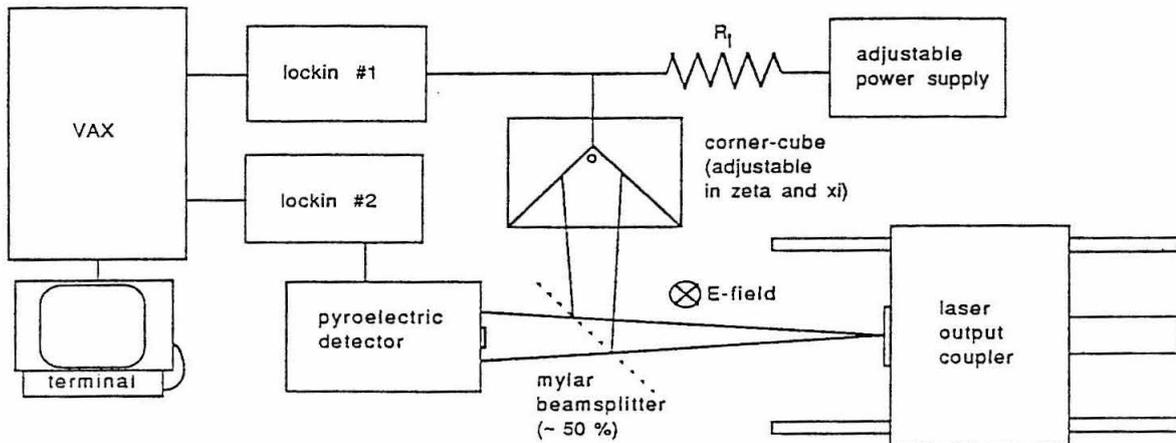


Figure 4.6 – Setup for corner-cube beam measurements.

The resulting corner-cube antenna pattern is shown in fig. 4.7, together with the theoretical beam pattern (given by equation 4-4) for an antenna of dimensions $L = 695 \mu\text{m}$, $d = 135 \mu\text{m}$. The whisker dimensions were measured with our optical microscope, fitted with a calibrated reticle. Thus, there are no free parameters in the theoretical beam. The main features of the measured beam are the following. Firstly, from an operational point of view, the beam is very good, with a main beam efficiency of 44 %, and half-power beamwidths of 14° and 16° in the E and quasi-H planes, respectively. Examining the pattern more closely, we find that the two best defined features are the angular positions, in the E-plane, of the main beam peak, and the first minimum beyond it. The measured separation of the two, $13.5 \pm 1^\circ$, is correctly predicted by theory. The first sidelobe is quite asymmetric about the E-plane. The symmetry of the main beam, however, is quite good, so it does not seem likely that the sidelobe's asymmetry can be attributed to the laser beamshape. The first sidelobe is highly elongated in the ξ direction, reaches a peak level of about 30 % of the main beam's, and has an angular size 2-3 times larger than the main beam.

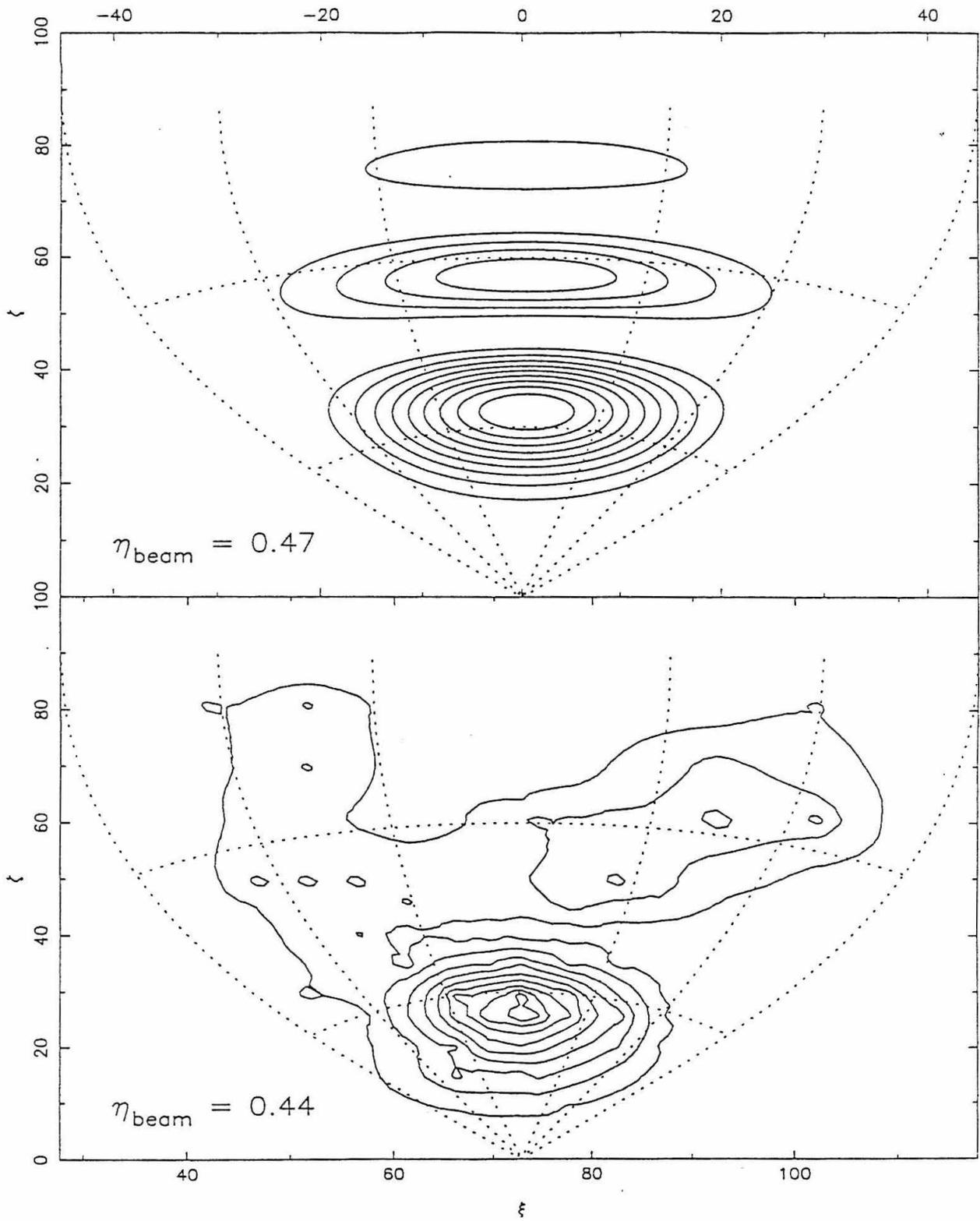


Figure 4.7 – Theoretical (top) and experimental (bottom) beams of the first version of corner-cube, measured at $\lambda = 214 \mu\text{m}$. Contours are every 10 % .

In comparison with the theoretical beam, the measured angles in the E-plane (i.e. ζ) of the main lobe, the first minimum, and the first sidelobe, are all about 5° too small. We attribute this to a slight skewness of the whisker, which was observable (*ex post facto*) during visual inspection beneath the stereo microscope. That is, the angle of the bend between the whisker's vertical and horizontal sections was $> 90^\circ$, so that the bend was closer to the dihedral than the diode. Accurate measurement of the angle was not possible, however. The half-power widths of the theoretical beam, in the E and quasi-H planes are 16° and 24° , respectively, the latter of which does not match experiment very well. Furthermore, the shape of the first sidelobe, and the existence of a second sidelobe (at $\zeta \sim 75^\circ$) are completely discrepant with experiment. As it turns out, however, the discrepancies partially cancel, (for example, the first theoretical sidelobe is higher, but also smaller in angular size, than experiment,) so that a comparison of theoretical and measured main beam efficiencies, 47 % vs. 44 % respectively, gives an illusory impression of success to the theory.

Unfortunately, a careful two-dimensional map of the entire beam has not been made at any wavelength besides $214 \mu\text{m}$. A map of the main lobe at $\lambda = 118 \mu\text{m}$, however, was made, and is shown in fig. 4.8. In this measurement, the far-infrared laser power was not separately monitored. The CO_2 laser was stabilized with the optoacoustic lock circuit. It was found, however, that oscillations of $\sim 80\%$ depth occurred in the signal as the corner-cube's orientation was varied. It was not determined whether this was due to standing waves between the laser output coupler and corner-cube or due to "pulling" of the far-infrared cavity. To compensate for them, the corner-cube was refocused (i.e. its position along the optical axis readjusted,) at each orientation so as to peak up the signal. The laser beam was focused onto the corner-cube by means of a 90° off-axis paraboloidal mirror, yielding a converging beam of 2.8° half-power diameter. The whisker length was measured to be $810 \mu\text{m}$ during these experiments. The corresponding theoretical beam pattern is also shown in fig. 4.8. There is marginal agreement between the half-widths of the theoretical beam ($11^\circ \times 16^\circ$ in ζ, ξ respectively,) and experimental beam ($8.4^\circ \times 10.6^\circ$, after Gaussian deconvolution.) The fact that the ellipticity of the measured beam's outer contours is oriented at $\sim 40^\circ$ from the principle planes is difficult to understand, however.

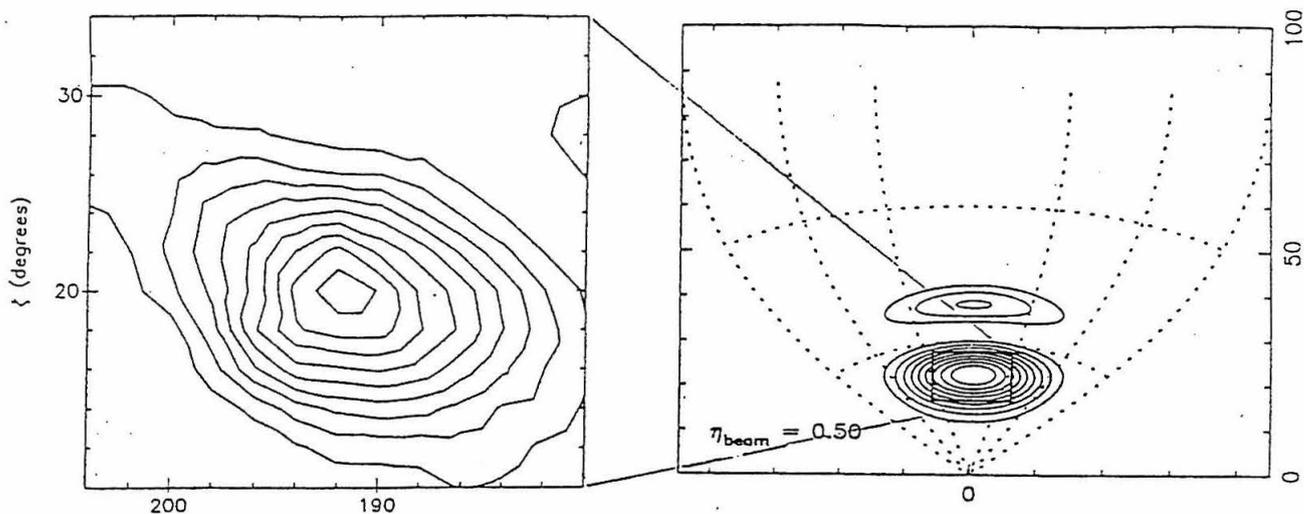


Figure 4.8 – (Left) Main beam of the corner-cube at $119 \mu\text{m}$, (right) Theoretical beam for $\lambda = 119 \mu\text{m}$, $L = 810 \mu\text{m}$, and $d = 120 \mu\text{m}$. Contours are every 10 %.

To summarize our work on corner-cube antenna patterns, we find that the standard theory of a traveling-wave, longwire current distribution plus images agrees qualitatively with experiment, at least in terms of the position and shape of the main beam. The strength and shape of the sidelobes are not well predicted by theory, but for the particular wavelength and whisker dimensions of our experiment, there is a (somewhat fortuitous) agreement between the measured and predicted main beam efficiencies. In addition, we find that, contrary to current dogma, the simple, standard theory does not predict maximum main beam efficiency at $L = 4\lambda$, $d = 1.2\lambda$. It only predicts minimum sidelobe level in the principal planes for that configuration. From the point of view of system design, we find our experimental antenna patterns to be adequate for our application, with a main beam efficiency of 44 % measured at $214 \mu\text{m}$. It does not seem that the shortcomings of the theory or the typical mechanical imperfections of the corner-cube and whisker are areas in which improvements would help the overall performance of the sideband generator much.

4.2 Theory of Schottky Diodes as Far-infrared Modulators

In this section we analyze the influence of the Schottky diode's electrical properties on its efficiency as a sideband generator. In order to make quantitative predictions, the values of various diode parameters must be known. Most of these are either well known material constants or else have been measured for

individual diodes by the group that supplied us with them (R. Mattauch's at the University of Virginia.) The diode's flat-band (or "built-in") potential, V_{FB} , was not known, however, and turns out to be important to performance because of the vanishing of the depletion region at voltages above it. In the course of our analysis of sideband generation efficiency, we discovered (theoretically) a simple and reliable technique for measuring V_{FB} using the bias dependence of the video responsivity. The technique was verified experimentally and the resulting value of V_{FB} used in the sideband generation model. Thus, we begin our analysis of the Schottky diode as a sideband generator by describing this method for measuring V_{FB} , which, to our knowledge, has not been recognized by previous workers.

We then go on to describe our analysis of the sideband generator proper, which we conceptually divide into three parts : the coupling of the optical beam into a traveling current wave on the whisker, the transmission of the current wave down the whisker, which for this step is viewed as a lossless transmission line, and the wave's partial reflection off the diode. The last of these processes is the most complicated, and contains most of the scope for improvement in the sideband generation efficiency. We address it in two parts. First, we use a fairly conventional model of the diode to predict its small-signal far-infrared reflectivity as a function of DC bias, which we denote $\Gamma(V)$. This model includes the high frequency parasitics due to carrier inertia, dielectric relaxation, and skin effect, as modeled by Champlin and Eisenstein (1978), and the vanishing of the depletion region above the flat-band potential, as modeled by Crowe and Mattauch (1986, 1987). In the second step of the analysis, we determine how to predict the overall single-sideband conversion efficiency, given $\Gamma(V)$, an optical coupling efficiency, and an assumed RF waveform, . We define the conversion efficiency ζ , as the power at the desired far-infrared sideband frequency, coupled into the optical output beam, divided by the power of the incident far-infrared carrier.

a. Determination of the Flat-band Potential

Physically, the barrier in a Schottky diode is created because the chemical bonding configuration in an n-type (p-type) semiconductor makes it energetically favorable for electrons (holes) to migrate across the interface into the metal. This sets up a "built-in" field in the region between the metal and a space-charge layer which accumulates in the semiconductor some distance (typically, say, $\sim 1000 \text{ \AA}$ at zero bias) from

the interface. Within this “depletion region”, no mobile carriers exist. The energy-band structure of the Schottky barrier is illustrated in figure 4.9.

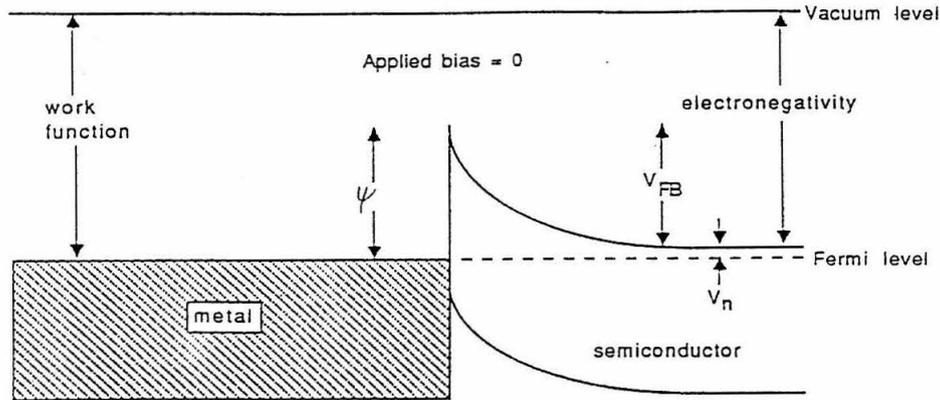


Figure 4.9 – Energy-band structure of the Schottky diode, illustrating the definitions of flat-band voltage V_{FB} , barrier height Ψ , and V_n . The Fermi level is indicated by the dashed line.

The video responsivity of Schottky diodes has been treated by many authors (e.g Torrey and Whitmer 1948, p. 335 ff. , Waksberg and Dreze 1984, Kreisler, Pyeé, and Redon, 1984). The zeroth-order equivalent circuit for the Schottky diode consists of an ideal Schottky barrier shunted by a voltage dependent, parasitic capacitance, C_b , the parallel combination then appearing in series with a “spreading resistance”, R_s . In reality, R_s is complex and frequency-dependent, and contains the effects of the high-frequency parasitics. As we shall show, under certain conditions, the video responsivity of the diode, that is, the change in its DC bias point when a high-frequency signal is incident, falls linearly with increasing bias voltage. The voltage at which the linearly extrapolated video response falls to zero is precisely V_{FB} .

We assume the current and voltage across the ideal Schottky barrier are related by the Richardson equation for pure thermionic emission :

$$i_b = A^* AT^2 \exp\left(\frac{e}{\eta kT}(V_d - \Psi)\right) \quad (4 - 5)$$

where A is the diode area, A^* the modified Richardson constant ($= 8.6 \text{ A cm}^{-2} \text{ K}^{-2}$ for n-type GaAs), T the absolute temperature, η the diode ideality factor, and Ψ the diode barrier height. The prefactor of the exponential, $A^* AT^2$, is often called the “saturation current”. The voltage-dependent capacitance that

shunts the ideal Scottky barrier is

$$C_b(V) = A \left(\frac{\epsilon_0 \epsilon N}{2(V_{FB} - V_b(V))} \right)^{1/2} \quad (4-6)$$

where N is the epilayer doping concentration in cm^{-3} , ϵ the relative permittivity of GaAs ($= 10.8$), and ϵ_0 the permittivity of free space ($8.84 \times 10^{-14} F/cm$). The square-root voltage dependence of the capacitance arises simply from a parallel plate capacitor model, with the depletion layer thickness given by the quadratic solution to Poisson's equation (Torrey and Whitmer, p. 72 ff.) applied to the depletion region. We ignore the fringing capacitance, which, for the 1E12 diode, should comprise about a 15 % correction, according to the expression of Copeland (1970). At DC, the voltage across the barrier, V_b , is simply related to the total diode voltage V by

$$\begin{aligned} V &= V_b + (Z_{skin}(DC) + Z_{sp}(DC))i_b \\ &= V_b + R_s(DC) i_b \end{aligned} \quad (4-7)$$

Thus, equations 4-5 and 4-7 may be solved iteratively to yield the diode's DC I-V curve. Using standard circuit theory and a Taylor expansion of the incident far-infrared waveform, it is not difficult to show (Waksberg and Dreze, 1984) that the rectified current due to an incident far-infrared signal of impedance R_a , (the antenna impedance,) and power P_{FIR} is given by :

$$\Delta i_b = \frac{P_{FIR}}{2V_0} \frac{R_a R_b}{(R_b + R_a + R_s)^2 [1 + \omega^2 R_{eq}^2 C_b^2]} , \quad (4-8)$$

where $V_0 = \frac{\eta kT}{e}$, R_b is the differential impedance of the ideal barrier, $\frac{\partial i_b}{\partial V_b} = \frac{\eta kT}{e i_b}$, and R_{eq} is the resistance of the parallel combination of R_b and $(R_a + R_s)$,

$$R_{eq} = \frac{R_b(R_a + R_s)}{R_b + R_a + R_s} . \quad (4-9)$$

If the small-signal approximation, $\frac{\Delta i_b}{i_b} \ll 1$ is valid, so that the diode is still operating in its square-law detection regime, then the video voltage signal is simply

$$\Delta V = R_b \Delta i_b . \quad (4-10)$$

This video signal has an impedance of $R_b + R_s$ as viewed by the video circuit (bias box, oscilloscope, lock-in, etc.). Thus, it will be shorted out by the video circuit unless the latter's impedance $R_L \gg R_b$.

We assume that the video frequency is low enough that the shunt capacitance of the video circuit may be neglected, which is an excellent approximation for our case. Thus, under current bias conditions, $R_L \gg R_b + R_s$, and for small signals, the voltage responsivity is just

$$S_V = \frac{1}{2V_0} \frac{R_a R_b^2}{(R_a + R_b + R_c)^2 [1 + \omega R_{e_q}^2 C_b^2]} \quad (4 - 11)$$

The well known "RC-type" rolloff at high frequency is apparent. It is not always appreciated, however, that the appropriate resistance for evaluating the 3 db rolloff frequency is $R_{e_q} \approx R_a + R_s$, not simply R_s . For typical antenna impedances of $\sim 150 \Omega$, this reduces the 3 db frequency by an order of magnitude, to 200 - 500 GHz for the small-area diodes generally in use. Thus, our entire range of operating frequency lies well into the 6 db/octave rolloff. For our 1E4 diode at 119 μ for example, $\omega R_{e_q} C_b = 8.1$.

The results of this theoretical treatment are shown in figure 4.9, as plots of S_V at three different frequencies. In this calculation, the correct value of $R_s = \Re(Z_{sp} + Z_{skin})$ at that particular frequency, including all parasitics, was used. It is now apparent why the video signal's dependence on bias provides such an easy method for determining V_{FB} . In the regime where $R_s, R_a \ll R_b \ll R_L$ and $\omega R_{e_q} C_b \gg 1$, the video signal is simply proportional to the inverse square of the the capacitance, due to the RC rolloff, while from (4-6), the capacitance is proportional to the inverse square root of $(V_{FB} - V_b)$. Therefore, over some range of bias voltage, (or strictly speaking, logarithm of bias current, since the diode is current biased,) the voltage responsivity will decrease with increasing voltage, in straight line fashion, with intercept equal to V_{FB} . The important question is then whether this voltage range is experimentally accessible, and whether it is large enough to permit an accurate determination of V_{FB} . Since R_b varies exponentially with diode voltage, it is by no means obvious that such a useful range of voltage exists. At the low voltage end, the limit is set by how high impedance a video circuit may be constructed. At the high end, it is certainly limited by the requirement of $R_b \gg R_a$, but it is also possible that deviations from ideal thermionic transport, and from the ideal inverse square-root dependence of capacitance (due, for example, to inhomogeneous epilayer doping,) both of which are expected to become increasingly significant at higher voltages, may limit the range of applicability of equation 4-11 much more severely.

In order to settle this question and (hopefully) to determine the true flat-band potential of our diodes,

Normalized Responsivity

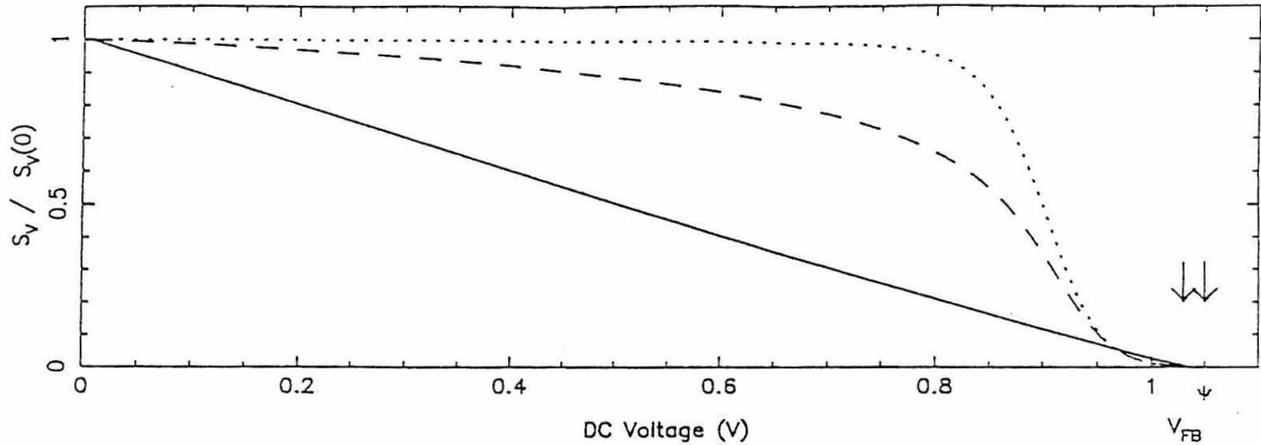


Figure 4.10 – Theoretical bias dependence of Schottky video responsivity, S_V , under current bias conditions. Dotted line is for $\frac{\omega_{FIB}}{2\pi} = 50$ GHz, dashed line for 250 GHz, and solid line for 1500 GHz. Curves have been normalized to their zero-bias values, which are 2330, 2090, and 345 V/W, respectively.

we performed an experiment. The setup is outlined in figure 4.10 . Except for the lens to focus the laser, the optics is identical to that used for the corner-cube beam measurements. Because the measurement does not involve changing the optical feedback, as adjustment of the corner-cube orientation did, there are fewer ambiguities than in the beam pattern measurement. The result is shown in at the right of figure 10. The error bars are the measured RMS dispersion, weighted by the signal strength, of several (typically 5-10) measurements at each bias point. The voltage range was covered with three values of load resistor, namely $R_L = 1M\Omega$, $100k\Omega$, and $10k\Omega$. As may be seen, the range over which the responsivity varies linearly is quite considerable – approximately 200 mV. It would certainly be possible to increase the video circuit's impedance to perhaps $100 M\Omega$, thereby extending the voltage range by another ~ 140 mV at the low end. This is not really necessary, however, since the voltage range available in the present experiment is already sufficient for quite an accurate extrapolation of the linear dependence to zero responsivity. Specifically, we find the intercept to be $1.03 \pm .02$ V, which we identify as the flat-band potential of the 1E4 diode.

Before entirely trusting the technique, however, it is necessary to ask whether the value of V_{FB} derived by it agrees with other known parameters of the diode. From the measured value of V_{FB} , we may use

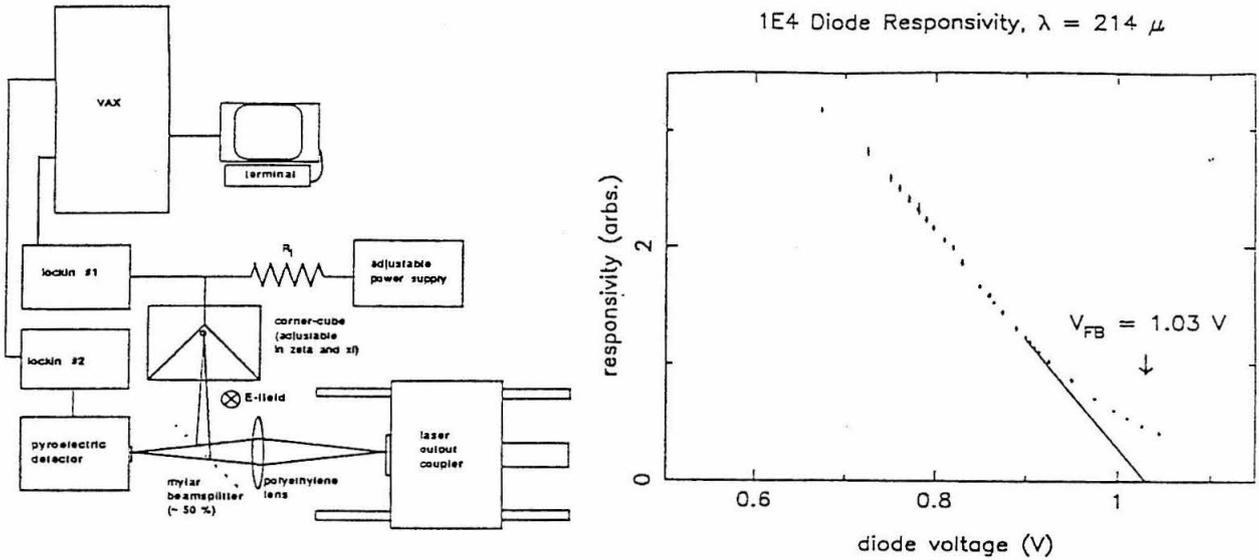


Figure 4.11 – (Left) Experimental setup for determination of V_{FB} , (right) measured voltage responsivity vs. bias voltage, under current bias conditions, showing the linear extrapolation to zero responsivity at $V_{FB} = 1.03$ V.

the known zero-bias capacitance and doping of the 1E4 diode (3.3 fF and $2 \times 10^{17} \text{ cm}^{-3}$, respectively, as privately reported by R. Mattauch's group,) to determine the effective electrical area of the diode by means of equation 4-6. This yields $A = 2.7 \times 10^{-8} \text{ cm}^2$, or a diameter of $1.8 \mu\text{m}$, to be compared with a best estimate of the geometrical diameter of $1.6 \mu\text{m}$ (T. Crowe, private communication). Furthermore, we may also use the theoretical relation between flat-band voltage and barrier height (Sze 1982, p. 16 ff.)

$$\Psi - V_{FB} \equiv V_n = \frac{kT}{e} F_{1/2}^{-1} \left(\frac{\sqrt{\pi} N}{2N_c} \right). \quad (4-12)$$

Here, $F_{1/2}^{-1}$ is the inverse Fermi function of order one-half, and N_c the density-of-states at the bottom of the conduction band. Using these determinations of A and Ψ , we may then use the Richardson equation to predict the current at any given voltage (within the exponential region of the diode's I-V curve.) Comparison with the measured current then yields a test of the self-consistency of the adopted diode parameters and of the accuracy of the equations used to derive them. Using $A = 2.7 \times 10^{-8} \text{ cm}^2$ and $\Psi = 1.03 + .02 = 1.05 \text{ V}$, the Richardson equation yields a current of $1 \mu\text{A}$ at 720 mV , versus a measured value of $1.2 \mu\text{A}$. This is to be compared with a factor of four discrepancy found by Crowe and Mattauch (1987) between the saturation current derived from fits to measured I-V curves, and that predicted from the Richardson equation and their adopted diode parameters. Discrepancies such as the latter are by no means uncommon (Kollberg

et al. 1986) when attempts are made to derive diode parameters by indirect methods, and underscore the importance of finding additional methods, such as this one, for directly determining diode parameters.

We conclude that measurement of the FIR voltage responsivity as a function of bias allows accurate determination of a Schottky diode's flat-band potential. The conditions required are that $R_s, R_a \ll R_b \ll R_L$ and $\omega R_{eq} C_b \gg 1$. Such measurements should yield a responsivity that falls linearly with voltage up to approximately ~ 100 mV below V_{FB} . An extrapolation of the linear falloff to zero responsivity then yields the flat-band potential. The method has been verified by measurement of the video responsivity at $\lambda = 214 \mu\text{m}$ for a 1E4 diode, and yields $V_{FB} = 1.03 \pm .02V$.

b. Far-infrared Reflectivity of the Schottky Diode

The efficiency of conversion of the incident FIR beam into a traveling current wave is described by a driving point impedance and Ohmic loss factor, R_{dr} and ϵ , for the antenna, and a normalized overlap integral of the antenna beam pattern, $\vec{E}_{ant}(\theta, \phi)$, and the incident beam pattern, $\vec{E}_{inc}(\theta, \phi)$.

$$\eta \epsilon P_{inc} = \frac{R_{dr}}{2} i_{inc}^2$$

$$\eta \equiv \frac{\left| \int_{4\pi} \vec{E}_{ant} \cdot \vec{E}_{inc} d\Omega \right|^2}{\left[\int_{4\pi} |E_{inc}|^2 d\Omega \right] \left[\int_{4\pi} |E_{ant}|^2 d\Omega \right]} \quad (4 - 13)$$

Frequently, when the incident beam is much narrower than the antenna beam, the overlap integral is divided into the product of a "source coupling efficiency" and a "main beam efficiency". As discussed earlier, the main beam efficiency is simply the fraction of power in the antenna beam which falls within the first minimum surrounding the peak. The source coupling efficiency is defined identically to η except that the integrations extend only over the main beam. In practice, it is most convenient to approximate the corner-cube's main beam and the incident far-infrared beam as Gaussian. In this case, the source coupling efficiency, η_s , may be expressed (Goldsmith, 1982, Kogelnick 1964) in terms of the opening angles of the two beams in the far field, θ_1 and θ_2 , and the defocussing, d , (i.e. the separation between the two beams' waists,) as

$$\eta_s = \frac{4}{\left(\frac{\theta_1}{\theta_2} + \frac{\theta_2}{\theta_1} \right)^2 + \left(\frac{\pi d}{\lambda} \theta_1 \theta_2 \right)^2} \quad (4 - 14)$$

As will be discussed in §4.3, estimates for η_{mb} and η_s may be made from the optical design of the polarizing Michelson interferometer and from our measured beam patterns. Comparison of theoretical and experimental values of the absolute video responsivity (in V/W) may be used to further constrain these efficiencies. For purposes of analyzing the diode's performance, however, it is sufficient to treat them as free parameters.

The next step in the sideband generation process is the transmission of the induced current wave down the whisker to the diode. The capacitance and inductance per unit length of the whisker antenna, and therefore its characteristic impedance when viewed as transmission line, can be calculated in a straightforward (though algebraically messy) way by the method of images. The result is

$$R_{ant} = 60 \Omega \left\{ \ln \left(\frac{s}{r_0} \right) + \ln \left(1 + \frac{(s - r_0)^2}{s^2} \right) + \ln \left(\frac{s}{2s - r_0} \right) \right\} \quad (4 - 15)$$

$$\approx 60 \Omega \ln \left(\frac{s}{r_0} \right) \quad \text{for } s \gg r_0$$

where r_0 is the radius of the whisker and s the distance between it and the dihedral reflector. For our corner-cube, $s = 1.2\lambda$, as per the conventional recipe (Krautle, Sauter, and Schultz 1977), and r_0 is $12.7 \mu\text{m}$, yielding an antenna impedance of approximately 145Ω .

Before moving on to the reflection of the current wave off the diode, we note that treatment of the whisker as a lossless transmission line for purposes of calculating its impedance is only an approximation. The correct antenna impedance to use for this problem would result from a calculation where the real part on R_{ant} were derived from an angular integral of the far-field distribution and the imaginary part from a volume integral of the near-field. However, we expect that the (real) R_{ant} derived from the transmission line calculation will be nearly correct since the total (radiation + Ohmic) loss per radian at the diode end of the whisker is small. (Most of the radiation loss occurs at the at the 90° bend at the opposite end.) This approximation is better for longer whiskers. In any case, however, it turns out that the dependence of reflectivity modulation on antenna impedance is very weak.

We are now ready to treat the third, and most complicated, sub-problem in the sideband-generation process, namely the issue of diode optimization. We assume first of all that a small-signal analysis is valid,

so that the FIR reflectivity may be written

$$i_r e^{i(\omega_{FIR}t - kz)} = \Gamma i_i e^{i(\omega_{FIR}t + kz)}$$

$$\Gamma = \frac{Z_d - R_a}{Z_d + R_a} \quad (4 - 16)$$

As illustrated in the equivalent circuit of figure 4.11, Z_d is the total FIR diode impedance including all parasitics, and R_a is the impedance of the whisker antenna, viewed as a transmission line.

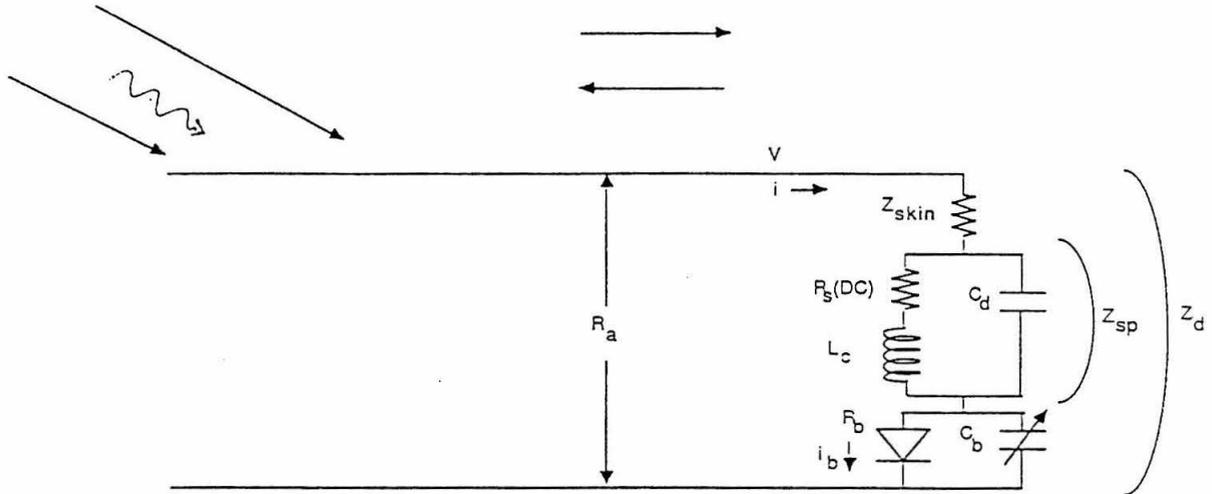


Figure 4.12 – Equivalent circuit representation of the FIR modulation process and of the high-frequency Schottky diode parasitics. DC bias and RF modulation drive are omitted.

“Small” signal in this case means that the fractional change in diode impedance is small over voltage excursions equal to the FIR wave amplitude. Under normal operating conditions, this is satisfied if and only if the diode’s video response is still “square-law”, i.e. if $\Delta V_{video} < \frac{2\eta kT}{e}$. The small-signal condition is tantamount to ignoring all mixing products ω_{mn} with $m \neq 1$, (where m indexes the FIR harmonic and n the RF harmonic). At $\lambda < 200 \mu\text{m}$ and an incident laser power $< 1 \text{ mW}$ the small-signal condition is (unfortunately) fairly well obeyed. For example, at $\nu = 2.52 \text{ THz}$, the typical video signal on our $1\text{E}12$ diode is approximately 20 mV , implying $\frac{e\Delta V}{2\eta kT} \sim .4$. Furthermore, in all sideband generation experiments to date, sideband power is found to vary linearly with incident FIR power, (Farhoomand, *et al.* 1985), indicating in a somewhat stronger way that a small-signal analysis is appropriate.

Near the flat-band voltage rather unusual behavior is possible. As V_{FB} is approached, the barrier capacitance formally approaches infinity, and infinitesimally small far-infrared waveforms are capable of producing reflected waveforms that are not just linearly related to the incident one by a factor of the reflectivity Γ , but are distorted as well. Physically, this is due to the fact that the barrier capacitance can hold only a finite amount of charge, so that the higher the bias voltage, and therefore the DC stored charge on the capacitor, the smaller a FIR current is required to saturate the capacitor and produce a “non-linear reflectivity”. The physical basis of the diode’s behavior near the flat-band voltage is discussed more fully by Crowe and Mattauch (1986) and by Torrey and Whitmer (1946).

Referring to the equivalent circuit of figure 4-11, the total diode impedance Z_d above and below the flat-band voltage is given by Crowe and Mattauch (1986 and 1987) :

$$\begin{aligned} Z_d &= Z_{skin} + Z_{sp} + (R_b \parallel \frac{1}{i\omega C_b}) \quad V < V_{FB} \\ &= Z_{skin} + Z_{sp} \quad V > V_{FB} \end{aligned} \quad (4 - 17)$$

Above flat-band, the depletion region vanishes and carriers are free to stream directly from the substrate to the anode with no hindrance but the series resistance. Although this model, summarized by equation (4-17), is expected to be a good approximation to reality well above and well below the flat-band voltage, it is obvious that the transition between the two regimes cannot really be perfectly abrupt. Equation (4-17) implies that there exists a discontinuity in the slope of the diode i-v curve at the flat-band voltage, whereas real diodes never exhibit such a discontinuity. In fact, both the finite temperature of the electron gas and the effect of electron tunneling, (and perhaps other effects as well,) may be expected to smear out the transition between the thermionic emission and free-streaming regimes.

Our next approximation is that the RF currents may be treated as DC. In other words, we assume

$$i = i_b \quad \text{at} \quad \omega = \omega_{RF} . \quad (4 - 18)$$

This approximation effectively eliminates one independent variable from the problem. Instead of the FIR reflectivity being a function of DC bias and RF power independently, equation (18) reduces it to being only a function of the sum voltage, i.e.

$$\Gamma_{FIR}(V(DC), P_{RF}) \rightarrow \Gamma_{FIR}(V = V(DC) + V(RF)) . \quad (4 - 19)$$

Relaxing this approximation makes the circuit problem considerably more complicated. It is easy to see that the first step in any more general treatment, namely determination of the barrier voltage as a function of time, requires solution of a first-order, but highly non-linear differential equation. Intuitively, however, we expect that the qualitative effect of $\omega_{RF} \neq 0$ will be that the barrier capacitance will partially shunt the RF current. Therefore, more RF power would be required to achieve the same reflectivity modulation. So long as sufficient RF power is available (which is the case in practice), this will only change the tuning.

Just as in our treatment of video detection, equations (4-5) and (4-7) may then be solved iteratively to determine the DC (i.e. DC + RF) i-v curve once the DC series resistance, $Z_{skin}(DC) + Z_{sp}(DC)$, which is directly measurable, is known. The standard treatment of carrier inertia, dielectric relaxation, and skin effect was developed by Champlin and Eisenstein (1978). Following them, we write

$$Z_{skin} = \mu_0 \left(\frac{i\omega}{2\pi k} \right) \ln \left(\frac{b}{a} \right) \quad (4-20a)$$

$$Z_{sp} = \mu_0 \left(\frac{i\omega}{2\pi a k^2} \right) \arctan \left(\frac{b}{a} \right) \quad (4-20b)$$

Here, Z_{sp} is the complex spreading resistance, $\mu_0 = \frac{377\Omega}{c}$ is the permeability of free space, b is the radius of the diode substrate, and k is the complex propagation constant in the epilayer :

$$k^2 = -\frac{\omega^2}{c^2} \epsilon + i\mu_0 \omega \sigma \quad (4-21)$$

In turn, σ is the frequency dependent conductivity of the epilayer :

$$\sigma = \frac{Ne\mu}{1 + i\omega\tau_s} \quad (4-22)$$

$$\mu = \frac{e\tau_s}{m^*}$$

Since k^2 has positive imaginary part, k lies in the first quadrant of the complex plane. Therefore Z_{skin} has positive imaginary part (i.e. it's inductive). The real part of k^2 is positive at low frequency and negative at high frequency, so Z_{sp} is inductive at low frequency and capacitive at high frequency, as it must be for a parallel LC circuit. In these equations, μ is the mobility of the epilayer and τ_s the scattering time in a Drude-type model of the mobility. The effective mass of the carriers ("carrier inertia") is denoted by m^* . It is independent of crystallographic direction since the constant energy surfaces in n-type GaAs are spherical.

Equations (4-20) - (4-22) represent a rather simple-minded model, neglecting for example, any dielectric dispersion (i.e. ϵ is assumed to be frequency-independent over the entire range of interest.) Nonetheless, we expect the most important effects of the parasitics to be evident in our results. Now, μ and τ_s are not directly observable quantities for our diodes. However, for numerical calculations, we have used the fact that the low-frequency limit of equation (4-20) is :

$$\left. \begin{array}{l} Z_{sp} \rightarrow \frac{1}{4a\sigma_{DC}} \\ Z_{skin} \rightarrow 0 \end{array} \right\} \omega \rightarrow 0 \quad (4-23)$$

Thus, from the measured value of $R_s(DC)$ and an estimate of the diode radius we calculate σ_{DC} , μ , $\sigma(\omega)$, and therefore k . From there, we have used the Champlin and Eisenstein model (equations (4-20a and b)), to calculate Z_{sp} and Z_{skin} , which are then substituted into the expressions for the total diode impedance and FIR reflectivity.

A modification to the Champlin and Eisenstein model has recently been proposed by Van Roos and Wang (1986). It is slightly more sophisticated in that the approximation $ka \ll 1$, which Champlin and Eisenstein assumed in deriving equation (4-20) is dispensed with. Physically, the additional effect which is included is the phase delay in the FIR waves that are reflected off the circumference of the depletion region when the dielectric wavelength becomes comparable to the diode radius. We have not added this modification to the Champlin and Eisenstein theory into our calculations. It may be shown that the expression Van Roos and Wang derive to replace equation (4-20b), differs from it by a factor of order $|ka|^2$. For our baseline 1E12 diode at 2.52 THz, $|ka| = .18$, so the error introduced into our calculations is expected to be negligible.

In order to derive the epilayer conductivity from $R_s(DC)$, it is necessary to know the diode radius. For the 1E12 diode, we have adopted the value $r = .67 \mu\text{m}$, which was determined from equation 4-6, using the zero-bias capacitance reported by those who supplied us with the diodes, and the flat-band voltage we measured for the 1E4 diodes. Since V_{FB} is a property of the epilayer only, and not of the geometry, it should be the same for the 1E4 and 1E12 diodes, which differ only in area. As with the 1E4 diodes, the value of r derived by this method agrees very well with the best estimate obtained from direct scanning

electron micrographs (SEM) of the diodes (T. Crowe, private communication). It also yields excellent agreement between the measured saturation current and the value predicted by the Richardson equation, A^*AT^2 . In short, we have considerable confidence in the values of diode parameters that we have adopted as inputs to our model. They are summarized in table 1.

c. Numerical Results for $\Gamma(V)$

Assuming that carrier inertia (modeled as the inductance L_c in the equivalent circuit of figure 4-12), dielectric relaxation (C_d), and skin effect impedance (Z_{skin}) may be neglected – which in fact is only valid at frequencies well below 1 THz – we may now examine the modulation of FIR reflectivity due to the variation in barrier impedance alone. This is shown in figure 4.13 as a locus of points on a Smith chart normalized to R_a as the bias ($V_{DC} + V_{RF}$) is varied from 0 to V_{FB} .

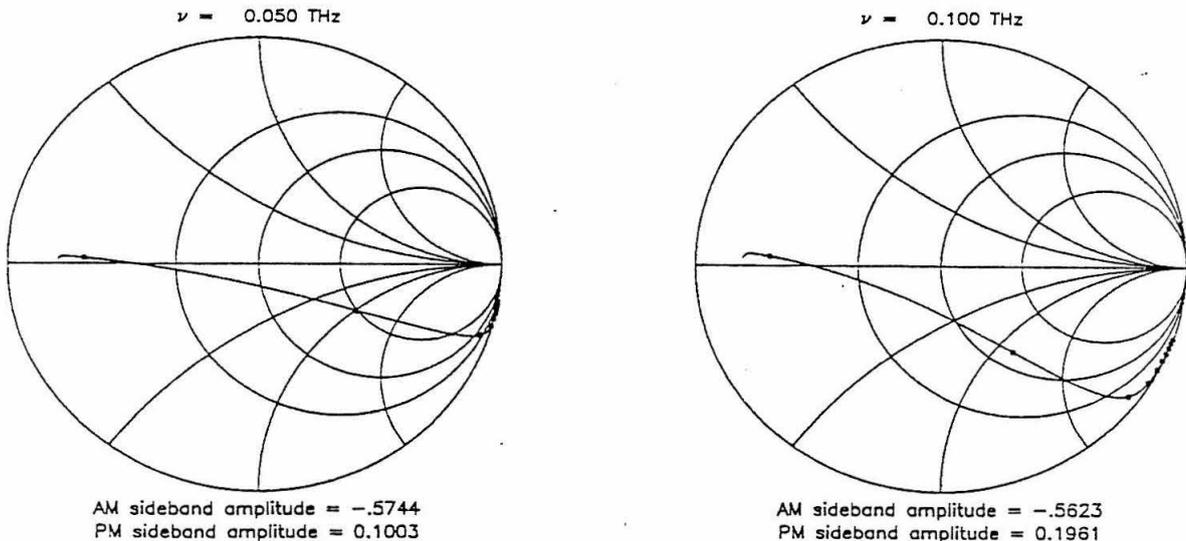


Figure 4.13. Reflectivity modulation in the low-frequency limit, where all parasitics except the barrier capacitance are negligible. For this computation, all parameters are those of our baseline 1E12 diode and corner-cube, but the carrier frequency $\frac{\omega_{FIR}}{2\pi}$ is 100 GHz and 50 GHz, as indicated.

Above V_{FB} , the reflectivity is constant because the I-V curve is, by hypothesis, Ohmic. At low bias, the barrier conductance is completely negligible. The barrier susceptance is also small, but only to the extent that the low-frequency limit is valid. As the bias is raised, the susceptance increases in accordance

with the square-root dependence of equation (4-6). At some voltage slightly (i.e. a few times $\frac{\eta kT}{e}$) below flat-band, the exponential increase in conductance finally overtakes the square-root increase in susceptance, and the reflectivity moves very rapidly along a curve of approximately constant susceptance. This occurs at $R_b \approx (\omega_{FIR} C_b)^{-1}$, which, we note in passing, is the conventional operating point for Schottky diode multipliers (Schneider 1982). The limiting, flat-band value of the reflectivity is determined by the (voltage-independent) series resistance. To the extent that $\omega_{FIR} = 0$, the series resistance and the limiting value of reflectivity are purely real, but as ν_{FIR} becomes non-zero, R_s becomes inductive. It is clear that in this low-frequency regime, both large amplitude ($\Delta|\Gamma|^2 \sim 1$) and large phase modulation ($\Delta \arg(\Gamma) \sim 1$ radian) are possible. As we shall now show, however, both are drastically reduced by the inclusion of the parasitic elements L_c , C_d , and Z_{skin} .

Figure 4.14 displays the results of the reflectivity versus bias calculation, for the baseline diode parameters listed in table 1 and a FIR frequency of 2.52 THz. Three features are immediately obvious. Firstly, the amplitude and phase modulation depths are much lower than they were at low frequency, and, in particular, the dip in $|\Gamma|$ at $R_b \sim (\omega C_b)^{-1}$ has nearly disappeared.

This is hardly surprising, since, as the frequency is raised, the minimum barrier resistance is fixed at $R_b(V_{FB})$ but the barrier susceptance continually declines; therefore, the condition of $R_b \sim (\omega C_b)^{-1}$ becomes farther and farther from being realized, and the available amplitude modulation continuously decreases. Secondly, the phase modulation dominates the amplitude modulation by a large factor – about 11 for these parameters. Note that the sideband strengths listed at the bottom of the figure are in amplitude, so in terms of power, FM sidebands would dominate AM sidebands by a factor of over 100.

Thirdly, the overall diode impedance is *inductive*, rather than capacitive, as it was at low frequencies, and as one might naively expect. This fact alone shows that a new physical mechanism is becoming important at high frequencies.

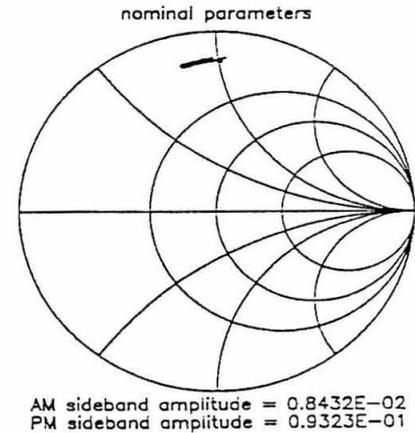


Figure 4.13 – Reflectivity modulation for the baseline 1E12 diode at 2.52 THz (119 μm).

Table 1

Baseline Parameters (assumed)		
Parameter	Value	Source
Dielectric constant	$\epsilon = 10.8$	Sze
Effective mass	$m^* = .068$	Sze
Richardson constant	$A^* = 8.6 \text{ A cm}^{-2} \text{ K}^{-2}$	Sze
Antenna impedance	$R_{ant} = 145 \Omega$ (real)	equation (4)
Temperature	$T = 300 \text{ K}$	room temperature
Baseline Parameters (measured)		
Flat-band voltage	$V_{FB} = 1.03 \text{ V}$	our measurement (see fig. 4.20)
Epilayer doping	$N = 2 \times 10^{17} \text{ cm}^{-3}$	Mattauch (private communication)
Zero-bias capacitance	$C_i(V=0) = 1.8 \text{ fF}$	Mattauch (private communication)
DC series resistance	$R_s(DC) = 9.3 \Omega$	Mattauch (priv. comm.) + our measurement
Ideality factor	$\eta = 1.20$	Mattauch (priv. comm.) + our measurement
Substrate radius	$b = 63 \mu\text{m}$	our measurement
Main beam efficiency	$\eta_{beam}(214 \mu\text{m}) = .44$ $\eta_{beam}(119 \mu\text{m}) \sim .3$	figure 4.6 estimate (see text)
Baseline Parameters (derived)		
Diode radius	$a = .67 \mu\text{m}$	$V_{FB}, C_i(V=0), + \text{eqn. 10}$
Barrier height	$\Psi = 1.05 \text{ V}$	$a + (740 \text{ mV @ } 1 \mu\text{A}) + \text{eqn. 6}$ or $N + V_{FB} + \text{eqn. 16}$
Mobility	$\mu = 1.25 \times 10^4 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$	$a + R_s(DC) + \text{eqn. 14}$
Scattering frequency	$\omega_s = \frac{eH}{m^*} = 1.8 \text{ THz}$	mobility
Dielectric relaxation frequency	$\omega_d = \frac{q}{4\pi\sigma} = 67 \text{ THz}$	mobility

$$\frac{e}{\mu m^*}$$

$$\frac{4\pi\sigma}{\epsilon}$$

Figure 4.15 shows the reflectivity modulation as a function of frequency in detail. At any given bias, the phase of the reflectivity cycles through 3π radians from DC to infinite frequency, indicating the existence of two resonances. The frequency of the first one (as indicated by a crossing of the real reflectivity axis) lies somewhere between .5 and 1.5 THz, depending on the bias. The frequency of the second lies just below 5 THz, and is almost completely independent of bias, i.e. the trace of reflectivity modulation becomes compressed to nearly a point at the second resonance. Both from this frequency behaviour and from the equivalent circuit of figure 4.12, it is clear that the new physical mechanism that becomes important at high frequencies is the carrier inertia, represented in figure 4.12 as the inductance L_c . The first resonance occurs due to the *series* combination of this effective inductance and the barrier capacitance. In the formulation of Champlin and Eisenstein, the inductance is given by $L_c = \frac{R_s(DC)}{\omega_s}$, ($= 4.5$ pH for the baseline parameters,) where $\omega_s = \frac{e}{m^* \mu}$ is the scattering frequency in the epilayer, i.e. the inverse of the momentum relaxation time. The frequency of the first resonance may then be written as

$$\omega_1^2 = \frac{4}{\pi m^* (e \epsilon \epsilon_0)^{1/2}} (V_{FB} - V)^{1/2} \frac{N^{1/2}}{a} \quad (4 - 24)$$

$$\frac{\omega_1}{2\pi} = 1.84 \text{ THz} \left(\frac{V_{FB} - V}{1.03 \text{ V}} \right)^{1/2} .$$

Thus, most of the reflectivity modulation that is occurring at frequencies of, say .5 to 3 THz, is due to the modulation of the barrier capacitance changing the frequency of this resonance. The fact that the series resistance in the resonant circuit, $R_s(DC)$, is so small compared to the antenna impedance guarantees that tuning through the resonance will produce nearly all phase modulation and hardly any amplitude modulation – the trace must be near the circumference of the Smith chart.

The second resonance is nothing but the well known plasma resonance in the undepleted epilayer, whose frequency is given by

$$\omega_p^2 = \frac{4\pi N e^2}{\epsilon m^*} \quad (4 - 25)$$

$$= 4.7 \text{ THz}$$

It is caused by the *parallel* combination of the effective inductance due to carrier inertia and the effective capacitance due to dielectric relaxation. Physically, it is caused by the interaction of the carrier inertia

with the electrostatic force due to the induced polarization charge in the dielectric. Because it is a parallel resonance in which neither of the voltage dependent elements participates, it is more detrimental to the total (amplitude + phase) reflectivity modulation than the first one. It is only a local minimum in reflectivity modulation, however. At higher frequencies than the plasma frequency, the modulation increases, peaks, and then monotonically falls to zero as the FIR frequency approaches infinity, and the total diode impedance approaches the skin effect resistance. The peak modulation possible above the plasma frequency is generally much lower than that possible well below the plasma frequency, however. This is clearly apparent in plots of the reflectivity modulation as a function of epilayer doping, at fixed frequency, shown in figure 4.16.

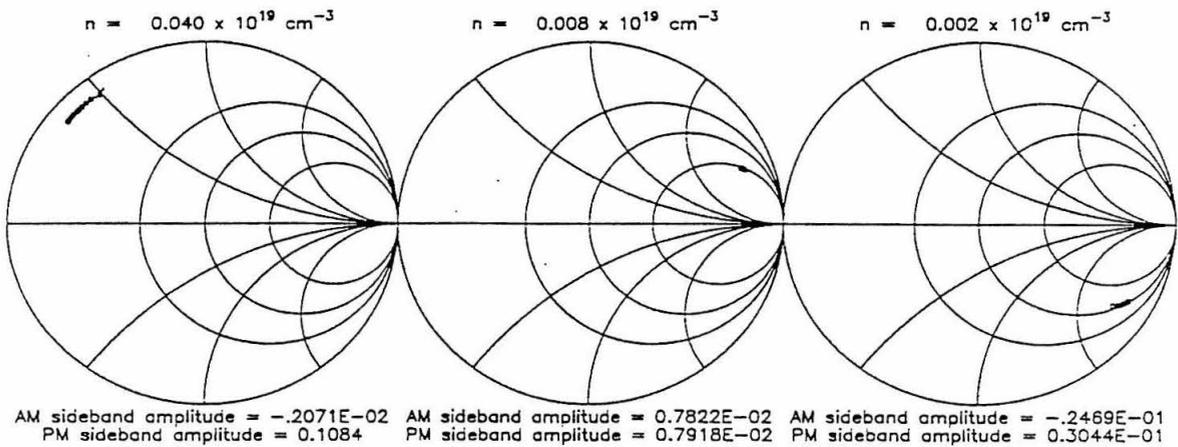


Figure 4.16 Reflectivity modulation for epilayer doping concentrations of 4×10^{17} , near where the conversion efficiency peaks, 8×10^{16} , where the plasma frequency passes through the operating frequency, and $2.5 \times 10^{16} \text{ cm}^{-3}$, well into the above-plasma frequency regime.

Starting from the baseline values listed in table 1, we have varied each of the diode parameters individually, while holding all the remaining parameters fixed. To quantify the results, we have used the method described in the next section to compute the sideband conversion efficiency due to amplitude and phase modulation of the reflectivity. Varying the antenna impedance from 50Ω to 300Ω changed the overall phase of the reflectivity from approximately 135° to about 45° , but the reflectivity modulation over

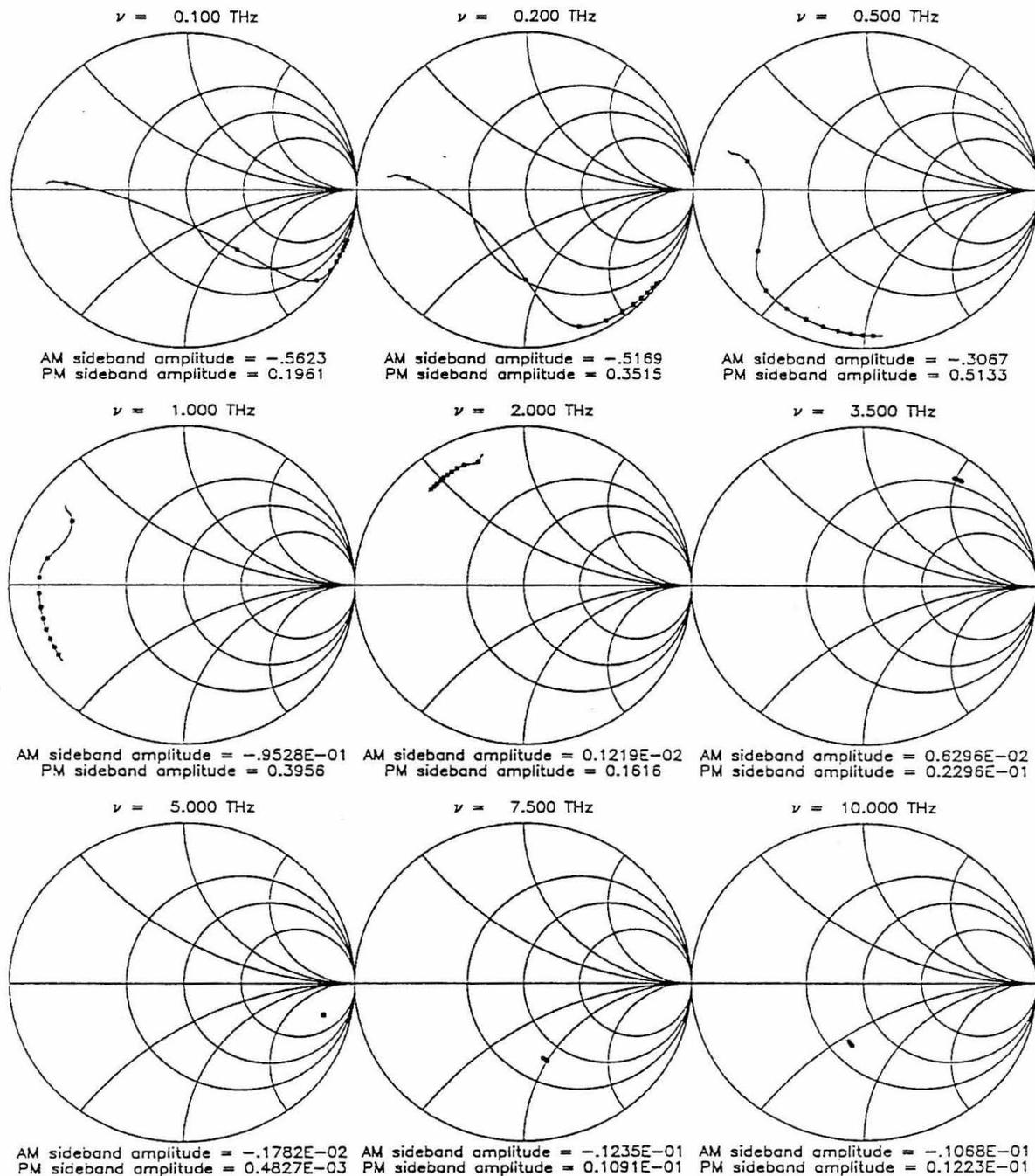


Figure 4.15 Reflectivity modulation as a function of FIR frequency. Each curve traces out the range from $V = 0$ to $V = V_{FB}$, with ticks every 100 mV. The sideband amplitudes α and β , have been computed using equation (4-28). The RF voltage was assumed to have a peak-to-peak amplitude of 1 V, and to be centered at $V_{DC} = 5$ V.

the bias range from 0 to V_{FB} varied very little. More specifically, there is a broad maximum in the total sideband amplitude at an impedance of $\sim 135 \Omega$. At 50Ω , it is reduced by about 30 % and at 300Ω by about 25 % from its peak value of .094 . Likewise, the dependence of reflectivity modulation on substrate size is also very weak. As the substrate radius is reduced, the modulation depth monotonically increases. An order of magnitude reduction in radius, from 1.5 mm to .15 mm (which is really a smaller chip than could conveniently be handled anyway,) only yields a 20 % improvement, though.

The dependence of reflectivity on Schottky barrier height is also relatively weak. Because the barrier height and flat-band voltage are coupled (via equation (4-12)), for this case we varied Ψ and V_{FB} by equal amounts around their baseline values. Our calculation shows a broad and gentle peak in the total sideband amplitude, (dominated, as usual, by the FM sidebands), at a barrier height of $\Psi = .85 \text{ V}$. The total sideband amplitude is only 20 % higher at the peak than at the actual barrier height of 1.05 V, however. Furthermore, the Schottky barrier height is not a parameter that can in practice be varied easily or over a very wide range. To zeroth order, it is simply the difference between the work function of the metal and the electronegativity of the semiconductor, i.e. about 1 V for Pd on GaAs. There are small deviations about this value, but they are never greater than, say, 20 %, and they depend on the details of the epilayer growth (defect density, doping uniformity, etc.) in an incompletely understood way.

Reflectivity modulation curves at 300 K and at two elevated temperatures are shown in figure 4.17. The essential result is that as the temperature is raised, the dip in $|\Gamma|$ which was prominent in the low-frequency reflectivity where $R_b \approx \frac{1}{\omega C_b}$, is partly restored. A rise in temperature raises the DC current and reduces R_b while holding all the other diode properties fixed. Thus, the condition of equal barrier resistance and reactance, which had become progressively more poorly satisfied at higher frequencies, can again be realized. The dependence of R_b is exponential, with a scale temperature of $T_0(V) = \frac{e(V_b - \Psi)}{\eta k}$. At the maximum voltage of V_{FB} , this scale temperature is about 200 K, but at a “typical” voltage level in the RF cycle, the scale temperature is huge – several thousand K. Therefore, the gain that might be realizable in practice is relatively modest. We find an increase in total sideband amplitude of $\sim 40 \%$ when the temperature is raised to 1000 (!) K.

These calculations of temperature dependence (as well as those of doping dependence) were done

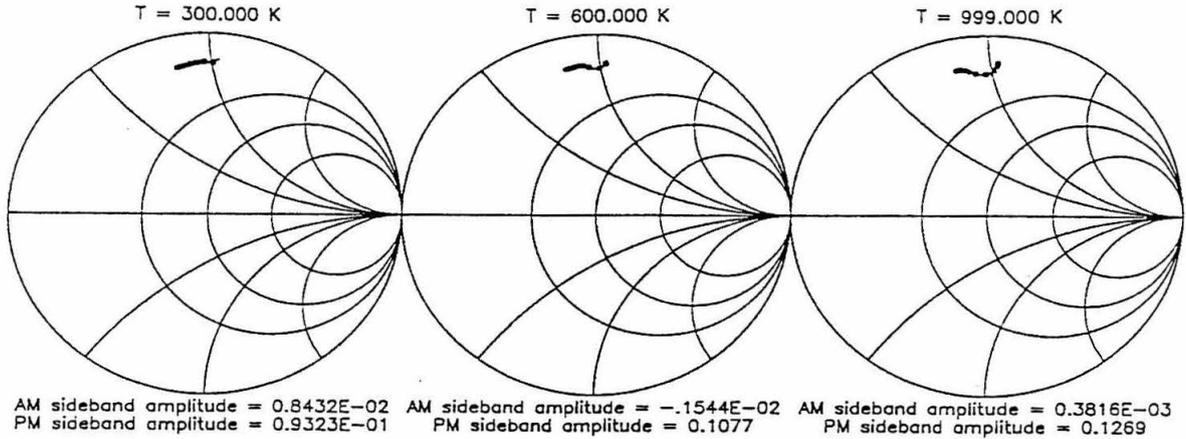


Figure 4.17 Reflectivity modulation at 300 K, 600 K, and 1000 K.

assuming the mobility of the epilayer to be constant. In fact, both acoustic phonon scattering and optical phonon scattering contribute significantly to the room-temperature mobility of GaAs; experimentally, the mobility is found to vary approximately as T^{-1} (Sze p. 28, ff.). The main effect of the temperature dependence, in terms of the equivalent circuit representation of the diode, is a linear increase in $R_s(DC)$ with temperature. The way in which $R_s(DC)$ enters into the reflectivity modulation in the .5 - 5 THz range is to set the real part of the total diode impedance near the lower resonance ω_1 . Reducing the mobility will increase the impedance at resonance (i.e. decrease the Q,) and therefore move reflectivity traces such as the 1 THz plot in figure 4.15 closer to the center of the Smith chart. The highest temperature at which one might conceivably operate the Schottky diode is only a factor of 2 - 3 greater than ambient, so $R_s(DC)$ will always be considerably less than R_{ant} , and the trace will not move far toward the center of the Smith chart. Therefore, we expect the effect of the lowered mobility to be very small.

The only diode parameters, therefore, which present much scope for optimization are the epilayer doping and the diode radius. The variation of the reflectivity over the full a-N plane will be discussed in the next section. Here we only describe the dominant effects on reflectivity in one-dimensional scans through the baseline parameter values. Firstly, there is, not surprisingly, a monotonic increase in both amplitude and phase modulation depth as the diode radius is reduced. Also, the balance between amplitude and phase modulation become more even as the diode radius is reduced. The lowest radius for which the

calculation was performed was $a = .2 \mu\text{m}$, our estimate of the smallest diode that could in practice be fabricated and contacted. In this case, the AM sideband amplitude is approximately thirty times higher, and the PM sideband amplitude approximately six times higher than for the baseline $a = .67 \mu\text{m}$. The total conversion efficiency is .38 (-4.2 db), which is virtually no worse than the low-frequency value.

The doping dependence was a little more complicated. Starting from the highest dopings, the modulation depth gradually increases as the doping is lowered due to the decreasing barrier capacitance. At a doping of approximately $4 \times 10^{17} \text{cm}^{-3}$, the total (mainly FM) sideband amplitude reaches a peak, and then falls extremely steeply as the plasma frequency passes through the operating frequency. At lower dopings still, the modulation increases as the diode begins to operate in the regime above plasma resonance. At the lowest doping for which we have performed the calculation, $N = 2.5 \times 10^{16} \text{cm}^{-3}$, the total sideband amplitude has risen back to about .4 times its baseline value (i.e. conversion efficiency 7.5 db lower.)

In this calculation of reflectivity at varying doping levels, we have ignored equation (4-12) and assumed the Schottky barrier height and flat-band voltages to be constant. In fact, this ought to be a fairly good approximation. When $\Psi - V_{FB} < 0$, the inverse Fermi function in equation 4-12 is approximately a logarithm, and the deviation of V_{FB} from constancy is quite slow. Furthermore, V_{FB} enters into the calculation primarily through the square-root dependence of the barrier capacitance on it (equation 4-6). Thus, we expect the variation of reflectivity on doping due to the change in V_{FB} to be very weak indeed, and to be completely dominated by a) the explicit dependence of barrier capacitance on doping, and b) the dependence of the plasma frequency on doping.

c. Optimization for Sideband Generation

We now consider the question of how to go from a calculated reflectivity $\Gamma(V)$ to the AM and FM sideband amplitudes and the single-sideband conversion efficiency, ζ . We have already used the results of this section to derive the numerical values listed at the bottom of our Smith charts of diode reflectivity. With the approximation that the RF frequency may be treated as DC vis á vis the diode, we formally write the reflected current wave in terms of the incident current wave as :

$$i_r(t) = i_i \Gamma(V_{DC} + V_{RF} e^{i\omega_{RF}t}) e^{i\omega_{PIR}t} \quad (4 - 26)$$

$$\bar{i}_r(\omega_{\pm s}) = i_i \bar{\Gamma}(\pm\omega_{RF}),$$

where

$$\omega_{\pm s} = \omega_{FIR} \pm \omega_{RF}$$

are the first upper and lower sideband frequencies and a tilde denotes a Fourier transform. The Fourier component of the reflectivity at the RF frequency, $\tilde{\Gamma}(\omega_{RF})$, is the basic figure of merit for the diode as a sideband generator. It is given by

$$\tilde{\Gamma}(\omega_{RF}) = \frac{1}{\pi} \int_{-\pi}^{\pi} \Gamma(V_{DC} + V_{RF} \cos x) \cos x \, dx. \quad (4-27)$$

This equation also describes the tuning of conversion efficiency with V_{DC} and V_{RF} . If the DC reflectivity, $\tilde{\Gamma}(\omega = 0) = \Gamma_{DC}$, were real, then the real and imaginary parts of $\tilde{\Gamma}(\omega_{RF})$ would be the amplitude and phase modulation depths respectively. Since Γ_{DC} is complex, the amplitude and phase modulation depths, α and β , are given by a rotation through $\arg(\Gamma_{DC})$ in the complex plane :

$$\alpha = \frac{\Re(\tilde{\Gamma}(\omega_{RF}))\Re(\Gamma_{DC}) + \Im(\tilde{\Gamma}(\omega_{RF}))\Im(\Gamma_{DC})}{|\Gamma_{DC}|} \quad (4-28)$$

$$\beta = \frac{\Re(\tilde{\Gamma}(\omega_{RF}))\Im(\Gamma_{DC}) - \Im(\tilde{\Gamma}(\omega_{RF}))\Re(\Gamma_{DC})}{|\Gamma_{DC}|}$$

The total single-sideband conversion efficiency ζ is then

$$\zeta = (\eta\epsilon)^2 |\tilde{\Gamma}(\omega_{RF})|^2 = (\eta\epsilon)^2 (\alpha^2 + \beta^2). \quad (4-29)$$

Applying equations (4-27) and (4-28) to our baseline calculation of $\Gamma(V)$ (figure 4.14), we find $\alpha = .0084$, $\beta = .093$, implying a conversion efficiency due to the diode alone of $(\alpha^2 + \beta^2) = -20.5$ db. We note, however, that the baseline 1E12 diode is in a regime where α and β are extremely sensitive to doping (i.e. it's on the wing of the plasma resonance). If the actual doping level were 1×10^{17} instead of 2×10^{17} , then the predicted diode losses would be -31 db.

It is clear how this computational procedure may be generalized to obtain conversion efficiencies for higher-order sidebands ($m \geq 2$). All that is necessary is to replace the $\cos(x)$ term in the integrand of equation (4-27) with $\cos(mx)$. However, an implicit assumption in our analysis has been that the RF waveform is sinusoidal, and this can lead to a serious underestimate of the conversion efficiency for the

higher-order sidebands. Non-sinusoidal variation of $\Gamma(t)$ can be produced either by non-sinusoidal (i.e. “spikey”) variation of $V(t)$, or by a non-linear dependence of Γ on V . As may be seen from figure 4.15, at the high FIR frequencies of interest ($\nu \geq 1$ THz), the non-linearity of $\Gamma(V)$ is not all that strong. Therefore, if the RF waveform is very spiky, the spikiness will probably be the main contributor to the higher-order sidebands, and the simple replacement of $\cos x$ with $\cos mx$ will seriously underestimate their power. In order to produce a highly spiked waveform, the diode’s barrier impedance R_b at the peak of the RF cycle must be much less than the RF source impedance (which is always 50Ω in our experiment). This condition can be achieved, but only marginally ($\frac{R_b}{50\Omega} \approx .4$), before diode burnout occurs.

Under the assumptions we have described, α and β have been calculated over a rectangular grid in the diode radius - epilayer doping plane. The calculations were performed for what was considered to be the full range of realistically accessible values, viz. $.2 < a < 1.5\mu\text{m}$, $2.5 \times 10^{16} < N < 2.5 \times 10^{18} \text{cm}^{-3}$. The results are presented as contour plots in figure 4.18, for frequencies of .5, 1.0, 2.0, and 4.0 THz.

The plots clearly show that diode optimization at high frequency is completely different than at low frequency, both in terms of diode radius and doping level. Examining the higher frequency plots first, we see that two basic features dominate the behavior of the conversion efficiency. The first is the precipitous drop in efficiency due to plasma resonance, which occurs, roughly independently of diode radius, at the doping level given by equation 4-25. The depth of the plasma resonance is not apparent in figure 4.18 because we have arbitrarily stopped contouring levels more than 30 db below the peak. Had this not been done, the areas of plasma resonance in the lower two figures would have been completely black with contour lines. In separate computer runs, we have determined that the conversion efficiency drops to at least -55 db at the plasma resonance. The other prominent feature of the high frequency plots is the slope of the contour lines at dopings well above the critical doping to produce plasma resonance. This region of the plot represents the usual operating regime, where the frequency is well below the plasma frequency. In this region, the efficiency contours are approximately lines of constant capacitance. This may be seen from their slope, which indicates $Na^4 \approx \text{constant}$ along a contour. Since $C_d \propto N^{1/2}a^2$ (see equation 4-6) this represents constant capacitance.

The two lower frequency plots show an additional feature. Each has a peak in the conversion efficiency

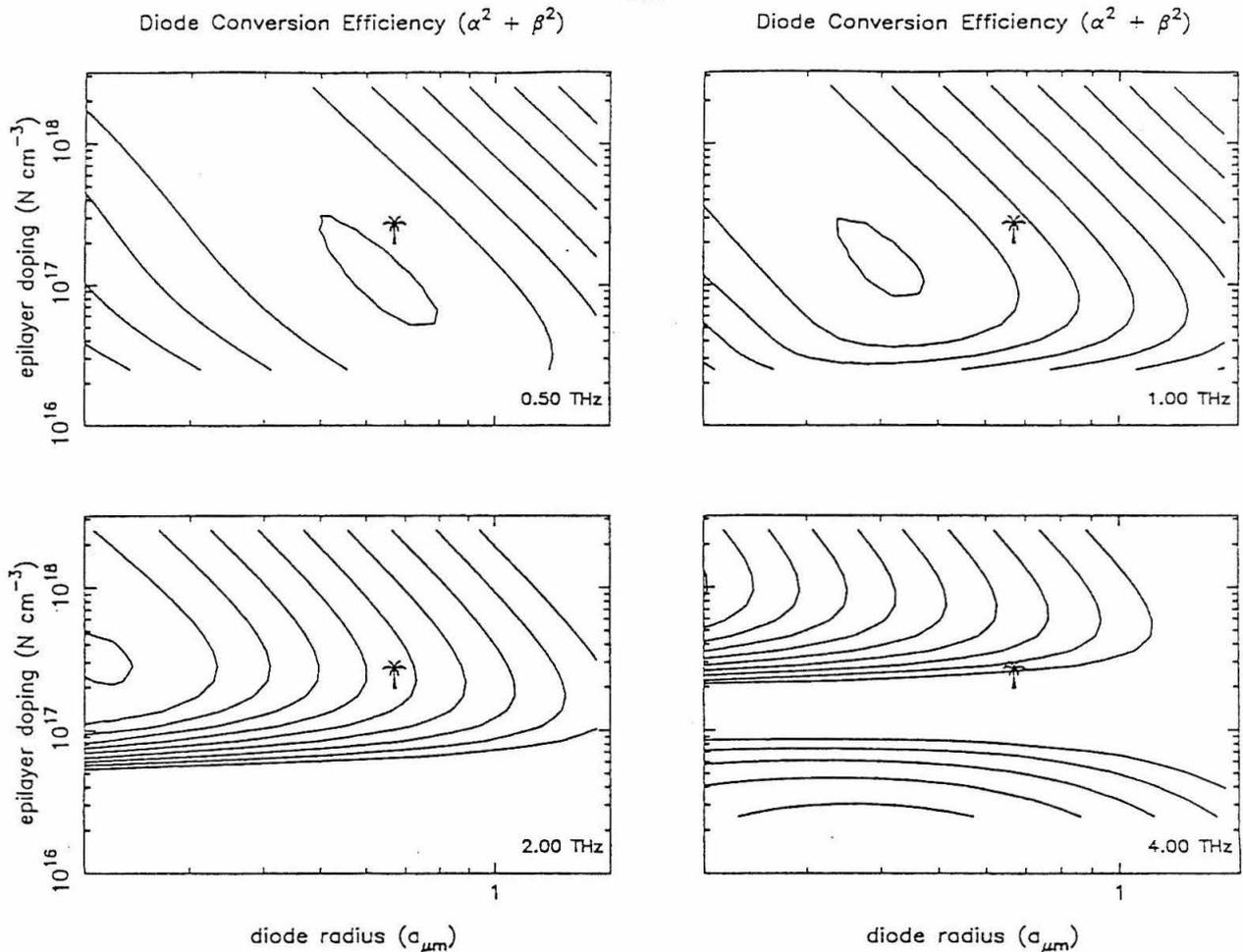


Figure 4.18 – Single-sideband conversion efficiency due to the diode alone, as a function of diode radius and epilayer doping. The base of the palm tree marks the actual parameters of our 1E12 diode. Each plot has been normalized to its peak value, which was approximately .4 (– 4 db) at 5 and 1 THz, .6 at 2 THz, and .2 at 4 THz. Contours are spaced every factor of two in sideband power; the highest contour is at 95 % of the peak.

that occurs (roughly) at some given value of capacitance, and at lower capacitance the efficiency drops. At first sight this is highly counter-intuitive. Examination of complete Smith chart traces of $|\Gamma(V)|$ in this region reveal there is actually a reasonable explanation, however. At the critical capacitance which produces the peak in conversion efficiency, reflectivity modulation of order unity is possible. That is, the $\Gamma(V)$ curve covers a significant fraction of the Smith chart as V varies from 0 to V_{FB} . At lower values of

capacitance, the same thing is still true. What changes is the *parametrization* of the reflectivity curve with voltage.

At the critical capacitance, the reflectivity modulation is largely due to the variation of susceptance. Since this has only a square-root dependence on voltage, the reflectivity moves over the complex plane relatively smoothly as the voltage is varied from 0 to 1 V (the assumed RF voltage swing in the conversion efficiency calculation). When the capacitance is much smaller than this, however, the reflectivity modulation is entirely due to variation in the barrier conductance. Therefore, the diode looks like an open circuit over nearly the whole range of the RF voltage swing, (0 to .9 V, say), and then it suddenly shoots across the Smith chart as the diode conductance exponentially passes through the antenna conductance. In both cases the reflected FIR wave has a peak amplitude comparable to the incident wave, but in the latter case, the reflectivity waveform is highly spiked. One gets a pulsed FIR wave with very low duty cycle.

Naturally, this latter wave has a smaller Fourier component at the fundamental ($m = 1$) sideband frequency than the former one. It is much richer in higher-order sidebands, however. In other words, the explanation of the paradox is simply that, when the diode capacitance drops below the critical value necessary to get reflectivity modulation of order unity, then the reflectivity waveform becomes spikey and the sideband power is progressively shifted into higher-order sidebands. It is possible to avoid this by proper tuning of V_{RF} and V_{DC} . In particular, one would simply reduce V_{RF} until the modulation of reflectivity over the Smith chart were fairly smooth as the voltage cycled over the reduced RF voltage range. In this sense, one can say that the falloff in conversion efficiency at capacitances lower than that at the peak in figures 4.18, is an artifact of holding the tuning fixed at $V_{DC} = V_{RF} = .5$ V. Had we made a plot of $(\alpha^2 + \beta^2)$ in which V_{DC} and V_{RF} were optimized at each individual point, we would have found the efficiency to plateau at the value of the peak in figure 4.18.

To summarize our results on diode optimization, we find that at low frequency (about 1 THz and below) the optimal doping level is $1 - 2 \times 10^{17} \text{ cm}^{-3}$, but that at higher frequency, it rises sharply. At 4 THz, the optimal doping level is approximately $1 \times 10^{18} \text{ cm}^{-3}$. The shift to higher doping is due to plasma resonance. When operating at a frequency below the plasma frequency, the efficiency is a very strong function of diode radius, due simply to the parasitic capacitance. At a frequency of 500 GHz, a radius of

.6 μm is sufficiently small to yield modulation of order unity, but at 2 THz, this radius is only .2 μm . Therefore, for all practical purposes, at high frequencies, ideal modulation performance is unachievable, and smaller diodes are always better. In this regime, it is a general fact that the phase modulation always dominates the amplitude modulation.

d. Conclusions regarding Schottky Diodes as Sideband Generators

We have theoretically analyzed the performance of a small-area Schottky diode mounted in a corner-cube antenna as a far-infrared modulator. We have proceeded by first calculating the small-signal far-infrared reflectivity $\Gamma(V)$, using the simple equivalent circuit model of Champlin and Eisenstein (1978) to describe the high-frequency parasitics. We then calculated the complex Fourier component of the reflectivity waveform at the RF frequency, in the limit of large RF drive. From it we determined the sideband strength due to amplitude modulation, the sideband strength due to phase modulation, and the total single-sideband conversion efficiency. In general, the phase modulation dominates the amplitude modulation by a large factor. We find that the sideband strengths vary extremely slowly with antenna impedance and diode substrate radius. They vary somewhat more quickly with Schottky barrier height and temperature, but the variation is still far too slow to yield much scope for sideband optimization. We find that by reducing the diode diameter to $\sim .4 \mu\text{m}$ and increasing the doping to $\sim 6 \times 10^{17} \text{cm}^{-3}$, an improvement of about 13 db in the diode losses may be obtained, at 2.52 THz. Above about 1 THz, the optimal doping level and diameter are strong functions of frequency, and are quite different than their low-frequency values.

4.3 – Polarizing Michelson Interferometer Optics

The optical configuration for the polarizing Michelson interferometer is shown in figure 4.19. It was designed assuming that all beams could be approximated as Gaussian, and using the formalism of Goldsmith (1982) for the propagation, focussing, and coupling of Gaussian beams. The design wavelength was 118.8 μm , the laser wavelength corresponding to our initial astronomical target, the ${}^2\Pi_{3/2}J = 5/2 \rightarrow 3/2$ rotational transition of OH. The input beam was assumed to have a beamwaist at the plane of the output laser mirror, with a 1/e - amplitude radius of $W_{00} = .77$ millimeter, corresponding to the measured FWHM angular size of 3.3° in the far field (figure 3.10).

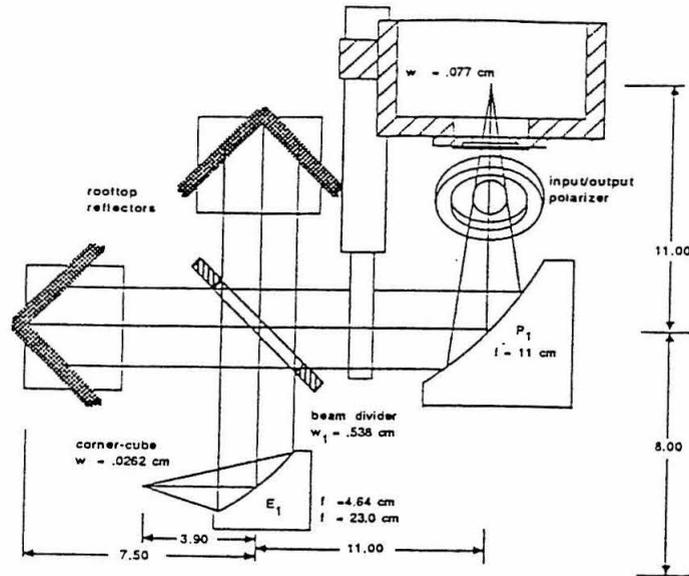


Figure 4.19 – (left) Optical configuration of the polarizing Michelson interferometer, corner-cube, and laser

In the Gaussian beam formalism, an elliptical mirror may be treated as equivalent to a pair of thin lenses, with focal lengths equal to the two focal lengths of the ellipsoid, and separated by zero distance. A paraboloidal mirror is treated as a single thin lens. P_1 and the first equivalent lens of E_1 therefore form a “Gaussian telescope”, since they are separated by the sum of their focal lengths. As a consequence, positioning the input beamwaist (i.e. the laser output coupler,) at the focus of P_1 ensures that a virtual beamwaist of width $W_{01} = \frac{f_1}{f_P} W_{00}$ will be located $f_1 = 23.00$ cm behind E_1 , independently of wavelength. The second equivalent lens of E_1 then refocuses this rather large beamwaist for matching onto the corner-cube. The final beam is quite fast, with a FWHM opening angle of 10° , implying that both the transverse positioning and the focussing of the corner-cube will be fairly critical. The position of the final beamwaist (i.e. the optimal position of the corner-cube’s phase center,) does vary with wavelength, but because the final equivalent lens, f_2 , is nearly in the geometric limit, (i.e. because it lies in the far-field of the input beamwaist, $|d_1| = |-23\text{cm}| \gg \frac{\pi w_{01}^2}{\lambda} = 6.9$ cm,) the wavelength dependence is weak.

A further consequence of the laser beamwaist being placed at the focus of P_1 is that the waist occurring inside the polarizing Michelson interferometer (PMI) is always located at the same position, (which we have chosen to be the PMI beamsplitter) independently of wavelength. The actual value of P_1 ’s focal length

determines the size of the beamsplitter waist. The latter is chosen so as to minimize the beam divergence within the arms of the interferometer, since this degrades the frequency resolution of the PMI. Considering only the divergence inherent in Gaussian beam propagation, this would favor the largest possible waistsize. Beyond a certain diameter, however, vignetting by the finite-sized beamsplitter, as well as the mechanical difficulty associated with moving the PMI farther back from the laser (because d_0 must equal f_{P_1}), militate for a smaller waistsize. The compromise adopted was to design the the beamwaist to just barely satisfy Goldsmith's criterion, $r > 2w_0$, for the projected aperture radius of the beamsplitter not to "significantly" vignette the beam, at $\lambda = 119 \mu\text{m}$. This is indicated on figure 4.19 by the -35 db contours. At longer wavelengths, e.g. $158 \mu\text{m}$, this criterion will be violated unless a larger aperture beamsplitter is used.

The polarizers we used were commercially supplied (Cambridge Physical Sciences, model IGP 224) gold wire grids evaporated on mylar. A potentially important flaw of such polarizers, (which is absent with more expensive, free-standing, grids,) is the fact that the mylar substrate can, depending on the manufacturing process, be birefringent. In that case, the polarizer will act to some degree as a phase retardation plate (or "waveplate") in addition to the polarizing action of the grids. We performed some measurements to evaluate the importance of this effect, even though complete ellipsometry was not possible. (We do not have any "true" polarizers that we can use as references.) A diagram of one such experiment is shown at the left of figure 4.20. It consists of simply blocking one of the arms of the interferometer, feeding the input port of the interferometer with a far-infrared beam that is linearly polarized normal to the corner-cube polarization, and measuring the signal on the corner-cube as the beamsplitter grid is rotated. It is not difficult to show that the theoretical power (not amplitude) efficiency for rotating the horizontally polarized beam into vertical should depend on angle as :

$$T \propto \sin^2(2\phi) \sin^4 \phi, \quad (4.31)$$

where

$$\phi = \arctan \left(\frac{\tan \theta}{\sqrt{2}} \right)$$

is the projected inclination of the wires from vertical, and θ the unprojected inclination, (i.e. the reading on the rotation stage.)

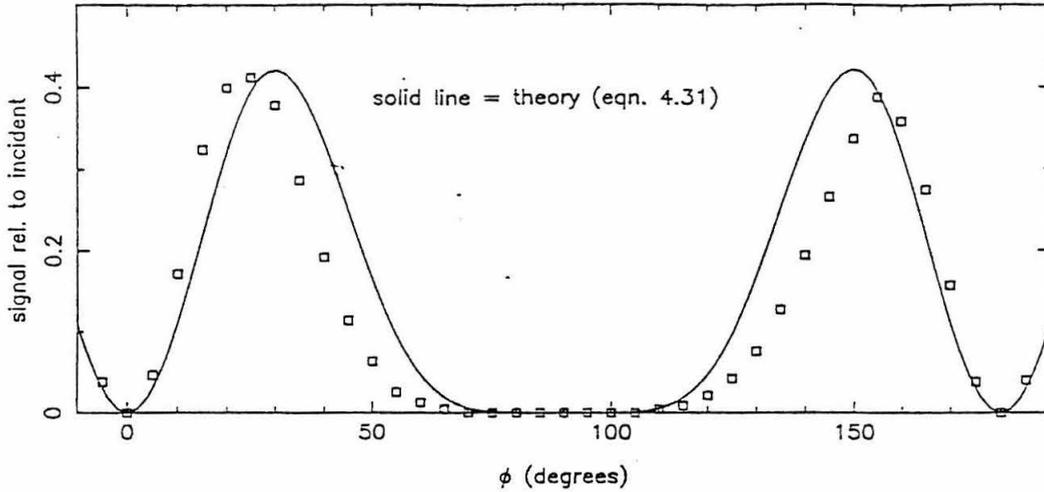


Figure 4.20 – First measurement of the retardation (“waveplating”) of our wire grid polarizers.

At the right of figure 4.20 the predicted dependence of equation 4.31 is plotted together with the experimental measurements. The experimental scan is substantially narrower than predicted. This could be due to polarization retardation (“waveplating”) by the polarizer, but it could also be due to two other effects : cross-polarized response of the corner-cube or elliptical polarization of the laser. As discussed in §4.1, the cross-polarized response of the corner-cube is expected to vanish at the peak of the mean beam (Vowinkel 1986,) and might therefore not have shown up in other measurements, but when the corner-cube is matched to a relatively fast beam, as it is in the PMI, then the cross-polarized response (if it exists) might conceivably be significant. In order to feed the interferometer with a linearly polarized beam for the polarizer measurement, the FIR laser was pumped with a linearly polarized CO₂ beam. This strongly breaks the degeneracy associated with a linearly polarized pump, and should yield a high polarization-purity beam.

To check these possibilities, another test of the retardation of the grids was made. It is illustrated in figure 4.21. The laser and corner-cube were coupled with a single off-axis paraboloid, and the signal was measured, relative to the signal with no grid inserted, when the grid was oriented for peak transmission and for minimum transmission. Then the signal was measured with the beam blocked, to determine the true depth of the minima. This was done for laser polarization both parallel and normal to the corner-cube

polarization, using the Babinet-Soleil compensator to rotate the CO₂ pump polarization by 90°. the entire measurement was made with the wire side of the polarizer facing the laser, and then with the wires facing the corner-cube. Two polarizers were tested. We note that all the off-axis mirrors we use are far too slow to produce significant cross-polarization themselves.

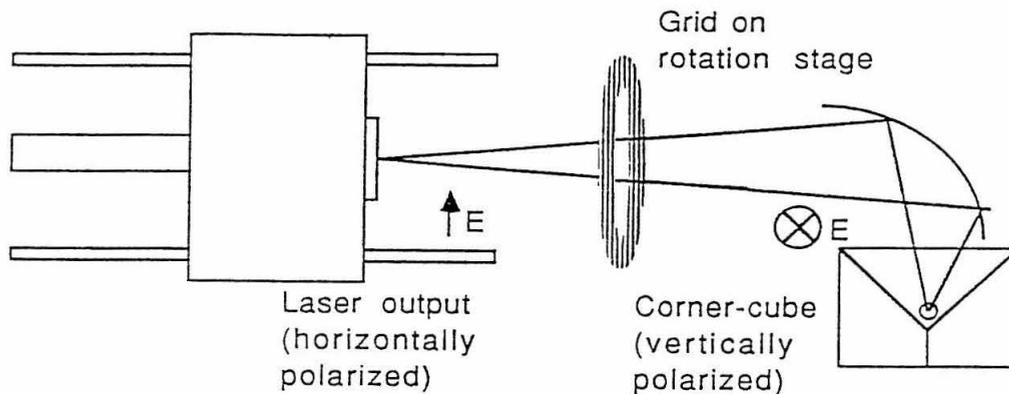


Figure 4.21 – Second measurement of retardation by wire grid polarizers

With the laser and corner-cube cross-polarized, and assuming the grid wires alone (without substrate) reject the parallel polarization component completely, the retardation of the substrate will allow signal transmission when the wires face the laser and are oriented normal to its polarization, or when they face the corner-cube and are normal to *its* polarization. The transmission should be exactly zero in the other two cases. The measured transmission for these various cases is shown in table 4.2. The fact that the measured power truly goes to zero (i.e. is the same as with the beam blocked) when the laser and corner-cube are cross-polarized and no grid is present eliminates the possibility that elliptical laser polarization or cross-polarized corner-cube response could account for the deviation between theory and experiment in figure 4.20. The data of table 4.2 are insufficient by themselves to solve for both birefringence parameters independently, namely the phase retardation $\Delta\phi$ and the angle between the principal axes and the corner-cube polarization, θ . It is not difficult to show that the expected transmission with cross-polarized laser

and corner-cube, in the favorable orientation (i.e. column 1 or 2 of table 4.2,) is given by

$$\tau = \frac{\sin^2 2\theta}{2} (1 - \cos \Delta\phi). \quad (4 - 32)$$

Assuming the orientation of the principal axes is the most favorable for rotating the polarization, namely that $\theta = 45^\circ$, we may invert 4-32 to derive a lower limit to the phase retardation. Using the mean values of columns 1 and 2 in the table, this yields $\Delta\phi \geq 9^\circ$ for grid 1 and $\Delta\phi \geq 2.5^\circ$ for grid 2.

Table 4.2

Waveplating by Wire Grid Polarizers on Dielectric Substrates

Laser \perp Corner-cube				
	Facing laser, \perp	Facing c.c., \perp	Facing laser, \parallel	Facing c.c., \parallel
Grid 1	0.60 %	0.64 %	< 0.02 %	< 0.02 %
Grid 2	0.02 %	0.08 %	< 0.02 %	< 0.02 %
No grid	< 0.02 %			
Beam blocked	< 0.02 %			
Laser \parallel Corner-Cube				
Grid 1	85.6 %	0.31 %	85.6 %	0.31 %
Grid 2	85.2 %	0.64 %	85.2 %	0.64 %

As discussed by Martin (1982), the need for accurate overlap of the beams in the two arms of the interferometer imposes a constraint on the amount by which the roof angles of the retroreflectors deviate from exactly 90° . As he shows, the roof angle ($90^\circ + \alpha$) must satisfy

$$\alpha \ll \frac{\lambda}{4D}, \quad (4 - 33)$$

where D is the beamwidth within the interferometer. For our interferometer this implies an accuracy better than ~ 4 arcminutes. Our rooftops are actually full three-plane corner-cubes (Precision Lapping Inc., Valley Stream, N.Y.) and are specified to an accuracy of 30 arcseconds, so this is not thought to be a significant problem. The off-axis mirrors, P_1 and E_1 , were fabricated on a conventional milling machine

(the Zen Machine Shop, Pasadena, CA) using the technique described by Erickson (1979). They were machined from brass, hand polished, and gold-plated in order to facilitate optical alignment with visible light.

Alignment of the PMI can be quite tricky. Aside from the tunable path-length difference, fine adjustment has been provided for the orientation of the rooftop reflectors about the two axes transverse to the rooftop dihedral. For the laboratory version of the interferometer, the two polarizing grids were mounted on rotation stages, allowing smooth and continuous adjustment of the grid orientation. In the airborne version, however, the grids are rigidly mounted in order to save space. Finally, the corner-cube mount is fully adjustable in five axes. We have found the most stringent test of the optical alignment to be the following : The corner-cube's angular orientation in the plane of the interferometer (i.e. the coordinate we have denoted ξ) is scanned and the profile of the corner-cube response recorded, first with one arm of the PMI blocked, then with the other blocked, and then with both transmitted. When the interferometer is aligned, the two single-arm scans should be equal in amplitude, center position, and shape, and the two-arm profile should reach a peak four times higher than the single-arm profiles. Furthermore, the profiles should be "well-behaved", i.e. symmetric, single-peaked, and smooth. It is remarkably difficult to align the various components so that all these desirable features are achieved. For example, some of the pitfalls associated with alignment are illustrated at the left of figure 4.22, which shows a set of these scans for one particular set of adjustments. The responses from the two arms are clearly not balanced, and in fact are offset from one another by several degrees. Note, however, that at the peak of the two-arm response, the responses of each individual arm are equal to within about 15 %, and the combined response is approximately 4.5 times the mean of the individual ones. Lacking the full scan in ξ , one would mistakenly conclude that the alignment was nearly perfect. Thus, simply peaking up on the two-arm response and checking the balance at that point can be seriously misleading. An additional check is that, when the interferometer is aligned, any small adjustment in the rooftops' orientations, in either of the two axes, should produce a well-defined interference pattern. Several fringes and nulls should be visible, and the fringe envelope decrease monotonically away from the peak.

The right hand side of figure 4.22 shows the best alignment scans we have achieved, after a thorough,

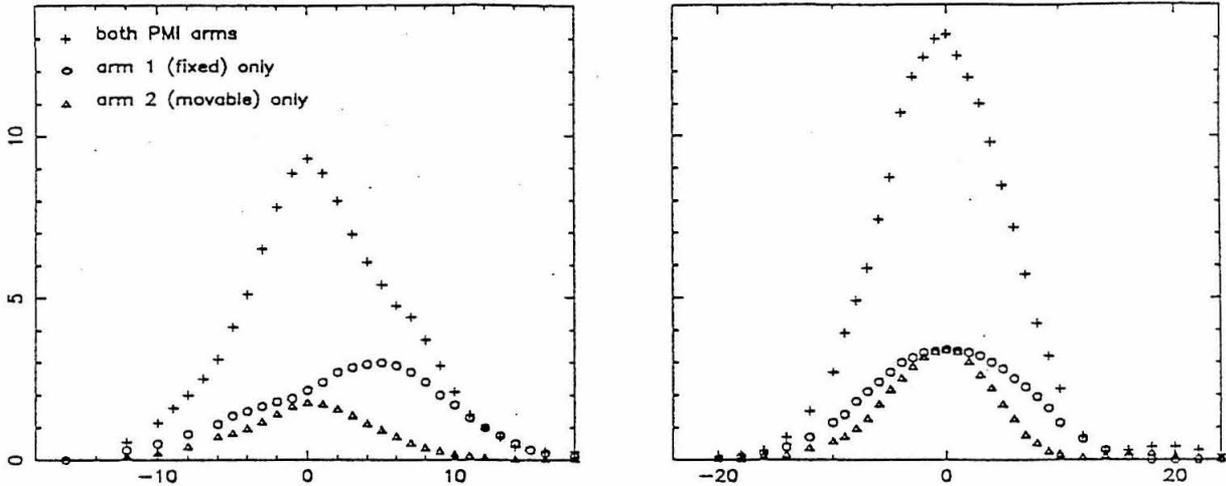


Figure 4.22 – Scans of the corner-cube signal as ξ is varied, a good diagnostic of PMI alignment.

(and quite laborious) process of optimization. The two arms are balanced in amplitude and coincident in angle, and the two-arm beam has exactly 4.0 times the power of the individual beams. The only problem is that the two arms' beamwidths are not equal, viz. 11.5° versus 17.5° . The reason for this is unclear. The collimation of the beam inside the interferometer was measured with an iris diaphragm placed at various positions inside it. The beam was found to be slightly converging, but not by enough to account for the size of the discrepancy in beamwidths. Perhaps it has to do with a small mispositioning of the beamsplitter which differentially vignettes the two beams. A second feature of figure 4.22 which is imperfectly understood is the fact that the balanced configuration which is displayed could only be achieved when the orientation of the grid wires to the vertical was $\theta = 69^\circ$, not $\theta = \arctan(\sqrt{2}) = 55^\circ$ as theoretically expected (see equation 4.31). This effect was most likely due to an elliptically polarized beam at the interferometer input, however. The elliptical polarization was probably partly due the fact that grid #1, whose phase retardation was later measured to be $\geq 9^\circ$, was used for the input/output polarizer. Probably, it was also partly due to ellipticity in the FIR laser output polarization when using a circularly polarized pump.

4.4 Sideband Generation Results

Our only reliable sideband generation experiments were made on the laboratory system mounted on

the optical table. Sidebands have been produced with the airborne version of the system, but they have neither been optimized nor systematically studied. The usual laboratory setup is illustrated in figure 4.23. The output beam of the PMI is focussed onto the detector with a single polyethylene lens. Most of our experiments have been carried out with an RF frequency of 6-8 GHz, corresponding to the required offset between the 119 μm methanol laser line and the ground state rotational transition of OH. Audio-frequency modulation may be applied either to the RF power, by means of a PIN diode switch, or to the Schottky diode's DC bias. The latter is provided by a homemade voltage-programmable current source, and is diplexed with the RF onto the corner-cube by means of a microwave bias tee (Hewlett-Packard Corp., model 33150). The DC and modulated detector signals are independently monitored. The usual procedure is to use the video response of the Schottky diode at a relatively low bias – typically $I_b = 1\mu\text{A}$, $V_b \approx 740\text{mV}$ – to align the PMI, to maximize the laser power and to optimize the corner-cube orientation. Then, the DC Schottky bias is raised to a typical operating point, say 100 μA , and an RF signal is applied at a power level of .1 - .5 mW, (a rough guess for optimal performance). The RF power typically reduces the DC diode voltage by several hundred millivolts. An audio frequency modulation is applied either to the RF power, while the DC current is held fixed, or to the DC current, while the RF power is held fixed. (If the sideband were used as an LO, this audio modulation would be removed, of course.) The polyethylene lens and the detector dewar are then positioned so as to maximize either the DC detector photocurrent or the demodulated photocurrent. With proper alignment these maxima should coincide. The final step, and the definitive proof of sideband generation, is then to insert the Fabry-Perot interferometer between the polyethylene lens and the dewar, and to scan its transmission peak through the lower sideband wavelength, the carrier, and the upper sideband.

In our initial experiment, we did not use the Fabry-Perot, and were therefore deceived into gross overestimates of the sideband power by an unforeseen, and somewhat insidious effect. In addition to coupling some of the incident radiation onto the whisker antenna, the corner-cube also acts as a simple retroreflector. Its reflectivity in this mode is virtually 100 %, independent of the incident polarization. This retroreflected beam is spatially distinct from the antenna-coupled beams, and is, of course, unaffected by audio or RF modulation applied to the whisker. The antenna-coupled beams include a strong re-radiated

the PMI. The ratio of this “audio homodyne” signal to the desired signal is very roughly given by the ratio of the conversion loss from the carrier to the RF sidebands to the frequency resolution of the PMI. Unfortunately, this ratio is much larger than unity, and therefore, nearly all the modulated photocurrent is due to the audio sideband of the carrier. When the Fabry-Perot is inserted between the PMI output and the detector and is tuned to the sideband wavelength, the modulated photocurrent is reduced by a large factor (> 10 db) compared to when it is tuned to the carrier wavelength. This may be seen at the left of figure 4.24, where we show our first Fabry-Perot spectrum demonstrating sideband generation. The error bars indicate the maximum fluctuations in the signal over several (5 to 10) integration times of the lock-in. This spectrum was taken with the same PMI alignment which yielded the corner-cube ξ -scans shown in figure 4.21. The expected positions of the RF ($\nu_{RF} = 7.4$ GHz) sidebands are shown. Only one sideband is present. Presumably, this is related to the fact that the waveplating of the PMI I/O polarizer is causing the PMI to be fed with an elliptically polarized beam – the same effect which was responsible for the best arm balance in the PMI being achieved with a non-nominal grid orientation for the beam divider. A curious aspect of the “audio homodyne” signal is that its phase is completely unpredictable. This is because the retroreflected carrier (the “LO”) and the re-radiated audio sideband (the “signal”) are spatially distinct beams, and therefore generally cover different path lengths (on the scale of an FIR wavelength) from the corner-cube to the detector, depending on the precise alignment and focussing of all the mirrors, the presence of stray reflections, etc. Because the path length from corner-cube to detector is many hundreds of wavelengths, it only takes a minute adjustment in the alignment or position of an optical component to produce many radians of phase change in “homodyne” signal. For this effect, the Fabry-Perot mirrors count as optical components, and it is therefore possible to get some rather bizarre looking “spectra” of the carrier, by monitoring the modulated photocurrent as the Fabry-Perot mirrors are scanned through the carrier frequency, as shown at the bottom of figure 4.24.

The signal-to-noise ratio in figure 4.24 is sufficient to be confident of the presence of an RF sideband, but is nonetheless rather low. This is characteristic of the relatively low value of load resistor used in the DC bias circuit for the detector, namely 10 K Ω . The detector used was LBL 108-17.7, the same used for the bulk of our heterodyne detector tests. (see §2.) The airborne version of the system used

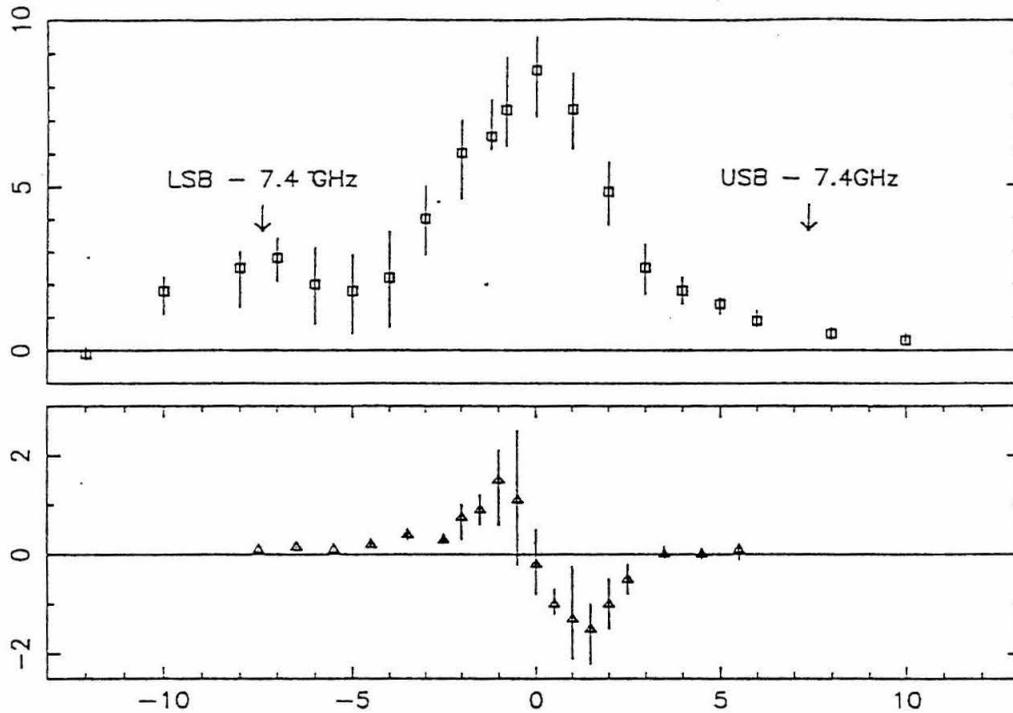


Figure 4.24 – Fabry-Perot spectra taken of the sideband generator output with detector LBL 108-17.7. (Top) First observation of RF sidebands, showing only one sideband, (bottom) Scan showing the “audio homodyne” effect, in this case manifested as a variation in the phase of the homodyne signal as the movement of the Fabry-Perot mirror changes the pathlength difference between the retroreflected carrier and the audio sideband.

a 5 k Ω DC load resistance, and had significantly more loss in the optical path from the PMI output to the detector. On the other hand, the bias circuit’s intrinsic noise was lower in that case, and the detector responsivity was higher, so the signal-to-noise for detection of the sidebands with the airborne system was similar to that in figure 4.24, although in that case, both sidebands had approximately equal amplitudes. As discussed in chapter 2, a relatively low value of load resistance is required when studying g.r. noise or optimizing heterodyne performance, in order that high LO levels not saturate the bias circuit. For simply studying the sideband generator, however, this is not a consideration. Therefore, all the further sideband measurements were made with a second dewar, optimized for low background, in which an ultra-high responsivity detector (LBL 583-4.6) was mounted in a bias circuit with a load resistance of 10 M Ω (at

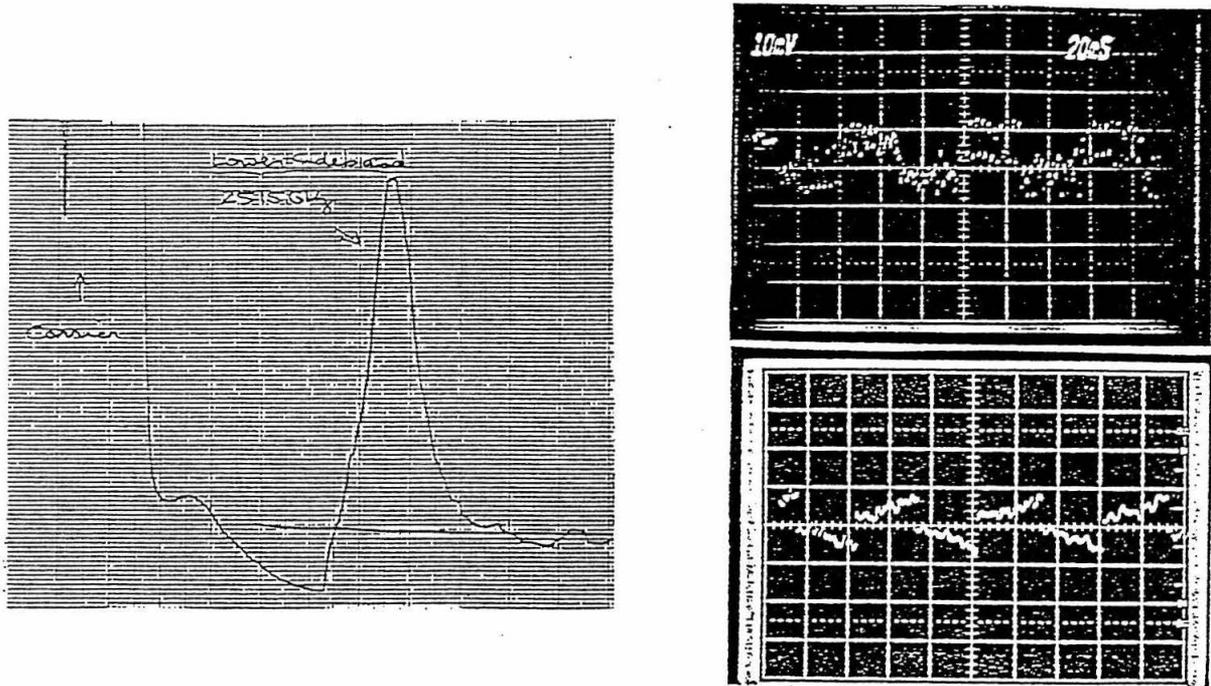


Figure 4.26 – (Left) Fabry-Perot spectrum of the sideband generator output taken with the ultra-high sensitivity detector LBL 583-4.6, (right) Oscilloscope traces of the sideband generator output when the Fabry-Perot was tuned to the sideband frequency and the RF power chopped

room-temperature) and a transimpedance amplifier. An improved FIR filtering arrangement was also used, which served the important purpose of reducing the room-temperature background radiation to level that did not saturate the photocurrent. The PMI was also realigned for these latter measurements, simply by using the He-Ne laser beam, and then checking the balance in the two arms. More importantly, the I/O polarizer of the PMI was oriented so that the wires faced the PMI rather than the laser, thereby ensuring that the PMI was fed with a linearly polarized beam.

The resulting Fabry-Perot scan is shown at the left of figure 4.25. The improved signal-to-noise ratio is obvious. Varying the RF frequency shifted the peak in fig. 4.25 in the proper way. Further verification that the peak is indeed an RF sideband was made by turning off the RF power to the Schottky diode and scanning the Fabry-Perot. No “sideband” peak was produced, demonstrating that the peak of fig. 4.25 is not, for example, just an artifact of a standing wave between the moveable Fabry-Perot mirror and any other

piece of optics varying the residual transmission of the audio-sideband of the carrier. The right-hand side of figure 4.25 shows the RF sideband signal as oscilloscope traces taken with the Fabry-Perot held fixed at the sideband frequency and the RF power chopped. The high-frequency oscillations in the upper trace are due to the dither in the CO₂ laser. Note that the approximately 10 mV (peak-to-peak) signals shown in the figure lie on top of an approximately 250 mV DC signal due to the room-temperature background incident on the detector.

Table 4-3

Conditions for Generation of Sidebands Shown in Figure 4.25

Corner-Cube

Whisker length	$L = 590\mu m (= 5.0\lambda)$
orientation	$\zeta_{peak} = 18^\circ$

Schottky diode (Mattauch 1E12)

DC Bias	$I_{DC} = 500\mu A$
	$V_{DC} = 920mV$ (RF off)
	$V_{DC} = 360mV$ (RF on)

RF Signal

Frequency	$\nu_{RF} = 8.0$ GHz
Incident power	$P_{RF} = 500\mu W$

The conditions of DC bias, RF power, etc. for producing these sidebands are compiled in table 4.3. It was found that the dependence of sideband power on Schottky bias, for fixed RF power, was relatively weak, as shown in figure 4.26. The sideband power monotonically increases with higher bias. Irreversible diode degradation occurred when the bias current was raised to 1 mA. Note that without RF power applied, we have generally found diode burnout to occur at DC currents of ~ 3 mA for the 1E12 diodes, although this figure can be lower for imperfect whisker contacts.

The absolute sideband power determined by calibrating the the ultra-high responsivity detector against the pyroelectric detector, assuming a responsivity for the latter of 4000 V/W, our standard value. (See chapter 3 for a discussion of our absolute power calibrations.) The saturation curve of the detector was

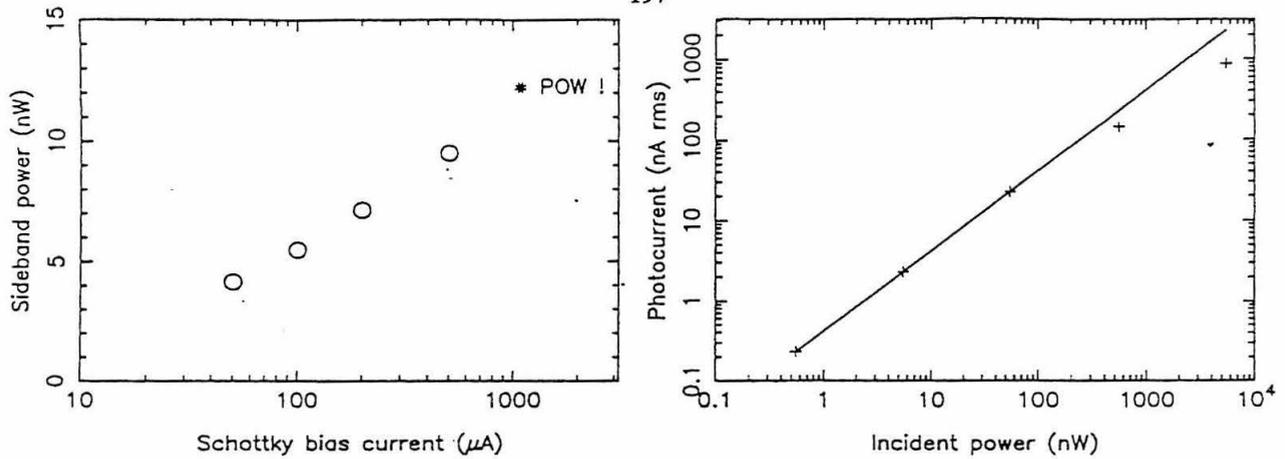


Figure 4.26 – (Left) Dependence of sideband power on DC Schottky bias, (right) Saturation curve for detector 583-4.6, used in the sideband measurements of figure 4.25

obtained by successive insertion of attenuators in the far-infrared beam, and is shown in figure 4.26. The response is reduced from its small signal value by approximately 2 db at an incident power level 550 nW, and is quite accurately linear below 55 nW. Using the detector in its unsaturated regime, the transmission of the Fabry-Perot was determined to be 7 % . The reason this is so low is that the usual 500 line-per-inch mirrors had been replaced for this measurement with 750 lpi mirrors in order to improve the Fabry-Perot's resolution, and therefore its rejection of the modulated carrier when tuned to the sideband frequency. Using the small-signal responsivity obtained from figure 4.26, we have listed in table 4.4 our best estimates of the FIR power levels available at various points in the system.

Table 4-4

Measured Sideband Power

Measurement Point	Measured Power	Comments
FIR laser output	1 mW	$\pm .5$ mW ?
Fabry-Perot input (single-sideband)	135 nW	$\rightarrow \zeta = -39 \pm 2$ db
Fabry-Perot output (single-sideband)	9.5 nW	
LO available to mixer (50 % beamsplitter)	5 nW	$> 2\mu W$ required

The highest single-sideband power achieved was 9.5 nW. Replacement of the Fabry-Perot mirrors would very likely provide an immediate improvement in the Fabry-perot's transmission to > 50 %, however,

yielding approximately 70 nW of sideband power, and 35 nW of usable LO power after a 50 % beamsplitter.

The power figures summarized in table 4-4 comprise our basic result for sideband generation performance. It is clear that even this power level, our best result, is between 15 and 20 db lower than the LO power required for our mixers to achieve g.r. noise-limited performance. This lack of local oscillator power is the primary problem with our system, as it stands in its present state. We therefore conclude with a summary of the reasons contributing to the lack of LO power and the prospects for reducing or eliminating them. Aside from the relatively straightforward issue of replacing the Fabry-Perot mirrors, there are two areas which offer significant scope for improvement : increasing the FIR laser power, and increasing the single-sideband conversion efficiency, ζ , of the Schottky diode/corner-cube combination. As discussed earlier, using our best estimates of the 1E12 diode parameters, our diode model predicts -20.5 db conversion loss due to the diode alone. That would imply that the measured value of $\zeta = -39$ db is approximately equally divided between diode losses and losses in optically coupling to the whisker. In other words, the product of Ohmic and antenna coupling losses, $\eta\epsilon$, must be approximately 10 %. As we have seen however, the results of detailed beam pattern measurements, along with our comparison of predicted and measured video responsivities, suggests an over coupling efficiency of 30 - 40 %. Furthermore, we have also seen that the diode losses are very sensitive to doping level, essentially because we are on the wing of the plasma resonance. A doping level of $1 \times 10^{17} \text{ cm}^{-3}$ instead of $2 \times 10^{17} \text{ cm}^{-3}$ would yield a predicted diode loss of -30 db rather than -19 db. Our best guess, therefore, is that the diode losses contribute -25 to -30 db, and the antenna losses -10 to -15 db, of the total conversion loss. Perhaps there is also a contribution due to non-optimal tuning of diode bias and RF power levels, but we have found experimentally that these dependences are weak, and therefore believe that their contribution to the overall conversion loss is unlikely to exceed ~ 5 db. DC bias and RF power tuning probably does not present much scope for improvement, therefore.

Obviously, the best hope for improvement in the conversion loss lies in diode optimization. The numerical results of our model indicate that an improvement of 10 to 15 db is possible by reducing the diode area and simultaneously increasing the epilayer doping. This improvement, though certainly the highest priority item in any future program of sideband generation, will probably be fairly laborious,

requiring repeated iterations of the cycle of diode fabrication, testing, and analysis. Furthermore, it is unlikely, by itself, to yield enough improvement in sideband power to realize the full potential of our mixers. In addition, it will be necessary to increase the incident laser power by roughly 10 db. As discussed in chapter 3, high power CO₂ lasers are commercially available (for a high price) that could achieve this, and that would be little more than “drop-in” replacements for our laser. Alternatively, it is perhaps possible that higher efficiency FIR cavities could be designed, that would produce more FIR power for the same CO₂ pump power.

In summary, the measured single-sideband conversion efficiency of the sideband generator we have designed and built is approximately -39 db. Including a realistic value for the Fabry-Perot transmission, and a 50 % beamsplitter loss, the laser and sideband generator can therefore provide only some 35 nW of usable local oscillator power at the mixer. This assumes perfect optical coupling between the sideband beam, the telescope beam, and the beam out of the detector dewar. Nonetheless, it is nearly 20 db too low for optimal mixer performance. We believe that we crudely understand the sources of this conversion loss, but that it cannot be improved without a fairly substantial effort to fabricate Schottky diodes of smaller area and higher epilayer doping.

References

- V.N. Abakumov, V.I. Perel', and I.N. Yassievich 1978 (APY), *Sov. Phys. Semicond.*, **12**, 1.
- N.B. Abraham, D. Dangoisse, P. Glorieux, and P. Mandel 1985, *J. Opt. Sci. Am. B*, **2**(1), 23.
- K.R. Armstrong and F.J. Low 1973, *Appl. Optics*, **12**, 2007.
- K.R. Armstrong and F.J. Low 1974, *Appl. Optics*, **13**, 425.
- G. Ascarelli and S. Rodriguez 1961, *Phys. Rev.*, **124**, 1321.
- J. Bardeen and W. Shockley 1950, *Phys. Rev.*, **80**(1), 72.
- A.L. Betz 1976, UC Berkeley Ph.D. dissertation.
- A. Betz and J. Zmuidzinas 1984, in proceedings of the Airborne Astronomy Symposium, Ames Research Center, NASA and the Astronomical Society of the Pacific.
- A. Betz and J. Zmuidzinas 1987, in presentation to L. Fisk on SOFIA, (Stratospheric Observatory for Infrared Astronomy), Ames Research Center.
- D.D. Bićanić, B.F.J. Zuidberg, and A. Dymanus 1978, *Appl. Phys. Lett.*, **32**, 367.
- D.D. Bićanić 1983, in *Infrared and Millimeter Waves* vol.7, K. Button (ed.).
- S.K. Bignell, F. Saiedy, P.A. Sheppard 1963, *J. Opt. Soc. Am.*, **53**, 466.
- M. Born and E. Wolf 1980 *Principles of Optics*, 6th ed., Oxford, Pergamon Press.
- P.R. Bratt 1977, chapter 2 in *Semiconductors and Semimetals* vol. 12, R.K. Willardson and A.C. Beer (eds.).
- H. Brooks 1951, *Phys. Rev.*, **83**, 879.
- E.R. Brown 1984, *Electronics Lett.*, **21**(10), 417.
- R.A. Brown and S. Rodriguez 1966, *Phys. Rev.*, **153**, 890.
- D.E. Burch 1968, *J. Opt. Soc. Am.*, **58**, 1383.
- C.M. Caves 1982, *Phys. Rev. D*, **26**, 1817.
- K.S. Champlin and G. Eisenstein 1978, *IEEE Trans. Microwave Theory and Techniques*, **MTT-26**, 31.

- T.Y. Chang and T.J. Bridges 1970, *Optical Comm.*, **1**, 423.
- G. Chin 1987, private communication.
- A.E-T. Chiou 1983, Calif. Inst. of Technology, Ph. D. dissertation.
- M.E. Cohen and P.T. Landsberg 1967, *Phys. Rev.*, **154**(3), 683.
- E. Conwell and V.F. Weisskopf 1950, *Phys. Rev.*, **77**(3), 388.
- J.A. Copeland 1970, *IEEE Trans. Electron Devices*, **ED-17**, 404.
- T.W. Crowe and R.J. Mattauch 1986, *IEEE Trans. Microwave Theory and Techniques*, **MTT-34**, 733.
- T.W. Crowe and R.J. Mattauch 1987, *IEEE Trans. Microwave Theory and Techniques*, **MTT-35**, 159.
- P.P. Debye and E.M. Conwell 1954, *Phys. Rev.*, **93**(4), 693.
- J.J. Degnan 1976, *Appl. Phys.*, **11**, 1.
- R. Densing, P.B. van der Wal, H.P. Röser, and R. Wattenbach 1985, Max-Planck Inst. f. Radioastron. preprint, submitted to *Astron. and Astrophys.*
- R.H. Dicke 1946, *Rev. Sci. Instrum.*, **17**, 268.
- N.R. Erickson 1979, *Applied Optics*, **18**, 956.
- N.R. Erickson 1985, *IEEE Trans. on Microwave Theory and Techniques*, **MTT-33**, 1179.
- N.R. Erickson 1987, speakin at the Terahertz Technology Conference, Lake Arrowhead, CA.
- H.O. Everitt, D.D. Skatrud, and F.C. DeLucia 1986, *Appl. Phys. Lett.*, **49**(16), 995.
- H.R. Fetterman, P.E. Tannenwald, B.J. Clifton, C.D. Parker, W.D. Fitzgerald, and N.R. Erickson 1978, *Appl. Phys. Lett.*, **33**(2), 150.
- J. Farhoomand, G.A. Blake, M.A. Frerking, and H.M. Pickett 1985, *J. Appl. Phys.*, **57**, 1763.
- R. Genzel and G.J. Stacey 1985, *Mitteilungen der Astronomische Gessellschaft*, **63**, 215.
- P.F. Goldsmith 1982, chapter 5 in *Infrared and Millimeter Waves*, vol. 6, Academic Press.
- E.N. Grossman 1987, *Intl. J. of IR. and MM. Waves*, **9**(11), in press.
- B.D. Guenther and P.W. Kruse 1986, *Intl. J. of IR. and MM. Waves*, **7**(8), 1091.
- E.G. Gwinn and R.M. Westervelt 1986, *Phys. Rev. Lett.*, **57**(8), 1060.

- N.M. Haegel, M.R. Hueschen, and E.E. Haller 1985, *Infrared Physics*, 25(1), 273.
- H. Haken 1975, *Phys. Lett.*, 53A(1), 77.
- J.J. Hall 1962, *Phys. Rev.*, 128(1), 68.
- E.E. Haller, N.P. Palaio, M. Rodder, W.L. Hansen, and E. Kreysa, 1984 *Proc. 4th Intl. Conf. on Neutron Transmutation Doping of Semiconductors*, R.D. Larabee ed., Plenum Press.
- A.I. Harris 1986, UC Berkeley Ph. D. dissertation.
- D.N. Held and A.R. Kerr 1978a, *IEEE Trans. on Microwave Theory and Techniques*, MTT-26, 49.
- D.N. Held and A.R. Kerr 1978b, *IEEE Trans. on Microwave Theory and Techniques*, MTT-26, 55.
- J. Heppner, C.O. Weiss, U. Hübner, and G. Schinn 1980, *IEEE J. Quantum Electronics*, JQE-16(4), 392.
- D.T. Hodges 1977, in *SPIE proceedings*, vol. 105, Far-infrared and Submm. Wave technology and Applications.
- D.T. Hodges, F.B. Foote, and R.D. Reel 1977, *IEEE J. Quant. Electr.*, JQE-13(6), 491.
- E.H.M. Hogenboom, W. Klische, C.O. Weiss, A. Godone 1985, *Phys. Rev. Lett.*, 55(23), 2571.
- B. Hughes, N.C. Fernandez, and J.M. Gladstone 1986, preprint submitted to *IEEE Trans. on Electron Devices*.
- R.L. Jones and P. Fisher 1970, *Phys. Rev. B*, 2(6), 2016.
- F. Julien and J-M. Lourtioz 1980, *Intl. J. of IR. and MM. Waves*, 1(2), 175.
- M. Kavaya 1982, Calif. inst. of Technology, Ph. D. dissertation.
- A.G. Kazanskii, P.L. Richards, and E.E. Haller 1977, *Appl. Phys. Lett.*, 31(8), 496.
- L. Kazovsky 1986, *J. Lightwave Technology*, LT-4, 415.
- C. Kittel 1976, *Introduction to Solid State Physics*, 5th ed., New York, Wiley.
- Knight 1986, in *CRC Handbook of Laser Science and Technology*, vol. II, M.J. Weber (ed.), Boca Raton, FL, CRC Press.

- S.H. Koenig, R.D. Brown, and W. Schillinger 1962 (KBS), *Phys. Rev.*, **128(4)**, 1668.
- H. Kogelnick 1964, *Polytechnic Inst. Brooklyn Symp. Quasi-Optics, 1964*, p 333.
- E.L. Kollberg, H. Zirath, and A. Jelenski 1986, *IEEE Trans. on Microwave Theory and Techniques*, **MTT-34(9)**, 913.
- H. Krautle, E. Sauter, and G.V. Schultz 1977, *Infrared Physics*, **17**, 477.
- A. Kreisler, M. Pyée, and M. Redon 1984, *Intl. J. of IR. and MM. Waves*, **5(4)**, 559.
- S.R. Kulkarni and C. Heiles 1986, chapter 3 in *Galactic and Extragalactic Radio Astronomy*,
K.I. Kellerman and G.L. Verschuur (eds.).
- M.L. Kutner and B.L. Ulich 1981, *Ap. J.*, **250**, 341.
- M. Lax 1960, *Phys. Rev.*, **119(5)**, 1502.
- M. Lefebvre, D. Dangiosse, and P. Glorieux 1984, *Phys. Rev. A*, **29(2)**, 758.
- D.F. Lester, J.D. Bregman, F.C. Witteborn, D.M. Rank, and H.L. Dinerstein 1986, preprint
accepted by *ApJ.*
- S.M. Liu and M.B. Das 1983, in Proc. UGIM Symposium, (Texas A&M Univ.), p. 169.
- S.M. Liu, M.B. Das, W. Kopp, and H. Morkoc 1985, *IEEE Electron Device Lett.*, **EDL-6(9)**,
453.
- E.N. Lorenz 1963, *J. Atmos. Sci.*, **20**, 130.
- D.K. Mansfield, G.J. Tesauro, L.C. Johnson, and A. Semet 1981, *Optics Lett.*, **6(5)**, 230.
- E.A.J. Marcatili and R.A. Schmeltzer 1964, *Bell Syst. Tech. J.*, **43**, 1783.
- D.H. Martin 1982, chap. 2 in *Infrared and Millimeter Waves*, vol. 6, New York, Academic
Press.
- L.M. Mattarese and K.M. Evenson 1970, *Appl. Phys. Lett.*, **17(1)**, 8.
- R.A. McClatchey, W.S. Benedict, S.A. Clough, D.E. Burch, R.F. Calfee, K. Fox, L.S. Rothman,
J.S. Garing 1973 *AFCRL Atmospheric Absorption Line Parameters Catalog*, (AFCRL
Environmental Research Paper no. 434), 1973, U.S. Dept. of Commerce, Springfield
VA.
- R.E. McMurray, N.M. Haegel, J.M. Kahn, and E.E. Haller 1986, preprint submitted to *Solid*

State Communications.

G. Melnick, G.E. Gull, M. Harwit, and D.B. Ward 1978, *Ap. J. Lett.*, **222**, L137.

G. Melnick, G.E. Gull, and M. Harwit 1979, *Ap. J. Lett.*, **227**, L29.

L.A. Newman and R.A. Hart 1986, in *Proc. of IEEE Conf. on Lasers and Electro-optics*,
Seattle, WA, appearing in *J. Opt. Soc. Am.*.

P. Norton and H. Levenstein 1972, *Phys. Rev. B*, **6**, 470.

A. Pais 1982, *Subtle is the Lord: the Science and the Life of Albert Einstein*, Oxford, Clarendon
Press.

N.P. Palaio 1983, UC Berkeley M.S. thesis.

I.S. Park, E.E. Haller, E.N. Grossman, and D.M. Watson 1987, in preparation.

W.C. Petersen, A. Gupta, and D.R. Decker 1983, *IEEE Trans. Microwave Theory and Tech-
niques*, **31(1)**, 27.

T.G. Phillips and K.B. Jefferts 1973, *Rev. Sci. Instrum.*, **44(8)**, 1009.

T.G. Phillips and D.P. Woody 1982, *Ann. Rev. Astron. and Astrophys.*, **20**, 285.

T.G. Phillips and D.M. Watson 1984, *The Large Deployable Reflector*, report of the LDR
Science Coordination Group Focal Plane Instruments Subcommittee, NASA.

T.G. Phillips and D.B. Rutledge 1986, *Scientific American*, **254(5)**, 97.

T.G. Phillips 1986, in proceedings of ESA Space Science Symposium, Segovia, Spain.

G.E. Pikus and G.L. Bir 1960, *Sov. Phys. Solid State*, **1**, 1502.

S.S. Prasad and W.T. Huntress 1980, *Ap. J. Suppl.*, **43**, 1.

S.S. Prasad and W.T. Huntress 1982, *Ap. J.*, **260**, 590.

R.A. Pucel, H.A. Haus, and H. Statz 1975, in *Advances in Electronics and Electron Physics*
vol. 38.

J. Pujol, R. Vilaseca, R. Corbalán, and F. Laguarda 1987, *Intl. J. of IR. and MM. Waves*, **8(3)**,
299.

A.K. Ramdas and S. Rodriguez 1981, *Rep. Prog. Phys.*, **44**, 1297.

R.P. Reed and A.F. Clark (eds.) 1983, *Materials at Low Temperatures*, Metals Park, OH,

American Society of Metals.

M.G. Richards, A.R. Andrews, C.P. Lusher, and J. Schratte 1985, *Rev. Sci. Instrum.*, **57**(3), 404.

H. Risken 1964, *Zeit. f. Phys.*, **180**, 150.

H. Risken and K. Nummedal 1968, *J. Appl. Phys.*, **39**(10), 4662.

H.P. Roeser, R. Wattenbach, E.J. Durwen, and G.V. Schultz 1986, *Astronomy and Astrophysics*, **165**, 287.

L.G. Rosengren 1975, *Appl. Optics*, **14**(8), 1960.

L.S. Rothmann, R.R. Gamache, A. Barbe, A. Goldman, J.R. Gillis, L.R. Brown, R.A. Toth, J.-M. Flaud, and C. Camy-Peyret 1983, *Applied Optics*, **22**, 2247.

Rubens and Nichols 1897, *Ann. der Physik*, **60**, 418.

R.W. Russell, G. Melnick, G.E. Gull, and M. Harwit 1980, *Ap. J. Lett.*, **301**, L57.

E. Sauter, G.V. Schultz, and R. Wohleben 1984, *Intl. J. of IR. and MM. Waves*, **5**, 451.

W. Schockley 1951, *Bell System Tech. J.*, **30**, 990.

M.V. Schneider 1982, chapter 4 in *Infrared and Millimeter Waves*, vol. 6, Academic Press.

C. Sparrow 1982, *The Lorenz Equations: Bifurcations, Chaos, and Strange Attractors*, New York, Springer-Verlag.

L. Spitzer 1978, *Physical Processes in the Interstellar Medium*, New York, Wiley.

G.J. Stacey, S.D. Smyers, N.T. Kurtz, and M. Harwit 1983, *Ap. J. Lett.*, **265**, L7.

G.J. Stacey, J.B. Lugten, and R. Genzel 1986, UC Berkeley preprint.

J.W.V. Storey, D.M. Watson, and C.H. Townes 1980, *Intl. J. of IR. and MM. Waves*, **1**, 15.

R. Stratton 1957, *Proc. Roy. Soc.*, **A242**, 355.

S.M. Sze 1981, *Physics of Semiconductor Devices*, New York, Wiley.

S.W. Teitsworth and R.M. Westervelt 1984, *Phys. Rev. Lett.*, **55**(27), 2587.

S.W. Teitsworth and R.M. Westervelt 1986, *Phys. Rev. Lett.*, **56**(5), 516.

J.J. Thompson 1924, *Philos. Mag.*, **47**, 337.

A.G.G.M. Tielens and D. Hollenbach 1985a, *Ap. J.*, **291**, 722.

- A.G.G.M. Tielens and D. Hollenbach 1985b, *Ap. J.*, **291**, 747.
- M.E. Tiuri 1964, *IEEE Trans. Antennas and Propagation*, **AP-12**, 930.
- H.C. Torrey and C.A. Whitmer 1948, *Crystal Rectifiers*, vol. 15 in the MIT Radiation Laboratory Series, S.A. Goudsmit, J.L. Lawson, L.B. Linford, and A.M. Stone (eds.), McGraw-Hill.
- C.H. Townes and A.L. Schawlow 1975, *Microwave Spectroscopy*, New York, Dover.
- W. Traub and M. Stier 1976, *Applied Optics*, **15**, 364.
- J.R. Tucker 1974, from *Intl. Conf. on Submm. Waves and their Applications*, Atlanta Ga.
- B.E. Turner and H.E. Mathews 1984, *Ap. J.*, **227**, 164.
- O. Van Roos and K.L. Wang 1986, *IEEE Trans. Microwave Theory and Techniques*, **MTT-34**, 183.
- B. Vowinkel 1986, *Intl. J. of IR. and MM. Waves*, **7**, 155.
- A. Waksberg and C. Dreze 1984, *Intl. J. of IR. and MM. Waves*, **5(10)**, 1349.
- D.B. Ward, B. Dennison, G.E. Gull, and M. Harwit 1975, *Ap. J. Lett.*, **202**, L31.
- D.M. Watson 1982a, UC Berkeley Ph.D. Dissertation.
- D.M. Watson 1982b, in proceedings of ESLAB Symposium XVI, Toledo, Spain.
- D.M. Watson 1984, in *Galactic and Extragalactic Infrared Spectroscopy*, M.F. Kessler and J.P. Phillips (eds.), pp. 195 - 219, Riedel.
- D.M. Watson 1985, *Physica Scripta*, **T11**, 33.
- S. Weinreb 1980, *IEEE Trans. Microwave Theory and Techniques*, **MTT-28**, 1041.
- C.O. Weiss 1985, *J. Opt. Soc. Am. B*, **2(1)**, 137.
- C.O. Weiss and J. Brock 1986, *Phys. Rev. Lett.*, **57(22)**, 2804.
- C.O. Weiss and W. Klische 1984, *Optics Comm.*, **51(1)**, 47.
- M.J. Wengler 1987, Calif. Inst. of Technology Ph. D. Dissertation.
- M.J. Wengler and D.P. Woody 1987, *IEEE J. Quantum Electronics*, **JQE-23**, 613.
- R.M. Westervelt and S.W. Teitsworth 1985, *J. Appl. Phys.*, **57(12)**, 5457.
- S.E. Whitcomb and J. Keene 1980, *Appl. Optics*, **19(2)**, .

B.H. Winters, S. Silverman, W.S. Benedict 1964, *J. Quant. Spectroscopy and Radiative Transfer*, **4**, 527.

A. Yariv 1975, *Quantum Electronics*, 2nd ed., New York, Wiley.

D.G. York and J.B. Rogerson Jr. 1975, *Ap. J.*, **203**, 378.

P. Zimmermann, 1987, speaking at the Terahertz Technology Conference, Lake Arrowhead, CA, 1987.

A. Zylberstejn 1962, *Phys. Rev.*, **127**(3), 744.